

Képfeldolgozók és Alakfelismerők

III. Konferenciája

Domaszék
2002. január 23–25.



Szerkesztette:
Kuba Attila, Máté Eörs, Palágyi Kálmán

NJSZT-KÉPAF, 2002.

Képfeldolgozók és Alakfelismerők III. Konferenciája

Domaszék
2002. január 23–25.



Szerkesztette:
Kuba Attila, Máté Eörs, Palágyi Kálmán

NJSZT-KÉPAF, 2002.

ISBN 963 00 9450 9

Nyomdai kivitelezés: SZTE Nyomda
Vezető: Nagy János

Előszó

Nagy öröm számunkra, hogy a Képfeldolgozók és Alakfelismerők Szakosztályának 3. Konferenciáját mi, a Szegedi Egyetem képfeldolgozással foglalkozó munkatársai rendezhetjük meg. Az orvosi képfeldolgozás Szegeden több mint 25 éves múltra tekinthet vissza, de nyilvánvaló, hogy szívesen látjuk és hallgatjuk az egyéb területen tevékenykedő kollégák előadásait. A beküldött előadásokból látszik, hogy ma Magyarországon a képfeldolgozást és (képi) alakfelismerést szinte minden egyetemi városban alkotó módon művelik. Új algoritmusok, rendszerek, elméletek és kutatási eredmények születtek, amelyeket most a Szakosztály tagsága számára is be lehet mutatni. Különösen fontos és jó, hogy számos fiatal van az előadók között, ami azt mutatja, hogy szakmánk a tudományos utánpótlás számára érdekes és vonzó.

A szervezők abban bíznak, hogy méltó keretet tudnak biztosítani ennek a rendezvénynek és a résztvevők nemcsak szakmailag fogják hasznosnak találni a Konferenciát, hanem kellemesen tudják eltölteni ezt a néhány napot Domaszéken.

A 3. KÉPAF Konferenciát támogatta a Neumann János Számítógép-tudományi Társaság és a Szegedi Tudományegyetem Informatikai Tanszékcsoportja.

A szervezők ezúton is köszönetet mondanak a KÉPAF vezetőség tagjainak, a beérkezett előadás-kivonatok bírálatában végzett munkájukért, valamint Balogh Emesének, Dudásné Nagy Mariannak, Nagy Antalnak, Nyúl Lászlónak, Palágyi Kálmánnak és az SZTE Alkalmazott Informatikai Tanszékének munkatársainak a rendezvényt kapcsolatos segítségükért.

Szeged, 2001. január 14.

Kuba Attila, Máté Eörs
szervezők

Tartalom

3D színtér színtér rekonstrukció

Robustifying the iterative closest point algorithm	1
<i>Dmitry Chetverikov, Dmitry Stepanov</i>	
Háromdimenziós színtér-rekonstrukció alapvető módszerei	10
<i>Jankó Zsolt</i>	
Sztereó látásban használt képillesztési algoritmusok	21
<i>Megyesi Zoltán</i>	
Calculating 3D sound-field using 3D image synthesis and image processing	31
<i>Barnabás Aszódi, Szabolcs Czuczor</i>	

Vegyes I

A sequential 3D curve-thinning algorithm	42
<i>Kálmán Palágyi</i>	
Content-based image retrieval algorithms at database server level	52
<i>Krisztián Veréb</i>	
Ecsetvonás transzformáció, mint sztochasztikus képszegmentáló, leíró és tömörítő algoritmus	62
<i>Szirányi Tamás, Tóth Zoltán</i>	
Opto-electronic CNN implementation provides a new and powerful tool for image processing applications	75
<i>Szabolcs Tótkés, László Orzó, Tamás Roska</i>	
Képfeldolgozás oktatása a mérnökképzésben (oktatói és hallgatói vélemények)	88
<i>Berke József, Hegedűs Károly</i>	

Orvosi alkalmazások I

Többdimenziós MRI képek feldolgozása	96
<i>Nyúl László, Jayaram K. Udupa</i>	
Virtual colonoscopy using direct rendering of surfaces from volumetric data	98
<i>Dávid Jocha, József Kolozsár</i>	

Virtual dissection of the colon	109
<i>Emese Balogh, Erich Sorantin, László G. Nyúl, Kálmán Palágyi, Attila Kuba</i>	

Színes képosztályozás alkalmazása szívátültetés során fellépő károsodást csökkentő anyagok hatásának vizsgálata során.....	118
<i>Csomain István, Szabó Gábor, Szabó József</i>	

A szívizom vérátfolyásának vizsgálatát támogató összehasonlító adatok előállítása matematikai módszerrel	123
<i>Máté Eörs, Dudásné Nagy Marianna, Kári Béla</i>	

Orvosi alkalmazások II

Digitális képtároló és képtovábbító rendszer (PACS) a Szegedi Tudományegyetemen.....	132
<i>Almási László, Nagy Antal, Alexin Zoltán, Nyúl László, Kuba Attila, Csernay László</i>	

Orvosi szoftver keretrendszer műtéti tervezéshez	140
<i>Hajdu András, Zörgő Zoltán</i>	

MedEdit: Műtéti tervezést segítő orvosi képfeldolgozó rendszer	152
<i>Ollé Krisztián, Kuba Attila, Varga Endre</i>	

Anyag vizsgálat, terep modellezés

Gravitáció és kristályosodás.....	153
<i>Pólska Csaba, Gácsi Zoltán, Réger Mihály</i>	

Sztereológia és képelemzés anyagtudományi alkalmazása	162
<i>Gácsi Zoltán, Sárközi Gábor</i>	

Acél széles szalagok szövetszerkezetének morfológiai jellemzése	175
<i>Csepeli Zsolt, Gácsi Zoltán</i>	

Háromdimenziós képi (terepi) adatokra épülő természetes folyamatok modellezése.....	184
<i>Berke József, Sisák István, Szabó József, Kelemen Dezső, Szeiler Gábor, Berkéné Várbíró Beáta</i>	

Vector-based map interpretation	192
<i>Endre Katona</i>	

Mozgás elemzés

Számítógépes spermium minősítő rendszer és alkalmazásai	201
<i>Kelemen Dezső, Gábor György, Szabó József</i>	

Pattern matching by Gabor filters for different views in a multi-camera system.....	202
<i>Tamás Szirányi, László Tizedes</i>	

Creating video animations combining stochastic paintbrush transformation and motion detection	210
<i>Levente Kovács, Tamás Szirányi</i>	
Digitális MOzgóképhelyreállító Rendszer Filmarchívumok számára (DIMORF)	218
<i>Czúni László, Szirányi Tamás, Licsár Attila, Hanis Attila, Schanda János, Kránicz Balázs, Farkas Péter</i>	
Supervised training based hand gesture recognition system	231
<i>Attila Licsár, Tamás Szirányi</i>	

Diszkrét tomográfia

Rekonstrukciós algoritmusok $h\nu$ -konvex 4- és 8-összefüggő diszkrét halmazokra	239
<i>Balogh Emese, Kuba Attila, Alberto Del Lungo</i>	
Kétirányú konvex bináris képek rekonstrukciója abszorpciós vetületekből ..	251
<i>Nagy Antal, Kuba Attila</i>	
Dinamikus SPECT vizsgálatok faktorképeinek rekonstrukciója diszkrét tomográfiai eljárással	252
<i>Dudásné Nagy Marianna, Kuba Attila</i>	
Körmetszeteket tartalmazó tárgyak rekonstrukciója néhány vetületből	253
<i>Kiss Zoltán, Kuba Attila, Rodek Lajos</i>	
Diszkrét tomográfiai képek VRML alapú megjelenítése	254
<i>Ruskó László, Kuba Attila, Balogh Emese</i>	

Vegyes II

An algorithm to register sets of 3D points to 3D lines for using arbitrary frame devices in image guided percutaneous therapies	255
<i>Attila Tanács, Gábor Fichtinger, Attila Kuba</i>	
Approximating the Euclidean circle using neighbourhood sequences	260
<i>András Hajdu, Benedek Nagy</i>	
Extracting feature vectors for character recognition by Walsh transformation	272
<i>Attila Fazekas, András Hajdu</i>	
Support Vector Machines for Face Detection	279
<i>Attila Fazekas, Ioan Buciu, Constantine Kotropoulos, Ioannis Pitas</i>	

Robustifying the Iterative Closest Point Algorithm

Dmitry Chetverikov and Dmitry Stepanov

Computer and Automation Research Institute,
Budapest, Kende u.13-17, H-1111 Hungary
csetverikov@sztaki.hu

Abstract. The problem of geometric alignment of two roughly pre-registered, partially overlapping, rigid, noisy 3D point sets is considered. A new natural and simple, robustified extension of the popular Iterative Closest Point (ICP) algorithm [1] is presented, called the Trimmed ICP (TrICP). The new algorithm is based on the consistent use of the Least Trimmed Squares (LTS) approach in all phases of the operation. Convergence is proved and an efficient implementation is discussed. TrICP is fast, applicable to overlaps under 50%, robust to erroneous measurements and shape defects, and has easy-to-set parameters. ICP is a special case of TrICP when the overlap parameter is 100%. Results of testing the new algorithm are shown.

1 Introduction

This paper addresses the problem of Euclidean alignment of two roughly pre-registered, partially overlapping 3D point sets in presence of measurement outliers and, possibly, shape defects. This problem has been mainly considered in 3D model acquisition (reverse engineering, scene reconstruction) and motion analysis, including model-based tracking. (See [13] for an overview of recent applications.) Given two 3D point sets, \mathcal{P} and \mathcal{M} , the task is to find the Euclidean motion that brings \mathcal{P} into the best possible alignment with \mathcal{M} .

The Iterative Closest Point (ICP) algorithm proposed by Besl and McKay [1] is a standard solution to the alignment problem. This iterative algorithm has three basic steps: 1. pair each point of \mathcal{P} to the closest point in \mathcal{M} ; 2. compute the motion that minimises the mean square error (MSE) between the paired points; 3. apply the motion to \mathcal{P} and update the MSE. The three steps are iterated; the iterations have been proved to converge in terms of the MSE.

Independently, Chen and Medioni [2] published a similar iterative scheme using a different pairing procedure based on surface normal vector. This formulation is only applicable to points on surfaces. In this paper, we prefer the formulation by Besl and McKay which is applicable to volumetric as well as surface measurements.

The idea of ICP proved very fruitful as it was followed by numerous applications, improvements and modifications. A comprehensive survey oriented towards range images is provided in the PhD thesis by Pulli [9]. Rusinkiewicz

and Levoy [12] give a fresh update of the variants of the ICP algorithm. They classify the variants according to the way the algorithms: (1) select subsets of \mathcal{P} and \mathcal{M} ; (2) match (pair) points; (3) weight the pairs; (4) reject some pairs; (5) assign error metric; (6) minimise the error metric.

Selection usually refers to random sampling of points when using a Monte Carlo technique, such as the Least Median of Squares (LMedS) [10, 13]. Pairs can be weighed or rejected based on the distribution of distances [14] or some geometric constrains [7]. Different cost functions and minimisation procedures are applied. For example, a recent paper by Fitzgibbon [5] presents an attempt of direct, rather than iterative, minimisation of the cost function (MSE) using the nonlinear Levenberg-Marquardt algorithm.

Most of the above modifications of ICP seek to improve robustness, convergence (speed) and precision. The most critical issue is probably that of robustness, as the original algorithm assumes outlier-free data and \mathcal{P} being a subset of \mathcal{M} , in the sense that each point of \mathcal{P} has a valid correspondence in \mathcal{M} . Numerous attempts have been made to robustify ICP by rejecting wrong pairs. In particular, robust statistics have been applied, such as LMedS or the Least Trimmed Squares (LTS) [11, 9]. Pajdla and Van Gool [7] proposed the Iterative Closest Reciprocal Point (ICRP) algorithm that exploits the ϵ -reciprocal correspondence: given a point $\mathbf{p} \in \mathcal{P}$ and the closest point $\mathbf{m} \in \mathcal{M}$, \mathbf{m} is back-projected onto \mathcal{P} by finding the closest point $\mathbf{p}' \in \mathcal{P}$. If $\|\mathbf{p} - \mathbf{p}'\| > \epsilon$, the pair (\mathbf{p}, \mathbf{m}) is rejected.

Often, different heuristics are combined, making the resulting ICP-variant efficient in cases when the underlying — sometimes, implicit — assumptions are met. Such heterogeneous combinations are difficult to analyse; in particular, convergence properties remain unclear.

Computational efficiency is another important issue, since some applications require fast real-time operation for medium-size datasets, such as range images [9]. Various data structures, like k-D tree [6] or spatial bins [14], are used to facilitate search of the closest point. To speed up the convergence, normal vectors are considered, which is mainly helpful in the beginning of the iteration process [9].

In this paper, we concentrate on the issue of robustness. A new natural and simple, robustified extension of ICP is presented, called the Trimmed ICP (TrICP). The new algorithm is based on the consistent use of the Least Trimmed Squares (LTS) approach in all phases of the operation. LTS [11] means sorting the square errors and minimising a certain number of smaller values; LMedS [10] minimises the median, that is, the value in the middle of the sorted sequence.

Previously, LTS has only been used in the context of randomised, Monte-Carlo type initial estimation of the alignment parameters [9], following the guidelines of the classical approach [10] to robust regression and outlier detection. In this approach, model parameters are repeatedly estimated as random samples are drawn whose size is sufficient for the estimation. After the initial estimation, outliers are detected and rejected, and the final least squares solution is obtained for inliers only.

LTS is preferred to LMedS because it has better convergence rate and a smoother objective function [11]. However, as robust statistics in the context of a randomised approach, LTS and LMedS have the same breakdown point of 50%. This means that the overlap between the two point sets has to exceed 50%.

Our basic observation is that LTS fits the original scheme of ICP without any significant modification. At each step of iteration, the optimal motion can be computed for trimmed squares in exactly the same way as it is done in ICP for all squares. (The median of squares does not facilitate this computation, rendering the LMedS variant [13] inapplicable to large point sets.) At the same time, trimming the squares makes the algorithm robust in the original deterministic framework, without randomisation. The resulting algorithm, the Trimmed ICP, is applicable to overlaps under 50%. As no additional heuristics are used, the convergence of the algorithm is easy to prove.

The paper is organised as follows. Section 2 formulates the problem and describes the framework of the new algorithm. Relevant details of implementation are discussed in section 3; convergence is proved in section 4. Results of tests are shown in section 5.

2 The new algorithm

Following the notation of [13], consider two sets of 3D points to align: the *data* set $\mathcal{P} = \{\mathbf{p}_i\}_1^{N_p}$ and the *model* set $\mathcal{M} = \{\mathbf{m}_i\}_1^{N_m}$. Usually, the numbers of points in the two sets are different: $N_p \neq N_m$. A large portion of the data points may have no correspondence in the model set. Assume the minimum guaranteed rate of the data points that can be paired is known; we will call this rate the *minimum overlap* and denote it by ξ . Then, the number of the data points that can be paired is $N_{po} = \xi N_p$.

If the value of ξ is unknown, one can run TrICP several times and select a result that combines a good MSE with the highest possible overlap. Our experience shows that a few runs are usually sufficient to set the parameter.

Similarly to most iterative algorithms, including ICP, our algorithm assumes that \mathcal{P} and \mathcal{M} have been *roughly pre-registered*, either manually or automatically. This can be done, for example, by aligning a few characteristic points or, in a controlled measurement setup, by calculating the sensor motion between the two views. It should be emphasised, however, that the initial alignment can be fairly rough: TrICP has been successfully applied to the initial relative rotations of up to 30° .

Also, it is assumed that the overlapping part of the two sets is characteristic enough to allow for unambiguous matching. In particular, this part should not be symmetric and ‘featureless’. This assumption is typical for most point set registration algorithms. On the other hand, we do not assume that the points are surface measurements: contrary to the methods relying on surface normal, TrICP can be applied to volumetric sets as well.

Under these assumptions, the problem is to find the Euclidean transformation that brings an N_{po} -point subset of \mathcal{P} into the best possible alignment with \mathcal{M} .

For an Euclidean motion with rotation matrix \mathbf{R} and translation vector \mathbf{t} , denote the transformed points of the data set by

$$\mathbf{p}_i(\mathbf{R}, \mathbf{t}) = \mathbf{R}\mathbf{p}_i + \mathbf{t}, \quad \mathcal{P}(\mathbf{R}, \mathbf{t}) = \{\mathbf{p}_i(\mathbf{R}, \mathbf{t})\}_1^{N_p} \quad (1)$$

Define the *individual distance* from a data point $\mathbf{p}_i(\mathbf{R}, \mathbf{t})$ to the model set \mathcal{M} as the distance to the closest point of \mathcal{M} :

$$\mathbf{m}_{cl}(i, \mathbf{R}, \mathbf{t}) = \arg \min_{\mathbf{m} \in \mathcal{M}} \|\mathbf{m} - \mathbf{p}_i(\mathbf{R}, \mathbf{t})\| \quad (2)$$

$$d_i(\mathbf{R}, \mathbf{t}) = \|\mathbf{m}_{cl}(i, \mathbf{R}, \mathbf{t}) - \mathbf{p}_i(\mathbf{R}, \mathbf{t})\| \quad (3)$$

We wish to find the motion (\mathbf{R}, \mathbf{t}) that minimises the sum of the *least* N_{po} square individual distances $d_i^2(\mathbf{R}, \mathbf{t})$.

The conventional ICP algorithm assumes that all data points can be paired: $\xi = 1$ and $N_{po} = N_p$. TrICP provides a smooth transition to ICP as $\xi \rightarrow 1$. In particular, the convergence of TrICP follows from the convergence of ICP.

The structure of TrICP is similar to that of ICP. The basic idea of is to consistently use the least trimmed squares (LTS) in all major aspects of operation: to cope with outliers, shape defects, or just partial overlap; to estimate the optimal transformation at each iteration step; and to form the global cost function which is minimised. The main steps of TrICP are as follows. These steps are iterated until any of the stopping conditions described below is satisfied. The iterations are started with $S_{LTS} = \text{huge_number}$.

Trimmed Iterative Closest Point Algorithm

1. For each point of \mathcal{P} , find the closest point in \mathcal{M} and compute the individual distances d_i^2 (eq.(3)).
2. Sort d_i^2 in ascending order, select the N_{po} least values and calculate their sum S'_{LTS} .
3. If any of the stopping conditions is satisfied, exit; otherwise, set $S_{LTS} = S'_{LTS}$ and continue.
4. Compute for the N_{po} selected pairs the optimal motion (\mathbf{R}, \mathbf{t}) that minimises S'_{LTS} .
5. Transform \mathcal{P} by (\mathbf{R}, \mathbf{t}) (eq.(1)) and go to 1.

We use the standard stopping conditions [13] related to the number of iterations N_{iter} and the mean square error for the N_{po} selected pairs:

- the maximum allowed N_{iter} has been reached, or
- the MSE S'_{LTS}/N_{po} is sufficiently small, or
- the relative change of the MSE $|S_{LTS} - S'_{LTS}|/S_{LTS}$ is sufficiently small.

Note that in [13] absolute, rather than relative, MSE change is tested.

3 Implementation details

Similarly to any variant of ICP, fast implementation of TrICP needs an efficient data structure supporting the closest point search in *step 1*. Currently, we use a simple boxing structure [3] that partitions the space into uniform boxes, cubes. (The paper [3] presents a 2D version of the structure; extension to 3D is straightforward.) Given a point in space, only the box containing this point and the adjacent boxes are to be considered during the search.

The heap sort [8] is used to efficiently sort the distances in *step 2*. Denote by $\{d_{si}\}_1^{N_p}$ the sorted individual distances:

$$d_{s1} \leq d_{s2} \leq \dots \leq d_{s_{N_{po}}} \leq \dots \leq d_{s_{N_p}}.$$

Then the sum of trimmed squares and the trimmed MSE are defined as

$$S'_{LTS} = \sum_{si=s1}^{s_{N_{po}}} d_{si}^2 \quad (4)$$

$$e = \frac{S'_{LTS}}{N_{po}} \quad (5)$$

In the stopping conditions of *step 3*, S'_{LTS} is compared to the trimmed squares of the previous step S_{LTS} . Initially, S_{LTS} is set to a huge value to avoid occasional stopping at the first iteration because of small relative change of MSE.

The optimal motion (\mathbf{R}, \mathbf{t}) in *step 4* is computed by the unit quaternion method due to Horn [4]. The same method was used in the original version of ICP [1]. There are different analytical ways to estimate the 3D rigid motion that minimises the sum of the square distances between the corresponding points. In [4], four such techniques were compared and the unit quaternion method was found to be robust with respect to noise, stable in presence of degenerate data ('flat' point sets) and relatively fast.

In *step 5*, \mathcal{P} is transformed according to the optimal motion. The transformation decreases (at least, does not increase) the *sum* of the N_{po} least square distances $\{d_{si}^2\}_1^{N_{po}}$. However, *some* of the individual distances, including the largest one, may occasionally increase. TrICP keeps track of the largest distance d_{max} and updates it after the transformation of step 5.

d_{max} falls sharply during the first few iterations. It is used to adaptively set the box size in step 1: at k^{th} iteration, we set $D_{box}(k) = d_{max}(k-1)$. This speeds up the search while ensuring that the N_{po} least distances will be obtained. Indeed, all the N_{po} pairs of the previous step will be found. If any other pair has to enter the list of the N_{po} least distances, then for this pair $d_i < d_{max}(k-1)$. Therefore, the pair will be found with the box size $d_{max}(k-1)$.

The box size and the search space normally decrease as the two sets get closer. The initial size D_{box} is a parameter of TrICP specifying the largest possible distance between the corresponding points. This parameter depends on dimensions and initial orientations of the two sets. Updating the box size is especially efficient at the beginning of the iterations.

4 Convergence

The difference between ICP and TrICP is that the latter minimises the MSE for N_{po} best pairs, while the former minimises the MSE for all pairs. This difference seems minor and only technical, but it is essential. Most previous attempts to robustify ICP incorporated some additional geometric or statistical heuristics which were not mathematically coherent with the original idea. The result was a combined algorithm whose properties, especially convergence, were difficult to analyse. TrICP incorporates a robust and statistically efficient approach, the least trimmed squares, in a way totally compatible with the philosophy and data structure of ICP. An important advantage of this natural extension is that the *convergence* of TrICP can easily be proved similarly to that of ICP.

Indeed, Besl and McKay [1] prove that the ICP algorithm always converges monotonically to a local minimum with respect to the MSE cost function. They do this by showing that the sequence of MSE values is nonincreasing and bounded below (by zero).

At each iteration step of ICP, there are two operations that affect the MSE: the optimal motion and the subsequent selection of the closest point. The optimal motion does not increase the MSE by definition: if it did, it would be inferior to the identity transformation, as the latter does not change the MSE. Besl and McKay point out that the selection of the closest point cannot increase the MSE either.

With TrICP, the situation is the same. The following theorem is valid.

Theorem: The Trimmed Iterative Closest Point algorithm always converges monotonically to a local minimum with respect to the trimmed mean-square distance objective function (5).

Proof: An iteration of TrICP includes three basic operations in the following order: (a) the optimal motion is computed, (b) the closest points are found and (c) the N_{po} least distances are selected. (In the first iteration, steps 1 and 2 just initialise the trimmed MSE.)

Let $e(k)$ be the trimmed MSE before the optimal motion. Similarly to the ICP, the optimal motion cannot increase the sum of trimmed squares S'_{LTS} : $0 \leq e_a(k+1) \leq e(k)$.

Operations (b) and (c) can also modify the least trimmed squares. Consider the list \mathcal{L} of the N_{po} pairs forming $e_a(k+1)$. View it as the starting point for updating the LTS. Updating the closest points for the data points of \mathcal{L} does not increase the MSE of \mathcal{L} , since no individual distance increases: $0 \leq e_b(k+1) \leq e_a(k+1) \leq e(k)$.

Now, consider those points of \mathcal{P} which are not in \mathcal{L} . When selecting the N_{po} least distances, any of these points can only enter \mathcal{L} if its pair substitutes in \mathcal{L} a pair with a larger d_i^2 . Consequently, the sum of the N_{po} least squares cannot increase:

$$0 \leq e(k+1) \doteq e_c(k+1) \leq e_b(k+1) \leq e_a(k+1) \leq e(k).$$

This proves convergence of the Trimmed ICP to a local minimum. *Q.E.D.*

Convergence to global minimum depends on the initial guess. To avoid local minima, ICP is usually run several times at different conditions. We have experienced that TrICP tends to avoid local minima as it is in a sense perturbed by permanent variation of the points that are active at each particular iteration. The analysis of how the LTS might improve the capability to reach the global minimum is beyond the scope of this paper; this is a topic of future research. At the same time, varying ξ one can run TrICP at different conditions and select the best result.

5 Tests

Figure 1 compares ICP and TrICP in aligning two partially overlapping and differently rotated measurements of Frog. Each of the two sets has about 3000 points. Some numerical results are shown in table 1, including number of iterations and execution time on an HP Visualize B132 workstation. (On a modern PC, the time would be 5–10 times shorter.) The TrICP alignment is much better and faster.

Table 1. Numerical results for Frog data

Method	N_{iter}	MSE	Exec.time
ICP	45	5.83	73 sec
TrICP 70%	88	0.10	19 sec

A systematic study is in progress, aimed to quantitatively compare TrICP to ICP, ICRP [7] and other methods for a large group of 2D and 3D shapes. Due to lack of space, a few results are only shown here. Figure 2 presents alignments for two 2D shapes, which are views of an ideal and a defective ferrite cores. The defective core is a distorted and incomplete version of the ideal one, rotated by 20.0° . The TrICP alignment is more precise. ICRP can be improved by running it a second time with modified parameters, using as input the output of the first run. However, TrICP has the advantage of proved convergence.

Acknowledgement

This work was partially supported by the EC grant ICA1-CT-2000-70002: MIRACLE - Centre of Excellence.

References

1. P. Besl and N. McKay. A Method for Registration of 3-D Shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14:239–256, 1992.

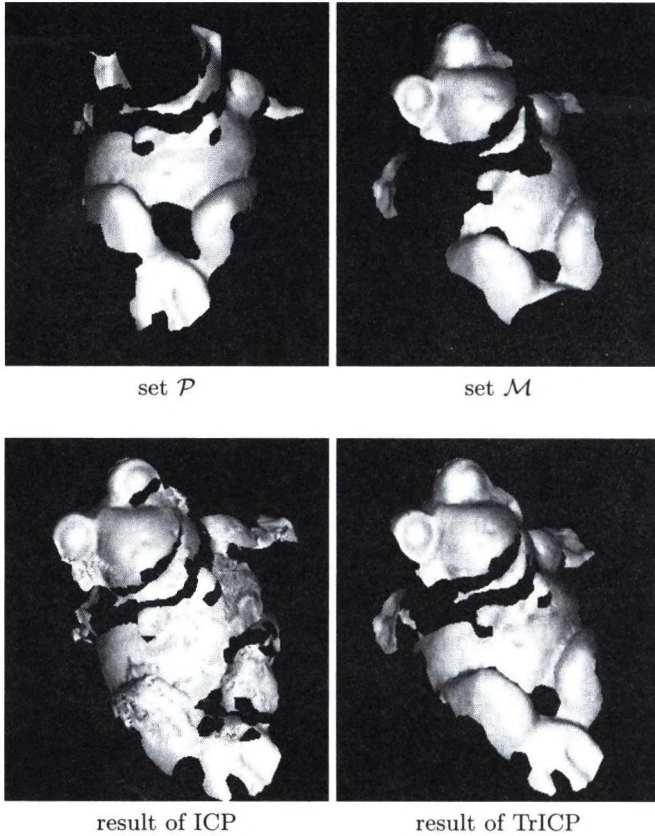


Fig. 1. Aligning two measurements of Frog.

2. Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10:145–155, 1992.
3. D. Chetverikov. Fast neighborhood search in planar point set. *Pattern Recognition Letters*, 12:409–412, 1991.
4. D. Eggert, A. Lorusso, and R. Fisher. Estimating 3-D Rigid Body Transformations: A Comparison of Four Major Algorithms. *International Journal of Machine Vision and Applications*, 9:272–290, 1997.
5. A. Fitzgibbon. Robust Registration of 2D and 3D Point Sets. In *British Machine Vision Conference*, 2001.
6. J. Friedman, J. Bentley, and R. Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Trans. on Mathematical Software*, 3:209–226, 1977.
7. T. Pajdla and L. V. Gool. Matching of 3-D Curves using Semi-differential Invariants. In *5th International Conference on Computer Vision*, pages 390–395, 1995.
8. W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992.

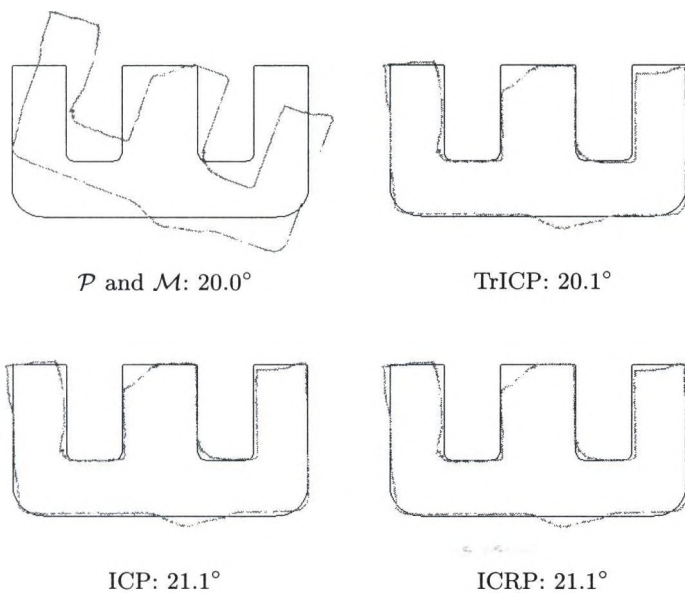


Fig. 2. Aligning ferrite cores.

9. K. Pulli. *Surface Reconstruction and Display from Range and Color Data*. PhD thesis, University of Washington, Seattle, 1997.
10. P. Rousseeuw and A. Leroy. *Robust Regression and Outlier Detection*. Wiley Series in Probability and Mathematical Statistics, 1987.
11. P. Rousseeuw and B. van Zomeren. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85:633–651, 1990.
12. S. Rusinkiewicz and M. Levoy. Efficient Variants of the ICP Algorithm. In *Third International Conference on 3D Digital Imaging and Modeling*, 2001.
13. E. Trucco, A. Fusiello, and V. Roberto. Robust motion and correspondence of noisy 3-D point sets with missing data. *Pattern Recognition Letters*, 20:889–898, 1999.
14. Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13:119–152, 1994.

Háromdimenziós színtér-rekonstrukció alapvető módszerei

Jankó Zsolt¹

MTA Számítástudományi és Automatizálási Kutatóintézete
janko@sztaki.hu

Absztrakt. A számítógépes látás egyik alapvető problémája, hogy hogyan lehet előállítani kétdimenziós fényképekből az eredeti színtér háromdimenziós modelljét. Amikor fényképet készítünk valamilyen háromdimenziós objektumról, adatot veszítünk (az ún. mélységet). Ezt akkor tudjuk visszanyerni, ha a kép mellett valamilyen további információval rendelkezünk. Mi azt az esetet vizsgáljuk, amikor ugyanarról a színterről kettő vagy több képünk van. A cikkben ismertetünk egy jelenleg általánosan alkalmazott, többlépéses módszert. Röviden szólunk az egyes lépésekről, a felmerülő problémákról, és a manapság alkalmazott megoldásokról, technikákról.

1. Bevezetés

A színtér-rekonstrukció célja egy rögzített színterről készült képfelvételek alapján az eredeti színtér egy háromdimenziós modelljének meghatározása. Amikor képfelvételeket készítünk, a háromdimenziós világot egy kétdimenziós síkra vetítjük le. Ezáltal természetesen adatokat veszítünk, amit úgy pótolhatunk, ha több kép áll a rendelkezésünkre. A képek alapján a megfelelő adatok kinyerése, és ezek összehangolása igen bonyolult feladat.

Megemlítjük, hogy amikor a világ síkra való levetítése a célunk, akkor *direkt problémáról* beszélünk. A színtér-rekonstrukció ennek a fordítottja, *inverz probléma*.

A cikkben legelőször a kamera egy matematikai modelljét adjuk meg. A kamera a 3D-s világ leképezésének alapvető eszköze. Ismernünk kell ahhoz, hogy megoldhassuk az inverz feladatot. Utána a rekonstrukció során felmerülő problémákat tárgyaljuk, majd ismertetünk egy többlépéses módszert, amit jelenleg általánosan alkalmaznak a probléma megoldására.

2. A kamera

A 3D-s színtér 2D-s képre való perspektív levetítését a következő egyenlettel szokás megadni:

$$\tilde{\mathbf{u}} \simeq P\tilde{\mathbf{X}}, \quad (1)$$

ahol P egy 3×4 -es, ún. projekciómátrix, $\tilde{\mathbf{X}} = [X, Y, Z, 1]^T$ és $\tilde{\mathbf{u}} = [x, y, 1]^T$ pedig a színtér, illetve a kép egy-egy pontja, homogén koordinátákkal megadva. A homogén koordináták miatt az egyenlőség akkor igaz, ha az egyik oldalt valamilyen (nem nulla) skalárral beszorozzuk. Ezt fejezi ki a \simeq jel.

Ha a vetítést kamerával végeztük, a projekciómátrixot felbonthatjuk a következőképpen:

$$P = [KR \mid -KRt]. \quad (2)$$

A képletben R és t a kamera térbeli helyzetét és irányát adja meg. R egy 3×3 -as forgatásmátrix, t pedig egy 3×1 -es eltolásvektor. Ezek elemeit hívjuk a kamera külső paramétereinek. K egy 3×3 -as felső háromszögmátrix, ami a kamera belső paramétereit tartalmazza.

A most bevezetett kameramodell igen általános, előfordul, hogy nekünk ennél egyszerűbb is megfelel. Ha például a színtér elég messze van a nézőponttól, a valóságban párhuzamos egyenesek a képen is párhuzamosnak tűnhetnek. Ekkor célszerűbb inkább affin kameramoddellel dolgozni, amit a perspektív modell egyszerűsítésével kapunk [1].

3. A rekonstrukciós probléma

Adottak tehát a képek egy rögzített színtérről. Ezek alapján szeretnénk meghatározni egy 3D-s modellt. Tegyük fel, hogy N képünk van. Legyen $\tilde{\mathbf{u}}_j^i \simeq P^i \tilde{\mathbf{X}}_j$ a színtér j -edik pontjának az i -edik képsíkra vett vetülete. A rekonstrukciós problémát röviden úgy fogalmazhatjuk meg, hogy az ismert $\tilde{\mathbf{u}}_j^i$ képpontok alapján keressük a P^i projekciómátrixokat és az $\tilde{\mathbf{X}}_j$ 3D-s pontokat, úgy, hogy az

$$\tilde{\mathbf{u}}_j^i \simeq P^i \tilde{\mathbf{X}}_j \quad (3)$$

egyenlőség minden esetben teljesüljön.

Ha egyetlen képünk van, akkor valamilyen heurisztikus módszerrel számolhatjuk ki a 3D-s pontokat. Ezzel az esettel mi a cikkben nem foglalkozunk, érdeklődők figyelmébe ajánljuk R. Hartley és A. Zisserman könyvét: [2].

Több kép esetén úgy juthatunk többlet-információhoz, ha meghatározzuk, hogy az egyes képeken melyik pontok felelnek meg ugyanannak a 3D-s pontnak, azaz melyek az összetartozó képpontok. A kezdeti megfeleltetések meghatározása a színtér-rekonstrukció első lépése.

A megfeleltetési probléma

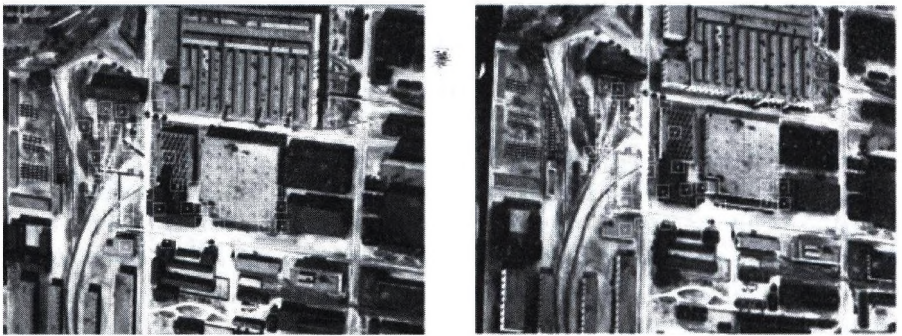
A megfeleltetések keresése során általában egyszerre csak két képet vizsgálunk. Olyan pontpárokat keresünk, amik nagy valószínűséggel ugyanannak a 3D-s pontnak a képei.

Alapvetően két különböző megközelítést ismerünk a megfeleltetések meghatározására.

Régió, vagy korreláció alapú megközelítés. Pontokat választunk ki az egyik képen (általában valamilyen jellemző alapján), majd ezeknek megfelelő pontokat keresünk a másik kép egy leszűkített részén (nem az egész képen). A pontok összehasonlítása a pont-környezetek korrelációs értéke alapján történik. A legnagyobb gondot az jelenti, hogy hogyan válasszuk ki azokat a pontokat, amelyek elég sajátosságok ahhoz, hogy a másik képen is megtaláljuk a nekik megfelelő pontokat. Elég sajátosságos pontok például a sarokpontok.

A módszer elsősorban akkor használható, ha a két kép nem tér el jelentősen egymástól. Komoly gondot okozhat a perspektív, vagy az affin torzulás, valamint a változó fényviszonyok. Nagy a hiba valószínűsége, ezért sok pontpár illesztése ajánlatos, hogy a rosszakat később ki lehessen szűrni. A módszer előnye, hogy egyszerű és könnyű implementálni.

Egy elég jó megvalósítás megtalálható J. Shi és C. Tomasi cikkében [3]. Az ennek alkalmazásával kapott eredmény az 1. ábrán látható.

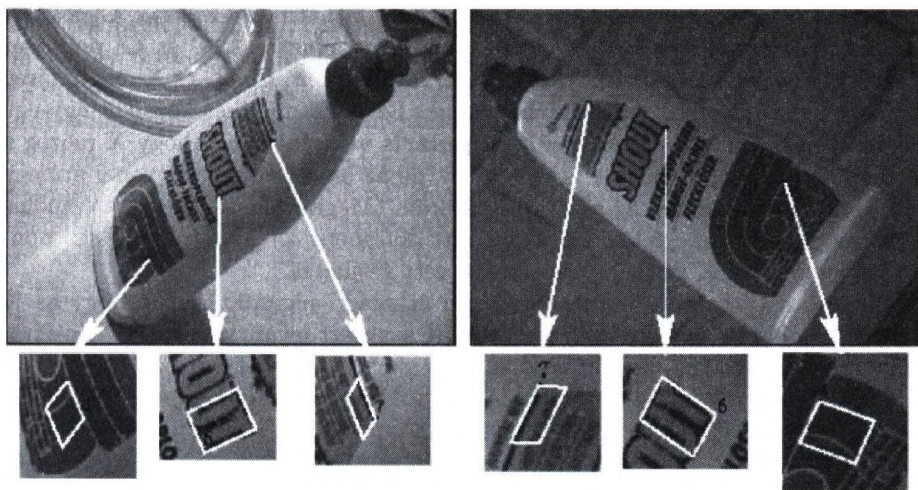


1. ábra: A sajátosságos pontok és megfelelőik (bekeretezve)

Tulajdonság alapú megközelítés. Ebben az esetben tulajdonsággal felruházható struktúrákat, területeket keresünk a két képen. A struktúrák lehetnek vonalak, paralelogrammák, ellipszisek, amelyekhez affin (esetleg projektív) invariáns jellemzőket rendelünk. A struktúrák párbaállítását a jellemzők segítségével történik. Itt is elsősorban az okoz problémát, hogy megtaláljuk azokat a területeket, amelyeket aztán párba tudunk állítani.

Ez a módszer, az előzővel szemben, kevésbé érzékeny a zajra és a fényviszonyok változására. A perspektív és az affin torzulást is jobban tudja kezelni, így még elég nagy elmozdulások esetén is használható. Hátrányként elsősorban a bonyolultsága említhető.

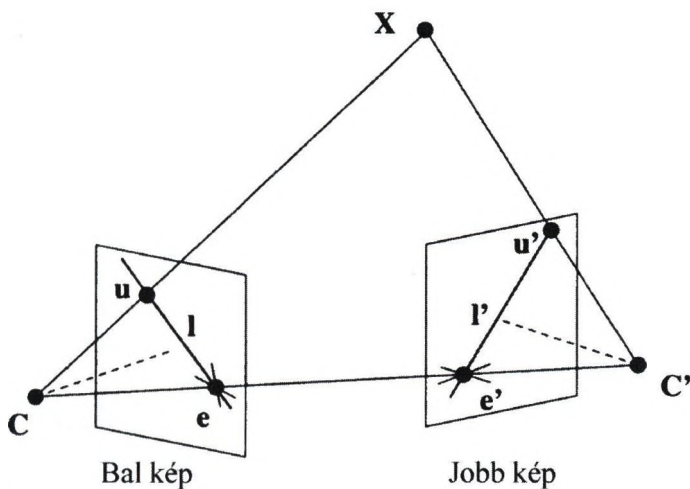
A probléma részletesebb elemzése megtalálható a következő cikkekben: [4, 5]. A 2. ábra affin invariáns területekre mutat példát.



2. ábra: Invariáns területek (forrás: [5])

Epipoláris geometria

Eljutottunk addig, hogy vannak a képeinken egymásnak megfelelő pontjaink. Mit is jelent ez pontosan? Mire tudjuk ezt az információt felhasználni? Ezekkel a kérdésekkel foglalkozik az epipoláris geometria.



3. ábra: Epipoláris geometria

Tekintsük a 3. ábrát. A képen két kamerát láthatunk, C illetve C' fókuszponttal, előttük a képsíkok. X a valós világ egy pontja. A C és C' pontokat összekötő egyenest **bázisvonalnak** hívjuk. Három pont, C , C' , és X meghatároz egy síkot, ezt nevezzük **epipoláris síknak**. Ez a sík mindkét képsíkot egy-egy egyenesben metszi, az l illetve az l' **epipoláris vonalakban**. Ahogy X mozog a térben, az epipoláris vonalak is változnak, de az összes epipoláris vonal átmegegy egy ponton, az e illetve az e' **epipólusokon**. (Megfigyelhető, hogy az epipólus a bázisvonalnak a megfelelő képsíkkal vett dőféspontja.) Az ábrán u és u' pontokkal jelöljük az X pontnak képsíkokra vett vetületét.

Nézzük a C pontból induló és az u -n áthaladó sugarat. Biztos, hogy az X pont valahol ezen az egyenesen lesz. A CX szakaszt levetítve a másik képsíkra, az l' szakaszt kapjuk. Ebből adódik, hogy az u' pont valahol az l' szakaszon található. Ez egy elég erős megszorítás (**epipoláris megszorítás**), ami a keresési teret kétdimenziósról egydimenziósra redukálja.

Mindezt algebrai módon a következő képlettel tudjuk megfogalmazni:

$$\tilde{u}^T F \tilde{u}' = 0, \quad (4)$$

ahol F egy 3×3 -as mátrix, a **fundamentális mátrix**. A fundamentális mátrix a kamerák belső paramétereit és a kamerák egymáshoz való elhelyezkedését (forgatás és eltolás) tartalmazza. Semmilyen információ nincsen benne a kamerák külső paramétereire vonatkozóan (ez érthető is, elvégre az összetartozó pontok alapján nehéz lenne meghatározni, hogy a kamerák pontosan hol helyezkednek el a világban).

Megjegyezzük, hogy abban az esetben, ha három, vagy annál több képünk van, az epipoláris geometria felépítése hasonlóan történik. Fundamentális mátrix helyett azonban trifokális, illetve multifokális tenzorokkal dolgozunk (a tenzor lényegében egy többdimenziós mátrix). A probléma bonyolultsága miatt ezt az esetet nem részletezzük jobban, érdeklődőknek R. Hartley és A. Zisserman könyvét ajánljuk: [2].

A szintér rekonstrukció második lépése tehát az epipoláris geometria felépítése. Két kép esetén ez azt jelenti, hogy meghatározzuk a fundamentális mátrixot. A fundamentális mátrix becslésére nagyon sok módszer van, egyszerű lineáris módszerektől kezdve bonyolultabb, robusztus módszerekig. Két hivatkozást említünk meg, ahol több módszer is össze van foglalva: R. Hartley és A. Zisserman könyvét [2] valamint Z. Zhang cikkét [6]. De ezen felül is számos cikk és könyv foglalkozik ezzel a témával.

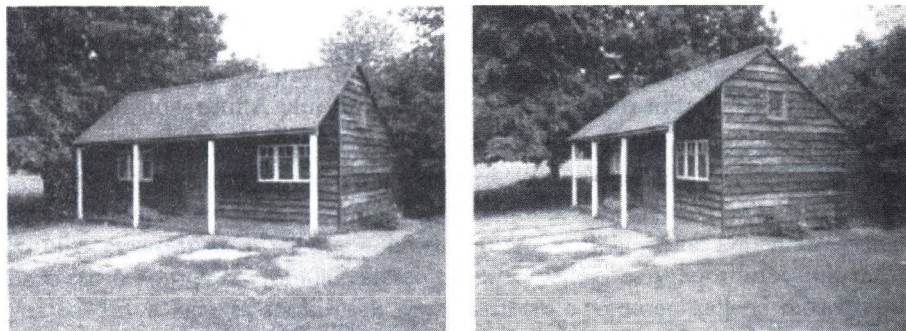
A projekciómátrixok meghatározása

Ha csak annyi ismeretünk van a képekről, hogy melyek az egymásnak megfelelő pontok, akkor a projekciómátrixokat mindössze egy projektív transzformáció erejéig lehet egyértelműen meghatározni. Ez annyit jelent, hogy ha P a projekciómátrix, H pedig egy tetszőleges projektív transzformáció 4×4 -es mátrixa, akkor PH is érvényes projekciómátrix. Ez könnyen látszik az alábbi képletből:

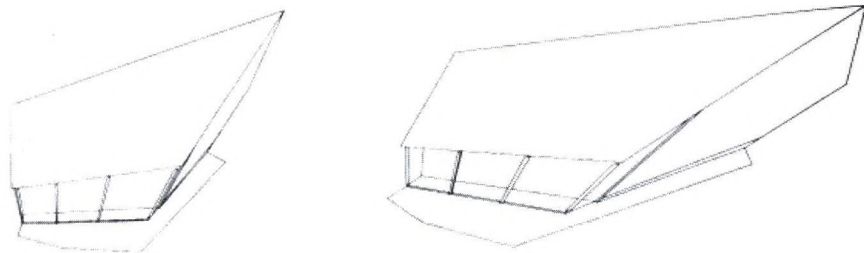
$$\tilde{u} \simeq P\tilde{X} = (PH)(H^{-1}\tilde{X}) = P'\tilde{X}' \quad (5)$$

Nem túl meglepő módon, hasonló a helyzet a fundamentális mátrixszal is. Azaz, ha H egy tetszőleges projektív transzformáció 4×4 -es mátrixa, akkor a P_1 , P_2 projekciómátrixokhoz tartozó fundamentális mátrix megegyezik a P_1H , P_2H projekciómátrixokhoz tartozó fundamentális mátrixszal.

Mint már említettük, mindez azt jelenti, hogy ha az egymásnak megfelelő pontokon kívül nincs más információnk (nem tudunk további megszorításokat tenni a projekciómátrixra), akkor csak egy projektív transzformáció erejéig tudjuk kiszámolni P -t. Mint a 4. ábra is mutatja, ez nem elegendő.



(a)



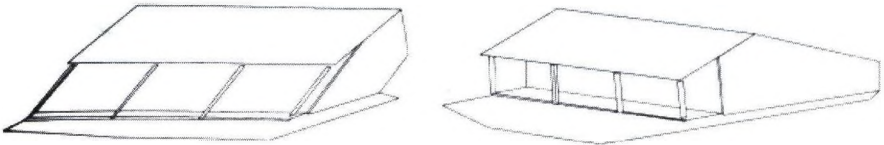
(b)

4. ábra: Projektív rekonstrukció: a) az eredeti képpár b) projektív rekonstrukcióval kapott modell (forrás: [2])

Finomítani szeretnénk a rekonstrukciónkat mindaddig, amíg el nem jutunk a metrikus rekonstrukcióig. Metrikus rekonstrukció esetén a kapott modell már csak egy hasonlósági transzformáció erejéig tér el az eredeti színtértől. A kamera külső paramétereinek pontos ismerete nélkül ennél jobbat nem tudunk elérni. (A külső paramétereket megbecsülhetjük, ha például tudjuk, hogy a képen látható egyik szakasznak a valóságban mekkora a hossza. Általában azonban erre nincs szükség, a metrikus rekonstrukció is megfelelő eredményt ad.)

Azt a folyamatot, amelynek során a projektív rekonstrukciót metrikussá finomítjuk, **rétegzett rekonstrukciónak** hívjuk. Ehhez természetesen további megszorításokat kell tennünk, az epipoláris megszorításon túl. Feltehetjük, hogy ismerjük a kamerák néhány belső paraméterét, vagy fixnek tekinthetjük őket. Egy elég gyakori eset például, hogy ha a képeket ugyanazzal a kamerával készítették, akkor a belső paramétereket állandónak vehetjük. M. Pollefeys, R. Koch és L. Van Gool cikkükben [7] részletesen leírják a további eseteket is, és azt, hogy az egyes esetekben hány kép szükséges a rétegzett rekonstrukcióhoz.

Előfordulhat, hogy az előző módszerrel nem tudunk elegendő megszorítást tenni (például, ha semmilyen információnk nincs a kamerák belső paramétereiről). Ilyenkor máshonnan kell információhoz jutnunk. Kihasználhatjuk mondjuk azt, hogy tudjuk, hogy a kép bizonyos egyenesei a valóságban párhuzamosak. Az is ugyanilyen jó, ha azt meg tudjuk mondani, hogy egy egyenesen felvett két szakasz hossza a valóságban milyen arányban van egymással. A projektív rekonstrukciót így affin rekonstrukcióvá finomíthatjuk. Affin rekonstrukciónál a projekciómátrixok már egy affin transzformáció erejéig egyértelműek. Az affin transzformációk megőrzik az arányokat, mint az az 5. ábrán is látható.



5. ábra: Affin rekonstrukció (forrás: [2])

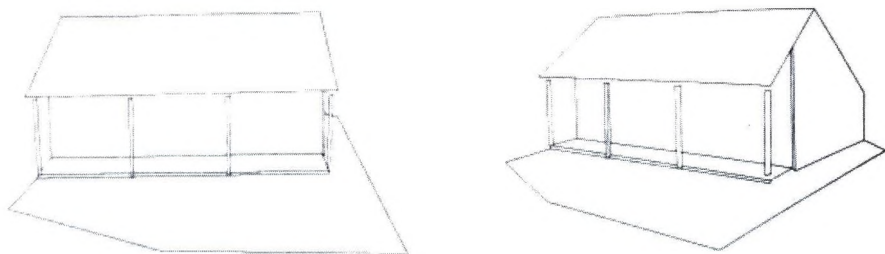
Ahhoz, hogy megkapjuk az affin rekonstrukciót, alapvetően az a fontos, hogy meghatározzuk az ideális síkot (a végtelenben vett síkot, lásd pl. [2]). Ezt lehet mondjuk a párhuzamos egyenesekkel, ahogy azt korábban említettük.

A következő lépés, hogy az affin rekonstrukciót tovább finomítjuk metrikussá. Amíg az előző lépéshez az ideális síkot kellett megbecsülnünk, addig most az ún. abszolút kúpszeletet. Az abszolút kúpszelet a $\sum_{i=1}^4 x_i^2 = 0$ felületnek és az ideális síknak a metszete. A probléma bonyolultsága miatt ennél többet nem írunk erről, inkább megadunk néhány hivatkozást, ahol az érdeklődők további részleteket is megtalálhatnak: [2, 7, 8].

A metrikus rekonstrukcióval kapott, 6. ábrán látható eredmény már megfelel az elvárásoknak.

Sűrű illesztés

A szintér-rekonstrukció negyedik lépése a sűrű illesztés. A kezdeti megfeleltetések keresése során már találtunk néhány pontpárt (vagy pont N -est, ha N képünk van). A 3D-s modellhez általában egy sűrű pontfelhő szükséges. Ahhoz, hogy



6. ábra: Metrikus rekonstrukció (forrás: [2])

ezt megkapjuk, minél több 3D-s pontot meg kell határoznunk, azaz a képeken további egymásnak megfelelő pontokat kell keresnünk. Lehetőleg minél többet, minél sűrűbben.

A kezdeti megfeleltetési problémához képest most már kihasználhatjuk az epipoláris geometria nyújtotta előnyöket. Az epipoláris megszorítás miatt nem kettő-, hanem csak egydimenzióban kell keresnünk a megfelelő pontokat. További technikai egyszerűsítést jelentene, ha el tudnánk érni, hogy két képen az egymáshoz tartozó epipoláris vonalak vízszintesek legyenek, és y koordinátájuk megegyezzen. Így a keresés során csak egy vízszintes egyenesen kellene mozognunk, ami jelentős gyorsítás. Az egyszerűsítő technikát **rektifikálásnak** hívjuk. A megvalósításra több módszer is ismert (lásd pl. [10, 11]), lényegük, hogy valamilyen invertálható transzformációt hajtunk végre a képeken.

Az illesztés viszont továbbra is igen bonyolult feladat, köszönhetően a képek közötti különbségeknek. A legnagyobb gondot a következők okozzák:

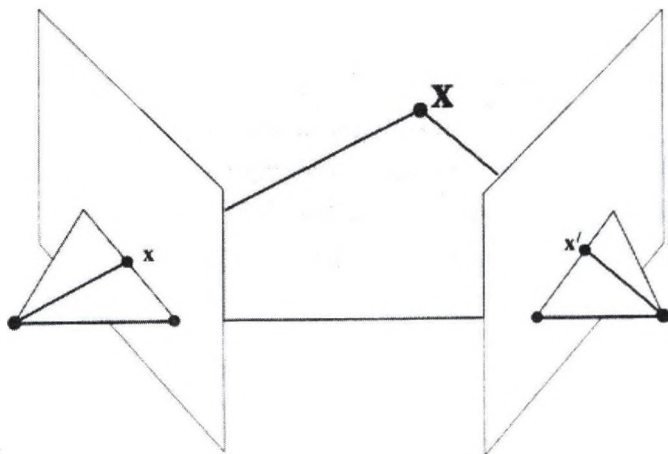
- takarások: lehet, hogy a keresett pont a másik képen nem is látszik, vagy fordítva, eddig eltakart területek kerülnek elő az új képen
- fotometrikus változások: változhat a megvilágítás, ennek következtében az árnyék is, esetleg ugyanaz a terület eltérő intenzitással jelenhet meg a két képen
- alakí torzulás: a változó nézőpont miatt azonos területek más-más alakot vehetnek fel a képeken

A párok illesztésére használt hasonlósági mérték jelen esetben egy korrelációs érték. Természetesen fontos, hogy ennek kiszámítása hatékonyan történjen, és hogy jól jellemezze a hasonlóságot. Gyakran előfordul, hogy az eredeti ponthoz nem a valójában neki megfelelő pont korrelál a legjobban. Általában igen erős heurisztikát kell alkalmaznunk a hibák csökkentése érdekében. A témával bővebben foglalkozik Megyesi Z. cikke [12].

A modell felépítése

A rekonstrukció befejező lépése a modell felépítése. Ismerjük a kamerák projekciómátrixait és az egymásnak megfelelő képi pontokat. Ezekből az adatokból már meg lehet határozni a modell 3D-s pontjait.

Vegyünk egy x pontot az egyik képen. A képhez tartozó projekciómátrix segítségével meg tudjuk adni azt az egyenest, amelyiken rajta kell lennie X -nek, annak a 3D-s pontnak, amelyet levetítve kaptuk x -et. Lényegében a kamera fókuszpontján és a x -en áthaladó egyenesről van szó (lásd 7. ábra).

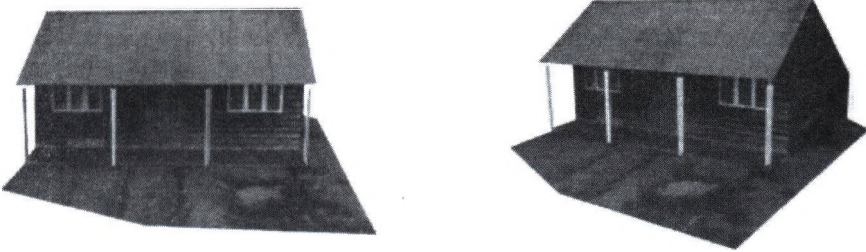


7. ábra: 3D-s pont kiszámítása

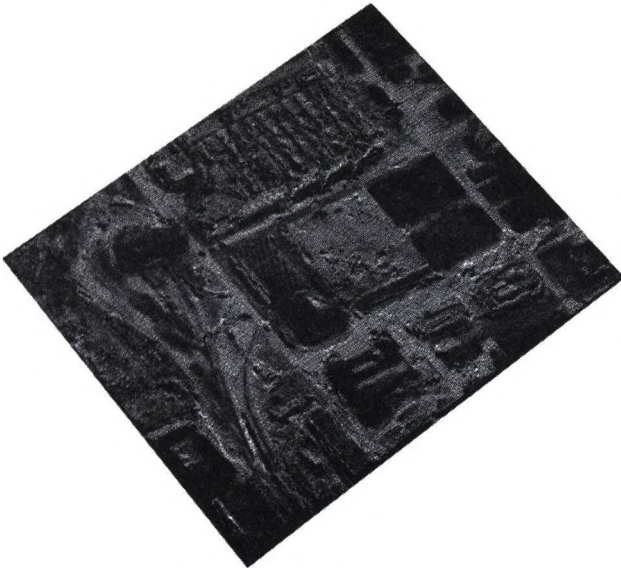
Ha több képünk van, X -nek több ilyen egyenesre kell illeszkednie. Vegyük tehát az egyenesek metszéspontját. Általában azonban, a numerikus hibák és a zaj miatt az egyenesek nem egy pontban metszik egymást, ha egyáltalán van metszéspontjuk. Ekkor a hibák minimalizálása nyújt elfogadható megoldást. Bővebben erről például E. Trucco és A. Verri könyvében [10], vagy R. I. Hartley és P. Sturm cikkében [13] olvashatunk.

Az így kapott pontfelhő már alkalmas a modell ábrázolására. Célszerű azonban a képek alapján a textúrát is ráhelyezni a pontfelhőre, ezáltal az sokkal szemléletesebb lesz. Megjegyezzük, hogy a jelenleg alkalmazott rekonstrukciós módszereknél ritkán alkalmazzák a sűrű illesztést. Ehelyett inkább kihasználják, hogy ismerik a színteret, tudják, hogy a képek mit ábrázolnak. Ha például egy házat, akkor elég, ha a ház sarkainak 3D-s koordinátáit számolják ki, ezeket összekötik, majd síkokat illesztenek rájuk (a ház oldalait), végül a textúrát levetítik az oldalakra. Az eredmény igen szemléletes lesz (lásd 8. ábra), ugyanakkor a módszer nem elég automatikus.

Városról készült légfelvételek esetén is kihasználható, hogy az épületeket sík lapok határolják. De ha mondjuk egy szobáról készülnek felvételek, vagy egyetlen, de bonyolult geometriájú alakzatról, akkor a sűrű illesztés látszik célravezetőnek. A hátrány viszont ekkor az, hogy a kapott modell nem elég látványos (9. ábra).



8. ábra: Texturált modell (forrás: [2])



9. ábra: Sűrű illesztéssel kapott, texturált modell (az 1. ábra képei alapján)

4. Összefoglalás

A cikkben bemutattunk egy lehetséges módszert, amellyel egy rögzített színtér-ről készült fényképfelvételek alapján elő lehet állítani a színtér 3D-s modelljét. A módszer több lépésből állt, ezek a következők voltak: kezdeti megfeleltetések keresése, epipoláris geometria felépítése, a projekciómátrixok meghatározása, sűrű illesztés, a modell felépítése. Az egyes lépéseket röviden elemeztük, beszélünk a felmerülő problémákról, és hogy jelenleg milyen megoldások ismertek.

A vizsgált módszerről elmondható, hogy nem ad tökéletes megoldást. Ennek legalapvetőbb oka a sűrű illesztés nehézsége. A rekonstrukció geometriájával kapcsolatosan számos kutatási eredmény született. Ezt a területet az irodalom nagyon részletesen tárgyalja, és az eredmények is elfogadhatóak. Más a helyzet az illesztéssel. Jelenleg még nem ismert olyan módszer, ami ezt a problémát elég jól megoldaná. Éppen ezért sok rekonstrukciós eljárás ezt a lépést át is ugorja. Viszont ahhoz, hogy valóban automatikus, és széles körben használható rekonstrukciós rendszert készítsünk, elkerülhetetlen a sűrű illesztés. Mindenképpen további kutatások szükségesek, hogy az eredmények igazán látványosak legyenek.

Irodalom

1. Shapiro, L. S.: *Affine Analysis of Image Sequences*. Sharp Laboratories of Europe, Oxford (1995)
2. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press (2001)
3. Shi, J., Tomasi, C.: *Good Features to Track*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR94) Seattle (1994)
4. Schaffalitzky, F., Zisserman, A.: *Viewpoint Invariant Texture Matching and Wide Baseline Stereo*. (2001)
5. Tuytelaars, T., Van Gool, L.: *Wide Baseline Stereo Matching based on local, affinity invariant regions*. British Machine Vision Conference (2000) 412-422
6. Zhang, Z.: *Determining the Epipolar Geometry and its Uncertainty: A Review*. The International Journal of Computer Vision (1997)
7. Pollefeys, M., Koch, R., Van Gool, L.: *Self-Calibration and Metric Reconstruction In spite of Varying and Unknown Intrinsic Camera Parameters*. International Journal of Computer Vision 32(1) (1998)
8. Fusiello, A.: *Uncalibrated Euclidean reconstruction: a review*. Image and Vision Computing (1999)
9. Faugeras, O.: *Stratification of three-dimensional vision: projective, affine, and metric representations*. Optical Society of America (1994)
10. Trucco, E., Verri, A.: *Introductory Techniques for 3-D Computer Vision*. Prentice-Hall, Inc. (1998)
11. Oram, D.: *Rectification for Any Epipolar Geometry*. (2001)
12. Megyesi Z.: *Sztereó látásban használt képillesztési algoritmusok*. KÉPAF (2002)
13. Hartley, R. I., Sturm, P.: *Triangulation*. Computer Vision and Image Understanding (1997)
14. Jankó Zs.: *Háromdimenziós színtér-rekonstrukció, 3D modell előállítás*. ELTE TTK, diplomamunka (2001)
15. Megyesi Z.: *Háromdimenziós színtér-rekonstrukció, képillesztési módszerek*. ELTE TTK, diplomamunka (2001)

Sztereó látásban használt képillesztési algoritmusok

Megyesi Zoltán

MTA Számítástudományi és Automatizálási Kutatóintézete
H-1111 Budapest, Kende u. 13-17
megyesi@sztaki.hu

Absztrakt. A számítógépes látás egyik szűk keresztmetszete a sztereó illesztés. Az illesztés egy objektumról készült sztereó képpár pontjainak minél tökéletesebb megfeleltetése, ami az objektum ismerete nélkül ma is a gépi látási folyamat ellentmondásos és kiforratlan része. Az illesztés szerepe egy elmozdulási térkép elkészítése, amely más kamera-információkkal együtt felhasználható háromdimenziós modellalkotásra. A cikkben a sztereó illesztés során felmerülő problémák ismertetése után módszereket mutatunk azok kezelésére, majd elterjedt algoritmusokat hasonlítunk össze.

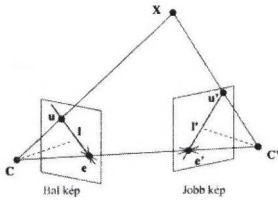
Bevezetés

A gépi látásnak számos formája és számos alkalmazása létezik. A *sztereó látás* feladata elsősorban egy helyszín mélységadatainak meghatározása a helyszínről készült fényképpárok alapján. A mélységadatokat a helyszín tárgyainak képek közötti elmozdulásából, torzulásából lehet kikövetkeztetni, ezért a feladat kulcsfontosságú része az objektumok vagy azok részeinek *megfeleltetése, illesztése*. A megfeleltetések megtalálása azonban számos problémát rejt. Legáltalánosabb esetben a képek között olyan eltérések is lehetnek, hogy komoly intelligencia szükséges a tárgyak azonosításához. Más feladatokhoz hasonlóan itt is mérlegelni kell, mennyi külső információt használhatunk fel, tudva hogy azzal árthatunk az általánosságoknak és automatizáltságoknak. Ésszerű megszorítások azonban jelentősen növelhetik a hatékonyságot.

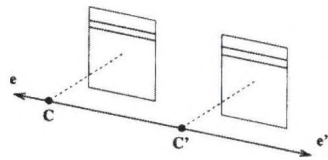
A legalapvetőbb megszorítások a képeket készítő kamerák paramétereire vonatkoznak. A klasszikus sztereó látás feladatát gyakran *rövid bázistávolságú sztereó*nak (short-baseline stereo) is nevezik, mivel itt a képeket készítő kamerák egymástól vett távolsága minimális. Az igazi problémák a *széles bázistávolságú sztereó* (wide-baseline stereo) esetében lépnek fel, amikor a kamerák közötti távolság jelentős, ezzel sok klasszikus illesztő algoritmus alkalmazhatatlanná válik. Számos módszer született, amely felhasználta és kiegészítette a klasszikus szemléletet [5, 6, 9], megtartva annak előnyeit az eredetinelő általánosabb esetben. Más megközelítések [4, 7, 8] magasszintű képi tulajdonságok keresésével ritka, de megbízható megfeleltetéseket adnak. Ezek gyakran erős heurisztikákat alkalmaznak, és sok becsült adatot eredményeznek. Ebben a cikkben a klasszikus illesztési algoritmusok kiegészítéseivel foglalkozunk.

Sztereoó illesztés

Mint említettük az illesztés célja a sztereoó látásban olyan adatok szolgáltatása, amelyek segítségével a képpáron látszó fizikai pontok mélysége kiszámítható. Ehhez olyan pontpárokat keresünk a két képen, amelyek egy látszó fizikai pont középpontos vetületei a két képsíkra. Mindehhez a matematikai háttér az *epipoláris geometria* [2, 5] adja, amely leírja a kamerák és képsíkok helyzetét. Két kép esetén a kamerákat összekötő egyenest *bázisvonalnak*, annak a képsíkokkal vett metszéspontjait *epipólusoknak* hívjuk. Az egyes *kameraközéppontokból* induló egyenesek másik kamera képsíkjára vett középpontos vetületei az *epipoláris vonalak*. Az epipoláris geometria ismeretében, az eredményként kapott pontpárokhoz hozzárendelhető a háromdimenziós fizikai pont. Ez a modell nemcsak a végeredmény szempontjából fontos. Ha vesszük egy X pont u vetületét az egyik képen, a másik képre vett v vetület csak az u -hoz tartozó epipoláris vonalon lehet. Sőt, a v -hez tartozó epipoláris vonal minden pontja is az u -hoz tartozó epipoláris vonalon van (lásd 1. ábra). Ez jelentősen csökkenti a keresési teret.



1. ábra: Epipoláris geometria



2. ábra: Kanonikus konfiguráció

A klasszikus sztereoó illesztés feltételei egy azonos helyszínről készült két kép, amelyeket hogy a két képsík egybeessen párhuzamos helyzetben levő azonos fókusztávolságú kamerákkal készítettek egy adott időpillanatban. Ez az úgynevezett *kanonikus kamera konfiguráció*. Ekkor az epipoláris vonalak párhuzamosak a bázisvonallal, és a megfelelő vonalak egybeesnek (lásd 2. ábra). Ha ezek a feltételek teljesülnek, a feladatunk az első kép pontjainak keresése a másik kép azonos sorában. Ebben az esetben, az epipoláris geometria teljes ismeretére csak a végső lépés, a mélység meghatározása miatt kell, az illesztés anélkül is hatékonyan elvégezhető.

Általánosabb esetben problémát jelent, hogy a teljes epipoláris modell nehezen számolható, és becslése több lépést igényel. A képek alapján megbecsülhető az epipoláris vonalak helyzetét leíró *fundamentális mátrix* ami az illesztés elvégzéséhez elegendő ugyan, de a mélység számításához szükséges a *kamerák kalibrálása*, ami automatikusan nehezen végezhető. A fundamentális mátrix birtokában *rektifikálással* a képek átalakíthatók a kanonikus esetre. Ez a módszer kiterjeszti a klasszikus illesztést általános helyzetből készült képekre, de ezzel új problémák is felmerülnek.

Az illesztés menete

Használva a kanonikus kamera konfigurációt, az egymásnak megfelelő epipoláris vonalak a két kép azonos sorai. Feladatunk tehát az, hogy az első kép sorain képpontonként haladva megkeressük a másik kép azonos sorában legjobb illeszkedő elemet. A képelemeket a környezetükkel együtt hasonlítjuk, hogy elkerüljük a kevés információból fakadó többértelmű és hibás találatokat. A környezet méretének megállapítása fontos, hiszen jelentősen befolyásolhatja az eredményt. A pont környezetek illeszkedésének meghatározására *hasonlósági mértéket* (correlation score) használunk. Minél magasabb a hasonlósági érték annál tökéletesebb az illeszkedés. Ez az érték felhasználható a találat *képelemnél pontosabb finomítására* (subpixel precision). A környező találat értékeire parabolát illesztve a maximum helye pontosabban megadható.

Az eredmény pontpárokat gyakran az első képhez viszonylagosan adjuk meg, úgynevezett *elmozdulási térképet* (disparity map) készítünk. Mivel a képpontok nagyrészéhez rendelünk elmozdulást, ezt a fajta illesztést *sűrű illesztésnek* (dense matching) is hívjuk.

Hasonlósági mértékek

A használt hasonlósági mértékeket nagyban befolyásolja a számítási kapacitás és képek között fellépő várható különbség. A hasonlóságot egy c korrelációs függvény határozza meg, amely egy \mathbf{u} képelem körüli w négyzetes ablakra, és t elmozdulásra

$$c_w(\mathbf{u}, t) = \frac{\psi_w(\mathbf{u}, \mathbf{u} + t)}{\sigma_w(\mathbf{u}, \mathbf{u} + t)} \tag{1}$$

alakú. Itt ψ_w valamilyen additív függvény, σ_w szerepe pedig az érték normálása. A leggyakrabban kétféle korrelációt használnak:

- **Kereszt-korreláció** (Cross-Correlation) [5, 6, 9] Legyen (1) -ben:

$$\psi(\mathbf{u}, \mathbf{v}) = \sum_{(k,l) \in w} [I_1(\mathbf{u} + \begin{bmatrix} k \\ l \end{bmatrix}) - \overline{I_1(\mathbf{u})}] \cdot [I_2(\mathbf{v} + \begin{bmatrix} k \\ l \end{bmatrix}) - \overline{I_2(\mathbf{v})}]$$

ahol $\begin{bmatrix} k \\ l \end{bmatrix}$ a w ablak elemei, és

$$\overline{I_k(\mathbf{u})} = \sum_{(k,l) \in w} \frac{I_k(\mathbf{u} + \begin{bmatrix} k \\ l \end{bmatrix})}{|w|} \tag{2}$$

az \mathbf{u} középpontú ablak intenzitás átlaga. Legyen továbbá $\sigma_w \equiv 1$. Ezesetben az (1) korrelációs érték nem invariáns az intenzitás skálázás változására,

ezért gyakran a

$$\sigma_w(\mathbf{u}, \mathbf{v}) = \sqrt{\sum_{(k,l) \in w} [I_1(\mathbf{u} + \begin{bmatrix} k \\ l \end{bmatrix}) - \overline{I_1(\mathbf{u})}]^2 \cdot \sum_{(m,n) \in w} [I_2(\mathbf{v} + \begin{bmatrix} m \\ n \end{bmatrix}) - \overline{I_2(\mathbf{v})}]^2}$$

értéket használják. Ekkor (1) a **normalizált kereszt-korreláció**, és ez már invariáns mind az intenzitás eltolódásra és skálázás változásra. Különösen normalizált esetben a keresztkorreláció számításigényes, de megbízható.

- **Különbségek négyzetösszege** (Sum of Squared Differences, [6, 1])¹ Legyen (1) -ben:

$$\psi(\mathbf{u}, \mathbf{v}) = - \sum_{(k,l) \in w} [I_1(\mathbf{u} + \begin{bmatrix} k \\ l \end{bmatrix}) - I_2(\mathbf{v} + \begin{bmatrix} k \\ l \end{bmatrix})]^2$$

Ekkor ez a korrelációs érték sem invariáns intenzitás eltolásra és skálázásra ($\sigma_w \equiv 1$), így sokszor

$$\psi(\mathbf{u}, \mathbf{v}) = - \sum_{(k,l) \in w} [(I_1(\mathbf{u} + \begin{bmatrix} k \\ l \end{bmatrix}) - \overline{I_1(\mathbf{u})}) - (I_2(\mathbf{v} + \begin{bmatrix} k \\ l \end{bmatrix}) - \overline{I_2(\mathbf{v})})]^2$$

függvényt használják a korreláció kiszámítására, mivel így (1) intenzitás eltolásra már invariáns lesz. Ez a módszer egyszerű és hatékonyan implementálható.

Megszorítások

Hogy a gyakorlatban is alkalmazható legyen ez a költséges keresés, a hatékonyság és a megbízhatóság javítására megszorításokat tehetünk a képekre. Ezek persze nagyon függenek a feladattól és a képektől, de gyakran jelentősen segítik a keresést. Néhány természetesen adódó és elvárható megkötés, amelyek szükségessé teszik a korrelációra alapuló illesztés hatékony működéséhez:

- **Epipoláris megkötés:** A keresett pontok az epipoláris vonalon vannak, elég az epipoláris vonalak mentén keresni.
- **Fotometrikus megkötés:** Pontnak és párjának intenzitása közel azonos, nem kell intenzitás változásra invariáns korrelációs függvényt alkalmazni.
- **Geometriai hasonlóság megkötés:** A két kép közötti torzulás nem túl nagy, így nincs szükség négyzetnél bonyolultabb ablakokra.
- **Maximális elmozdulás megkötés:** Az elmozdulások egy bizonyos határon belül mozognak, a teljes epipoláris vonalnak csak egy részén kell keresni.

Hasznos, de nem mindig teljesülő megkötések, amelyek kihasználása növelheti a hatékonyságot, vagy segítik a megbízható illesztést:

¹ A név tulajdonképpen nem a korrelációs értékre, hanem az ellenkező előjelű hibafüggvényre vonatkozik, amit minimalizálni szoktak.

- **Egyediségi megkötés:** Minden pontnak legfeljebb egy pont felel meg a másik képen. Ha teljesül, az objektumhatárokon elkerülhetők a hibás illesztések.
- **Folytonossági megkötés:** Az elmozdulások nem változnak gyorsan. A jó találatokból kiindulva tovább csökkenthető a keresési tér.
- **Sorrendiségi megkötés:** Pontok sorrendje a megfelelő epipoláris vonalakon nem változik, azaz nincsenek hirtelen kiugró objektumok. Az ilyen objektumokat nehéz jól illeszteni.
- **Jobb - bal konzisztencia megkötés:** A pontpárok jobb-bal konzisztensek, azaz a képeket fordított sorrendben feldolgozva is azonos pontpárokat kapunk. Ellenőrizve kiszűrhetők a rossz párok.
- **Alaki változás megkötés:** Élnek élre kell esnie. Kihasználható pontosabb keresés megvalósítására.

Hatékonyság

A korrelációs értékek megadása a megszorítások alkalmazása esetén is számítási-gényes. Ennek javítására kihasználhatjuk, hogy az additív függvények számítása könnyen megvalósítható az úgynevezett *futószűrőszerű* megoldással. Ilyenkor a már meglévő információkat felhasználva, a soron következő ablakban elhagyott képelemre számolt korrelációs értékeket kivonjuk és az újakét hozzáadjuk a kiinduló ablak korrelációs értékéből. Ezzel jelentősen csökkenthető a számítási-gény.

Problémák

A fizikai világ leképezésekor sok információt veszítünk, amelyeknek tökéletes pótlása lehetetlen. Emellett a sztereó látás a képek közötti különbségekre épül, és az illesztés feladata a különbségek feldolgozása. Akármilyen feltételeket is teszünk, bizonyos problémák valamilyen szinten megmaradnak, mint például az öntakarás, a torzulás, a fotometrikus különbségek és az alacsony textúráltságú területek kezelése. A legjobb képeken is előfordulhatnak zajból fakadó hibás találatok. A hibák mennyiségétől függően érdemes:

- A képekre illesztés előtt *adaptív hisztogram kiegyenlítést* alkalmazni. Ezzel a fotometrikus egyenletlenségek csökkenthetők.
- Gyakori öntakarások esetén *többablakos módszert* (lásd később) alkalmazni.
- Az illesztés után zajszerű hibákat szűréssel csökkenteni. A határok megtartása végett érdemes *mediánszűrőt* alkalmazni.

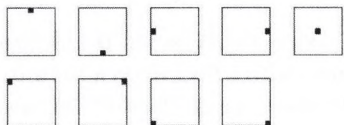
Más módszer a hibás adatok kezelésére a bizonytalan találatok elhagyása majd pótlása *interpolációval*. Hibás pontpárok lehetnek:

- Alacsony korrelációs értékű pontpárok.
- A keresési tartomány határán talált pontok, mivel ezek nem lokális extrémumok.
- Nem jobb-bal konzisztens pontpárok.

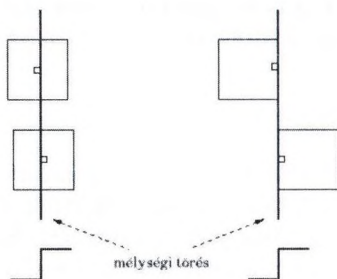
A hiányzó pontok pótlására szűrés is alkalmazható.

Többablakos módszer (SMW algoritmus, [1])

Ez a módszer megpróbál reális értéket rendelni az eltűnő pontokhoz is, figyelembe véve az objektum-határokat. A fő gondolat, hogy egy folytonos felület határán lévő pontot jobban lehet olyan ablakokkal illeszteni, amelyek teljesen azon a felületen vannak. Tehát az elmozdulás megállapításához többféle ablakot vizsgálunk (lásd 3. ábra), és azt az ablakot választjuk amelyik legjobban korrelál, mivel az nagy eséllyel azonos felületen van a ponttal. (Lásd 4. ábra. Minden mélységi törés csak növeli a hibát a korrelációs ablakon belül.)



3. ábra: Ablakok különböző középponttal



4. ábra: Mélységi törés többablakos módszer esetén. Bal oldalon a hagyományos korrelációs ablakok, jobb oldalon a többablakos módszer korrelációs ablakjai.

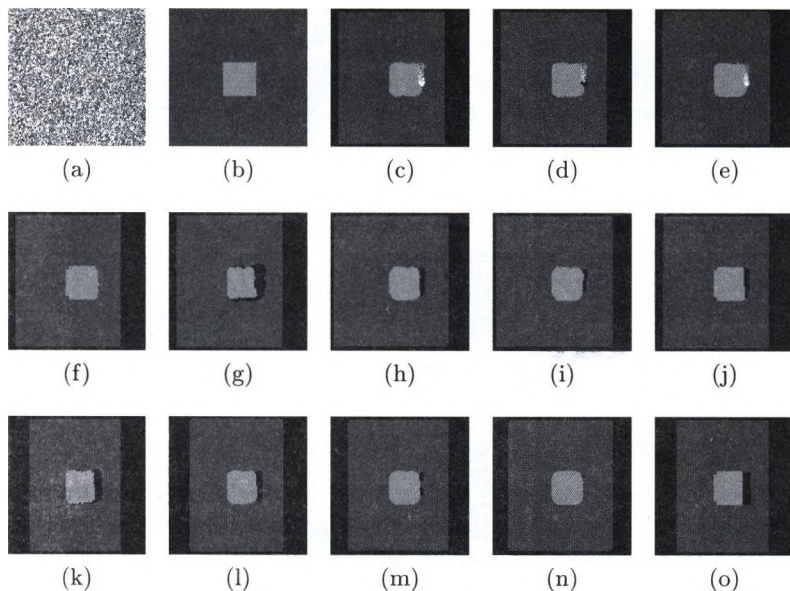
A megvalósításhoz nem kell többször elvégeznünk a keresést, hiszen elég már kiszámolt szomszédos korrelációs értékek közül választanunk legjobbat. Megfontolandó viszont új korrelációs érték megadása a középsőtől eltérő ablakokhoz, jelezve hogy a természetestől eltérő módon kaptuk a pontpárt. A módszer hatására jól megmaradnak az objektumhatárok, viszont keskeny tárgyakat teljesen el is veszíthetünk. Fontos a jó ablakméret választás.

Eredmények, tapasztalatok

Sajnos minden képre általánosan jó megoldást adni szinte lehetetlen. A különböző szinterekről készült képeket nemcsak a kamerák távolsága, hanem a felmerülő mélység, a fényviszonyok és a tárgyak textúráltsága is jellemzi. Ezekhez kell megválasztani az illesztés szépszámu paramétereit. Most mutatunk néhány eredményt mesterségesen generált véletlen pont sztereogrammon [3], ideális és nehezen kezelhető sztereó képekre alkalmazott módszerről.

A *véletlen pont sztereogram* első képe teljesen véletlen pontokból áll, melynek bizonyos részeit különböző eltolásokkal eltoljuk, és így kapjuk a második képet.

Először megmutatjuk a véletlen képet, majd a valós eltolást, végül a különböző módszerekkel kapott elmozdulási képeket. Az utóbb említett ábrákon minden szín egy elmozdulásnak felel meg (a feketének a hibás pontok), ami bár nem azonos a mélységgel, de érzékelteti azt.



5. ábra: Véletlen pont sztereoagramm illesztési eredményei a különböző módszerekkel. Véletlen kép (a); Valós elmozdulási kép (b); Illesztés eredménye 7×7 -es korrelációs ablakkal (c); Lokális minimumok (d); Az eredmény mediánszűrővel szűrve (e); Az eredmény többablakos módszerrel (f); Jól korreláló pontok (g); Jól korreláló pontok 7×7 -es mediánszűrővel, 2 iteráció (h); Jól korreláló pontok 7×7 mediánszűrővel, 3 iteráció (i); Többablakos módszer jól korreláló pontjai (j); Jobb-bal konzisztens pontok (k); Konzisztens pontok 5×5 -ös mediánszűrővel, 1 iteráció (l); Konzisztens pontok 7×7 -es mediánszűrővel, 2 iteráció (m); Konzisztens pontok 7×7 -es mediánszűrővel, 3 iteráció (n); Többablakos módszer konzisztens pontjai (o).

Irodalom

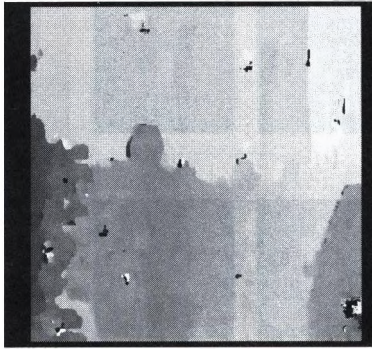
1. A. Fusiello, V. Roberto, E. Trucco. Symmetric Stereo with Multiple Windowing. *International Journal of Pattern Recognition and Artificial Intelligence*, Vol.14, No 8. 2000.
2. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2001)
3. B. Julesz. Binocular depth perception of computer generated patterns. *Bell Systems Technical Journal*, 39, 1990.



Baloldali sztereó kép



Jobboldali sztereó kép

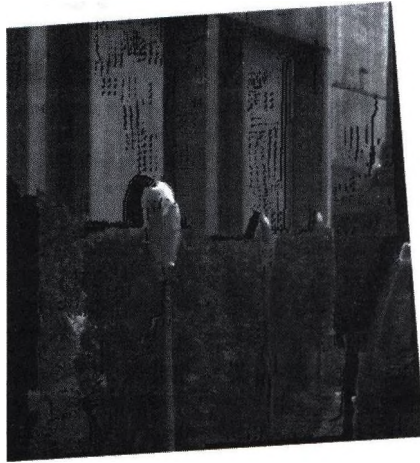


Illesztés eredménye, 15x15-ös ablak

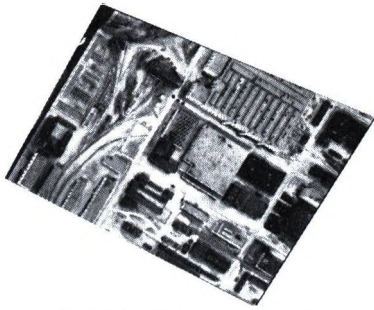


Többablakos módszer

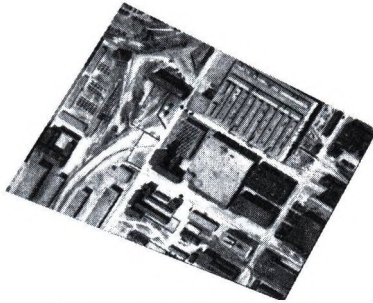
6. ábra: A módszerek alkalmazása



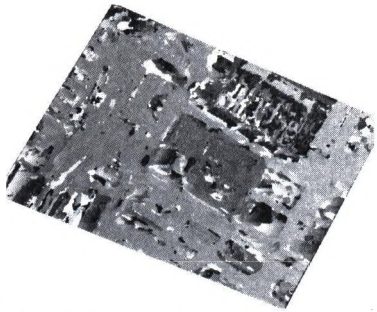
7. ábra: 3D modell megjelenítve



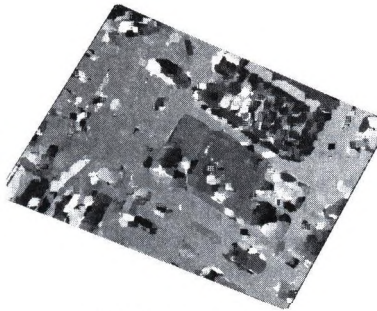
Rektifikált baloldali kép



Rektifikált jobboldali kép

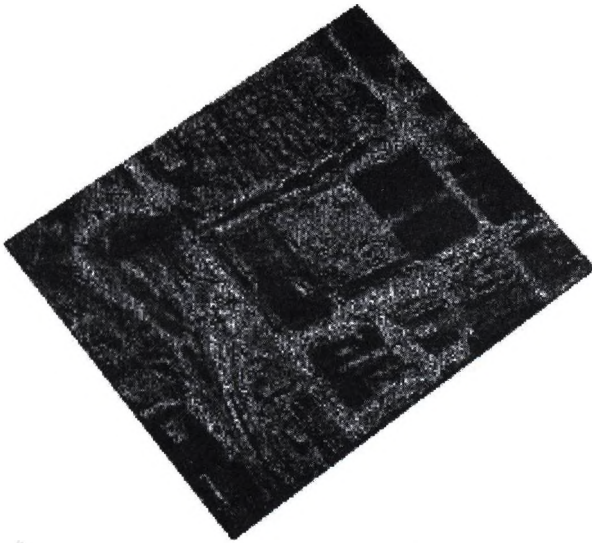


Illesztés eredménye, 7x7-es ablak



Többablakos módszer

8. ábra: A módszerek alkalmazása



9. ábra: 3D modell megjelenítve

4. Schaffalitzky, F., Zisserman, A.: Viewpoint Invariant Texture Matching and Wide Baseline Stereo. (2001)
5. M. Sonka, V. Hlavac, R. Boyle. *Image Processing, Analysis, and Machine Vision. Second Edition.* 448-486, 1998.
6. E. Trucco, A. Verri. *Introductory Techniques for 3-D Computer Vision.* 150-175, 1998.
7. T. Tuytelaars, L. Van Gool: Wide Baseline Stereo Matching based on local, affinely invariant regions. *British Machine Vision Conference*, pp. 412-422, 2000.
8. T. Tuytelaars, L. Van Gool: Content-Based Image Retrieval Based on Local Affinely Invariant Regions. *International Conference on Visual Information Systems*, pp. 493-500, 1999.
9. Z. Zhang, R. Deriche, O. Faugeras, Q. T. Luong. A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry. INRIA No2273 1994.

Calculating 3D Sound-Field using 3D Image Synthesis and Image Processing

Barnabás Aszódi¹ and Szabolcs Czuczor²

¹ Budapest University of Technology and Economics, Hungary Department of Control Engineering and Information Technology
ab011@ural2.hszk.bme.hu

² Budapest University of Technology and Economics, Hungary Department of Control Engineering and Information Technology
cs007@ural2.hszk.bme.hu

Abstract. This paper proposes several algorithms to compute the 3D sound-field in complex environments. The basic idea is to explore the similarities of light and sound propagation and take advantage of efficient rendering algorithms and hardware to compute the sound field. First we consider only the direct effect of sound sources. Then single reflections of the sound are also computed using the analogy of local illumination models. The main problem in establishing such an analogy is that the effects of different sources emitting different sound waves cannot be added. Instead we have to keep the information from which source this sound was emitted.

1 Introduction

At the dawn of the history of movies, the sound was mono then became stereo. Nowadays the sound is surround, which is either recorded or computed and finally restored by directed speakers. Meanwhile, images rendered by computer graphics algorithms have become more attractive and more efficient algorithms and image synthesis hardware have come to existence [1]. The main objective of this paper is to propose methods for realistic sound computation using computer graphics and image processing techniques. This approach has two merits. On the one hand, efficient image synthesis algorithms and hardware can be utilised for sound synthesis. On the other hand, the correct correspondence of the animation and the sound can be obtained automatically. The sound has similar properties as the light. The sound is also a wave that originates at sources, travels in space and scatters on surfaces as light does. If the sound waves are simulated by light, we can visualise the sound field. So every sound-source and sound-reflecting surface is interpreted as a light-source and light-reflecting surface, respectively. The intensity of the sound wave is proportional to the intensity of the light ray. Generating images from the location of the observer, we can describe the sound field of his environment. Taking into account the sound direction and spread, and computing only single reflections, these images allow sufficiently realistic sound generation.

In order to compute the sound field, the light equivalent of the sound is determined for every pixel or pixel-group in the image. The volume of the sound wave is proportional to the intensity of the pixels. The sound-samples which are coming from different pixel-directions are weighted according to the direction of the speakers. We have examined two methods for the waveform generation. The first method uses rendered images and obtains the required information only from these images. The advantage of this approach is that we can use off-the-shelf image synthesis algorithms and should use only simple image processing to find the sound field. Consequently, this method is fast and easy to implement. The second proposed method, on the other hand, uses the 3D definition of the scene and implements the complete light and sound rendering pipeline simultaneously. The second method is also able to produce Doppler effect and is more realistic. However, the price of the higher realism is the increased computational cost.

2 Importance functions of the sound channels

The sound wave is equivalent to the time function of voltage amplitude. The waveform is stored in a WAV file in digitized PCM (Pulse Code Modulation) format [15]. The loudness is depending on the power of signal (P), consequently it is directly proportional to the square of voltage amplitude (U^2):

$$I \propto P = \frac{U^2}{Z} \quad (1)$$

where U is the voltage amplitude and Z is the impedance of the system. Directing a sound wave in a stereo system into a specific direction, the signal amplitude is weighted according to the direction of the speakers. The multipliers of direction weighting are called importance-functions. The latent range of importance-functions is between zero and one. If there are more than two channels, i.e. we have at least a quadraphonic system (see figure 1), then we can represent the whole sound field that surrounds the listener. In this case, we have to define the importance functions of the other channels. The importance functions must be interactively controllable because the position of the speakers can be altered at any time.

3 Simulating the sound with light

In this section we examine, which visual rules shall apply to get the information of the sound from the rendered images. The primary questions are which and how many colours are used, and how we can mix them in the bitmaps to identify the sound sources unambiguously. We consider two rendering models. In the first model, no light (i.e. sound) reflection is allowed and the anti-aliasing is turned off. This means that we do not take sound reflection into account. In the second model this limitation is lifted.

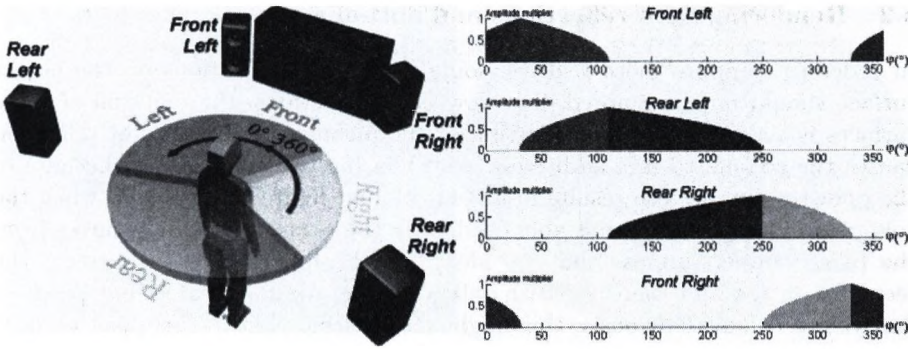


Fig. 1. Quadraphonic sound system and its importance functions.

3.1 Rendering with no reflection and without anti-aliasing

Suppose that we use different colours to identify the sound sources, mute objects and sound absorbers. If no light reflection is allowed and the anti-aliasing is turned off, then the image can contain only the own colour of the objects, thus they can be trivially recognised in different pixels (see figure 2).

The problem of these settings is that only the sound source objects can be recognised and thus heard. Since other objects do not reflect the sound, the result is not realistic. Moreover, the minimal size of the small or far distant objects are limited by the pixel size. Objects having smaller projected size may appear and disappear in an unpredictable way.

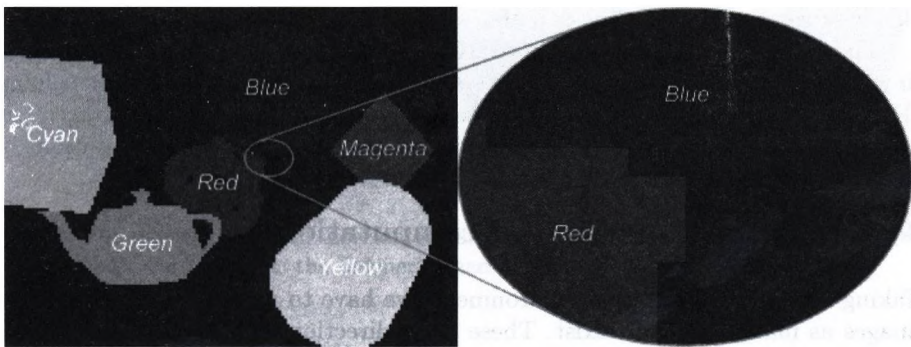


Fig. 2. Using colours to identify sources. If no reflection and anti-aliasing are allowed, the colours distinguish the sources from each other.

3.2 Rendering with reflections and anti-aliasing

In order to compute more realistic sound fields, the reflections on the object surface should not be ignored. To allow light reflections, the material of these surfaces is set to reflect some portion of the incoming energy. Light reflection causes the colours to mix additively. Anti-aliasing, which is needed because of the poor resolution, also results in mixed colours (figure 3). However, when the colours are mixed, we are not able to identify the respective sound sources from the pixel values. Suppose that we identified the first source with green, the second with red and the third with yellow colour. Assume that in one pixel we find yellow colour. Obviously, this might mean either that in this pixel we hear both the first and second sources or that only the third source can be heard here.

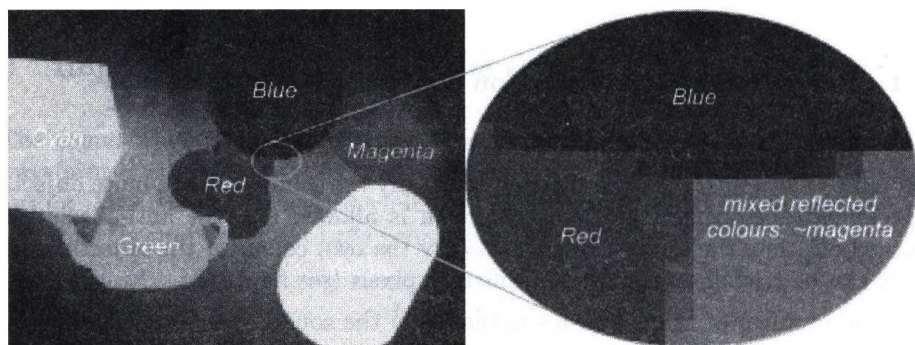


Fig. 3. Using colours of the objects when reflection and anti-aliasing are enabled.

There are two ways to separate the sources. The first considers only one sound source at the same time and renders as many images as sources exist. In this case we work in grey-scale mode.

The other possibility is to turn to RGB mode and follow three sound sources in parallel. This time we identify the sound sources with the three base colours (Red, Green, Blue) (see figure 4). These colours can only be used in true colour mode. We prefer the second method, because this requires less computation.

4 Image based sound field computation

Taking a photo of our entire environment, we have to use as many perspective images as main directions exist. These main directions are usually classified as front, back, left, right, top and bottom. These perspective images should be mapped over a sphere surface, viewed from the centre of the sphere. On the other hand, we use six cameras looking at the main directions, their aspect ratio is 1:1 and their “angle/field of view” (FOV) is 90 degrees. The six views (the six rendered images) are on the six sides of a cube that is in the same position

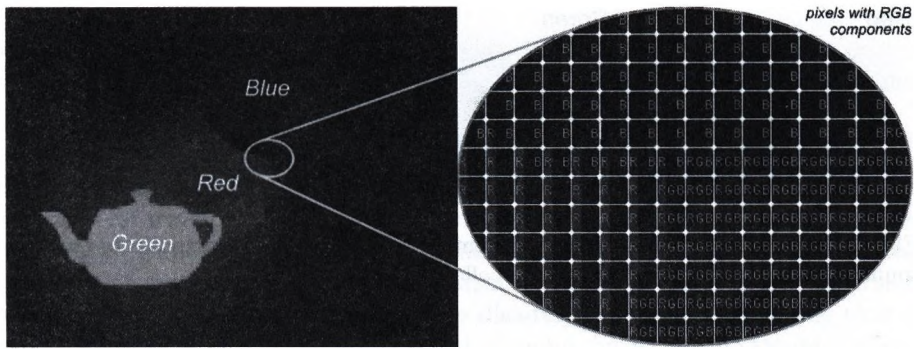


Fig. 4. Using three colours (R, G, B), reflection and anti-aliasing. The effects of only three sound sources can be separated.

as the sphere. The surround speakers are usually in one plane (in the level of the listener's head), so we can distinguish only the "left-right" and the "front-back" sound directions. Using this sound system, we can feel the "top-bottom" direction only by special distortion of the sound. Therefore we use only the four horizontal images of the "view cube" and the top and bottom images are ignored (figure 5). Of course, this is a big loss, because we have thrown away the third axis of the sound field description. However, we can correct this error by increasing the horizontal view of the cameras. But in this case the aspect ratio of the images will be changed.

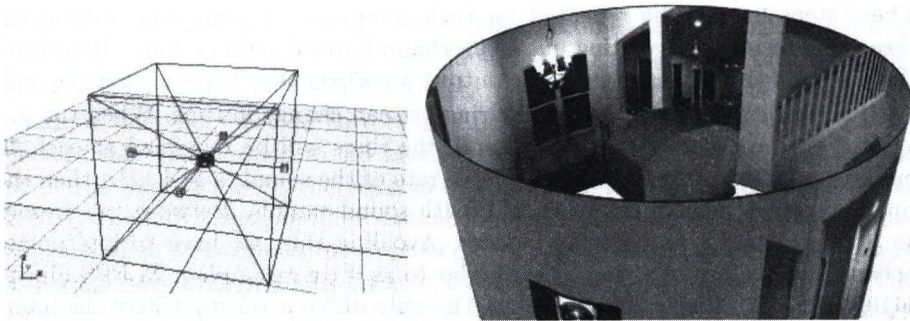


Fig. 5. On the left: the four horizontal cameras in the listener's position make the "view cube". On the right: the listener's head and the view of his environment.

After separating the RGB components the columns are collapsed, to obtain three ring-like lists. (See figure 6) This corresponds to the reality, because the speakers are placed along a big ring on the same level.

The length of the rings is equal to the sum of the columns of the images. The elements of rings contain the sum of pixel intensities in an image column.

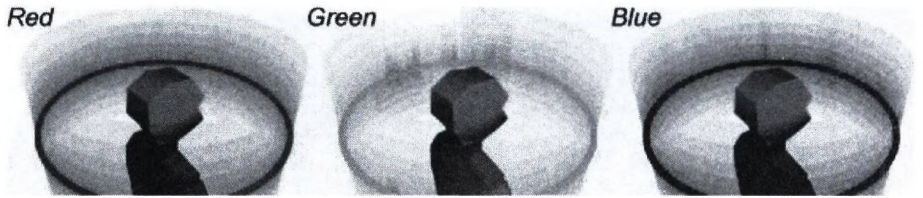


Fig. 6. View of the listener's environment separated to RGB components (to three sound sources) where the columns are collapsed.

For example, if the resolution of the four horizontal images is 90×90 , the size of each ring is $4 \times 90 = 360$. Moreover, each element contains the sum of the intensities of 90 pixels. Thus if 360 elements exist, then we can represent the sound intensity of a sound source at each degree in a one-dimension list. After this we have to normalise the list to $[0..1]$ and weight for the channels of the speakers, from which the output waveforms can be generated. The output is predefined: 44.1 kHz, 16-bit, the number of channels is changeable.

Let us first assume that the objects are still. The output file is built sample-by-sample, and is written in packages of the actual signal amplitudes of each channel. Each sample contains the weighted ring-value of the (maximum three) sound sources. The weighting (importance-function) depends on the position and the number of the speakers (figure 1). Let us consider the generation of a single sample. We have a ring-list assigned to a sound source. Each element of the list is multiplied by the channel weight depending on the direction of the element, and the actual signal amplitude. These products are added for each speaker. These steps have to be repeated for each sound source (ring-list). Adding up these ring-results and writing out to its channels result in a package. Repeating these steps 44100 times, we get the output waveform that takes a single second.

If the objects can also move, the ring-list can change in time. When the image refresh rate is 25 frame per second, the rings will be redefined at each 40 milliseconds. It means that if the sample rate of the sound is 44.1 kHz, then the rings are given new values at each 1764th sound sample. Between two frames an abrupt sound transition may occur. Avoiding this, we have to interpolate between two values of a ring. It is similar to as if we resample a 25 FPS film to 44100 FPS with linear interpolation. The only difference is that here the interpolation uses not the whole image, but only the rings containing the collapsed columns of the image.

The basic advantage of the image based sound generation is that the size of objects in a perspective view is inversely proportional to the square of its distance from the listener (equation 2). The sound intensity behaves exactly in the same way. Thus, scaling the loudness according to the image areas makes sense, because bigger viewable object surface means more intensive sound:

$$I = \frac{P}{4\pi \cdot r^2} = \frac{U^2}{4\pi \cdot Z \cdot r^2} \quad (2)$$

where I is the intensity of an omnidirectional sound source of P homogeneous sound power at distance r , U is the voltage amplitude and Z is the impedance of the system.

The other advantage of the method is that no special image-synthesis algorithm is needed, because the used images can be rendered any existing rendering software (for example, Kinetix - 3D Studio MAX [4] [5] [6] [11]). Taking advantage of the similarity of sound and light spreading, this method handles the problems of area sound sources and partial occlusions. Using light dispersion we can model the sound dispersion. The disadvantage of the method is that it cannot handle the effects caused by speed of sound, including the Doppler effect, echo, phase shifting, etc. To simulate these effects, we also need the distance and the motion information of objects and the phase of sounds at any time, and not only at each 1764th sample.

In order to compute the distances, we can, for example, use one of the base colours (e.g. Blue). We place an point light source in the listener's position with distance depending intensity to get the distance information. The nearer objects are illuminated intensively and the further ones are dark. Thus the blue pixel intensities of the images are proportional to the distance. However, the object surfaces are usually shaded according to the illumination and shading models of computer graphics [1] [2], thus the visible colours will not be identical to the colour of the light source. Moreover, the distance information can be handled only in a single frame, but not between two frames. We do not have information about what happens between two frames. If an object is moving, its loudness is interpolated from the pixel information of the actual and the previous (or next) frames. For example, when a red object is moving from left to right, the red object on the left of the image becomes black in the next frame, because it moves to the right (volume fade out on the left). On the right side, the object turns gradually red from black (volume fade in on the right). This technique does not provide correct distance information, but it would rather generate something which would mean that the red object moves away on the left and it approaches on the right in one frame time. The loudness simulates properly the movement, but the distance information is wrong. This can cause problems when we calculate the Doppler effect. The object on the left moves away with deep voice and it comes back on the right with a high-pitched tone. The problem of this method can be solved by some kind of shape recognition, which identifies the moving objects. The other disadvantage of this method is only one kind of distance (object-listener or object-light source) can be represented with the selected colour.

If we want to use reflection(s), we have to consider all object distance types, including object-listener, object-object and object-light source. Thus we need a database that also stores the shape, extension, position, orientation and movement information. Such a complete database can be visualised by a special OpenGL program that renders the reference images. On these images we can identify the objects. Suppose that this time we do not use the collapsing operation and work directly with the 2D images. Finding the centre of mass of the

projections of the identified objects, we can interpolate the centre and simulate the movement of the object. This is a working solution, but this is not free from problems either. One of them is how to calculate the centre of mass of the objects that surround us. The other one is how to handle the complicated objects having points with very different distances from the viewer.

5 Sound field computation using the motion vectors of each pixel

In this method we still work with 2D images and assume that the images are rendered in real-time. The method consists of the following steps. First it renders the images whose pixels contain the sound intensity information. The motion is analysed by following the change of the pixel in which an object point is projected. Such a way, we can talk about the movement of pixels. Then the method assigns a record to each moving pixel, which contains the intensity of the pixel and two motion vectors which represent the 2D movement of the pixel from the previous and to the next frame, respectively. The record also contains the direct or indirect (reflected) distance of the sound source from the listener through the pixel. (See figure 7)

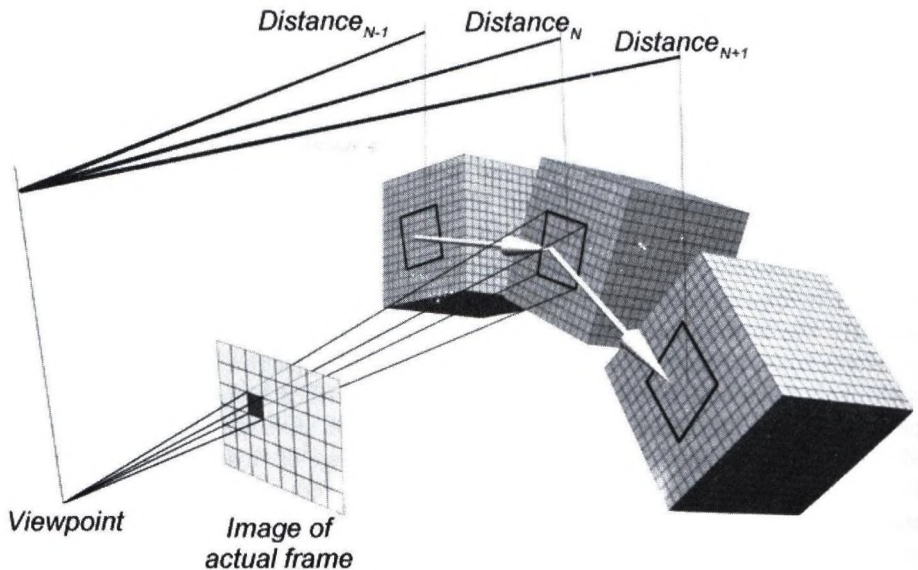


Fig. 7. Calculating the record of a pixel. The cube in the middle is the sound source in the actual (N) frame, the left one is the same sound source in the previous ($N-1$) frame and the right one is the same sound source in the next ($N+1$) frame. The distances are direct distances and the two big white arrows are the motion vectors of the examined pixel.

The method calculates the record fields not only in the actual, but also in the previous and the next frames. The intensity of a moving pixel is set to zero in the previous and in the next frames. It is required, because in the neighbouring frames there is no guarantee that the examined pixel is visible or does not coincide with an another pixel. With this we make a fade-in and a fade-out effect. Summarising, each pixel of every image has the following record:

Table 1. Record of a pixel

	frame $N - 1$ (prev.)	frame N (actual)	frame $N + 1$ (next)
Intensity	0	Given	0
Distance	Calculated	Calculated	Calculated
Motion vector	Calculated	X	Calculated

The pixel intensities of the actual images are given. The distances and the motion vectors are calculated only by a 3D transformation. When we want to know the position of a pixel in the neighbouring frames (and the distance and the motion vectors from these positions), we have to determine the pixel's owner object and the position of the pixel in the coordinate system of its owner. This operation can be accomplished by inverting the viewing pipeline. The position and the orientation of the owner object determine the coordinates of the pixel, which move with its owner in the neighbour frames. From these coordinates we can calculate the remained record fields.

5.1 Generating the output sound waveform

The output waveform is generated like in the previous section, but here we also use a frame-time buffer. It means that each 40 ms waveform is generated in two parts. Let us examine how the algorithm works. After rendering the first image (without previous image) the algorithm calculates the record fields of each pixel by the data of next frame. The next step is calculating the signal amplitude, phase shift and direction by an interpolation of record fields at each sound samples (one sound sample represents 1/44100 second). Superposing the next frame's sound material that belongs to the current frame with the actual 40 ms sound material, we get the first 40 ms of the output waveform. (See figure 8)

6 Conclusions

This paper proposed several algorithms to compute the 3D sound field using rendering and image processing algorithms. The most sophisticated method uses the motion vectors of the pixels. The main advantage of this method over the method based just on the 2D images is that it also takes the movement of each

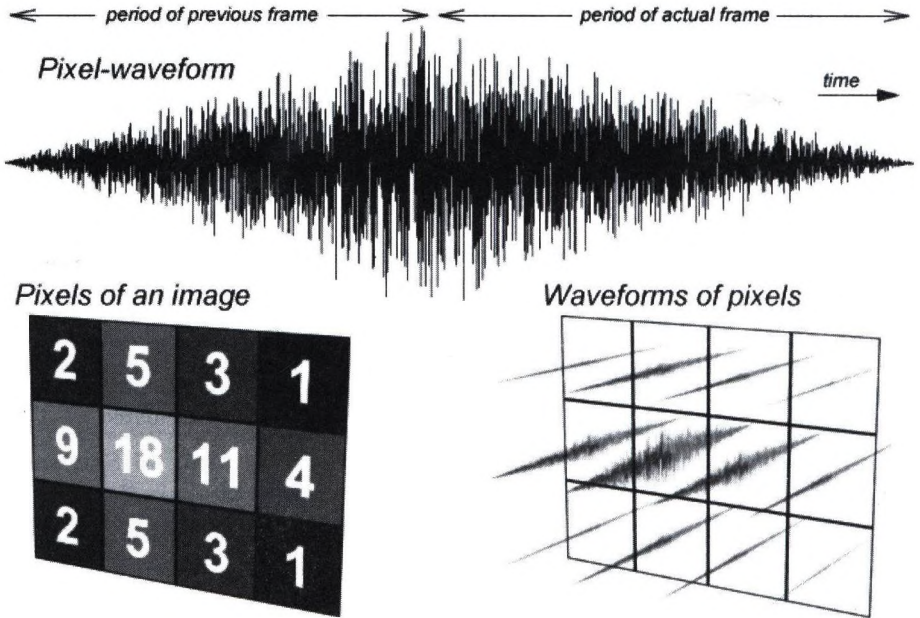


Fig. 8. Calculating the waveform of a pixel.

pixel into account, thus it can handle the effects caused by the speed of sound (phase shift, Doppler effect, echo). Unfortunately, neither this method is really perfect, because in an unfortunate case interpolating, fading and superposing the sound of pixels can cause interference, which can distort the results. The other problem is the propagation of quantization error caused by too low amplitude (volume or loudness) of the added sound samples of the pixels. Using enough quantization levels we can avoid these errors and reach lower signal to noise ratio (SNR). The bigger quantization (bit depth) needs more operative memory. The database required by the motion vector method needs smaller storing space than the image lists used by image based approach. This database is specific, so the algorithm can handle limited input data formats, which cannot be interpreted some 3D modeller applications. Its other advantage is that its steps can be supported by the current graphics hardware. The 3D transformation of the pixels and the interpolation of the sound samples by frame and by pixel may slow down the computation. Choosing the right image resolution, we can find a reasonable compromise between production quality and the system requirements.

7 Acknowledgements

This work has been supported by the National Scientific Research Fund (OTKA ref. No.: T029135).

References

1. Dr. Szirmay-Kalos László (editor): Theory of three-dimensional computer graphics, Akadémia Kiadó, 1995.
2. Dr. Szirmay-Kalos László: Számítógépes grafika, ComputerBooks, 1999.
3. Füzi János: 3D grafika és animáció IBM PC-n, ComputerBooks, 1995.
4. Aurum-Boca: 3D Studio MAX, Aurum DTP Stúdió Kiadó, Kereskedelmi és Szolgáltató Kft, 1997
5. Ted Boardman, Jeremy Hubbell: Inside 3D Studio MAX - volume II: Modeling and Materials, New Riders, 1998
6. George Maestri, et al: Inside 3D Studio MAX - volume III: Animation, New Riders, 1998
7. Juhász Mihály, Kiss Zoltán, Kuzmina Jekatyerina, Sölétormos Gábor, Dr. Tamás Péter, Tóth Bertalan: Delphi - út a jövőbe, ComputerBooks, 1998
8. László József: Hangkártya programozása Pascal és Assembly nyelven, Computer-Book, 1995
9. László József: A VGA-kártya programozása Pascal és Assembly nyelven, ComputerBook, 1996
10. Dr. Budó Ágoston: Kísérleti fizika I. (mechanika, hangtan, hőtan), Nemzeti Tankönyvkiadó, 1970
11. <http://www.ktx.com> (Kinetix Home Page)
12. <http://www.goldwave.com> (GoldWave Home Page)
13. <http://www.dolby.com> (Dolby Laboratories, Inc. Home Page)
14. <http://www.wareing.dircon.co.uk/3daudio.htm> (Audio and Three Dimensional Sound Links)
15. <http://www.borg.com/~jglatt/tech/wave.htm> (WAVE File Format)
16. <http://www.soundblaster.com> (Sound Blaster Home Page)
17. <http://www.creative.com> (Creative Technology Ltd. Home Page)
18. <http://www.thx.com> (Lucasfilm's THX Home Page)

A Sequential 3D Curve-Thinning Algorithm

Kálmán Palágyi

Department of Applied Informatics, University of Szeged, Hungary
palagyik@inf.u-szeged.hu

Abstract. Skeleton is a frequently applied shape feature to represent the general form of an object. Thinning is an iterative object reduction technique for producing a reasonable approximation to the skeleton in a topology preserving way. This paper describes a sequential 3D curve-thinning algorithm for extracting medial lines of objects in (26,6) pictures. The algorithm has been successfully applied to "tubular" medical images.

1 Introduction

The notion of *skeleton* was introduced by Blum [2] as a region-based shape descriptor which summarises the general form of objects/shapes. A very illustrative definition of the skeleton is given using the prairie-fire analogy: the object boundary is set on fire and the skeleton is formed by the loci where the fire fronts meet and quench each other. This definition can be naturally extended to arbitrary dimensions. In discrete spaces, the *thinning* process is a frequently used method for producing an approximation to the skeleton in a topology-preserving way [5]. It is based on digital simulation of the fire front propagation: border points of a binary object that satisfy certain topological and geometric constraints are deleted in iteration steps. The entire process is repeated until only the "skeleton" is left.

In this paper, a sequential 3D curve-thinning algorithm is described for extracting medial lines from elongated objects.

2 Basic Notions and Results

Let p be a point in the 3D digital space \mathbb{Z}^3 . Let us denote $N_j(p)$ (for $j = 6, 18, 26$) the set of points j -adjacent to point p (see Fig. 1). The sequence of distinct points $\langle x_0, x_1, \dots, x_n \rangle$ is a j -path of length $n \geq 0$ from point x_0 to point x_n in a non-empty set of points X if each point of the sequence is in X and x_i is j -adjacent to x_{i-1} for each $1 \leq i \leq n$. (Note that a single point is a j -path of length 0.) Two points are j -connected in the set X if there is a j -path in X between them. A set of points X is j -connected in the set of points $Y \supseteq X$ if any two points in X are j -connected in Y .

The 3D binary (m, n) digital picture \mathcal{P} is a quadruple $\mathcal{P} = (\mathbb{Z}^3, m, n, B)$ [5]. Each element of \mathbb{Z}^3 is called a *point* of \mathcal{P} . Each point in $B \subseteq \mathbb{Z}^3$ is called a

black point and value 1 is assigned to it. Each point in $\mathbb{Z}^3 \setminus B$ is called a *white point* and value 0 is assigned to it. Adjacency m belongs to the black points and adjacency n belongs to the white points. A *black component* (or *object*) is a maximal m -connected set of points in B . A *white component* is a maximal n -connected set of points in $B \subseteq \mathbb{Z}^3$. We are dealing with (26,6) pictures. It is assumed that any picture contains finitely many black points.

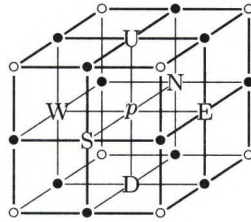


Fig. 1. The frequently used adjacencies in \mathbb{Z}^3 . The set $N_6(p)$ contains the central point p and the 6 points marked U, N, E, S, W, and D. The set $N_{18}(p)$ contains the set $N_6(p)$ and the 12 points marked “•”. The set $N_{26}(p)$ contains the set $N_{18}(p)$ and the 8 points marked “o”

A black point in a (26, 6) picture is called *border point* if it is 6-adjacent to at least one white point. A border point p is called *U-border point* if the point marked by “U” in Fig. 1 is white. We can define N-, E-, S-, W-, and D-border points in the same way. A black point is called *end-point* if it has exactly one black 26-neighbor (i.e., the set $N_{26}(p) \cap (B \setminus \{p\})$ is singleton).

A black point is called *simple point* if its deletion does not alter the topology of the picture. The simplicity is a local property (for (26,6) pictures): it can be decided in view of the set $N_{26}(p)$ (i.e., the $3 \times 3 \times 3$ neighborhood of the investigated point p) [8]:

Theorem 1. *Black point p is simple in the picture $\mathcal{P} = (\mathbb{Z}^3, 26, 6, B)$ if and only if all of the following three conditions hold:*

1. *The set $(B \setminus \{p\}) \cap N_{26}(p)$ contains exactly one 26-component.*
2. *The set $(\mathbb{Z}^3 \setminus B) \cap N_6(p)$ is not empty.*
3. *Any two points in $(\mathbb{Z}^3 \setminus B) \cap N_6(p)$ are 6-connected in the set $(\mathbb{Z}^3 \setminus B) \cap N_{18}(p)$.*

3 Skeletonization Techniques

During the last two decades skeletonization (i.e., skeleton extraction from digital binary images) has become a challenging research field. There are two major requirements to be complied with [15]. The *geometrical* requirement means that the skeleton must be in the “middle” of the object and invariant under some

relevant geometrical transformations including translation, rotation, and uniform scaling. The second requirement is *topological* requiring that the produced skeleton must be topologically equivalent to the original object.

There are three major skeletonization methods for discrete objects: the distance transform method, the method based on Voronoi diagrams, and thinning.

The first method is to find local maxima or ridges in the distance map calculated from the border points of an object. The distance map is a (non-binary) picture in which each point has a value that approximates the distance to the nearest border point [3]. The result of the distance transformation depends on the selected distance and the ridge extraction is a rather difficult task. The distance map based method fulfils the geometrical requirement if a good approximation to the Euclidean distance is applied, but the topological correctness is not guaranteed.

The Voronoi diagram of a discrete set of points (called generating points) is the partition of the given space into cells so that each cell contains exactly one generating point and the locus of all points which are nearer to this generating point than to other generating points. It has been shown that the skeleton of an object which is described by a set of border points can be approximated by a subgraph of the Voronoi diagram of that generating points. Both the geometrical and the topological requirements can be fulfilled by the skeletonization based on Voronoi diagrams but it is regarded as an expensive process, especially for large and complex objects [9].

The thinning is a frequently used method for producing an approximation to the skeleton in a topology-preserving way [5]. Border points of a binary object that satisfy certain topological and geometric constraints are deleted in iteration steps. The entire process is repeated until only the "skeleton" is left. The topologically oriented thinning pays less attention to the metric properties of the object, therefore, the invariance under rotation (object orientation) and scale change is not guaranteed.

The skeleton represents some kinds of *local object symmetries* [15]. The skeleton of a 3D object can contain some surface patches (representing mirror symmetry and/or rotational symmetry) and some line segments (representing axial symmetry).

In case of "near tubular" 3D objects (e.g., airway, blood vessel, and gastrointestinal tract), it is generally desired to determine their central paths. It is rather difficult to extract "surface-free" skeletons by applying the distance transform based method or even the Voronoi-based one. Thinning has a major advantage over the other two methods since two kinds of approximations to the skeleton can be produced: *curve-thinning* produces the *medial lines* while *surface-thinning* produces the *medial surfaces* of objects. The results of the surface thinning algorithms are closer to the skeleton than the "skeletons" produced by curve thinning algorithms. Axial symmetry is emphasised by curve thinning and other kinds of symmetries are suppressed.

Note that some authors (for example Tsao and Fu [16]) proposed a 2-phase curve thinning process for extracting medial lines. After the first phase, the

medial surface is extracted from the original object. After the second phase that intermediate result is converted to medial lines. In our opinion, the direct extraction of the medial lines is much better than the 2-phase method [10].

Since the fire front propagation is by nature parallel, most of the existing thinning algorithms are parallel (i.e., all border points satisfying the deletion condition of the actual phase of the process are simultaneously deleted) [1, 7, 10, 16]. Despite of this fact, some sequential (or partly parallel) thinning algorithms have been proposed [4, 12-14] and there is a hybrid one (i.e., marking deletable points in parallel followed by a sequential re-checking phase) [6].

This paper presents an effective sequential 3D thinning algorithm is that directly extracts medial lines from elongated binary objects. It is an improved version of the one presented in [11].

4 The 3D Curve-Thinning Algorithm

Let $(\mathbb{Z}^3, 26, 6, B)$ be a 3D finite picture to be processed. Since set B is finite, it can be stored in a finite 3D binary array X (each voxel being not in X is looked on 0). The pseudocode of the sequential 3D thinning algorithm is given as follows:

procedure SEQUENTIAL_THINNING (X, Y)

1. $Y = X$;
2. *border_list* =< new empty list >;
3. **for** each point p in Y **do**
4. **if** IS_BORDER_POINT (Y, p) **then**
5. INSERT_LIST (*border_list* , p);
2. **repeat**
3. *modified* = 0;
4. *modified* = *modified* + SUBITERATION (*border_list*, Y, U);
5. *modified* = *modified* + SUBITERATION (*border_list*, Y, E);
6. *modified* = *modified* + SUBITERATION (*border_list*, Y, S);
7. *modified* = *modified* + SUBITERATION (*border_list*, Y, D);
8. *modified* = *modified* + SUBITERATION (*border_list*, Y, W);
9. *modified* = *modified* + SUBITERATION (*border_list*, Y, N);
10. **until** *modified* > 0

The two parameters of the procedure SEQUENTIAL_THINNING are the binary array X representing the picture to be thinned and the binary array Y storing the result. At first, all the border points are inserted in the list *border_list*. Since the minor part of voxels in a usual binary picture belongs to the objects, manipulating a list containing 1's is much faster than scanning an array. The kernel of the **repeat** cycle corresponds to one iteration step of the thinning process. Each iteration step is composed of six successive subiterations corresponding to the six kinds of border points. Some U-border points can be deleted in the first subiteration and certain W-border points are deleted in the

sixth one. In this way, the elongated objects are shrunk uniformly in each direction. Function SUBITERATION returns the number of deleted points. Variable *modified* is to accumulate the number of deleted points. The thinning process is completed when no points are deleted (i.e., no further changes occur).

Note that choosing another order of the deletion directions yields another algorithm. The proposed order (see Fig. 2) shows a kind of symmetry, therefore, the thinned objects are in their geometrically correct position (i.e., in the "middle" of the original objects).

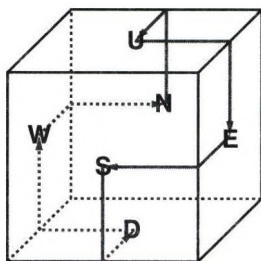


Fig. 2. The ordered list of the six deletion directions (corresponding to the six kinds of border points) is proposed $\langle U, E, S, D, W, N \rangle$

function SUBITERATION (*border_list*, *Y*, *direction*)

1. *modified* = 0;
2. *dir_list* = < new empty list >;
3. **for** each point *p* in *border_list* **do**
4. **if** IS_DIR_BORDER_POINT (*Y* , *direction*, *p*) **then**
5. **if** IS_SIMPLE_AND_NOT_ENDPOINT (*Np*) **then**
6. INSERT_LIST (*dir_list* , *p*);
7. **for** each point *p* in *dir_list* **do**
8. **if** IS_SIMPLE_AND_NOT_ENDPOINT (*Np*) **then**
9. SET_ZERO (*Y* , *p*);
10. *modified* = *modified* + 1;
11. DELETE_LIST (*border_list* , *p*);
12. **for** each object point *q* in the actual *Y* being 6-adjacent to *q* **do**
13. INSERT_LIST (*border_list* , *q*);
14. **return** *modified*;

The work of function SUBITERATION is composed of two phases. All the border points of the actual type (corresponding to the actual direction) that are simple but not end-points are inserted in the linked list *dir_list* in the first phase (see the first **for** cycle). This phase (i.e., marking points for deletion) is

followed by a sequential re-checking procedure (see the second **for** cycle): each point in the list is removed if it remains simple and not end-point in the actual (modified) image. The sequential deletion of simple points does not alter the topology of the picture, therefore, the topological correctness of our thinning algorithm is guaranteed.

Not only deletion is performed in the re-checking phase, but the *border_list* is updated, too. Note that it is to be assured that *border_list* stores only one copy of each border point in the actual picture.

function IS_SIMPLE_AND_NOT_ENDPOINT (Y, p)

1. $Np = \text{COLLECT_26_NEIGHBORS} (Y, p);$
2. **if** LUT[Np] **then**
3. **return** *true*;
4. **else**
5. **return** *false*;

Function IS_SIMPLE_AND_NOT_ENDPOINT uses an array of 2^{26} binary digits. The additional data structure LUT is a preset look-up-table (requiring 8 megabytes storage space in memory) that can be addressed by a 26-bit integer Np , where the i -th bit of Np corresponds to the neighbor marked " i " in Fig. 1 ($i = 0, \dots, 25$) of the investigated point p . The variable Np is set by function COLLECT_26_NEIGHBORS. The value LUT[Np] is 1 if the neighborhood assigned to Np belongs to a simple but not end-point (else it is 0). We generated the array LUT by the help of Theorem 1.

The proposed sequential algorithm differs from others in the following regards:

- it is "fully sequential" (algorithms in [4] and [6] are partly parallel),
- different order of the six deletion directions are proposed (algorithm in [6] uses the order $\langle U, D, N, S, W, E \rangle$),
- no distance information is used (algorithms in [4], [12], and [14] are distance driven),
- no multiple scans (i.e., traversal of the entire image array) in a (sub)iteration nor different phases of the thinning process are needed (algorithm in [13] proposed a thinning approach consisting of two phases namely primary-thinning and final-thinning and each iteration step is completed in three successive scans).

5 Results

Different shapes of objects have been tested by the proposed curve-thinning algorithm. Here we present some examples. Thinning for two symmetric and regular (synthetic) objects can be seen in Fig. 3.

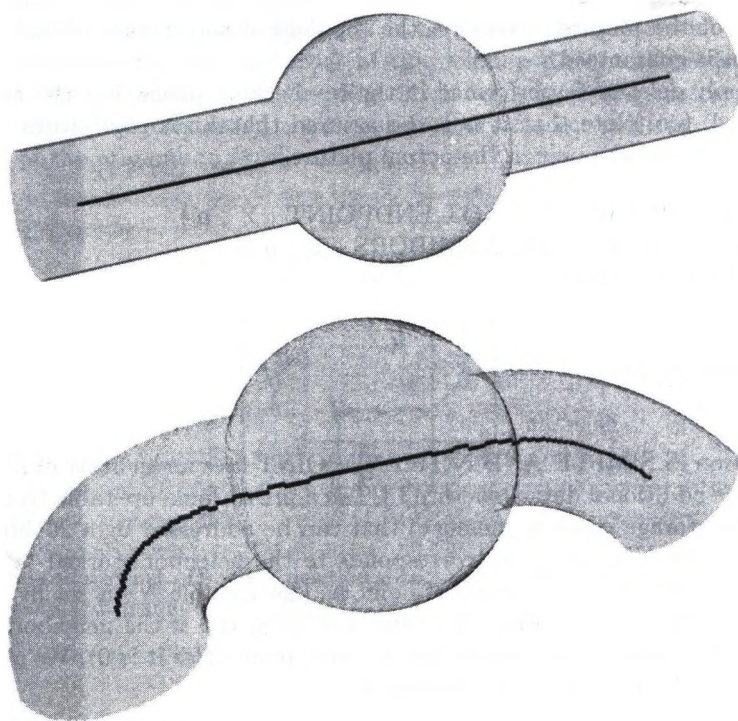


Fig. 3. Two synthetic objects and their central paths extracted by the proposed curve-thinning algorithm

Additional figures present the central paths of some kinds of segmented medical objects including infrarenal aorta (Fig. 4), upper respiratory tract (Fig. 5), cadaveric phantom (Fig. 6), and pulmonary phantom (Fig. 7).

Note that the proposed algorithm (using linked lists and LUT) is fairly fast: our implementation takes only 4.5 sec. for a $512 \times 512 \times 400$ image containing 125000 object points (on a Pentium III PC).

Acknowledgment

This work was supported by the CEEPUS A-34 Grant.

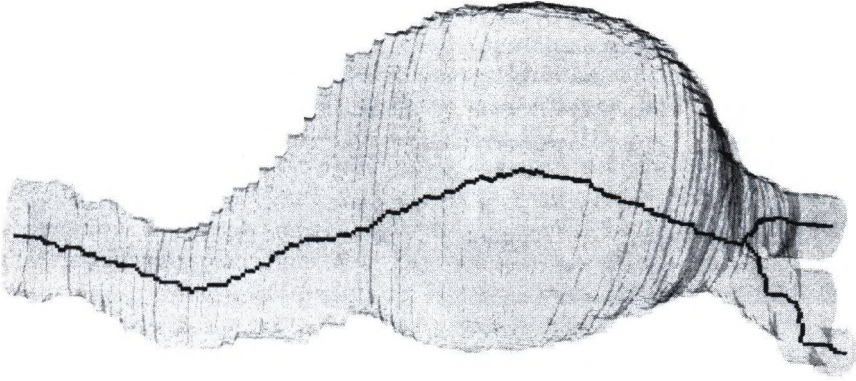


Fig. 4. Central path of a segmented part of an infrarenal aorta produced by the proposed curve-thinning algorithm. In order to get longer central paths, the endpoints were automatically identified before the thinning and those points were regarded as "anchors" during the thinning process (i.e., their deletion were prohibited). In that way, the topologically correct thinning algorithm was urged to connect them

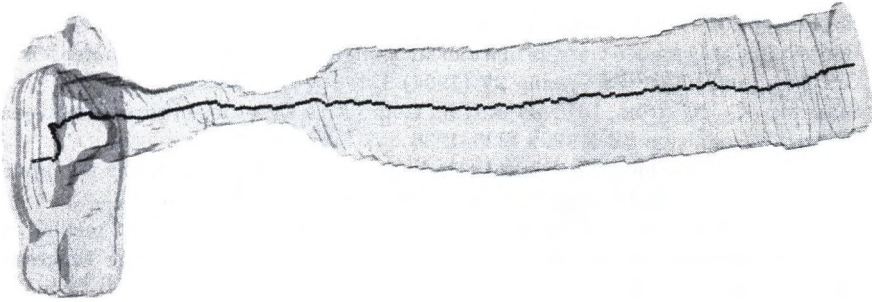


Fig. 5. Central path of a segmented upper respiratory tract produced by the proposed curve-thinning algorithm

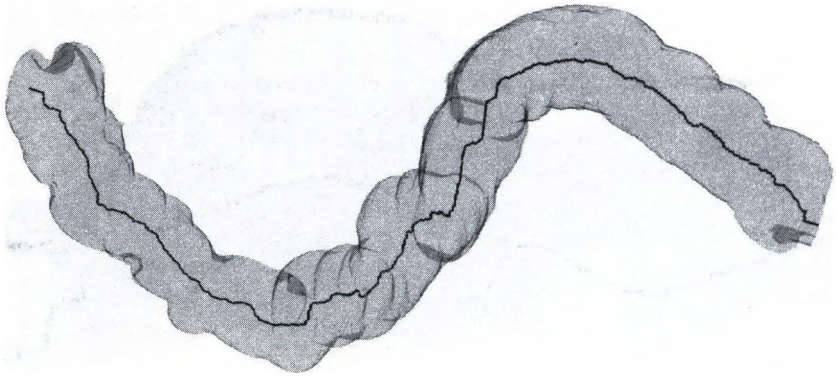


Fig. 6. Central path of a segmented cadaveric phantom produced by the proposed curve-thinning algorithm

References

1. Bertrand, G.: A parallel thinning algorithm for medial surfaces. *Pattern Recognition Letters* **16** (1995) 979–986
2. Blum, H.: A transformation for extracting new descriptors of shape. *Models for the Perception of Speech and Visual Form*, MIT Press, (1967) 362–380
3. Borgfors, G.: Distance transformations in arbitrary dimensions. *Computer Vision, Graphics, and Image Processing* **27** (1984) 321–345
4. Borgfors, G., Nyström, I., di Baja, G.S.: Computing skeletons in three dimensions. *Pattern Recognition* **32** (1999) 1225–1236
5. Kong, T.Y., Rosenfeld, A.: Digital topology: Introduction and survey. *Computer Vision, Graphics, and Image Processing* **48** (1989) 357–393
6. Lee, T., Kashyap, R.L., Chu, C.: Building skeleton models via 3-D medial surface/axis thinning algorithms. *CVGIP: Graphical Models and Image Processing* **56** (1994) 462–478
7. Ma, C.M., Sonka, M.: A fully parallel 3D thinning algorithm and its applications. *Computer Vision and Image Understanding* **64** (1996) 420–433
8. Malandain, G., Bertrand, G.: Fast characterization of 3D simple points. In: *Proc. 11th IEEE International Conference on Pattern Recognition* (1992) 232–235
9. Náf, M., Székely, G., Kikinis, R., Shenton, M.E., Kübler, G.: 3D Voronoi skeletons and their usage for the characterization and recognition of 3D organ shape. *Computer Vision, Graphics, and Image Processing* **66** (1997) 147–161
10. Palágyi, K., Kuba, A.: A parallel 3D 12-subiteration thinning algorithm. *Graphical Models and Image Processing* **61** (1999) 199–221
11. Palágyi, K., Sorantin, E., Balogh, E., Kuba, A., Halmai, Cs., Erdőhelyi, B., Hausegger, K.: A sequential 3D thinning algorithm and its medical applications. In: *Information Processing in medical Imaging, IPMI 2001, Lecture Notes in Computer Science*, Vol. 2082. Springer (2001) 409–415
12. Pudney, C.: Distance-ordered homotopic thinning: A skeletonization algorithm for 3D digital images. *Computer Vision and Image Understanding* **72** (1998) 404–413

13. Saha, P.K., Chaudhuri, B.B., Majumder, D.D.: A new shape-preserving parallel thinning algorithm for 3D digital images. *Pattern Recognition* **30** (1997) 1939–1955
14. Saito, T., Toriwaki, J.: A sequential thinning algorithm for three dimensional digital pictures using the Euclidean distance transformation. In: *Proc. 9th Scandinavian Conf. on Image Analysis, SCIA'95* (1995) 507–516
15. Székely, G.: Shape characterization by local symmetries. Habilitationsschrift, Institute for Communication Technology, Image Science Division, ETH Zürich, 1996
16. Tsao, Y.F., Fu, K.S.: A parallel thinning algorithm for 3-D pictures. *Computer Graphics and Image Processing* **17** (1981) 315–331

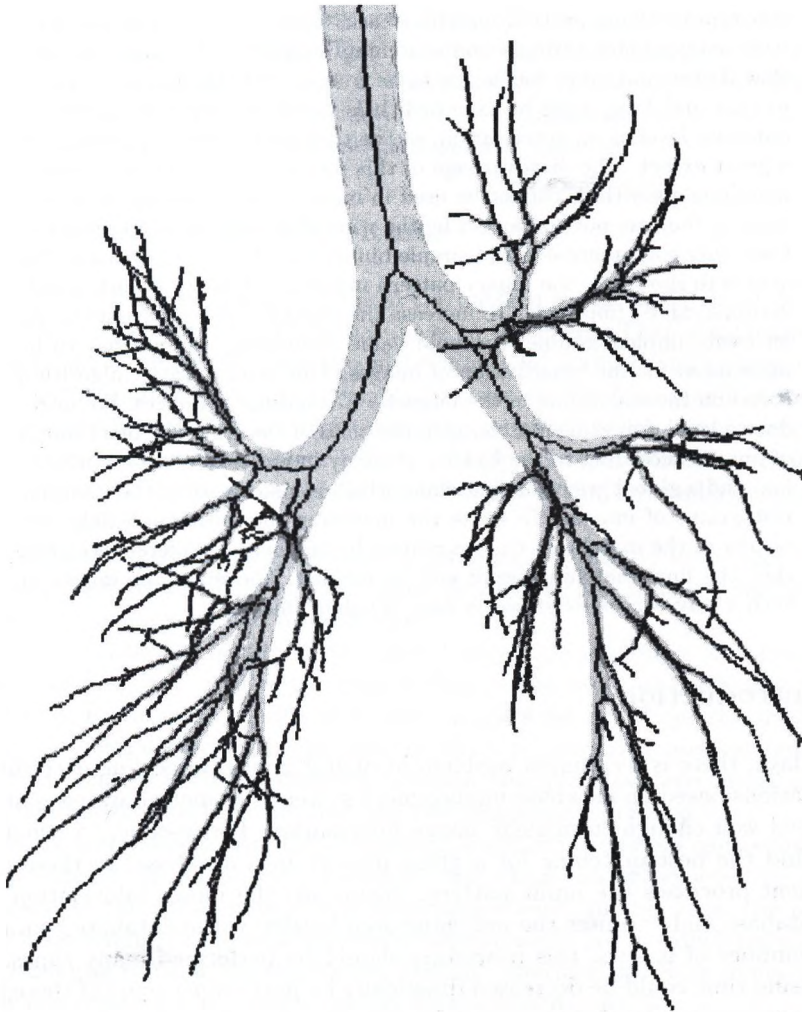


Fig. 7. Central path of a pulmonary phantom produced the proposed curve-thinning algorithm

Content-based Image Retrieval Algorithms at Database Server Level

Krisztián Veréb¹

Department of Information Technology, Institute of Mathematics and Informatics,
University of Debrecen
sparrow@math.klte.hu

Abstract. Doing pattern matching tasks in images stored in the database management systems and searching images in databases are very slow if each and every image has to be transported through the network so that matching could be executed. It is worth executing operations at database level as an acceleration, so the transport of images decrease to a great extent. The disadvantage of this method is that rather complex matching algorithms cannot be used in most database management systems as they are not supported by the wide spread systems. In most systems only comprehensive and simple binary matchings can be used. Our goal is to show how the binary pattern matching strategy of such simple systems can be improved. To increase the efficiency of binary matchings, an easily implementable threshold value computing method has to be introduced for the binarization of images. This is an iterative algorithm based on the smoothing of the intensity histograms of images. We introduce a local greyvalue matching on the basis of the greyscale pixel values of images according to the Eakins' strategy in pictorial database matching, and a global greyvalue matching which is also based on the intensity histograms of images. To make the matching the most applicable, the results of the matchings can be refined by the help of different weights, thus the final matching result can be defined. The matching improved such a way can be extended in case of colour images.

1 Introduction

Nowadays, there is a common problem in digital image processing according to applications based on database management systems. Namely, databases are not prepared well enough to manage image information. For instance, a usual task is to find the best matching for a given pattern in a database. In these cases, the client processes the input pattern, downloads the image information from the database and executes the matching step locally. If the database contains a large number of images, this procedure should be performed many times. The processing time could be decreased drastically by performing some of these steps at database server level. Unfortunately, there is an additional problem with implementing algorithms at database level. Database management systems do not support programming very much, since they only contain simple programming

tools and languages. Algorithms with universal solution purposes not at all, or just to a certain extent can be adopted. Hence, we cannot take advantage of most mathematical methods if we want to implement an algorithm at database server level.

Another important point is to implement algorithms in a way that they be economic and fast since beside running these algorithms, the database server have other server tasks to perform, like managing divided or multi user transactions, locking etc. The above mentioned concepts are taken into consideration.

We give a parametrisable matching method to execute the content-based image retrieval. The method has met Eakins' [2] multilevel matching strategy. It operates with the global and local greyvalues and makes a threshold value computing to improve the binary matchings. To compute the threshold value we give an iterative algorithm. This iterative algorithm smoothes the intensity histogram of an image, to make it easier to find values for thresholding. We show how this procedure can be implemented at database server level using the programming tools of the database management systems.

2 Concepts of the Algorithms

To minimize the data transfer, we post the pattern image to the database rather than the stored images from the database to the client. Then a matching algorithm could be done. We give two algorithms and a thresholding algorithm to improve the matching. We give weights to compute the final result as well. The first two method results are two measuring numbers which are not very large to be stored. The original algorithm is a binary pattern matching, that is the reason why the performance could be increased if the thresholded images are stored as well.

To threshold the images, we have to compute a threshold value based on the intensity histograms of the given images. Smoothing the intensity histogram of an image, binarization can be performed.

We compute the results of the global greyvalue matching, the local greyvalue matching and — after the binarization — the result of the binary pattern matching. From these results with some weights we could get the final result applying a simple formula. The final result is a measuring number between 0 and 100. This is a virtual distance between the query pattern image and the matched image. The 0 means that the two images are identical.

As we mentioned before the processing speed of the system can be improved by decreasing the amount of transferred data. This can be achieved by making the most of the routines that run at database level. In this case, the routines have to be simple, easily implementable because we do not usually have advanced programming tools at database level. Difficult routines based on many floating point operations are not supported by database systems.

In the next sections we show an example how a routine used in a digital image processing application can be implemented at database level. The aim of this routine is to perform binarization by smoothing the intensity histogram.

By smoothing, the peaks and valleys of the histogram can be calculated in a more reliable way. This binarization method can be inserted into the database management system before the binary pattern matching step. Then we detail the global and the local greyvalue matching as well.

3 Thresholding

Before presenting our smoothing method, at first we recall the concept of thresholding [3] [4] [7]. Let $f(x, y)$ be a K -level digital image over a digital set, with the size of $M \times N$. Our aim is to transform the image into two or more intensity levels. The simplest way is to use $\frac{k}{2}$ as the threshold value, by supposing that the distribution of the intensity values is uniform. Thus, the resulting binary image $g(x, y)$ is given by

$$g(x, y) = \begin{cases} 0, & \text{if } f(x, y) \leq \frac{k}{2} \\ 1, & \text{otherwise} \end{cases}$$

This calculation is rather trivial and for most types of images we get poor results (for example, if the image is too dark or light, or the contrast is low). There can be an improved procedure by choosing the mean of the intensity values to threshold the image:

$$g(x, y) = \begin{cases} 0, & \text{if } f(x, y) \leq \overline{f(x, y)} \\ 1, & \text{otherwise} \end{cases}$$

where

$$\overline{f(x, y)} = \frac{1}{M \cdot N} \sum_{i=0}^M \sum_{j=0}^N f(i, j).$$

This method can be applied a bit more effectively than the previous one but only for binarization and not for multilevel thresholding. To obtain more accurate threshold values, we can analyse the intensity histogram of the image, which characterizes the frequency of the intensity values. In the next section, we present an iterative process for smoothing this histogram in order to find suitable threshold values.

4 Thresholding by histogram analysis

Let

$$h(x) = \sum_{i=0}^M \sum_{j=0}^N f_x(i, j),$$

be the intensity histogram of the image $f(x, y)$, where

$$f_x(i, j) = \begin{cases} 1, & \text{if } f(i, j) = x \\ 0, & \text{otherwise} \end{cases}$$

It is well-known that $h(x)$ should meet the condition

$$\sum_{i=0}^K h(i) = M \cdot N$$

Intuitively, the histogram $h(x)$, $0 \leq x \leq K$ can be considered as a bar diagram with K bars, where the heights of the bars are between 0 and $M \cdot N$ (If one of the bars has the height $M \cdot N$, then the others have zero height, so the image is constant.) The construction of the histogram is not a very time consuming process, it can be calculated by scanning through the image data once.

After producing the intensity histogram, it is an obvious idea to compose a function that matches the histogram well. To get the threshold values, we can search for the local minimum values of the function, which reside between local maximum values. This way, we can perform multilevel thresholding by designating local minimum values as threshold values. To find such an approximating function is rather a time consuming procedure. To save computation time, we should choose another method to smooth the histogram.

A general smoothing method is based on averaging the values of the intensity histogram. We can repeat this procedure iteratively, thus we obtain a sequence of histograms, which contains smoother and smoother histograms. This procedure is fast and can be implemented by simple programming tools.

Let $(h_i(x) : \{0, 1, \dots, K-1\} \rightarrow \mathbb{N})_{i=0, \dots, p}$ be a finite sequence of histograms, where p is a positive integer. The input of the algorithm is the intensity histogram $h(x)$ of the image, the output is the smoothed histogram $h_p(x)$. The resulting histogram $h_p(x)$ has q local maximum places, where q is a positive integer, given by the programmer. So, if we want to threshold the image into l levels, then we need to choose q to be equal with l . Now we present the iterative steps of our algorithm.

- (i) Let $h_0(x) = h(x)$, (initialisation)
- (ii) Let

$$\begin{aligned} h_i(0) &= h_0(0) \\ h_i(j) &= \left[\frac{h_i(j-1) + h_{i-1}(j)}{2} \right] \end{aligned}$$

where $i = 1, \dots, p$, $j = 1, \dots, K$ and $[x]$ is for the integer part of x .

In the end, the threshold values can be indicated as the local minimum places between the local maximum places. These minimum values are shifted by p , since the approximating histogram $h_i(x)$ is shifted to the right by 1 at every iterative step. So, for example, if the histogram $h_p(x)$ is bimodal where the local maximum places are $M1$ and $M2$, the local minimum value is m with $M1 < m < M2$, the threshold value for binarization is $m - p$.

To determine when to stop the iterative process, we have to count the number of the local maximum places, after every iteration. If this number is equal with q after the i^{th} iterative step, then we stop the iteration, thus $p = i$.

Our method is a very simple one. This feature eases the implementation. More complex methods can be used with preprocessing as well [7], [9], but the time of the processing should be longer if the methods are not simple enough.

5 Global Greyvalue Matching

This algorithm operates on K level greyscale images like the thresholding algorithm. The global value matching is a simple histogram comparison. If the two images are f and g , the matching constructs the two appropriate intensity histograms and computes the result. The result is a number between 0 and 100. This matching studies only the distribution of greyvalues in the images f and g . It will give us the similarity value in global greyvalues.

To compute the global greyvalue matching result we have to rescale the two images into the same size. Let the dimension be $M \times N$. Let us construct the two histograms $h_f(x)$ and $h_g(x)$ for the two images f and g . Then we have to apply the following formula:

$$G(f, g) = \sum_{i=0}^K \frac{|h_f(i) - h_g(i)|}{2 \cdot M \cdot N} \cdot 100$$

where K is the maximum number of the bars in the histogram h_f or h_g .

As we mentioned above, the result G is a number between 0 and 100. The 0 means that the two histograms are identical.

6 Local Greyvalue Matching

This algorithm also operates on K level greyscale images. The local greyvalue matching is a total matching on greyscale images. It is very noise sensitive but it has small computing time. A similar but more complex algorithm is in [10]. It could be implemented in PL/SQL in a very easy way.

To compute the local greyvalue matching, the two images have to be scaled into the same size. So let the dimensions of the two scaled images be M and N .

$$L(f, g) = \sum_{i=0}^M \sum_{j=0}^N \frac{|f(i, j) - g(i, j)|}{M \cdot N \cdot K} \cdot 100$$

where K is the level of the images, namely the number of the maximum possible bars in their histogram. The result L is between 0 and 100. The 0 means that the two images are identical in the local greyvalues.

7 The Final Result

To get the final result (the virtual distance between the query image f and the matched image g), we have to compute a weighted sum from the local and global greyscale matching and the original binary matching.

We have to use our new thresholding algorithm to construct the two threshold values to f and g . Let the values be t_1 and t_2 , and denote the two thresholded image by f_{t_1} and g_{t_2} . Denote the original binary pattern matching in the database

by B . We assume that the result of B is a number between 0 and 100. If it is not true, we could transform the result into a normalized result between 0 and 100 in an easy way. (If $a \leq B \leq b$ for all possible B , the normalized value is $\frac{B-a}{b} \cdot 100$).

We have to construct three weights w_1 , w_2 and w_3 where $0 \leq w_1 \leq 1$, $0 \leq w_2 \leq 1$, $0 \leq w_3 \leq 1$, $w_1, w_2, w_3 \in \mathbb{R}$. Thus, the final result F can be computed in the following way:

$$F(f, g) = w_1 \cdot G(f, g) + w_2 \cdot L(f, g) + w_3 \cdot B(f_{t_1}, g_{t_2})$$

The final result is also a number between 0 and 100, it is the virtual distance between f and g . If all of the weights are greater than 0, the result will be 0 iff the two images are identical. If $w_1 = 0$ and $w_2 = 0$, we give the original method.

These modifications could increase the efficiency of the original matching and improve the database image retrieval applications.

8 Implementation solution

As we can see from the description above, our computation algorithm based on averaging is not a very sophisticated one. The main advantage of this algorithm is its simple implementability at database level. Our algorithm is based on averaging and the division operator is implemented in almost every database management system as a basic operator.

The Oracle *interMedia* 8.1.7 module [6], and the PL/SQL language [8] are wide-spread tools and have great importance in implementing digital image processing algorithms at database level. Thus, we use a PL/SQL-like pseudo language to present the first few steps of our algorithm.

In our example the level of the images is 256 ($K = 255$).

As the first step, we have to define a "histogram" type to start working with histograms.

```
CREATE OR REPLACE TYPE histogram AS OBJECT
{
  intensity    v_of_num, //like an ARRAY 0..255 OF NUMBER,
  value       v_of_num, //like an ARRAY 0..255 OF NUMBER,

  MEMBER PROCEDURE process( step IN NUMBER ),
  MEMBER FUNCTION threshold RETURN NUMBER,
  MEMBER FUNCTION modal RETURN NUMBER
};
```

The intensity values of the histogram are stored in the array `intensity`, while the number of the pixels with the intensity values are stored in the array `value`. We divide the whole procedure into two additional member methods. Procedure `process` contains the iterative step of the histogram smoothing algorithm, while

the `threshold` member method calculates the threshold values. We have a simple function `modal`, which returns the modality of the approximating intensity histogram.

```
MEMBER PROCEDURE process( step IN NUMBER ) IS
  oldvalues ARRAY 0.255 OF NUMBER;
BEGIN
  FOR J IN 1..step
  LOOP
    oldvalues := value;
    FOR I IN 1..255
    LOOP
      value[ I ] := ( value[ I - 1 ] + oldvalues[ I ] ) / 2;
    END LOOP;
  END LOOP;
END;
```

The `threshold` method returns the minimum place between two maximum places in case of binarization. So, the whole binarization process can be implemented in the following way:

```
WHILE NOT modal = 2 LOOP
  process( 1 );
END LOOP;
```

After this routine, the threshold value can be obtained by the `threshold` function. The two other methods (the local and the global greyvalue matching) can easily be implemented as well.

9 Experimental results

Our experiments prove that in case of a common greyscale image the histogram smoothing algorithm provides a nice approximation of the original intensity histogram within a small number of iterative steps. So, for example, if we want to binarize an image, we only need a small number of steps to reach the bimodal histogram and find the threshold value according to this histogram.

The following figures show the results of our experiments. We can see an original intensity histogram (Fig. 1) and the approximating histogram after five and ten iterations (Fig. 1 and Fig. 2).

With predefined weights the matching gives the following final results in an Oracle 8i database management system using a self developed binary matching

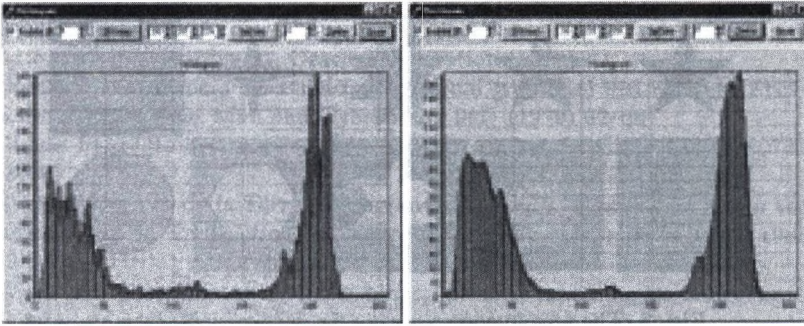


Fig. 1. The original intensity histogram and the approximating histogram after 5 steps.

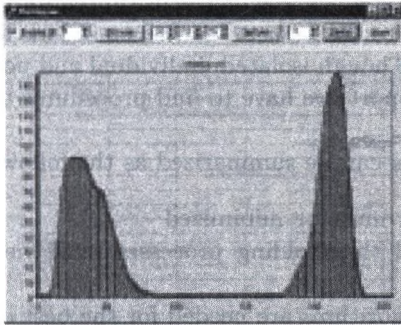


Fig. 2. The bimodal histogram after 10 steps.

algorithm. The query image was the image i_1 . (Fig. 3).

i_2	3.88
i_3	16.46
i_4	11.64
i_5	48.23
i_6	45.14
i_7	50.47
i_8	37.00
i_9	44.65
i_{10}	32.32

10 Conclusion

This paper shows an example how digital image processing algorithms can be implemented at database level and how the original database binary pattern matching can be improved. These procedures can be studied in literature but now we need a new approach, since we do not have that rich programming support at

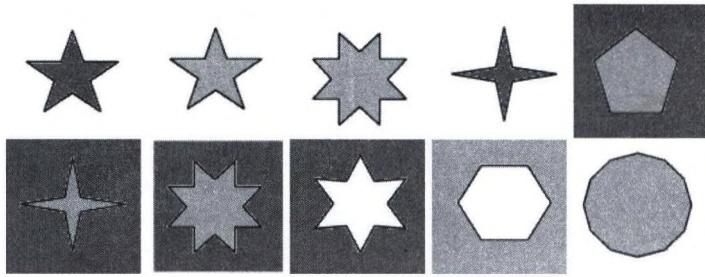


Fig. 3. The test images $i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8, i_9, i_{10}$.

this level. Nowadays, systems based on the use of the Internet and database management systems come to the front, so mathematical models should be aligned to this tendency as well. Though isolated, individual and non-reproducible systems may give us better support, we have to find procedures that can also be used in globalised systems.

The global concepts can be summarized as the following:

- The image transfer must be minimized
- The matching and thresholding processes must be executed at database server level
- Fast and simple algorithms are needed for the best efficiency
- Thresholded images must be stored after the analysis
- Local- and global greyvalue techniques have to be used to improve the binary pattern matching
- Some weights have to be used to refine the algorithms for special tasks

In future work we would like to refine and enhance the efficiency of the matching and test more. Since the thresholding algorithm is extensible to multimodal histograms and the local- and global greyvalue matchings meet our requirements, we would like to extend the whole strategy to true-colour images.

References

1. Chang, C. C., Lee, S. Y.: Retrieval of similar pictures on pictorial databases. *Pattern Recogn.* 24(7) (1991) 675–681
2. Eakins, J. P.: Automatic image content retrieval: Are we going anywhere? In *Proceedings of the 3rd International Conference on Electronic Library and Visual Information Research* (1996)
3. Glasbey, C. A.: An analysis of histogram-based thresholding algorithms. *CVGIP-Graphical Models and Image Processing* 55 (1993) 532–537
4. Gonzales, R. C., Woods, R. E.: *Digital image processing*. Addison-Wesley, Reading, Massachusetts (1992)
5. Grosky, W. I., Mehrotra, R.: Index-based object recognition in pictorial data management. *Comput. Vision Graph. Image Process.* 52(3) (1990) 416–436

6. Oracle interMedia Audio, Image, and Video User's Guide and Reference. Release 8.1.5, Part No. A67299-01, Oracle Corporation (1999)
7. Otsu, N.: A threshold selection method from grey-level histograms. *IEEE Transactions on Systems, Man and Cybernetics* 9(1) (1979) 62-66
8. Oracle PL/SQL 101. McGraw-Hill Professional Publishing, ISBN: 007212606X (2000)
9. Rosenfeld, A., P. de la Torre: Histogram concavity analysis as an aid in threshold selection. *IEEE Transactions on Systems, Man and Cybernetics* 13(3) (1983) 231-235
10. Veréb, K.: Generalization of the binary pattern matching in image processing. *Computers and Mathematics with Application* (2001) to appear

Ecsetvonás transzformáció, mint sztochasztikus képszegmentáló, leíró és tömörítő algoritmus

Szirányi Tamás^{1,3}, Tóth Zoltán^{2,3}

¹ Analogikai és Neurális Számítások Laboratórium,
A Magyar Tudományos Akadémia Számítástechnikai és Automatizálási
Kutatóintézete,
1111 Budapest, Kende u. 13-17.

sziranyi@sztaki.hu

² Elosztott rendszerek osztály,
A Magyar Tudományos Akadémia Számítástechnikai és Automatizálási
Kutatóintézete

sac@samson.aszi.sztaki.hu

³ Veszprémi Egyetem, Képfeldolgozás és Neuroszámítógépek Tanszék, 8200
Veszprém, Egyetem u. 10.

Absztrakt. Jelen cikk bemutatja a módosított ecsetvonás transzformációt, illetve az ecsetvonás transzformációval készített szegmentálási eredményeket. Mesterségesen létrehozott képre készített szegmentáláson megadja a szegmentálás hatásfokát. Összehasonlíttja az átalakított ecsetvonás transzformációt egy Markovi véletlen mező matematikai modellt használó Simulated Annealing sztochasztikus képszegmentáló algoritmussal. Mesterségesen létrehozott képeken összeveti a szegmentálások eredményét. Elsősorban a tömörebb képkódolás érdekében többszintű ecsetvonás-generálást vezet be. A két, illetve három szintű multiscale ecsetvonás transzformáció gyorsítására piramidális módszert mutat be, illetve bemutatja az így nyert futási eredményeket, összeveti azokat a hagyományos multiscale ecsetvonás transzformáció futási eredményeivel. Sztochasztikus optimalizációs módszert mutat be az ecsetvonások hatékonyabb generálására, elfogadására, az algoritmus futási sebességének javítására.

1. Bevezetés

Az ecsetvonás transzformáció olyan általunk kifejlesztett sztochasztikus képfeldolgozási eljárás, melynek során egy bemeneti képből festményszerű hatást keltő képet hozunk létre (1. ábra). Az algoritmus futása során kapott eredménykép homogén területeket (magukat az ecsetvonásokat) tartalmaz, illetve jó közelítéssel megőrzi a bemeneti képen található éleket. Ez a tulajdonsága adja az ötletet, hogy az eredeti algoritmusunkból kis módosítással olyan szegmentáló algoritmust lehet készíteni, mely hasonlít egy, a Markovi véletlen mező matematikai modellt használó képszegmentálásra is használható Simulated Annealing optimalizáló algoritmusra. A 2. fejezetben áttekintjük a felhasznált algoritmusokat (ecsetvonás transzformáció, MRF-SA, módosított ecsetvonás transzformáció). A

3. fejezetben pixel alapú, illetve folt alapú szegmentálási eredményeket vetünk össze. A 4. fejezetben módszert mutatunk be az eredeti ecsetvonástranszformáció futási sebességének valószínűségi számítási alapokon nyugvó javítására. Az 5. fejezetben megemlítünk egy multiscale-módszert az ecsetvonástranszformáció végrehajtására.

2. A felhasznált algoritmusok

2.1. Az eredeti ecsetvonás transzformáció

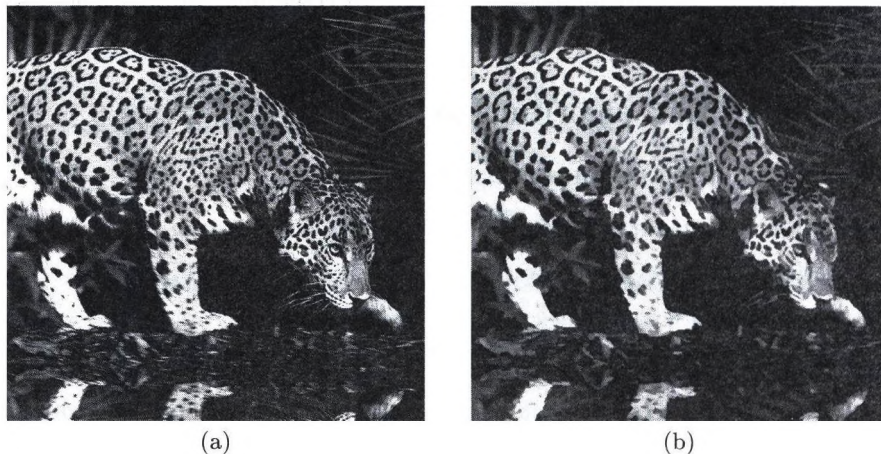
Az eredeti Ecsetvonás transzformáció [1, 2] menete a következő:

1. algoritmus. *Ecsetvonás transzformáció (X : az aktuális állapot)*

1. lépés: *Kezdés a legnagyobb méretű ecsetvonással.*
2. lépés: *Ha már nem az első iterációnál járunk, akkor a következő, kisebb méretű ecsetvonás kiválasztása (δ).*
 - *Ha megkaptuk a legfinomabb ecsetvonással kapott eredményt is, akkor lépés 14-re.*
3. lépés: $C_{\varphi\delta}$: *Az I bemeneti kép konvolúciója a φ orientációjú és δ méretű ecsetvonásokkal. Valamennyi δ vonásméretnél 8 db. $C_{\varphi\delta}$ konvolvált képet kapunk a 8 ecsetirányultságnak megfelelően. Ezekre a képekre az ecsetvonások színének becsléséhez van szükség.*
4. lépés: D : *Különbségkép képzése I és a jelenlegi X festett részeredmény között.*
5. lépés: A : *Négyzetes hiba kép létrehozása D -ből.*
6. lépés: *Ha az előző iteráció óta A normája nem változott szignifikánsan,*
 - *lépés 2-re.*
7. lépés: E : *Az A elmosása δ ecsetvonásméretnek megfelelő mértékben.*
8. lépés: *Az E elmosott hibakép értékeiből hisztogram készítése, küszöbérték meghatározása, melynél nagyobb hiba előfordulási valószínűsége ϵ .*
9. lépés: *(x, y) pozíció és φ ecsetirányultság véletlen választása.*
10. lépés: *Ha a hibaérték E -n az (x, y) pozícióban nagyobb, mint a 8-ban meghatározott ϵ -hoz tartozó küszöb, akkor*
 - *Rendeljük a $C_{\varphi\delta}(x, y)$ szint egy (x, y) pozíciójú φ irányultságú δ méretű ecsetvonáshoz.*
 - *Vagy, ha nagy méretű ecsetvonásokkal festünk, akkor az ecsetvonás területe alatt I -n található leggyakrabban előforduló szint rendeljük az (x, y) középpontú ecsetvonáshoz.*
11. lépés: *Helyezzük el a vonást a képen.*
12. lépés: *Ha a jelenlegi állapot jobb eredményt ad négyzetes hibában, mint az előző,*
 - *Véglegesítsük az ecsetvonást.*
 - *Különben pedig vonjuk vissza az ecsetvonás kirajzolását.*
13. lépés: *Ha az ecsetvonások száma nem halad meg egy képméretfüggő küszöbértéket,*
 - *Lépjünk 9-re.*

– Egyébként lépjünk 4-re.

14. lépés: Vége az eljárásnak.



1. ábra: Az eredeti 'leopárd'-kép (a), a 'leopárd'-kép festett közelítése (b).

2.2. Az MRF-SA algoritmus szegmentálásra, az algoritmus matematikai modellje

Az eredeti SA algoritmussal jelentős mennyiségű cikk, konferencia-kiadvány, könyv foglalkozik ([4, 5, 7]). A képfeldolgozás területén Geman és Geman [7] óta alkalmazzák szélesebb körben a Markovi véletlen mezők (MRF) modellt. Ebben, az MRF segítségével Bayesi matematikai modellt adtak a szerzők a kezünkbe képfeldolgozási problémák megoldására. A Simulated Annealing (SA) algoritmust Černý [8], Kirkpatrick [9] és munkatársaik publikálták először. Hammersley és Clifford (1971) hatékony matematikai eszközt adott a kezünkbe, amikor bizonyították az MRF és a Gibbs-eloszlás közötti ekvivalenciát [10].

Zajjal terhelt bemeneti kép esetén célunk az \mathcal{F} bemeneti kép valamennyi f_s pixelének besorolása valamely osztályba. Ilyen pl., amikor légifelvételből színes túristatérkép készítünk. A kimeneten a bemeneti képpel megegyező méretű képet kapunk (ω : Címke mező). A szegmentálásnál szóba jöhető bemeneti képmo-
dellnek megfelelő osztályok a μ_λ átlagértékekkel, illetve σ_λ szórással vannak jellemezve. Az MRF-SA futása során a pixelek besorolásánál a következő energiagombokat vizsgálja meg:

$$U = U_1(\omega, \mathcal{F}) + U_2(\omega) \quad (1)$$

$$U_1(\omega, \mathcal{F}) = \sum_{s \in \mathcal{S}} \left(\ln(\sqrt{2\pi}\sigma_{\omega_s}) + \frac{(f_s - \mu_{\omega_s})^2}{2\sigma_{\omega_s}^2} \right) \quad (2)$$

$$U_2(\omega) = \sum_{c \in \mathcal{C}} V_2(\omega_C) \tag{3}$$

ahol $V_2(\omega_C) = V_{\{s,r\}}(\omega_s, \omega_r) = \begin{cases} -\beta & \text{if } \omega_s = \omega_r \\ +\beta & \text{if } \omega_s \neq \omega_r \end{cases}$ amikor is $\beta \geq 0$ (4)

A 3. egyenletben szereplő β felelős a kialakuló régiók homogenitásáért. A 4. egyenlet egy lokális műveletet definiál, $V_{\{s,r\}}$ ugyanis csak azon s, r pixelekre van értelmezve, amik egymás 4-es szomszédságában vannak.

Az eredeti SA algoritmus a következő képpen nézett ki:

2. algoritmus. MRF-SA szegmentálás

1. lépés: ω véletlen inicializálása, megfelelően magas kezdőhőmérséklet választása ($T \approx +\infty$).
2. lépés: Olyan η próbakonfiguráció készítése, amely a jelenlegi ω állapottól csak egyetlen elemében különbözik.
3. lépés: (Metropolis-kritérium) $\Delta U = U(\eta) - U(\omega)$ kiszámítása és η elfogadása, amennyiben $\Delta U < 0$, egyébként elfogadása $\exp(-\Delta U/T)$ valószínűséggel:

$$\omega = \begin{cases} \eta & \text{ha } \Delta U \leq 0, \\ \eta & \text{ha } \Delta U > 0 \text{ és } \xi < \exp(-\Delta U/T), \\ \omega & \text{egyébként} \end{cases} \tag{5}$$

ahol ξ egyenletes eloszlású véletlen valószínűségi változó $[0, 1]$ -en.

4. lépés: Ugrás 2-re az egyensúlyi helyzet eléréséig.
5. lépés: A T hőmérséklet megfelelően kicsi lépésekben (lásd [4] további részletekért) történő csökkentése a rendszer befagyásáig.

Az algoritmus különféle hűlési folyamatok analógiájaként jött létre. Futása során időfüggetlen, homogén Markovi lánc generálódik a hőmérsékletcsökkentések között (elterjedtek az inhomogén módszerek is, melyek gyakorlatilag minden lépésben csökkentik a hőmérsékletet).

Az így definiált algoritmusról bebizonyítható, hogy eredménye elméletileg 1 valószínűséggel konvergál a globális energiaminimumhoz (optimális elérhető szegmentált eredményhez).

2.3. A módosított ecsetvonás transzformáció

Az előző két szakaszban tárgyalt sztochasztikus algoritmus működési elvét megfigyelve több hasonló vonást fedezhetünk fel. Az egyik legfontosabb, hogy mind a kettő lokális energiák számítását végzi, és próbál valamiféle globális energiaminimumra törekedni. Az ecsetvonás transzformáció futása során kapott eredmény, bár az algoritmus alapvetően nem szegmentálásra lett kitalálva és optimalizálva, az ecsetvonások homogén volta miatt homogén képszegmenseket tartalmaz (2. kép).



(a)



(b)



(c)



(d)

2. ábra: Az eredeti "Léna"-kép (a), az eredeti kép élképe (b), 37x9-es ecsetvonásokkal festett kép (c), illetve a festett kép élképe (d).

Az eredeti ecsetvonás algoritmust egyszerűsítettük, illetve az MRF-SA szegmentálásnál használt képleteket használtuk az energiatagok számításánál annak érdekében, hogy a két modell egyszerűbben összehasonlíthatóvá váljon. Az így kapott algoritmus a következő:

3. algoritmus. Módosított ecsetvonás transzformáció szegmentálásra

1. lépés: ω inicializálása (nem besorolt állapot), megfelelően magas kezdőhőmérséklet választása ($T \approx +\infty$).

2. lépés: Véletlen pozíció és véletlen ecsetvonás (δ) választása az ecsetvonás-készletből.

3. lépés: Olyan η próbakonfiguráció készítése, amely a jelenlegi ω állapottól csak a homogén δ területe alatt tér el.

4. lépés: (\sim Metropolis-kritérium) $\Delta U = U(\eta) - U(\omega)$ kiszámítása és η elfogadása, amennyiben $\Delta U < 0$, egyébként elfogadása $\exp(-\Delta U/T)$ valószínűséggel:

$$\omega = \begin{cases} \eta & \text{ha } \Delta U \leq 0, \\ \eta & \text{ha } \Delta U > 0 \text{ és } \xi < \exp(-\Delta U/T), \\ \omega & \text{egyébként} \end{cases} \quad (6)$$

ahol ξ egyenletes eloszlású véletlen valószínűségi változó $[0, 1]$ -en, normálva az alkalmazott ecsetvonás területével.

5. lépés: Ugrás 2-re az egyensúlyi helyzet eléréséig.

6. lépés: A T hőmérséklet megfelelően kicsi lépésekben (lásd [4] további részletekért) történő csökkentése a rendszer befagyásáig.

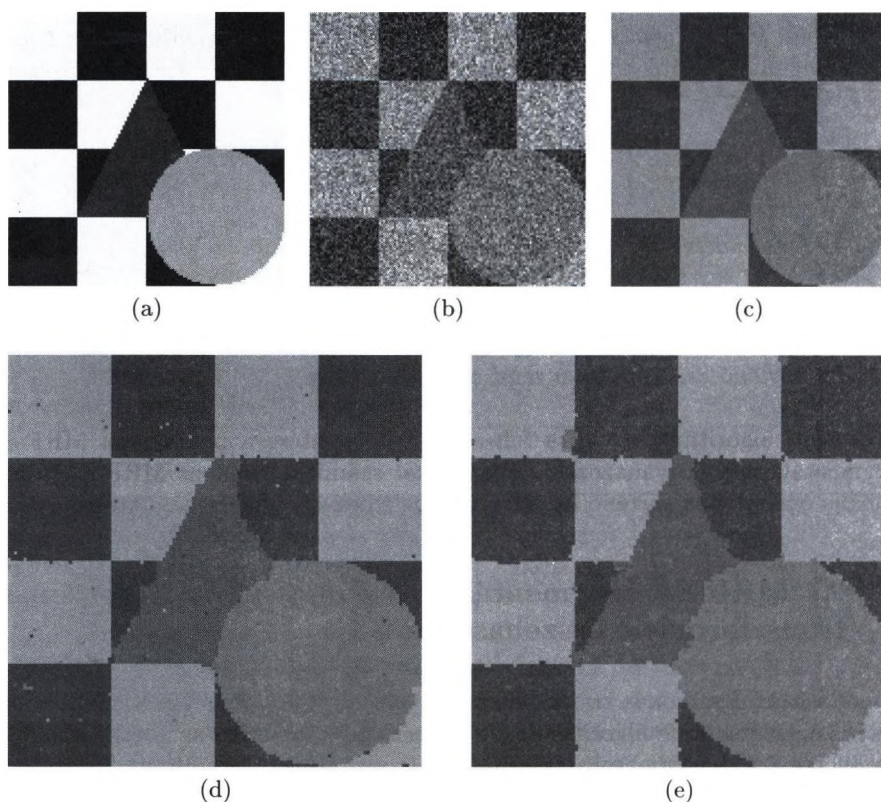
A fenti algoritmus egyfajta folt-alapú kiterjesztése a pixel-alapú MRF-SA szegmentálásnak (ugyanazon energiatagokat számítja, mint az MRF-SA), de a globális optimumba történő konvergencia jelenleg még nem bizonyított.

3. Az MRF-SA szegmentálás és a módosított ecsetvonás transzformáció összehasonlítása

A módosított ecsetvonás transzformáció 1×1 -es vonások esetén valójában egy MRF-SA szegmentáló algoritmust. A teszteredmények ezért az 1×1 -es vonásokat, illetve 7×3 -as vonásokat alkalmazó algoritmusokat hasonlítanak össze.

A 3. képen a bemeneti, mesterséges ábra látható, illetve a pixel és a folt alapú megközelítéssel kapott végeredmények optimális β paraméterek esetén. A bemeneti képen 0 várhatóértékű 0.1 szórású Gaussi-zaj volt ($snr = 3dB$). A szegmentálás eredménye hasonló minőségű volt (1.096%-a, illetve 0.731%-a a pixeleknek lett nem megfelelő osztályba sorolva a szegmentálás során 7×3 -as illetve 1×1 -es ecsetvonások használata esetén), de a konvergencia 1×1 -es esetben harmadannyi időt vett igénybe, mint a folt alapú megközelítés esetén. Ennek oka az, hogy egy rosszul megválasztott ecsetvonás annak nagyobb szabadsági foka miatt nehezebben javítható ki, mint egy rosszul megválasztott (nem megfelelő osztályba sorolt) pixel. A sztochasztikus, globális optimumra törekvő relaxáció viszont megköveteli a hiba, illetve az energia növekedését is. Az összehasonlításnál további problémát jelent a régiók homogenitását szabályozó β paraméter értelmezése is (lásd 3-4. egyenletek). Az ecsetvonások ugyanis homogenitásukból kifolyólag magukban hordoznak egy, a végeredmény homogenitását befolyásoló tényezőt.

Mind a folt, mind a pixel alapú megközelítés esetén a legnagyobb problémát a szegmentálás lassúsága jelenti. Hogy folt alapú esetben a már említett problémát, a rossz osztályba történő besorolás esélyét csökkentjük, a matematikai modellek hasonlóságának rovására, további módosítást hajtottunk végre az ecsetvonás transzformáció algoritmusán. Az ecsetvonások elfogadásánál az U_1 tagban (2. egyenlet) nem engedjük meg az energia romlását, vagyis ha az elfogadott ecsetvonásra számított U_1 energia (mennyire hasonlít az ecsetvonás a bemeneti

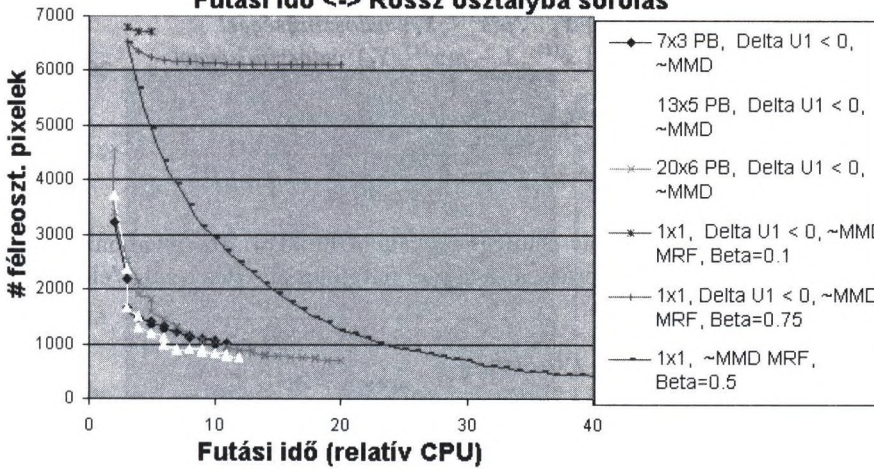


3. ábra: (a) eredeti ábra, (b) zajos bemeneti ('Trinois') kép, (c) tapasztalati átlag alapján készített "ideálisan" szegmentált kép, (d) végeredmény 1×1 -es ecsetvonások alkalmazása esetén a fenti modellben, (e) végeredmény 7×3 -as ecsetvonások alkalmazása esetén

képen a területe alatt található pixelekre) romlik (növekszik), akkor az ecsetvonást nem fogadjuk el. Összevetve a pixel-szintű modellel, a festés során a területi hiba növekedése azt jelentené, hogy az MRF-ekvivalensben a pixelek többsége rossz címkét kap, miközben az átlagos szín (osztályok többsége) helyes. Ennek elhanyagolható a valószínűsége, ezért nem is engedjük meg, ami belefér az ekvivalenciába.

A futási eredményeket megfigyelve (4. ábra) azt tapasztalhatjuk, hogy nagyarányú sebességnövekedést érünk el a gyakorlatban elérhető képminőség effektív változása nélkül. Ez esetben az algoritmus kb. fele/harmad annyi idő alatt konvergált a módosított ecsetvonás transzformációval a végeredményhez, mint a pixel alapú megközelítés algoritmus. Az ábráról az is kitűnik, hogy a döntési szabály ilyen jellegű változása a pixel-alapú szegmentálás esetén viszont az elért képminőség durva romlásához vezetne.

Szegmentálás 4 színnel (Trinois 128*128)
Futási idő <-> Rossz osztályba sorolás



4. ábra: A konvergencia menetének ábrázolása több esetmére olyan esetekben, amikor megengedtük az U_1 energiateg romlását, illetve, amikor tiltottuk azt ($\Delta U_1 \neq 0$).

4. Az eredeti ecsetvonás transzformáció gyorsítása

Az ecsetvonás transzformáció futása során egy új ecsetvonás pozíciójának meghatározása teljesen véletlenszerűen történik. Az ecsetvonások elfogadása csak a jelenlegi állapottól függ, az algoritmust tárgyalhatjuk, mint Markovi-lánc folyamatot. Ezen folyamat konvergenciáját felgyorsíthatjuk egy megfelelően választott Monte Carlo Markov Lánc módszer bevezetésével a döntési lépésekben.

Az eredeti ecsetvonás transzformáció esetében a célunkat megfogalmazhatjuk a következő formában: közelítsük a bemeneti képet homogén ecsetvonásokkal oly módon, hogy a kapott eredmény minden ecsetvonása és a bemeneti kép között számított négyzetes hiba egy előre meghatározott érték alatt maradjon. Jól leírható ez a követelmény egy úgynevezett cél eloszlás megadásával.

A következőkben az alábbi jelöléseket fogjuk használni. Az elemi esemény ($X^{(t)}$) esetünkben egy ecsetvonás generálása az $S^{(t)}$ pozícióra $C^{(t)}$ színnel. A képi hiba $X^{(t)}$ és az eredeti kép $X^{(t)}$ területe alatti pixelek között $E^{(t)} = E(X^{(t)})$. $S^{(t)}$ véletlenszerűen generált, vagy függhet az előző állapottól, illetve a bemeneti kép paramétereitől. $C^{(t)}$ szint meghatározza a vizsgált ecsetvonás területe alatti pixelek többségének színe a bemeneti képen. A keresett eloszlás f , esetünkben $E(X^{(t)})$ célfüggvénye: $f = f(E^{(t)})$.

Felírhatjuk tehát a következő Metropolis-Hastings [11] szabályokat az $X^{(t)}$ -k generálására:

4. algoritmus. PB-MH

Adott $x^{(t)}$, a t -edik lépésben kapott ecsetvonás.

1. lépés: Generáljuk: $Y_t \sim q(y|x^{(t)})$

2. lépés: *Vegyük:*

$$X^{(t+1)} = \begin{cases} Y_t, & \rho(x^{(t)}, Y_t) \text{ valószínűséggel} \\ x^{(t)}, & 1 - \rho(x^{(t)}, Y_t) \text{ valószínűséggel} \end{cases} \quad (7)$$

ahol

$$\rho(x, y) = \min \left\{ \frac{f(y)q(x|y)}{f(x)q(y|x)}, 1 \right\} \quad (8)$$

Az algoritmus f eloszlás célfüggvényként a hibakorlátot alkalmazva indul. Az S^{Y_t} véletlenpozíció generátora, a $q(y|x)$ feltételes eloszlás a hiba/él-kép és az $X^{(t)}$ előző vonás függvénye.

5. algoritmus. *Független PB-MH*

Adott $x^{(t)}$.

1. lépés: *Generáljuk (a bemeneti kép hibaképéből és/vagy élképéből):*

$$Y_t \sim g(y)$$

2. lépés: *Vegyük:*

$$X^{(t+1)} = \begin{cases} Y_t, & \min \left\{ \frac{f(Y_t)g(x^{(t)})}{f(x^{(t)})g(Y_t)}, 1 \right\} \text{ valószínűséggel} \\ x^{(t)}, & \text{egyébként} \end{cases} \quad (9)$$

Ez esetben a javasolt sűrűségfüggvény, $g(y)$ az $S^{(t+1)}$ véletlenpozíció generátora, függetlenül az előző $S^{(t)}$ pozíciótól. A $g(y)$ függvény azonos lehet az egész képre.

6. algoritmus. *Véletlen séta PB-MH*

Adott $x^{(t)}$.

1. lépés: *Generáljuk: $Y_t \sim g(y - x^{(t)})$*

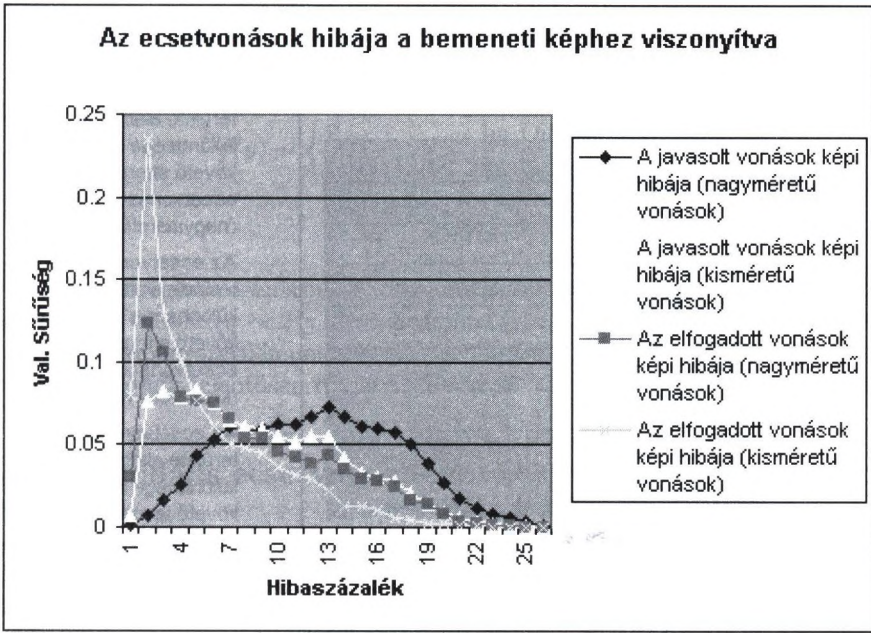
2. lépés: *Vegyük:*

$$X^{(t+1)} = \begin{cases} Y_t, & \min \left\{ \frac{f(Y_t)}{f(x^{(t)})}, 1 \right\} \text{ valószínűséggel} \\ x^{(t)}, & \text{egyébként} \end{cases} \quad (10)$$

Ez esetben a javasolt sűrűségfüggvény, $g(y - x^{(t)})$ az $S^{(t+1)}$ véletlenpozíció generátora, $S^{(t)}$ pozíciótól függően.

Az eredeti algoritmust tekinthetjük úgy, hogy az elfogadást/elutasítást meghatározó f függvény egy egyszerű küszöbszámítás E -n ($f(X) = f(E(X))$), ahol $f(x^{(t)})$ és $f(Y_t)$ az adott ecsetvonáspozíción vizsgáljuk, és $X^{(t)}$, illetve $Y^{(t)}$ a teljes eredményképet jelenti a festés aktuális állapotában.

Lévé, nem találtunk olyan valószínűségi elfogadás/elutasítás stratégiát, amely ebben a környezetben jobb eredményt adott volna, mint a hagyományos modellünk (a kapott képminőség vagy/és a konvergencia sebessége volt jelentősen rosszabb), továbbléptünk egy teljesen sztochasztikus MCMC-alapú modell felé. Olyan algoritmust kerestünk, amely egyfajta önkalibráló módon határozza meg az ecsetvonások elfogadásának módszerét.



5. ábra: A képi hiba valószínűségi sűrűségfüggvényei a javasolt (Y_t), illetve az elfogadott ($X^{(t+1)} = Y_t$) vonások és a bemeneti kép között a 'Barbara'-képre készített teszt futtatása során. Először nagy méretű (20×5), majd kisebb (7×2) vonásokat generáltunk.

A következőkben jelöljük az állapottérben $X^{(t)}$ (jelenlegi állapot), illetve Y_t (javasolt, új állapot) állapotokhoz tartozó generált kimenetet S^t -vel, illetve S^{t+1} -gyel (az adott kimenet tartalmazza az összes addig legenerált ecsetvonást).

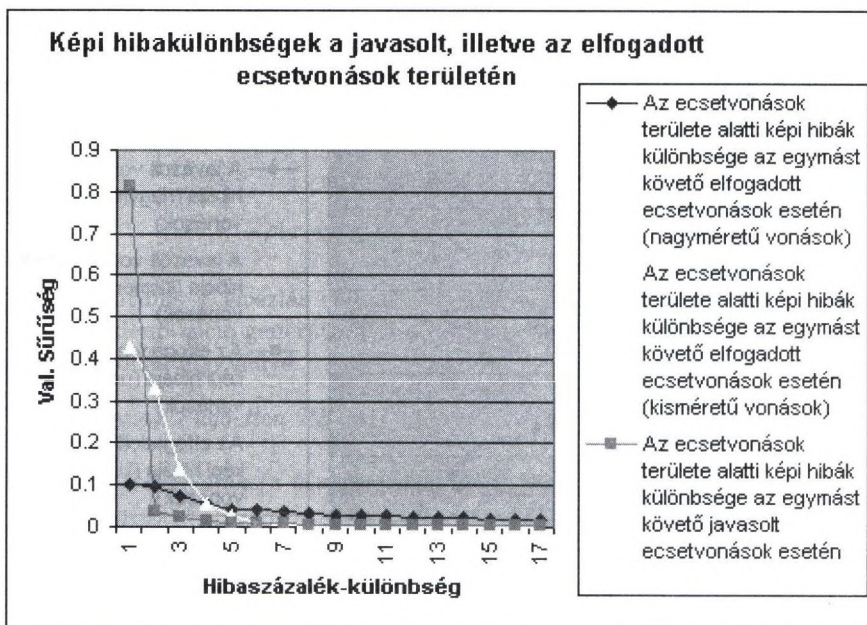
A fellépő képi hibákat a következő formában jelölhetjük:

- $E(x^{(t)}) = D(X^{(t)}, I)$ a képi hiba a bemeneti kép és az előző állapot között.
- $E(Y_t) = D(Y_t, I)$ a képi hiba a bemeneti kép és a javasolt Y_t állapot között.
- $E(x^{(t-1)}) = D(X^{(t-1)}, I)$ a képi hiba a bemeneti kép és az előző előtti állapot között.

$D(.,.)$ négyzetes hibaként van definiálva, és normalizálva van az ecsetvonás területével a $[0.0, 1.0]$ intervallumra.

Megvizsgáltuk a javasolt és elfogadott $D(Y_t, I)$ értékeket. Azt tapasztaltuk, hogy az $f(D(Y_t, I))$ eloszlás a nullában a Dirac-delta függvényhez konvergál. Egy mérés során nyert eloszlásfüggvény (5. ábra) azonban nem alkalmazható általánosan célfüggvényként (az elérhető képminőség kép, illetve ecsetvonásfüggő). Figyelembe kell továbbá venni, hogy ha elfogadunk olyan ecsetvonásjavaslatokat is, amelyek növelik a képi hibát, az eredménykép könnyen struktúrálatlanná válik.

Ezen megfontolások alapján a következő korlátozó feltételeket kell szem előtt tartani:



6. ábra: A képi hibakülönbségek valószínűségi sűrűségfüggvényei a javasolt, illetve az elfogadott egymás után következő vonások esetén ($\text{Diff}(Y_t)$). A statisztika a 'Barbara'-képre készített teszt futtatása során készült. Először nagyméretű (20×5), majd kisebb (7×2) vonásokkal közelítettük a bemenetet.

- Csak a képi hibát csökkentő vonásokat szabad elfogadni.
- Az előző elemi eseményt és a jelenlegi különféle javasolt ecsetvonáspozíciókat kell összevetni a javasolt ecsetvonás területén található képi hiba vizsgálata helyett.
- az MCMC sűrűségfüggvény-becslés dinamikus, követi a változó célfüggvényt.

Több kísérletet elvégezve arra a megfontolásra jutottunk, hogy nem is annyira a $D(\cdot, \cdot)$ tagok, hanem inkább azok különbségei a fontosak az optimalizációs folyamat során. Definiáljuk a következő különbségeket:

$$\text{Diff}(Y_t) = E(x^{(t)}) - E(Y_t)$$

$$\text{Diff}(x^{(t)}) = E(x^{(t-1)}) - E(x^{(t)})$$

Tesztek futása során vizsgáltuk a fenti hibakülönbségeket (pl.: 6. ábra). Ezen vizsgálatok, illetve további elméleti megfontolások során jutottunk a következő ecsetvonás elfogadó/elutasító szabályra:

$$X^{(t+1)} = \begin{cases} Y_t, & \rho(x^{(t)}, Y_t) \text{ valószínűséggel} \\ x^{(t)}, & 1 - \rho(x^{(t)}, Y_t) \text{ valószínűséggel} \end{cases} \quad (11)$$

ahol

$$\rho(x, y) = \min \left\{ \frac{f(y)q(x|y)}{f(x)q(y|x)}, 1 \right\} \quad (12)$$

$$f(y) \sim \begin{cases} 0 & , \text{ ha } \text{Diff}(Y_t) < 0 \\ (\text{Diff}(Y_t))^q & , \text{ egyébként} \end{cases} \quad (13)$$

$$f(x) \sim (\text{Diff}(x^{(t)}))^q \quad (14)$$

Teszteket végezve ezen szabály alkalmazásával azt kaptuk, hogy jól beállított paraméterek esetén ($q \sim 2 \dots 3$) a hagyományos ecsetvonástranszformáció $\epsilon = 0.01$ szép, festményszerű képet generáló beállításához képest nem sokkal rosszabb eredményt kapunk jelentősen rövidebb idő alatt (kb. kétszeres sebességjavulás).

5. Piramidális ecsetvonás transzformáció

A hatékonyság növelésének az előző fejezetben leírt módszertől eltérő megközelítése, amikor valamely multiscale megoldást használunk az ecsetvonások generálásánál. Ez esetben ugyanis nem a vizsgálandó ecsetvonások számát próbáljuk csökkenteni, hanem az egy ecsetvonás vizsgálatára fordított időt. Eddigi tesztjeink során is alkalmaztunk már egymás után különféle méretű ecsetvonásokat ugyanazon kimeneti kép generálása során. E mögött az az elgondolás rejlik, hogy az algoritmus a kezdeti szakaszban néhány nagyméretű vonással lefedheti a nagy, homogén régiókat, kiváltva ezzel több kisméretű ecsetvonás kirajzolását. A módszer eredménye elsősorban a kapott kód kompaktságában volt lemérhető.

Az új elgondolás az, hogy ne az ecsetvonások, hanem a kép méretét változtassuk az eredmény generálásának különböző szakaszaiban. Az ecsetméret-váltásoknál azután mindig eggyel nagyobb méretosztályba tartozó ecsetvonással rajzoljuk ki a méretváltásig legenerált kódot a megnövelt területű kimeneti képre (az ecsetvonások koordinátáit is természetesen arányosan meg kell változtatni) Ez csekély mértékű romlást eredményez ugyan az elért képminőség-kódtömörség arányt tekintve (a kicsinyített képen egy kis méretű ecsetvonás csak közelítése a visszagyújtott kép nagy ecsetvonásának, illetve az ecsetvonások lehetséges pozícióinak száma is lecsökken), de ez a változás elfogadható mértékű az elért sebességnövekedéshez viszonyítva. Jól prezentálja ezt a 1 táblázat. A táblázatban szereplő teszt során két alkalommal léptünk ecsetmérettartományt, mindannyiszor kétszeresére növelve a kezdetben lekicsinyített méretű kép, illetve részeredmény szélességét és magasságát.

Köszönetnyilvánítás

Josiane Zerubia-nak, a francia INRIA kutatóintézet Ariana project vezetőjének a franciaországi kutatómunka lehetőségéért, illetve Kató Zoltánnak, amiért PhD téziseit rendelkezésünkre bocsátotta.

1. táblázat: A hagyományos ecsetméretváltással, illetve piramidális módszerrel a 'Lena'-képre 256×256 készített tesztek eredményeinek összehasonlítása. Mindkét teszteredmény három ecset-, illetve képméretet használt a végeredmény legenerálásához.

	'Lena'-teszt hagyományos ecsetméretnöveléssel	'Lena'-teszt a képméret változtatásával
A végeredmény minősége PSNR-ben (dB)	29,05	28,99
A végeredmény kódjának hossza (kbyte)	59,1	60,7
Optimalizált kód méret (kbyte)	36,9	37,9
Futási idő (sec)	1061	168

Irodalom

1. Szirányi Tamás, Tóth Zoltán, Stochastic Paintbrush Transformation, Képfeldolgozók és Alakfelismerők 2. Konferenciája, Noszvaj, 2000. január 20-22.
2. T. Szirányi, Z. Tóth: Random Paintbrush Transformation, 15th ICPR, Barcelona, IAPR & IEEE, V.3, pp.155-158, 2000
3. Z. Kató: Multiscale Markovian Models in Computer Vision, Application to SPOT Satellite Image Segmentation (PhD thesis), INRIA and University of Nice-Sophia Antipolis, France, 1994.
4. P.J.M van Laarhoven and E.H.L. Aarts: Simulated Annealing: Theory and Applications, D.Reidel, 1987.
5. S.Z. Li: Markov Random Field Modeling in Computer Vision, Springer, 1995.
6. T. Szirányi, Z. Tóth: Optimization of Paintbrush Rendering of Images by Dynamic MCMC methods, EMMCVPR-2001, Lecture Notes on Computer Science, Springer Verlag, Vol. LNCS 2134, pp.201-215, 2001.
7. S. Geman and D. Geman: Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, **6**:721-741, 1984.
8. V. Černý: Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm, J.Opt. Theory Appl., 45(1):41-51, January 1985.
9. S. Kirkpatrick, C. Gellatt, and M. Vecchi: Optimization by simulated annealing, Science **220**, pp 671-680, 1983.
10. G. Winkler: Image Analysis, Random Fields and Dynamic Monte-Carlo Methods, Springer Verlag, 1995.
11. C.P. Robert, George Casella: Monte Carlo Statistical Methods, Springer Verlag, 1999.

Opto-electronic CNN implementation provides a new and powerful tool for image processing applications

Szabolcs Tókécs, László Orzó, and Tamás Roska

Analogical and Neural Computing Laboratory,
Computer and Automation Institute,
Hungarian Academy of Sciences, PO box H-1111
1111 Kende u.11, Budapest, Hungary
{Tokes, Orzo, Roska}@sztaki.hu
<http://www.labanalogic.sztaki.hu/index.html>

Abstract. An opto-electronic CNN implementation, as it combines the huge inherent computational capabilities of a massively parallel optical system with programmability of the CNN-UM systems, provides an efficient frame for diverse image processing applications. Our optical CNN implementation is based on a new, semi incoherent optical correlator architecture, which is superior from several points of view than any alternative optical correlator. This architecture provides the possibility of real time reprogramming of the correlator like in the joint Fourier transform correlators, but preserves the speed of VanderLugt type of systems. Programming is done through fast holographic read-out by B-templates. With combination of a high speed CNN-UM sensor chip it forms a programmable opto-electronic analogic computer (POAC) architecture. This device is especially useful in those image-processing algorithms, which requires a great number of pattern matching tasks.

1 Introduction

Image understanding (pattern recognition, classification etc.) is important task in image processing applications. Moreover, its realization requires several imageprocessing steps. Optical correlators can fulfill these tasks efficiently because the computation is accomplished by physics in a vastly parallel way. Our optical correlator with an embedded visual CNN-UM chip forms a programmable opto-electronic analogic computer (POAC) architecture what is intended to compare and process images with several TeraOps speed.

Demands for fast identification and tracking of targets in surveillance systems has been increasing dramatically during the last years [1], [2]. Although the known optical processors can fulfil these task with the required speed, they are usually not flexibly programmable [3], [4], [5]. Another problem of optical processors is that they require considerable pre- and post-processing (spatial filtering, adaptive thresholding, analog storage, further nonlinear processing). As these computations are usually accomplished by a DSP system, the serial

download (upload) of the gained results restricts overall performance. An optical processing systems is required, in which a large number of templates can be stored and rapidly retrieved, hybridized with a parallel electronic device having built-in programmable high sensitivity optical sensors and possible pre- and post-processing capabilities [6], [7], [8], [9]. We developed an efficiently programmable opto-electronic computer architecture (POAC), estimated its possible capabilities, investigated its feasibility and constructed a working model of it [10]. Combining our specially designed, programmable optical correlator — which has high speed and massive parallelism — and a CNN-UM chip [8] — with flexible stored programmability and local analog memories — we have constructed a system that overcame the main technical bottlenecks both of the optical and VLSI implementations. In this paper we have defined the goals and demands we were facing and introduce the suggested hybrid architecture. We describe the solution that has been chosen and characterize the proposed and realized programmable opto-electronic analogic computer (POAC) system. After a brief summary of the accomplished results we delineate the POAC system's so far achieved and in the near future achievable performance and compare it with that of other existing technologies and implementations. A number of possible applications are proposed, where our hybrid system is more efficient than any other existing technology. In one of these applications (in target tracking) we provide simulation and measurement results and propose the ways of further improvements of this POAC algorithm. Finally, we outline the possible perspectives of our POAC system.

2 Method

We were looking for a hybrid (opto-electronic) system, which contains a stored (and/or re-) programmable optical processor and a high-speed programmable sensor array [9]. As the sensor array can fulfill all the necessary pre- and post-processing tasks, it can alleviate the serial output transfer (feedback) bottleneck of the common optical and opto-electronic processors. From several point of views (see [5]/pp.258) a CNN-UM chip [11], [12], [13], [14] with special optical input is an optimal sensory array for our purposes. It provides an efficient frame, possessing all the required characteristics: It has highly sensitive, programmable optical sensors; local connectivity; built in non-linearity; local analog memory etc.. It can also fulfil all the required pre- and post-processing tasks in a parallel way.

In the POAC framework, the major part of processing can be done at the speed of light with a programmable optical correlator and the rest of the processing on fast, parallel and stored programmable CNN-UM chips. Another key aspect of our endeavor was to find and use special analog memories for high-density dynamic optical storage [15], [16], we developed and used a special version of dry bacteriorhodopsin films as a dynamic holographic store. With all these components it becomes possible to confidently detect, identify and track several targets with our system within milliseconds.

In recent years, several studies have demonstrated that a cellular neural/non-linear network (CNN) based computer architecture, the CNN Universal Machine (CNN-UM), a massively parallel nonlinear array processor, is a powerful solution for solving a number of complex image processing tasks e.g. [17], [18], [19]. While emulated digital and mixed-signal analog implementations are emerging, there have been only a few attempts to build an optical (or opto-electronic) implementation of the CNN [20], [21], [22], [23]. As the basic operation of the CNN is convolution (or correlation), optical implementation seems to be plausible and feasible. This way, almost all the known constraints of a VLSI implementation (only nearest neighborhood templates, limited resolution) can be alleviated. Additionally, the analogic CNN-UM paradigm provides a programming frame to optical computations, as well.

3 Optical correlator architecture

Based on the fundamental functional characteristics of the traditional joint Fourier transform correlator (JTC) system, a new type of correlator architecture (the t_2 -JTC) was invented [10], [23]. Within this architecture the correlation operation is executed in two consecutive steps. First a hologram of the input image is recorded in bacteriorhodopsin (BR) film, which is a fast dynamic holographic material. In the following step (steps) we reconstruct the recorded hologram with specially adjusted parallel laser beams. Angular coding of the template pixels is accomplished by a 2D acoustooptical deflector followed by a beam expander. This way, each beam reconstructs an appropriately shifted version of the original image and forms the correlogram. These beams are coherent but mutually incoherent due to the AO deflector caused different Doppler shifts and time difference. Within this architecture we can apply reference templates either in the hologram-recording step or in the readout-reconstructing step, or in both. Despite both the holographic recording and reconstruction are coherent, the shifted images are summed incoherently. It is especially beneficial, when different kinds of phase errors are present in the system (it is frequently the case in customary liquid crystal displays). This way the majority of the otherwise always-existing coherent noise is eliminated, while errors of shadow casting correlators are not present in the system. In this setup we record reflection type holograms, but transparent holograms have also been tested.

After every hologram recording steps we can fulfill a great number of read-out, correlation generation (measurement) steps. Read-out can work at the speed of light, as there is no needs of photon-electron conversion, just only diffractive interactions occur. Thus, only the detector and the displaying time of the template limit its pace. Exact positioning of the input and the reference (t_1) beams are not required in this amended architecture. A fractional Fourier setup is also operable and this way we can avoid saturation of the recording material, caused by the Fourier spectrum's highly inhomogeneous intensity distribution that requires high dynamic range of the holographic media. With a dual axis (fractional) joint Fourier correlator architecture applying a high-resolution dynamic holographic

material as an analog memory ("square law" detector: Bacteriorhodopsin film in our system) and with the use of very high-speed reconstruction (t_2 template) we demonstrated that the POAC principle is realizable.

Reconstructing the hologram by coherent but mutually incoherent template pixel beams we decreased the coherent noise considerably. The most important advantages of incoherent correlators are preserved, without their main hindrances (diffraction is only a nuisance in a shadow casting system).

3.1 Final POAC architecture

Using the above methods we constructed and optimized a final POAC architecture. It can be seen in Fig. 1.

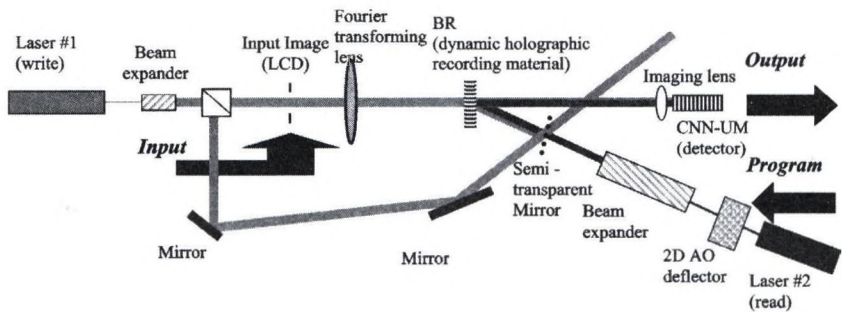


Fig. 1. Our POAC architecture, which consists of a new programmable optical correlator and a VLSI CNN-UM chip with special optical input and on-chip processing capabilities. Correlation is determined in two consecutive steps: first a hologram of the input image (600x800) is recorded (indicated by light gray beam) in a dynamic holographic medium (Bacteriorhodopsin film). In the consecutive step(s), suitably modulated laser beams (dark gray beam), according to angular coded t_2 template pixels, reconstruct the appropriately shifted versions of the input image and produce correlation by incoherent summation. A high-speed two-dimensional acousto-optical deflector accomplishes this angular coding. This way a large number of CNN B-template operations (1000 frames/second with 32x32 templates) can be executed on the recorded input image.

Our model device computes correlation in two consecutive steps. First a hologram of the input image is recorded in bacteriorhodopsin (BR) film, which is a fast dynamic holographic material [16]. The BR samples with the desired characteristics has been developed and fabricated by a world-leading laboratory¹ of the field. In the following step (steps) we reconstruct the recorded hologram with specially adjusted parallel laser beams. Angular coding of the B-template pixels is accomplished by a 2D acousto-optical deflector followed by a beam expander. This way, each beam reconstructs an appropriately shifted version of

¹ Institute of Biophysics, Biological Research Center, H-6701 Szeged, Hungary

the original image and forms the correlogram. These beams are coherent but mutually incoherent due to the AO deflector caused different Doppler shifts and time difference. The applied architecture was robust and less sensitive for misalignments of the elements than in the case of other architectures. Only a few key parameters had to be set carefully [24].

3.2 Achieved and Predictable performance after modifications and developments

In our current setup we were able to compute correlation of 32x32 binary templates on 500x500 and on 600x800 input images. The speed of the correlation measurement was limited by the speed of the data acquisition card to 20-25 frames/second. By the application of a high-speed sensor and on chip post processing (ACE16K) this speed can increase up to 1000 frames/second. Development of an improved AOD driver electronics is just completed, so we can exploit the physical limits of the AOD ($1\mu\text{s}/\text{template pixel}$). An 800x600 LCD has been purchased and tested. Up to 1000x1000 pixel input image size there is no need of considerable modification of the current optical setup. From these data it is evident that 10^{13} gray-scale operations per seconds is easily achievable by the current system after the above mentioned modifications.

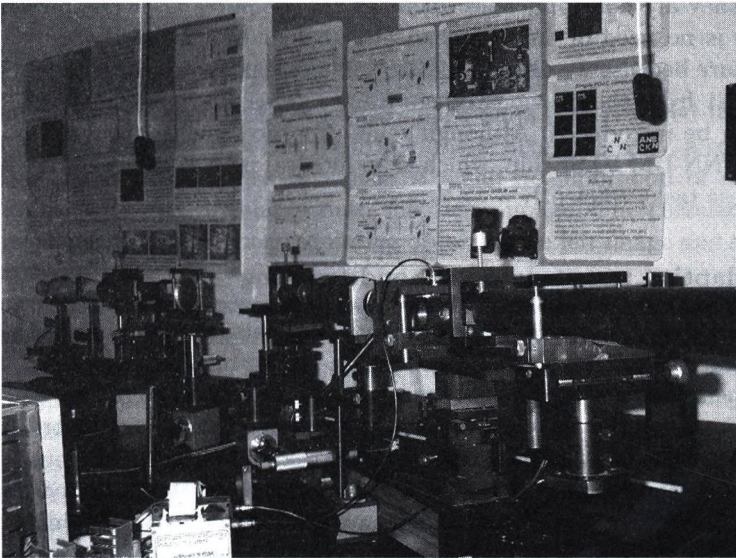


Fig. 2. Photograph of our semi-incoherent optical CNN setup.

Input SLMs with resolution of up to 3500x4500 (and even 4500x4500 OASLMs) pixels are commercially available. With such a high-resolution microdisplay we will considerably improve the systems overall performance. Different types of 2D

AOD are available, which can provide much higher speed (down to 20 ns/pixel). Alternative "scanning" methods like the application of appropriate microlaser diode arrays can overcome the problem caused by the template's serial scanning. Utilizing the above mentioned elements even 10-100 TeraOps (6-8 bit gray scale operations) (frame size in pixels/elementary reconstruction time) is attainable with an enhanced POAC system.

4 Possible image processing applications

The POAC can have extremely wide field of applications. It will be superior to other types of computers in cases where large images, data matrices have to be processed at high speed. Among these, image understanding and processing seems to be the most fitting tasks to the POAC capabilities.

In our work we were looking for image processing tasks where the application of a high-speed parallel optical computer has great advantages and fast (re) programmability is also required. Optical computer implementations are especially superior to digital signal processing counterparts in those tasks, where great amount of correlation has to be completed with large templates on large images.

The spatial frequency filtering based VanderLugt type, correlators are fast, due to their inherent massive parallelism. However, because of the time consuming process of complex optical filter (matched filter) design and their slow display, they are not applicable when the task requires fast (re-) programming. As there is no need of any matched filter calculation in our advanced t_2 -JTC architecture because we display spatial domain templates, we can apply the data gained just from the previous image frames. This way a certain kind of adaptivity can be introduced. Although some preprocessing is always required, the parallel CNN-UM VLSI chip, which is an integral part of the POAC system, can fulfill these tasks efficiently.

4.1 Adaptive target detection and tracking

Correlation based recognition and tracking of targets is required in many applications, because alternative tracking methods (based on other special characteristics like heat emission, color etc.) are relatively easy to decoy. However, distance, coverage and view of the targets alter in a hectic way during the tracking procedure. Therefore, to accomplish a target-tracking task an enormous library of templates is required in all of the previously known correlation-based implementations. As large numbers of templates have to be correlated on all image frames, these non-adaptive algorithms become too slow for any practical application.

Within the POAC framework we can implement an adaptive target-tracking algorithm. Here we consider only the adaptive detection of the target and do not concern any type of position approximation. In this algorithm we assume that the target image changes continuously between the consecutive image frames. So a target image section of the previous frame can be used in the algorithm as template (reference object) besides a standard template library. Applying our

adaptive target-detection and tracking algorithm we can decrease the number of the necessary templates and so the speed of the algorithm increases considerably. Other, more refined tracking procedures can use the position data provided by our algorithm.

Digital implementations should apply the same algorithm. Computing correlation, however, even if a DSP system applying fast Fourier transformation-based algorithms requires considerable computer power. Therefore, due to the limitations of the currently available DSP systems, digital implementations are not appropriate. It is especially the case in high-speed applications.

Although, other optical correlator architectures (like VanderLugt correlators) can have the required speed, but they would require an appropriate set of adaptive Fourier plane filter (matched filter). Design of these "matched" filters can require even more computation than the direct determination of the correlation. Furthermore, conventional JTC systems' speed is limited by the pace of hologram recording. Our t_2 optical correlator architecture overcomes this limitation.

As the whole tracking algorithm is fulfilled within the POAC framework some additional, more complicated algorithmic steps (extraction of local image features to find the optimal templates, motion estimations etc.) can be implemented by the in-built CNN-UM chip. Necessary pre-processing steps, like high pass filtering (with antagonistic center-surround B template operation), can be executed also by the VLSI CNN-UM chip. This way a POAC system provides an extremely efficient frame for this application.

4.2 Stereo image processing algorithm

To obtain depth information from stereo image pairs it is required to find and compare the exact position of the according image pixels. Usually, a correlation-based algorithm accomplishes this task. A small neighborhood of each pixel of the "right" image is correlated onto the "left" image. The correlation peak indicates the position of the corresponding object in the left image and so defines its disparity.

Several digital signal processing-based implementations exist, but due the huge amount of the necessary correlation measurements these are slow. Although, displacements of the according pixels are relatively small and usually confined to one dimension, due to the pixel misalignments, 2D-correlation determination seems to be still advantageous. Other optical correlator (VLC) architectures are not adequate for this algorithm as it require extremely fast change of the templates. Furthermore, it is more efficient to implement this algorithm in our optical correlator than in other JTC systems, as the pace of the hologram recording does not limit its speed.

Our programmable optical correlator can fulfill this task easily. The required neighborhood does not exceed our POAC system limits. Usually some pre and post processing steps are also required but they can be fulfilled within a POAC system. Although the correlation determination step is only one (early) step of a depth identification algorithm it can provide the necessary data at a high speed for further processing. As the digital (manual) stereo image processing

algorithms are slow, the application of a fast optical correlator can speed up notably the 3D calculations.

4.3 Motion estimation algorithm for image compression

In motion image compression algorithms the most computation is usually consumed when the procedure tries to find the relative position of the according image blocks in the consecutive image frames. The available computing power usually delimits the block size and the possible search range in digital signal processing implementations.

Our optical correlator based CNN implementation provides an efficient frame to solve this part of the algorithm. As fast change and reprogramming of templates (image blocks) is required, a t_2 -JTC type joint Fourier correlator setup seems to be the adequate optical implementation. Our correlator can find the relative position of blocks of the previous (consecutive) frame from the current positions (distance) and provides also a similarity measure. A VLSI CNN-UM chip can solve the required further image segmentation steps and so we can apply all components of the POAC system efficiently.

4.4 Identification systems

We can apply our optical CNN implementation efficiently in all the applications where great amount of fast correlation, pattern-matching operation is required. It is the case for example in palm, fingerprint, and face recognition, or iris identification algorithms. Usually these algorithms are based on the determination of a huge set of local features (image primitives) and their relative position (distances).

Our opto-electronic device can help both in preprocessing of the images and in the fast size- and rotation-invariant identifications of the image primitives. In the majority of recognition tasks high overall speed is essential. As our POAC system can provide the required data at much higher pace than any competitive DSP technology, the application of this device seems to be advantageous. The size range of these applications' usual input images (FBI standard) fits to that of the presently realizable POAC system. Even the available larger template size of the POAC suits to the requirements of these applications. Although other optical correlator architectures (VLC, JTC) are also applicable, because there is no real need of fast programming (a fix set of primitives are required in these applications), our architecture is simpler and much more flexible. The proposed POAC ensures extremely high frame rate for small templates and it is not limited by the hologram recording speed (in contrast to conventional JTC systems).

5 The Adaptive Target Detection and Tracking Algorithm

In moving image sequences a target appears to be extremely variable, due to the frantically changing viewing angle, distance, illumination and coverage. However,

the target usually shows up relatively small changes between consecutive image frames. Based on this feature we constructed and tested a simple target-tracking algorithm, which has been efficiently implemented in our POAC system. In the case of large displacements of the targets, fast movements, it is more efficient than any other optical or DSP implementation.

This algorithm is based on our programmable optical correlator that determines peak positions in the correlation of the appropriate template on the input image frame. The template is the high-pass filtered and windowed portion of the target object that has been found in the previous frame. The (global or local) maxima of this correlation define the position of the targets on this frame. According to the found location of the correlation peak, we extract a new template² what the algorithm will use in the next frame or frames. This way we can track the target adaptively. The sequence of the gained templates might be used as an input of a further refined recognition algorithm.

We have made some numeric simulations to explore the robustness and possible deficiencies of this simple algorithm. It has been found that although this algorithm works efficiently, we have to be cautious about the used input image sequences. In a few test sequence our algorithm missed the target just because of an input image sequence error. That is, the tracked part of the target got outside of the image frame. It is not the error of our adaptive algorithm, but caused by the erroneous manual tracking of the target. If we apply several templates, overlapping on the same target we can decrease the probability of this type of error. Additionally, relative positions of these templates, defining different sections of the target, can provide data for algorithms which aims to define the target movement (rotation, time to collision etc.).

On the other hand, our algorithm has to ensure that the auto-correlation functions of the adaptively changing templates have distinct peaks. Otherwise, it can miss the targets. A template does not have a distinct auto-correlation peak, when it is not a real 2D image, but instead, a one-dimensional vector or a scalar value can fully describe it. If the template changes (quasi) along only one-dimension, the correlogram has triangular shape (perpendicular to this dimension) and gives not a real 2D peak. So the noise and alteration of the target — as we are measuring cross correlation — can cause misplaced detection of the correlation peak. Fortunately, this type of error usually does not cause the overall miss of the target.

5.1 Simulation and measured results

Simulation demonstrates that, in spite of the considerable change of the targets viewing angle, our algorithm can track the target. If the same template was applied all along the whole simulation, after a short transient interval, the

² The extraction of the new template is extremely fast because its size is relatively small (32x32). To cut-out and download the required part of the input image can be fulfilled within 50 μ sec. All the required further processing of the templates (windowing) can be accomplished well within the time-window of the template operations.

correlation based nonadaptive algorithm lost the target. Comparing the correlation peak amplitudes of adaptively updated and non-adaptive templates we demonstrated the effectiveness of our algorithm.

A simplified version of the adaptive target detection algorithm is already implemented on the experimental POAC system. In spite of the limitations of the current setup we can demonstrate that a POAC system can fulfill the basic operations of the algorithm (all of these limitations can be (will be) overcome by building more advanced elements into the system). The input is a test image sequence of a moving and rotating airplane. For the sake of demonstration in this printed document, the time sequence of the airplane images is displayed simultaneously in the input plane of the POAC. Positions of the correlation peaks show that the translating and rotating aircraft was tracked along its path.

The POAC system is able to find and track up to sixteen targets (with 32×32 size templates) in every frame (800×600) of a 60-frames/sec video sequence. By use of more enhanced building elements (and/or by the application of smaller templates) even this performance can be improved considerably (input image: size 1000×1000 pixels, 60-180 frame/second; template size 32×32 gray-scale; correlation time 1ms.). In Fig 3 we can compare the digitally simulated and measured tracking algorithm.

6 Conclusion

We proved that the POAC is a fruitful and promising concept and our optical correlator based architecture is new and innovative. Due to its massive parallelism this solution provides a new algorithmic frame and produces extremely high-speed operations for image processing, target detection and tracking tasks. Appropriate key elements and devices are required to reach a final product stage.

We have developed the appropriate architecture and working principles, but further improvements of the key-elements and building blocks are still necessary to achieve all the inherent possibilities of the POAC. These include the application of the followings. High power, actively Q-switched lasers (frequency doubled YAG lasers). An enhanced 2D AOD or a laser diode (VCSEL) matrix for template input. Improved LCD or OASLM devices for input. A high-speed CMOS detector matrix or/and visual CNN-UM chip for detection and processing of the output. A CNN-UM chip with optical input and output. An enhanced dynamic holographic material (e.g. bacteriorhodopsin) with improved parameters. The size of the system can be finally decreased to 30 by 40 by 20 cm. This way a POAC will become a commercially successful device. The performance of its kernel processor, the semi-incoherent optical correlator; is superior to any known optical correlators.

We demonstrated that there are image processing tasks which can not be fulfilled in real time with any contemporary technology but with the computational power of our opto-electronic device. Our POAC system can be efficiently used in adaptive target tracking tasks. Furthermore, we propose the utilization of our POAC system for iris, fingerprint and face recognition algorithms as well, as an

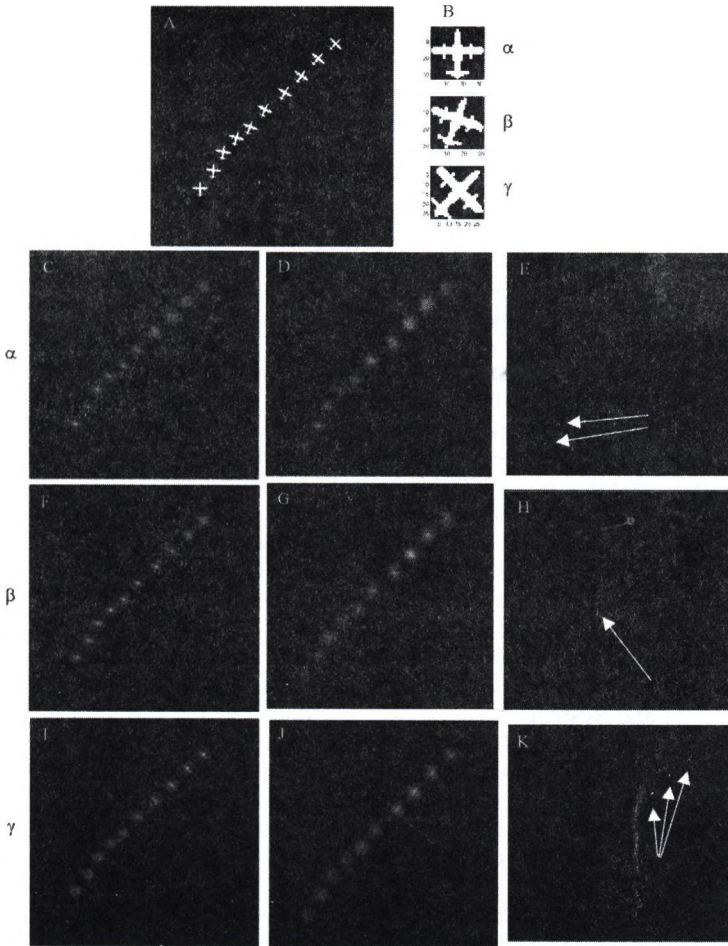


Fig. 3. Demonstration of the POAC system's applicability for target detection and tracking tasks. Here we show the input image (500x500) (A) and a set of the applied templates (B). You can compare the numerically computed (C, F, I) and the optically measured (D, G, J) correlations. After applying an appropriate edge enhancing filter and a threshold, tracking detection is possible (E, H, K). (White arrows indicate the measured correlation peaks)

extremely fast and rapidly reprogrammable image primitive extractor unit. It can be also useful in image segmentation algorithms, which are based on the identification of relatively big texture elements and in motion picture compression. However, we are still looking for further key applications for our new, powerful hardware.

References

1. Javidi, B., Horner, J.L.: Real-time optical information processing. Academic Press, Inc. London. (1994)
2. Yu, F.T S., Jutamulia, S.: Optical Pattern Recognition, Cambridge University Press. (1998)
3. Alam M.S.: Selected Papers on Optical Pattern Recognition Using Joint Transform Correlation, SPIE Milestone Series Vol. MS 157, SPIE-The International Society for Optical Engineering, Bellingham, Washington USA, (1999)
4. Tanida, J., Ichioka, Y.: Digital Optical Computing, (in ed. Wolf, E.: Progress in Optics, Vol. XL. (2000)
5. VanderLugt, A.: Optical Signal Processing, John Wiley & Sons, Inc., (1992)
6. Chua, L.O., Yang, L.: Cellular Neural Networks: Theory and Applications. IEEE Transactions on Circuits and Systems. 35. (1988) 1257-1290
7. Chua, L.O., Roska, T.: The CNN Paradigm, IEEE Trans. on Circuits and Systems. 40. (1993) 147-156
8. Roska, T., Chua, L.O.: The CNN Universal Machine: An Analogic Array Computer. IEEE Transactions on Circuits and Systems-II, 40. (1993) 163-173
9. Roska, T.: Computer-Sensors: Spatio-temporal Computers for Analog Array Signals, Dynamically Integrated with Sensors. Journal of VLSI Signal Processing Systems. 23(2/3). (1999) 221-238
10. Tökés, Sz., Orzó, L., Rekeczky, Cs., Roska, T., Zarándy, Á.: An optical CNN implementation with stored programmability", Proc. IEEE International Symposium on Circuits and Systems, Geneva. Vol. 2. (2000) 136-139
11. Cruz, J.M., Chua, L.O., Roska, T.: A Fast, Complex and Efficient Test Implementation of the CNN Universal Machine. in Proc. of the third IEEE Int. Workshop on Cellular Neural Networks and their Application (CNNA-94), Rome, (1994) 61-66
12. Paasio, A., Kananen, A., Halonen, K. and Porra, V.: A QCIF Resolution Binary I/O CNN-UM Chip. Journal of VLSI Signal Processing Systems, 23.(2/3). (1999) 281-290
13. Roska, T., Zarándy, Á., Zöld, S., Földesy, P., Szolgay, P.: The Computational Infrastructure of Analogic CNN Computing - Part I: The CNN-UM Chip Prototyping System. IEEE Transactions on Circuits and Systems. 46(2). (1999) 261-268
14. Espejo, S., Domínguez-Castro, R., Linán, G., Rodríguez-Vázquez, A.: A 64x64 CNN Universal Chip with Analog and Digital I/O. in Proc. of 5th Int. Conference on Electronics, Circuits and Systems (ICECS'98), Lisbon, September (1998) 203-206
15. Vsevolodov, N.N., Dyukova, T.V.: Retinal-protein complexes as optoelectronic components. Trends in Biotechnology. 12. (1994) 81-88
16. Tökés, Sz., Orzó, L., Váró, Gy., Dér, A., Ormos, P. and Roska, T.: Programmable Analogic Cellular Optical Computer using Bacteriorhodopsine as Analog Rewritable Image Memory, NATO Book: "Bioelectronic Applications of Photochromic Pigments" (Eds.: L. Keszthelyi, J. Stuart and A. Der) IOS Press, Amsterdam, Netherlands (2001) 54-73

17. Rekeczky, Cs., Roska, T., Ushida, A.: CNN-based Difference-controlled Adaptive Nonlinear Image Filters. *International Journal of Circuit Theory and Applications*. 26. (1998) 375- 423
18. Rekeczky, Cs., Chua, L.O.: Computing with Front Propagation: Active Contour and Skeleton Models in Continuous-time CNN. *Journal of VLSI Signal Processing Systems*. 23(2/3). (1999) 373-402
19. Szatmári, I., Rekeczky Cs., Roska, T.: A Nonlinear Wave Metric and its CNN Implementation for Object Classification. *Journal of VLSI Signal Processing Systems*. 23(2/3). (1999) 437-448
20. Slot, K., Roska, T., Chua, L.O.: Optically Realized Feedforward Only Cellular Neural Networks. *International Journal of Electronics and Communications*. 46(3). (1992) 158-167
21. Fruehauf, N., Lueder, E., Bader, G.: Fourier Optical Realization of Cellular Neural Networks. *IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing*. 40(3). (1993) 156-162
22. Jankowski, S., Buczynski, R., Wielgus, A., Pleskacz, W., Szoplik, T., Veretennicoff, I., Thienpont, H.: Digital CNN with Optical and Electronic Processing. in *Proc. of 14 European Conference on Circuit Theory and Design*, Stresa, Italy, Vol. 2. (1999) 1183-1186
23. Tőkés, Sz., Orzó, L., Török, L., Ayoub, A., Roska, T.: An advanced joint Fourier transform correlator (JTC). *Proceedings of DO'01 Topical Meeting on DIFFRACTIVE OPTICS 2001*, Budapest, (2001)
24. Tőkés, Sz., Orzó L. and Roska T.: Design Aspects of an Optical Correlator Based CNN Implementation. *Proc. IEEE International Symposium on Circuits and Systems*, Helsinki, Vol. 1. (2001) 353-356

Képfeldolgozás oktatása a mérnökképzésben (oktatói és hallgatói vélemények)

Berke József¹, Hegedűs Károly²

Veszprémi Egyetem, Georgikon, Mezőgazdaságtudományi Kar,
¹Gazdaságmatematika, Statisztika és Informatika Tanszék, Keszthely
berke@georgikon.hu
www.georgikon.hu/digkep.htm

²Oktatási és Akkreditációs Csoport, Georgikon Hallgatói Egyesület, Keszthely
hegedusk@georgikon.hu

Abstract. Az információs társadalom megvalósítása során fontos szerep jut a vizuális információknak tömörsége és függetlensége miatt. A vizuális információ megszerzése és feldolgozása során kiemelt jelentőséggel bír a képfeldolgozás. Cikkünkben bemutatjuk a VE Georgikon karon folyó képfeldolgozás oktatás célját, szerepét és az oktatás megvalósításának néhány mozzanatát. Ismertetésre kerül a képfeldolgozás alapú tantárgyak rövid tartalma. A cikket igazán az teszi érdekessé, hogy mindezt az oktató és a hallgatók szemszögéből is megismerhetjük. Független (egyetemi, főiskolai) felmérések, oktatói és hallgatói vélemények alapján az egyes kurzusok értékelésébe is bepillantást kapunk.

1. Bevezetés

Egyre több egyetemen és főiskolán kerül be a számítógépes képfeldolgozás tantárgy a hallgatók által választható tárgyak közé. Így a Veszprémi Egyetem Georgikon Mezőgazdaságtudományi Karán is már több mint 10 éve választhatják ezt a tárgyat a hallgatók. A továbbiakban a tananyag felépítését, az oktatási anyagot és az oktatással kapcsolatos tapasztalatokat ismertetjük, oktatói és hallgatói vélemények alapján, összehasonlítva a véleményeket a Gábor Dénes Főiskolán készült oktatói és hallgatói értékelésekkel.

2. Háttér

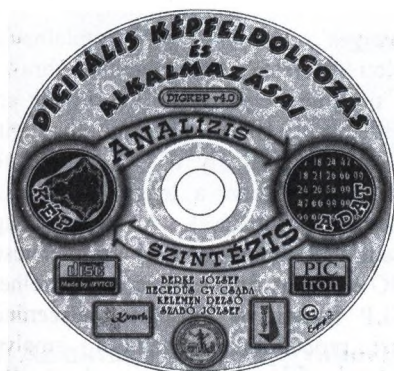
A „Digitális képfeldolgozás” tantárgyat a hallgatók fakultatív tárgyként választhatják. A tananyag elsajátítása elsősorban a kontakt órákon zajlik, amelyek 14 héten keresztül tartanak, heti 2 órás (2 x 40 perc) elfoglaltságot jelentve a hallgatóknak. Ebben természetesen nincs benne a házi feladatok, beadandó anyagok elkészítésének ideje. Az oktatás során főleg az alább részletesebben ismertetett tankönyv ismereteire

támaszkodik az oktató, valamint a tankönyv mellékleteként szereplő CD-n található információkra (1. ábra). A hallgatók a CD segítségével tudják elkészíteni házi feladataikat, beadandó anyagaikat.

3. A tananyag felépítése

Bár két magyar nyelvű bevezető jellegű könyv is létezik [1, 2], de ezek inkább egyetemi posztgraduális hallgatók részére készültek, így meglehetősen részletes matematikai összefoglalót tartalmaznak az elméleti részek megértéséhez. A Gábor Dénes Főiskolán jól meghatározott és azonos matematikai alappal rendelkező hallgatók részére kellett egy tankönyvet készíteni, amely távoktatásból eredő speciális elvárásoknak is megfelel. További igény volt, hogy a képfeldolgozás gyakorlati alkalmazásáról is adjon a könyv áttekintést, hogy a hallgatók, akik többnyire különböző területeken dolgoznak, saját szakterületükhöz megfelelő kapcsolódási pontokat találjanak. A felmerült igények miatt új tankönyv készült [3, 9, 1.ábra]. A tananyag, illetve a tankönyv készítésénél a korábban megfogalmazott célkitűzésekből [4] indultunk ki, amely a „TANKÉP” rendszer egy közelítésének is tekinthető. Az elkészített tananyag ugyanis nem multimédia megoldásban, hanem hagyományos könyv formában jelent meg, de a könyvhöz csatolásra került egy program, amely az otthoni gyakorlást és ezzel az ismeretek mélyebb megértését tette lehetővé.

A legújabb oktatási tananyag [9] a felhasznált matematikai fogalmakat vagy az előzetes stúdiumok alapján ismertnek tételezi fel, vagy ha szükséges az első előfordulási helyen röviden ismerteti. Így matematikai alapok áttekintését tartalmazó fejezetet nem tartalmaz a könyv. Miután bevezető jellegű könyvről van szó - nem elméleti szakembereknek szól - a felhasznált matematikai ismeretekben igyekeztünk mértékletességet tanúsítani és a felmerülő fogalmakat szemléletes példákon keresztül megvilágítani.



1. ábra „Digitális képfeldolgozás és alkalmazásai” oktatási anyag

Az alapszintű ismeretanyagot nyolc fejezetben foglaltuk össze, ezek a következők:

1. AZ EMBERI LÁTÁS
2. A KÉPFELDOLGOZÁS ESZKÖZEI
3. DIGITÁLIS KÉPALKOTÁS
4. KÉPJAVÍTÁS
5. GEOMETRIAI KORREKCIÓ
6. SZEGMENTÁLÁS
7. OSZTÁLYOZÁS
8. KÉPKÓDOLÁS és TÖMÖRÍTÉS.

A következő fejezetek különböző alkalmazási területeket ismertetnek. Ezekben a fejezetekben általában először az adott alkalmazási terület jellegzetességeit emeltük ki és foglaltuk össze, majd néhány konkrét alkalmazás ismertetésével kíséreltük meg érzékeltetni a felmerülő problémákat és a megoldási variációkat.

Az alkalmazásokkal foglalkozó fejezetek:

9. TÁVÉRZÉKELÉS
10. ORVOSI ALKALMAZÁSOK
11. IPARI KÉPFELDOLGOZÓ RENDSZEREK
12. MEZŐGAZDASÁGI ALKALMAZÁSOK
13. IRODAI ALKALMAZÁS – OCR
14. BŰNÜGY- ÉS BIZTONSÁGTECHNIKA
15. DIGITÁLIS FÉNYKÉPEZÉS ÉS SZKENNELÉS

Ezek a fejezetek, szándékaink ellenére nem fedik le az összes lehetséges alkalmazási területet és nem teljesen azonos felépítésűek. Ennek magyarázata, hogy a szerzők által művelt, így jobban ismert területek óhatatlanul nagyobb súllyal szerepelnek, valamint az, hogy a könyv megírására rendelkezésre álló idő nem tette lehetővé, hogy az alkalmazási területek módszeres irodalmi feldolgozásával egy rendszerezettebb áttekintés készüljön. [11]

Az utolsó fejezet pedig 16. VIZUÁLIS ADATFORMÁTUMOK egy összefoglaló áttekintést ad az ismert és a gyakorlatban elterjedt vizuális adatformátumokról.

Az egyes fejezetek után ellenőrző kérdések találhatóak, ráirányítva a figyelmet a fejezetekben található ismeretek súlyponti témáira, ezáltal is elősegítve az önálló felkészülést.

A könyv számos gyakorlati példát tartalmaz, melyek a tananyagba építve találhatóak meg, közvetlenül a feladathoz kapcsolódó elméleti ismeretek közlése után (2. ábra).

A könyv feladatait a CD mellékleten található program segítségével készíthetik el a hallgatók. A program a PICTRON Kft által fejlesztett TULIP (TUtoriaL for digital Image Processing) [5, 6.] program oktatási változata, amely a CD mellékleten (DIGKEP v4.0) található közel 1000 képpel lehetővé teszi a leggyakoribb és a tankönyvben szereplő képfeldolgozási funkciók kipróbálását. A képeket úgy válogattuk össze, hogy az egyes funkciók hatása eredményesen és szemléletesen jelenjen meg a hallgatók számára. Bár WINDOWS alatt futó egyszerűen kezelhető programról van szó, és a használatot a HELP funkció is elősegíti, célszerűnek tartottuk az egyes fejezetekhez feldolgozási procedúrák megadását, amelyek alátámasztják a könyvben leírtakat. Ezek a feladatok tehát feldolgozási szekvenciákat adnak meg és a javasolt mintaképek neveit amelyekkel egy-egy képfeldolgozási fogalom vagy funkció jól demonstrálható (2. ábra).

Gyakorlatok a világoasságkód-transzformációkhoz

6. Gyakorlat:

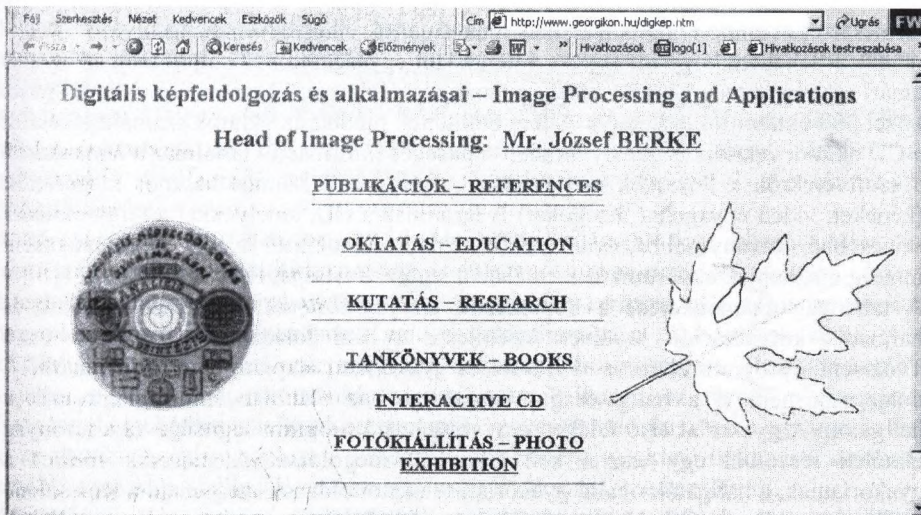
Hisztogram Indít Szín fekete/fehér Képforrás Tesztkép Kép2 Számol

Ismételjük meg a gyakorlatot úgy, hogy először egy kis részletet (20*20 pontnál kisebb nem lehet) jelölünk ki a háttérben, majd a munkadarabon! Ismételjük meg a gyakorlatot a Kép3, Kép4, Kép6 és Kép1 képekkel! Ne felejtjük a Kép1 esetén a színt RGB-re állítani!

2. ábra Képfeldolgozás gyakorlatok egy példája

A legújabb tankönyv, valamint a hozzá kapcsolódó CD melléklet kiegészül internetes támogatással is, amely az alábbi Internet címen érhető el: <http://www.georgikon.hu/digkep.htm>. Ezen a webcímen a tananyagon kívül megtalálhatóak a témában készült szerzői publikációk, kutatási projektek listája, interaktív CD bemutató, virtuális fotókiállítás, védett hallgatói segédanyagok, valamint számos érdekesség (3. ábra).

A tankönyv ezen kívül kiegészül a további tájékozódást elősegítő irodalomjegyzékkel, valamint WWW címlistával.



3. ábra <http://www.georgikon.hu/digkep.htm>

4. Az oktatás menete, gyakorlati lebonyolítás

A „Digitális képfeldolgozás” című tantárgyat minden páros félévben, míg a „Távérzékelés” tantárgyat minden páratlan félévben hirdeti meg a Gazdaságmatematika, Statisztika és Informatika Tanszék a Georgikon Karon. A tárgyakat félévente közel 50 fő veszi fel és hallgatja le. A tárgyak 14 hetes bontásban

heti 2 órában kerülnek oktatásra. Az órák szerkezete minden óra esetében közel azonos. Az óra első részében elméleti áttekintés, vagy szoftverismertetés folyik az anyagnak megfelelően. Ez után gyakorlati feladatok megoldása következik, vagy az ismertetett szoftver gyakorlati alkalmazása. Végül minden alkalommal mintegy fél óra áll rendelkezésre a diákoknak egyéni feladatok megoldására, vagy a szükséges anyagok letöltésére az Internetről. A tananyag számonkérése a beadandó feladatok útján valósult meg. A hallgatóknak év közben négy-öt egyéni, illetve páros feladatot kellett elkészíteni, és az oktatóhoz eljuttatni. A tantárgyak (változtatások melletti) oktatása nemcsak a mérnökképzésben folyik, hanem a Kar felsőfokú szakképzéseinek keretében is (pl.: Multimédia fejlesztő, Oktatásinformatikus, stb.), valamint a Georgikon Kar Akkreditált Iskolarendszerű Felsőfokú Szakképzésében (AIFSZ) is. A Gábor Dénes Főiskolán a tantárgyat mint kötelező tárgyat közel 6000 hallgató vette fel három országban, magyar anyanyelven. A korábbiakban ismertetett tankönyv és program mellett egy részletes tantárgyi útmutató, internetes támogatás, videó oktatási anyag és házi feladatok formájában ellenőrző kérdések is segítik a hallgatók felkészülését.

A tantárgy több, korábban oktatott informatikai tárgyhoz is szervesen kapcsolódott (pl. Matematika, Számítógépes grafika, Multimédia eszközök és szoftver, Kommunikációs eszközök, stb.). Két szakirányon is kötelező tárgyként szerepel, egyrészt a számítógép-alkalmazási szakirányon, valamint a multimédia szakirányon, természetesen eltérő tematikával, más-más ismertekre helyezve a hangsúlyt.

Annak érdekében, hogy országos szinten közel egységesen valósuljon meg az oktatás tanári egységcsomag készült, amely a tananyag vázlatos bemutatását is tartalmazta közel 200 oktatói fóliavázlaton. A fent említettek mellett az oktatók számára elkészült a CD oktatói verziója is, amely minden szükséges információt tartalmaz a tematikáról, a szoftvekről, a feladatok megoldásáról. Ezen kívül számos hasznos kiegészítést (képeket, videó anyagokat, hangokat) is tartalmaz a CD, amelyekkel a tárgy oktatása színesebbé, látványosabbá, érdekesebbé tehető. Az oktatók a CD folyamánként mindig megkapják elektronikus formában a legújabb frissítéseket, verziókat. [13]

A tananyag számonkérése a következők alapján történik. Az első, második és harmadik konzultáción kerülnek kijelölésre a házi feladatok a tankönyv és a Tanársegéd program ellenőrző kérdései és gyakorlatai alapján. Nyomtatott zárthelyi dolgozatot nem ír a hallgató. Vizsgajegy az alábbiak szerint kaphatnak a hallgatók: A gyakorlat első felében egy vizsgáztató program segítségével a tananyag elméleti részeiből egy tesztet kell sikeresen megoldani. A második felében a gyakorlatnak, a hallgatók önálló gyakorlati feladatot oldanak meg, majd a kapcsolódó kérdésekre szóban válaszolnak. A feladat megoldásához a tankönyv és a mellékelt CD lemez is használható. A vizsgajegy csak akkor ajánlható meg, ha a hallgató az elméleti részt és a gyakorlóprogrammal elvégzendő gyakorlati feladatot is legalább elégséges szinten teljesítette.

A „digitális képfeldolgozás és alkalmazásai” tantárgy oktatása konzultációk és gyakorlati foglalkozások keretében történt, távoktatási elemeket tartalmazó módszerek alapján. A konzultáción a tankönyv ismeretanyagának feldolgozása történt, előadások formájában. Az előadásokat néhány alkalmazást bemutató videofilm, valamint a TULIP programmal történő feldolgozási bemutatók tették változatosabbá. Természetesen az előadások kapcsán az igények szerinti tényleges konzultációra is sor került. A gyakorlati foglalkozáson a hallgatók tanári segítséggel

dolgozhattak több képfeldolgozási programmal is. A Silicon Graphics hazai képviselőjének (Silicon Graphics Hungary Ltd.) és forgalmazójának (Silicon Computers Hungary Kft.) támogatásával a tantárgy zárásaként bemutatóval és gyakorlati lehetőséggel egybekötött konzultáción ismerkedhetett meg a hallgatóság a vizuális informatika éllavasának [8] korszerű hardver és szoftver termékeivel.

5. Tapasztalatok, vélemények

Az oktatási tapasztalatok főleg a hallgatóság által kitöltött kérdőívekből, valamint azok kiértékeléséből származnak. E mellett egyéni vélemények is megfogalmazódtak mind a hallgatóság, mind az oktatók részéről. [12]

A Veszprémi Egyetem Georgikon Mezőgazdaságtudományi Karán a hallgatói visszajelzésekből főképp az szűrhető le, hogy a tárgyat választó diákok hasznosnak ítélték az elméleti alapokra épülő gyakorlatorientált képzést. A pozitív tapasztalatokat az is hitelesen támasztja alá, hogy folyamatosan emelkedik azon hallgatók száma, akiknek diplomadolgozatában megjelennek a képfeldolgozás gyakorlati alkalmazásai.

A Veszprémi Egyetemen a hallgatói véleménynyilvánítás egyik lényeges eszköze az Oktatói Munka Hallgatói Véleményezése (OHV) elnevezésű kérdőív. Ez által a hallgatóknak lehetőségük nyílik arra, hogy véleményükkel minősítsék oktatóik munkáját, kifejezzék meglátásaikat, esetlegesen javaslatokat tegyenek bizonyos változtatásokra.

A kérdőívet a Veszprémi Egyetem Hallgatói Önkormányzata állította össze. Ezt minden oktatónak ki kell töltenie minden egyes oktatott tárgy esetében a hallgatókkal. A válaszadás teljesen anonim, a hallgatónak nem kell a nevét ráírnia az értékelőlapra.

A kérdőíven 13 kérdés szerepel, minden kérdésre egytől ötig adható válasz, a megfelelő négyzet bejelölésével. Az egyes válasz a legelmarasztalóbb, az ötös válasz a leginkább dicsérő vélemény. Nem kötelező minden kérdésre válaszolni, az üresen hagyott kérdések nem számítanak bele a véleményezésbe. A kérdések zöme (11 kérdés) az oktatóra vonatkozik, hangsúlyt helyezve felkészültségére, „hallgató-barát” voltára, előadási, órátartási szokásaira, stílusára, hallgatókhoz fűződő viszonyára. Ezen kívül egy kérdés a hallgató óralátogatási szokásait méri fel, illetve egy kérdés tér ki az oktatási segédanyagokra. A kérdőíven egy kérdést az oktató is feltehet, amelyet később visszakap a hallgatók válaszaival, amit leválasztanak a kérdőívről.

A kérdőívek feldolgozása a következők szerint történik. A feldolgozás során első lépésként meghatározzák, hogy melyik kérdőív értékelhető. Ennek eldöntése az első kérdés alapján történik. Aki az első kérdésre az egyes, vagy a kettes választ jelölte meg, az csupán az órák olyan csekély részén vett részt, hogy a véleménye nem lehet teljesen megbízható. Ezen kérdőívek nem kerülnek kiértékelésre. Az értékelhető kérdőíveket ezután kérdésenként dolgozzák fel. A kérdésekre adott válaszok pontszámát kérdésenként összeadják, majd elosztják az értékelhető kérdőívek számával. Miután minden kérdést kiértékeltek az átlagok számtani átlagát veszik, ami az oktató végleges minősítését adja. Ha ez az átlag két féléven keresztül alacsony (kettő egész körüli), akkor az oktatót figyelmeztetésben részesíthetik.

Az elmúlt félévben kitöltött kérdőívek kiértékelése után a következő megállapítások tehetőek. Az értékelés során az oktató, valamint ezáltal a tárgy, jó minősítést kapott, a diákok hasznosnak ítélték a tárgyat, az oktatót felkészültnek, azonban az oktató előadási szokásaival nem voltak teljesen elégedettek.

Itt azonban azt is meg kell említenünk, hogy a kérdőív összeállítása meglehetősen egyoldalú, túlzottan az oktatóra van kielevezve [14]. Hiányoznak az oktatási segédanyagokat minősítő, a tankönyvet értékelő, esetleg a vizsgáztatást megítélő kérdések. Ezen kívül a kérdések közül sok szubjektív elemet tartalmaz, valamint nagyban függ a hallgató óralátogatási szokásaitól.

A Gábor Dénes Főiskolán a hallgatók visszajelzéseit összegezve megállapítható, hogy azon hallgatók akik rendszeresen dolgoznak, vagy részt vesznek grafikai, képfeldolgozási vagy videotechnikai jellegű alkalmazásokban, fejlesztésekben kimondottan hasznosnak ítélték az elméleti megalapozást és az egyes eljárások gyakorlásának lehetőségét. Meglepő volt, hogy ezen hallgatók száma több mint 15 %-a volt a teljes létszámnak.

A csupán „érdeklődő” hallgatók elsősorban az egyes alkalmazásokat - szakirányhoz illeszkedően - tartották kiemelten fontosnak.

Az előadásokat és a gyakorlatot meglepően magas számú (kitartó) hallgatóság látogatta. Az oktatásban részt vevő kollégák elismeréssel és örömmel nyilatkoztak a tantárgy felépítéséről, szerepéről és a segédletekről. Természetesen számos az oktatást és a tananyag elsajátítását tovább segítő javaslat is elhangzott, mind a hallgatóság, mind az oktatók részéről.

A tárgy népszerűségét mutatja, hogy évente egyre több TDK munka témáját a képfeldolgozás alkalmazásai adják, évente közel 10 ilyen dolgozat születik. E mellett minden évben három-négy diplomamunka is készül ilyen témában.

A GDF hallgatók által kitöltetett kérdőívet a Gábor Dénes Főiskola állította össze. Fő célja az volt, hogy lemérje a hallgatók elégedettségét, valamint felmérje, hogy az oktatási anyagok mely része szorul korrekcióra, tökéletesítésre. A kérdőív 8 kérdést tartalmazott – több alponttal –, melyekre egytől tízig terjedő skálán elhelyezett választ kellett adni. Ebben az esetben az egyes a legrosszabb, a tízes pedig a legjobb véleményt jelentette.

A kérdőívek feldolgozása során az egyes kérdésekre, alkérdésekre adott pontszámokat összeadták, majd elosztották a válaszadók számával, és így egy átlagot képeztek.

Az előző évi kérések kiértékelése során a következő eredmények születtek. A hallgatók legkevésbé a tantárgyi útmutatóval voltak megelégedve, és mindannyian az oktató munkáját, oktatási stílusát tartották a legjobbnak. Az összes többi segédanyag (könyv, oktatócsomag), illetve a foglalkozások (előadás, gyakorlat, konzultáció, vizsga) jó minősítést kaptak.

Ki kell emelnünk azt is, hogy a Főiskolán kitöltetett kérdőív sokkal objektívebb képet ad a tárgy oktatásáról, mint az egyetemi kérdőív. Hiszen a fent említett nyomtatvány kitöltése során a hallgatók nemcsak az oktatót minősítették, hanem a kiadott segédanyagokat is, valamint a foglalkozásokat is.

Irodalomjegyzék

- [1] ÁLLÓ, G - HEGEDŰS, GY. CS. - KELEMEN, D. - SZABÓ, J. (1989): A digitális képfeldolgozás alapproblémái. Műszaki Tudományok, Az elektronika újabb eredményei. Akadémiai kiadó, Budapest.
- [2] ÁLLÓ, G - FÖGLEIN J. -HEGEDŰS, GY. CS. - SZABÓ, J. (1993): Bevezetés a számítógépes képfeldolgozásba. BME Mérnök-továbbképző Intézet. Egyetemi jegyzet Javított kiadás.
- [3.] BERKE, J.- HEGEDŰS, GY. CS. - KELEMEN, D. - SZABÓ, J. (1996): Digitális képfeldolgozás és alkalmazásai. Keszthelyi Akadémia Alapítvány, ISBN 963 04 7466 2.
- [4] SZABÓ, J.- HEGEDŰS, GY. CS. - KELEMEN, D. - BODROGI, H. - BERKE, J. (1996): „TANKÉP” Számítógépes oktató, gyakorló és vizsgarendszer a képfeldolgozás tanításához. Informatika a felsőoktatásban '96-Networkshop'96 Konferencia kiadvány. Debreceni Universitas, 682-688.
- [5] HEGEDŰS, GY. CS. (1993): CADI: Computer Assisted educational package for Digital Image Processing, 5th International Conference CAIP'93 Budapest. Springer-Verlag, Lecture Notes in Computer Science, 719:770-774.
- [6] HEGEDŰS, GY. CS. (1995): Képfeldolgozási célú oktató-fejlesztő rendszer, Multimédia alkalmazása az oktatásban, Keszthely, 64-69.
- [7] SZABÓ, J. (1995): Vizsgáztatás multimédia eszközökkel, Multimédia alkalmazása az oktatásban, Keszthely, 82-85.
- [8] BERKE, J. - HALÁSZ, G. (1997): Vizuális informatika a XXI. században. Networkshop '97, Keszthely.
- [9] BERKE, J. - HEGEDŰS, GY. CS. - KELEMEN, D. - SZABÓ, J. (2001): Digitális képfeldolgozás és alkalmazásai. Georgikon – Szeiler – Kvarc, Keszthely - Pictron Kft., Budapest, ISBN 963 00 5744 1.
- [10] BERKE, J. – HEGEDŰS, K. - PAPP, ZS. - VARGA, S. (2001): Képfeldolgozás oktatása az agrármérnök képzésben /oktatói és hallgatói vélemények/. Média - Informatika - Kommunikáció, Veszprém, 2001. szeptember 13-15.
- [11] BERKE, J. - KÁRPÁTI, L. - GYÖRFFY, K. - FISCHL, G. (1995): Applied Image Processing as a Subject of Scientific Instruction for Hungarian and Foreign Students, A2-Science and Technology in the Alpine-Adriatic Region, 6:10-11.
- [12] BERKE, J. (1999): Vizuális informatikai tantárgyak oktatási tapasztalatai. GEORGIKON-MÉDIA '99 konferencia, Keszthely, 1999. Május 27-28., Computer Panoráma 10:99/7.
- [13] BERKE, J. – KUDEC, L. (2000): Interaktív vizuális informatikai oktatási médiumok fejlesztése, SZÁMALK MÉDIA 2000, multimédia konferencia, Budapest, 2000. Május 25-27., Computer Panoráma 11:2000/7.
- [14] BACSI, ZS. – KOVÁCS, E. (2001): Mit ér az ember, ha tanít?, Georgikon 1:2001/1., <http://www.georgikon.hu/georgikon/georg4/2.htm>.

Többdimenziós MRI képek feldolgozása

Nyúl László¹, Jayaram K. Udupa²

¹ Alkalmazott Informatikai Tanszék, Szegedi Tudományegyetem
nyul@inf.u-szeged.hu

² Medical Image Processing Group, Department of Radiology,
University of Pennsylvania, Philadelphia PA, USA
jay@mipg.upenn.edu

A többdimenziós képfeldolgozás célja, hogy a különböző modalitásokkal felvett 3D képekből az azokon ábrázolt objektumokról, szervekről mennyiségi és minőségi jellemzőket vonjunk ki, ezek elemzésével normális és beteg szövetek között különbséget tegyünk, az elváltozásokat jellemezzük és így a diagnózist és a betegellátást segítsük. A 3D képkalkító modalitások között megtaláljuk az anatómiai részleteket mutató CT, MRI és ultrahang, valamint a funkcionális jellemzőket leképező PET, SPECT, fMRI illetve doppler ultrahang képkalkító módszereket.

A mágneses rezonancián alapuló technika (MRI) az egyik közkedvelt módszer lágy szövetek (pl. az agy) vizsgálatára, mivel nagy felbontású és anatómiailag részletes képeket készít, ugyanakkor nincs kimutatható káros mellékhatása. MRI felvételi protokollok széles választéka áll rendelkezésre, melyek alkalmazásával a különböző szövetekről változatos kontrasztviszonyokkal készíthetők felvételek. A protokollok sokasága azonban egy nagyon komoly problémával is együtt jár. Az MRI-ben a képpontok intenzitás értékei nem rendelkeznek konkrét jelentéssel (ellentétben pl. a számítógépes tomográfiával (CT)), még akkor sem, ha ugyanazon a szkenneren ugyanazon protokoll szerint készítünk felvételeket ugyanazon beteg ugyanazon testrészéről. Ez komoly problémákat jelent a képek megjelenítésekor, szegmentálásakor és azok kvantitatív analízisében is. A képfeldolgozási módszerek szinte kivétel nélkül paraméterek megfelelő beállítását igénylik. Ezek megadása lényegesen nehezebb, ha a képpontok értékei nem bírnak jól meghatározott protokoll- illetve szövet-specifikus jelentéssel.

Bemutatunk egy olyan előfeldolgozó lépést, amellyel az MRI képek egyszerűen transzformálhatók úgy, hogy az eredmény képek lényegesen hasonlóbb intenzitás jelentéssel bírnak [1, 2]. Ez egy két lépéses, hisztogram alapján történő transzformáció. Az első, tanulási lépésben meghatározzuk a transzformációs paramétereket a tanulóképek hisztogramjain beazonosított pontok alapján. A második, transzformációs lépésben minden egyes kép hisztogramját transzformáljuk a megtanult paraméterek szerint. Ez az új transzformáció egy standard skálát rendel minden egyes protokollhoz és a vizsgálat tárgyát képező testrészhez. A standardizált képek megjelenítésekor előre rögzített paraméterek használhatók, nincs szükség azok képenkénti állítgatására, valamint a módszer használatával a képszegmentáló eljárások is jobban automatizálhatók.

Az előfeldolgozó eljárások közül az egyik legfontosabb, és egyben legnehezebb is, a képek szegmentálása. Ez tekinthető két egymással összefüggő feladat, a

felismerés (detection) és a körülhatárolás (delineation) együttesének. A felismerés a magasabb szintű feladat, amelyben az objektum hozzávetőleges elhelyezkedését határozzuk meg a képen. A körülhatárolás az alacsonyabb szintű feladat, amelyben az objektum pontos kiterjedését és pontonkénti átmenetes összetételét határozzuk meg. A legtöbb felismerési feladatban egy képzett szakember lényegesen jobban teljesít, mint a számítógépes algoritmusok. Ezzel szemben léteznek olyan számítógépes algoritmusok, amelyek az embernél precízebben, pontosabban és hatékonyabban elvégzik az objektumok illetve azok határainak kijelölését. A cél tehát olyan módszerek kidolgozása, melyek az ember és a számítógép erősségeit hatékonyan kombinálják.

Az utóbbi években a fuzzy összefüggőséget [3] alkalmazó szegmentálási technikát számos jelentős feladatban alkalmazták. [4, 5] Ezzel a módszerrel azonban volt egy komoly probléma, annak nagy számításgigénye. A módszer felgyorsítására új, hatékony algoritmusokat dolgoztunk ki, melyeket modern, gyors PC-ken alkalmazva, interaktív szegmentálást érhetünk el a fuzzy összefüggőség módszerét alkalmazva. [6, 7] A megbízható felismerés (melyet képzett szakértő segít) és a pontos, hatékony és összetett körülhatárolás (melyet a számítógép automatikusan elvégez) lényegében egyetlen interaktív folyamatba sűrítendő. Amennyiben olyan képeket használunk, melyek a szövetekre jellemző intenzitás értékeket tartalmaznak (pl. CT vagy standardizált MR képek), a szegmentáló eljárás paramétereinek többsége előre rögzíthető, minden köztes adat előre automatikusan kiszámítható és az interaktív lépésben már lényegesen több információt és segítséget nyújthatunk a szakértő felhasználónak. Bemutatjuk egy új, lényegesen megbízhatóbb és hatékonyabb, interaktív képszegmentáló rendszer lehetőségét.

Irodalom

1. Nyúl, L.G., Udupa, J.K.: On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine* **42** (1999) 1072–1081
2. Nyúl, L.G., Udupa, J.K., Zhang, X.: New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging* **19** (2000) 143–150
3. Udupa, J.K., Samarasekera, S.: Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. *Graphical Models and Image Processing* **58** (1996) 246–261
4. Udupa, J.K., Wei, L., Samarasekera, S., Miki, Y., van Buchem, M.A., Grossman, R.I.: Multiple sclerosis lesion quantification using fuzzy-connectedness principles. *IEEE Transactions on Medical Imaging* **16** (1997) 598–609
5. Udupa, J.K., Nyúl, L.G., Ge, Y., Grossman, R.I.: Multiprotocol MR image segmentation in multiple sclerosis: Experience with over 1000 studies. *Academic Radiology* **8** (2001) 1116–1126
6. Nyúl, L.G., Falcão, A.X., Udupa, J.K.: Fuzzy-connected 3D image segmentation at interactive speeds. In Hanson, K.M., ed.: *Medical Imaging 2000: Image Processing*. Volume 3979 of SPIE Proceedings., San Diego, CA (2000) 212–223
7. Nyúl, L.G., Udupa, J.K.: A protocol-independent brain MRI segmentation method. In Sonka, M., Fitzpatrick, M.J., eds.: *Medical Imaging 2002: Image Processing*. Volume 4684 of SPIE Proceedings., San Diego, CA (2002)

Virtual Colonoscopy using Direct Rendering of Surfaces from Volumetric Data

Dávid Jocha¹ and József Koloszar²

¹ Computer Graphics Research Group,
Department of Control Engineering and Information Technology,
Budapest University of Technology and Economics
jd202@hszk.bme.hu

² Computer Graphics Research Group,
Department of Control Engineering and Information Technology,
Budapest University of Technology and Economics
kj212@hszk.bme.hu

Abstract. Virtual colonoscopy is a non-invasive computerized medical diagnostic method for examining the interior of the human colon, aiding the detection of polyps. Cross-sectional image scans of the patient's abdomen obtained via a CT scanner are used to construct a 3D volumetric dataset, views of which can be rendered using various volume visualization techniques. In this paper we present our real-time direct volume rendering based visualization technique, which can render iso-surfaces directly from volumetric data without extracting the surface data itself. We found this approach efficient enough to yield interactive frame-rates on a consumer-end personal computer system, and implemented it in our interactive virtual colonoscopy software *ColVis*.

1 Introduction

According to US figures, colon and rectum cancer (colorectal carcinoma) is the second leading cause of cancer deaths. In order to efficiently diagnose this form of cancer in its early developing stage, a cost-effective patient-comfortable procedure is needed. Current methods are neither of the above. Computerized visualization, however, may provide a viable alternative: Virtual Colonoscopy.

1.1 Non-computerized methods

Colonoscopy and barium enema are the two classic diagnostic methods for detecting polyps or carcinoma of the colon [3]. Colonoscopy is an invasive procedure involving the introduction of optics into the patient's rectum, thus carrying a small risk of perforation. Intravenous sedation is required, and the entire procedure may last up to 40 minutes.

Barium enema involves filling the colon with contrast material, and positioning the patient in front of an x-ray screen. This can be very cumbersome, and a

great deal depends on the experience of the doctor interpreting the results. The efficiency of barium enema is also subject to debate [16].

There are also no accepted means of examining the small intestines of a patient.

1.2 Virtual Colonoscopy

There are already a number of applications in medical imaging today, which aid diagnostics by rendering images of organs in 3D. These are based on data obtained from Computer Tomography (CT) or Magnetic Resonance Imaging (MRI) [23]. Virtual Colonoscopy is one such application aiming to deliver images of the interior surface of the colon.

The scanner obtains a set of 100–200 2D cross-sectional images (Figure 1, Left) of the subject's abdomen, which are then reconstructed into a volumetric data set, a 3D array of voxels. The obtained dataset may be subjected to various image enhancing algorithms and segmentation, and is finally visualized in a manner easily interpreted by medical personnel. In this case an interior view of the colon (Figure 1, Right) is rendered.

We found that specific aspects of virtual colonoscopy allowed for simplification of some volume based rendering techniques. We made an effort to see how virtual colonoscopy could be implemented in an operator-friendly manner on a consumer-end personal computer instead of high-end graphics workstation.

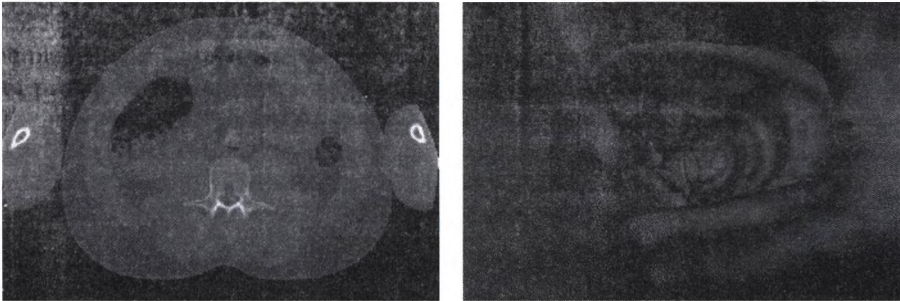


Fig. 1. Left: Cross-sectional image obtained from a CT scan. Right: Virtual view of the human colon

2 The Visualization Problem

Virtual colonoscopy uses algorithms of volume rendering. In the dataset built from CT scans, voxel values represent density measured at points of the body. (Note that one of the drawbacks of virtual colonoscopy is that no color information is obtained.) CT provides good air-to-tissue contrast, which is why it is

preferred over MRI in colonoscopy. Visualization is usually subdivided into two stages: off-line pre-calculations, and interactive on-line rendering.

Off-line preprocessing involves procedures such as algorithmic segmentation, calculating potential fields, etc. It is possible to process all the information off-line if interaction is not a requirement. An entire fly-through of the colon may be generated, yielding video-type output, which can be viewed and reviewed at different speeds. Calculations for such a task, however, may take days to complete.

All calculations with variables depending on interaction with the operator are performed in the on-line stage. User input typically repositions the camera, and the scene has to be re-rendered. Real-time response is one of the key aspects to keep in mind when evaluating algorithms of on-line processing. Frame rate should be as high as possible in order to immerse the operator and shift his/her focus to the medical evaluation of visuals.

We suggest that off-line calculations should be minimized, in order to cut time and technological resources required by the examination as a whole. Instead, real-time interaction should be emphasized.

2.1 Basics of Volume Visualization

Methods of volume visualization usually fall into two categories: direct or indirect [1].

Indirect methods transform the 3D volume model into some structure that is "easier" to render. It is typical to obtain surfaces from the volume and represent them as polygon grids, which can be efficiently rendered using hardware acceleration. However, these methods usually involve a great deal of off-line pre-calculation. The number of triangles may raise memory issues, algorithmic data structures must be used to aid visibility checking for rendering in real time, and image quality becomes an issue, especially when the camera is close to the colon wall. In order to generate images consistent with physical reality, very careful segmentation is needed in most cases [6]. Also note, that extraction parameters cannot be tuned in real-time.

As the name suggests, direct methods work on the voxel array directly to generate images. The two approaches are volume projection, and volume ray tracing [1]. The former projects slices parallel to the viewing window onto the screen, while the latter casts rays for key pixels of the screen into the volume, evaluating luminosity and transparency along the way.

In the early phases of our research we found volume ray tracing to be the most viable of the choices above. After implementing a raw, brute force volume ray tracer, we made simplifications to the algorithm permitted by the following aspects crucial to optimizing algorithms for virtual colonoscopy:

- It is obvious that the colon interior should be treated as empty (lowest density), and seeing through the colon wall is not of importance, as it is the inner surface of the colon wall we wish to visualize in detail.

- As for illuminating the volume, we wish to simulate the "headlight" effect of classic optics employed in colonoscopy, meaning that we only deal with a single source of light with the same location (and orientation) as the camera.

2.2 Volume Ray-tracing

The first algorithm implemented in *ColVis* was based on the following variant of volume ray tracing.

Steps in the iteration along the ray are the following [1]. The ray is cast from the pixel ($trace = 0$), assuming that we have processed distance s ($trace = s$) along the ray $L(s)$ being the accumulated radiance and $A(s)$ the opacity used to multiply radiance from beyond s . Incrementing s by Ds along the ray, the following formulae apply:

$$L(s + Ds) = L(s) + (1 - A(s)) * C(s) \quad (1)$$

$$1 - A(s + Ds) = (1 - a(s))(1 - A(s)) \quad (2)$$

$C(s)$ and $a(s)$ are the color and the opacity values of the encountered voxel, respectively. Iteration is carried out until $1 - A(s)$ drops below the sensitivity of the camera. At that point, radiance coming from beyond s can no longer significantly influence resulting pixel color. In order to maximize performance, the following conditions should be met:

- The inside of the colon is "empty", meaning $C = 0$ and $a = 0$ for all voxels inside the colon. (This can be achieved through proper contrasting.)
- The ray should not traverse deep beyond the colon wall, as it is the inner surface of the colon that should be rendered. (Again, contrasting and proper tuning of camera sensitivity are needed.)

The above conditions simplify calculations to be made at points sampled along the ray. The camera is always positioned inside the colon. Theoretically we may skip along the ray until we hit the colon wall, as there is nothing to calculate for points inside the colon. At points of interest we determine the normal needed by the shading method to define the color of the pixel.

The shading method of choice for *ColVis* is a simplified version of Phong shading. As illumination is based on the simple headlight setup, color can be calculated as a linear function of the scalar product of the normalized gradient and direction of the ray.

One of the most important choices we made, was to implement a relatively new method based on 4D linear regression [7] to estimate the gradient (which can be normalized, and used as the surface normal). The concept being that a linear function

$$f(x, y, z) = Ax + By + Cz + D \quad (3)$$

is used to estimate density in the close proximity of a voxel, where x, y, z are distances from the voxel along the corresponding axis. Minimizing the error-square based on values of the 26 closest neighboring voxels A, B, C , and D can

be calculated very efficiently. The (A, B, C) vector is a good estimate of the gradient, and D can be used as the filtered density value for the pixel. For homogenous regions the gradient is always zero. In our case the gradient inside the colon is not relevant, but for non-empty homogenous regions, the gradient should be assumed to be pointing at the camera.

2.3 Brute Force Ray-tracing

We chose to estimate the gradient for the eight closest voxels to the sampling point, and perform tri-linear interpolation to determine the filtered density and the gradient for the sampling point (Figure 2). We found that tri-linear interpolation was needed to obtain desired image quality, but it presented a significant hit to performance (as gradient estimation had to be performed on eight voxels for every sampling point). We introduced a “low-detail” mode, which omitted all interpolation, thus providing frame rates sufficient for interactive browsing of the colon, at the expense of image quality. Other interpolation methods suggested in [7] might also be used at the appropriate computational cost.

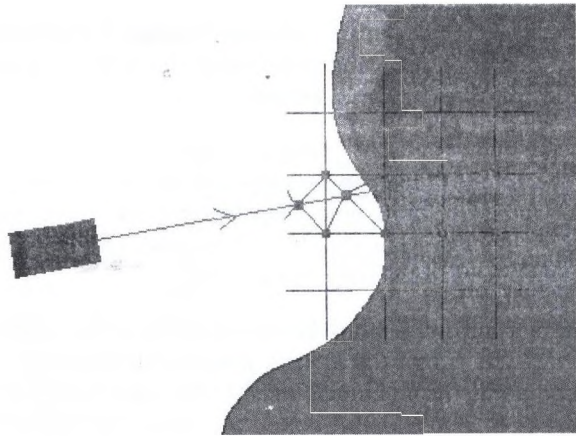


Fig. 2. Sampling points (red dots on the ray), and corresponding voxels (blue dots on the grid) where the gradient is evaluated

There are a number of ways to accelerate the algorithm described above. With sufficient resources (time and memory) the potential field as well as gradients and filtered density values for voxels may be calculated off-line in advance, greatly increasing performance. Assuming 8 bits/voxel for storing value (50–100MB for a typical dataset) we found the following. For all voxels inside the colon, this amount is at least doubled if the potential field is calculated. For voxels close to the colon wall this value is at least five-folded if (A, B, C) and D are calculated. The resulting memory requirement for linearly structured arrays quickly exceeds the reasonable scope of current PC hardware.

Because of tri-linear interpolation, gradients and the filtered voxel values are calculated for eight voxels per sampling. As suggested by Figure 2, at least a double recalculation of most variables is performed during rendering of an entire frame. Caching gradients at frame level for voxels already "touched" may significantly improve performance. We will address this issue in future research.

Hardware specific accelerations may also be implemented. Utilizing SIMD capabilities provided on recent PC processors is a promising aspect. However, the aim of our project was not to optimize code at assembly level.

Empty-space traversal (skipping over the interior of the colon) is another issue to be addressed. For the time being we implemented the following. At every sampling point a single neighboring voxel is checked to see if we are still inside the colon (a simple value check against a threshold). If the result is positive, the ray is incremented without performing costly shading calculations.

A more efficient method we have considered is potential-field assisted empty-space traversal [6]. In the potential field, for every pixel inside the colon the distance to the colon wall is pre-calculated and stored along with the voxel value. When the proximity of the voxel is encountered during the trace, the distance stored in the potential-field can safely be skipped ahead. Though this algorithm has been shown to provide dramatic acceleration, we propose exploiting frame-to-frame coherence instead, because of the off-line pre-calculations involved in setting up the potential field.

Assuming that the camera only moves or rotates little at a time, we find that information gained from rendering the previous frame can aid us in rendering the current frame. If between frames the camera

- is either rotated by a constant angle around one of its three major axis (roll, pitch, yaw),
- or moves forward or backward a constant distance,

frame-to-frame acceleration becomes easy to implement. Though the conditions above may seem to be rough limitations (as real physics-based simulation of dynamics works all six degrees of freedom of motion), we suggest that in our application, where the goal is careful exploration of a 3D environment, these limited controls are sufficient.

In case the camera moves forward, we may assume that distances along the rays to the colon wall (which can be skipped ahead as we trace the ray) in the new frame will roughly be those in the old frame minus the distance the camera moved (Figure 3). Because of perspective projection (implemented with a field of view of 90°), this is trivial for rays at the center of the viewing frame, but gradually fails to apply as we near the edges of the frame. With this in mind, a simple "shortening function" may be devised that returns the distance that could be safely skipped ahead depending on information from the previously rendered frame and pixel coordinates. Every Nth frame should be rendered from scratch (key frames), to make sure errors are not accumulated to cause artifacts, or inconsistent images. The case of rotation by a constant angle is even easier to handle. The jump-skip distance for a ray in the new frame can be derived from those obtained in the previous frame for its neighbors corresponding to the

angle and axis of rotation (Figure 3). This type of frame-to-frame acceleration is a viable and very fast alternative to potential-field aided ray tracing described earlier in the section.

We did implement the case of moving forward and backward with limited success, but as we found the issue was (not yet) the bottleneck, we started focusing on sampling instead.

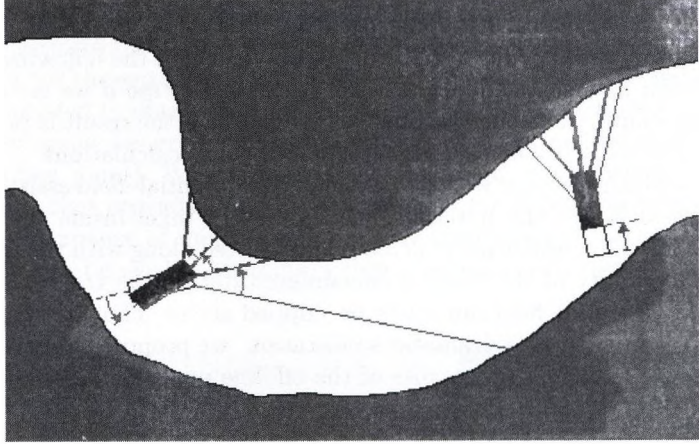


Fig. 3. How rays in the new frame relate to those in the old in case the camera is rotated (left) or moved forward (right)

2.4 Hit-Based Ray-tracing

To obtain acceptable visual quality, about three to five samples in the proximity of the colon wall had to be evaluated along each ray. Considering that the primary objective is to render the colon wall, which, in the most basic case is represented by an iso-surface in the volume, seemed to be an oversight.

We found that the linear regression based gradient estimation could be generalized to obtain the gradient value on a specific iso-surface in a single step. The idea is based on the fact that the function

$$f(x, y, z) = Ax + By + Cz + D \quad (4)$$

actually approximates the density in the close proximity of a voxel. Thus we can solve this equation coupled with the ray equation for a specific $f(x, y, z)$ value (the value of the iso-surface) for the (x, y, z) point where the ray penetrates the specified surface (in the close proximity of the voxel). We modified the ray tracing algorithm to exploit this feature. (We will refer to unit cubic sub-volumes cornered by eight neighboring voxels as cells.)

We traverse the interior of the colon until we arrive at a cell where the possibility of hitting the iso-surface exists. At the simplest level of approximation we check if any cornering voxels are above the threshold value presented by the wall. A negative result means the ray can be incremented without further calculations. If the check result is positive we proceed as follows:

- Calculate A_i, B_i, C_i, D_i for all $i = 1..8$ cornering voxels.
- Solve the system of ' $SurfaceValue = A_ix, B_iy, C_iz, D_i$ ' and the ray equation for $(x, y, z)_i$ for the cornering voxels.
- Calculate w_i , the inverse of the distance between $(x, y, z)_i$ and $voxel_i$ for all cornering voxels
- Calculate the weighted average of $(A, B, C)_i$ -s, using weights w_i .

The resulting (A, B, C) can be normalized to obtain the normal vector to be used for shading.

Hardware specific and empty-space traversal accelerations discussed in the previous section can be applied without significant modifications.

2.5 Results

We found the hit-based algorithm to be three to four times faster than brute-force ray-tracing without tri-linear interpolation, and up to ten times faster than brute-force ray-tracing with tri-linear interpolation with no significant loss to visual quality compared to the latter (!). On our current developing platform:

- AMD AthlonXP 1600+ processor
- 256MB RAM
- Windows XP Professional / (modified) Linux Mandrake 8.1 operating systems

hit-based rendering algorithm performs at up to three to four frames a second.

Window and event handling was implemented using the SDL cross-platform library. The core algorithms were implemented in ANSI C++. The compilers used to build the application on the two platforms were Microsoft Visual C++ and GNU gcc respectively.

3 Conclusions

We presented a direct volume based rendering algorithm with great potential for practical application. Note that applicability of the technique is not limited to virtual colonoscopy, or more generally virtual endoscopy, but can be applied to a number of problems involving iso-surface rendering.

Results from the implementations we have done are promising. We suggest that requirements of near real-time imaging (10–15 frames a second) can be met if hardware specific and low-level optimizations, such as SIMD arithmetic and cache-hit optimized data storage are implemented.

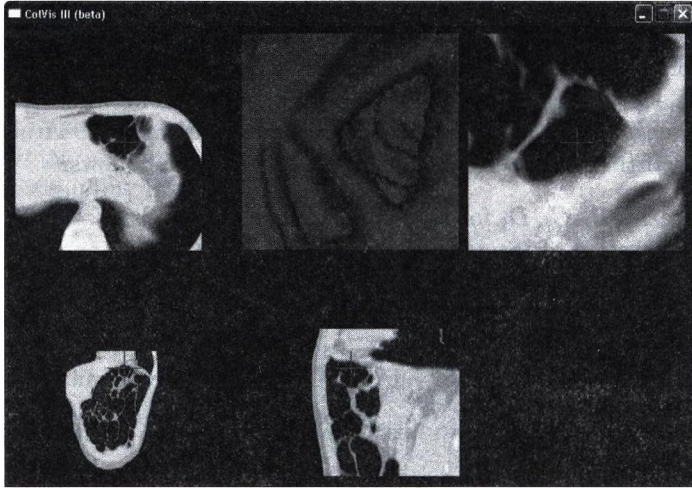


Fig. 4. The current version of *ColVis* in action (low detail mode)

This, as well as continued development and practical testing of *ColVis* is where we focus our current efforts. Based on our findings we also suggest that visualization of virtual colonoscopy will soon be manageable on mainstream hardware.

Based on the material we have researched, we found no reliable algorithm for automatic segmentation (identification of the colon wall, and colon interior) of the volume dataset obtained from CT scans. Though direct volume rendering suggests no need for segmentation while working on actual patient data, we found artifacts (for example little “bridges“ arcing across the colon) in the dataset. They could not be removed by means of simple contrasting, thus we associated them with colonic residue, or scanning flaws. Artifacts like these can severely confuse path-building algorithms, and because of the lack of color information may also hinder polyp detection.

The relative vulnerability of the colon to cancer mandates the continued exploration of virtual diagnostic techniques as both a supplement to and eventual replacement for traditional non-virtual diagnostic modalities.

Acknowledgements

The authors are grateful to Prof. László Szirmay-Kalos and Dr. Zsolt Tarján for their invaluable support.

References

1. Dr. Szirmay-Kalos L.: Számítógépes Grafika. Computerbooks, 1999.

2. Jocha D., Kolozsár J.: Interactive Virtual Colonoscopy. Center European Seminar on Computer Graphics, 2001.
3. L. Hong, A. Kaufman, Y. Wei, A. Viswambharan, M. Wax, Z. Liang: 3D Virtual Colonoscopy. IEEE Symposium on Biomedical Visualization, October 1995.
4. L. Hong, S. Muraki, A. Kaufman, D. Bartz, T. He: "Virtual Voyage: Interactive Navigation in the Human Colon". SUNY.
5. R. Chiou, A. Kaufman, Z. Liang, L. Hong, M. Achiotou: Interactive Path Planning for Virtual Endoscopy
6. M. Wan, Q. Tang, A. Kaufman, Z. Liang, M. Wax: Volume Rendering Based Interactive Navigation within the Human Colon
7. L. Neumann, B. Csébfalvi, A. König, E. Gröller: Gradient Estimation in Volume Data using 4D Linear Regression
8. T. Galyean. Guided Navigation of Virtual Environments. ACM Symposium on Interactive 3D Graphics, ACM, April 1995.
9. M. Gleicher and A. Witkin. Through-the-Lens Camera Control. Computer Graphics (SIGGRAPH 92 Conference Proceedings), ACM SIGGRAPH, July 1992.
10. N. Greene, M. Kass, and G. Miller. Hierarchical Z-Buffer Visibility. SIGGRAPH 93 Conference Proceedings, Annual Conference Series. ACM SIGGRAPH, August 1993.
11. J. Latombe. Robot Motion Planning. Kluwer Academic Publishers, 1991.
12. W. Lorensen and H. Cline. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. Computer Graphics (SIGGRAPH 87 Conference Proceedings), ACM SIGGRAPH, July 1987.
13. W. Lorensen, F. Jolesz, and R. Kikinis. The Exploration of Cross-Sectional Data with a Virtual Endoscope. In R. Satava and K. Morgan (eds.), *Interactive Technology and New Medical Paradigm for Health Care*, IOS Press, 1995.
14. G. Rubin, C. Beaulieu, V. Argiro, H. Ringl, A. Norbash, J. Feller, M. Dake, R. Jeffrey, and S. Napel. Perspective Volume Rendering of CT and MR Images: Applications for Endoscopic Imaging. *Radiology*, vol. 199, May 1996
15. T. Saito and J. Toriwaki. New Algorithms for Euclidean Distance Transformation of an N-Dimensional Digitized Picture with Applications. *Pattern Recognition*, vol. 27, no. 11, 1994
16. C. Morosi, G. Ballardini, and P. Pisani. Diagnostic Accuracy of the Double-Contrast Enema for Colonic Polyps in Patients with or without Diverticular Disease. *Gastrointest Radiology*, vol. 16, 1991.
17. K. Shoemake. Animation Rotation with Quaternion Curves. Computer Graphics (SIGGRAPH 85 Conference Proceedings), vol. 19, ACM SIGGRAPH, July 1985.
18. S. Teller and C. Sequin. Visibility Preprocessing For Interactive Walkthroughs. Computer Graphics (SIGGRAPH 91 Conference Proceedings), vol. 25, ACM SIGGRAPH, July 1991.
19. R. Turner, F. Balaguer, E. Gobbetti, and D. Thalmann. Physically-Based Interactive Camera Motion Control Using 3D Input Devices. *Computer Graphics International '91*. Springer-Verlag, June 1991.
20. D. Vining, D. Gelfand, R. Bechtold, E. Scharling, E. Grishaw, and R. Shifrin. Technical Feasibility of Colon Imaging with Helical CT and Virtual Reality. Annual Meeting of American Roentgen Ray Society, 1994.
21. C. Ware and S. Osborne. Exploration and Virtual Camera Control in Virtual Three Dimensional Environments. ACM Symposium on Interactive 3D Graphics. ACM, March 1990.
22. M. Wan, W. Li, K. Kreeger, I. Bitter, A. Kaufmann, Z. Liang, D. Chen, M. Wax: 3D Virtual Colonoscopy with Real-time Volume Rendering

23. Tarján Zs., Makó E., Szirmay-Kalos L., Csébfalvi B., Székely G.: Virtual CT Enteroscopy. 84th Scientific Assembly and Annual Meeting of the Radiological Societies of North America RSNA 98 Conference, Chicago 1998. Abstr: Radiology Vol 209, 1998.
24. Tarján Zs., Makó E., Szirmay-Kalos L., Csébfalvi B., Székely G.: Early experiences with virtual CT enteroscopy. European Congress of Radiology, Vienna 1999. Abstr: European Radiology Vol 9. 1999.

Virtual Dissection of the Colon

Emese Balogh¹, Erich Sorantin², László G. Nyúl¹,
Kálmán Palágyi¹, and Attila Kuba¹

¹ Department of Applied Informatics, University of Szeged, Hungary
{bmse, nyul, palagyi, kuba}@inf.u-szeged.hu

² Department of Radiology, University Hospital Graz, Austria
erich.sorantin@kfunigraz.ac.at

Abstract. Virtual dissection refers to a display technique for polyp detection, where the colon is digitally straightened and then flattened using multirow detector Computed Tomograph (CT) images. As compared to virtual colonoscopy where polyps may be hidden from view behind the folds, the unraveled colon is more suitable for polyp detection, because the entire inner surface of the colon is displayed in a single view. The method was tested both on artificial and cadaveric phantoms. All polyps can be recognized on both phantoms. This technique for virtual dissection requires only a minimum of operator interaction.

1 Introduction and Medical Background

Colorectal cancer is the third most commonly diagnosed cancer for older adults, and colonic polyps are known precursors of this particular cancer. By early detection and treatment it is possible to cure this disease in most cases by simply removing the polyp.

Colonoscopy is regarded as the gold standard for detecting colorectal cancer. This method is invasive, assumes a certain risk for the patient, and it can be completed in only about 85 to 95% of the cases. Another problem with this technique is that only the retrograde viewing direction is possible, which makes the detection of small polyps difficult.

If screening of colorectal cancer is considered, the chosen investigation should allow assessing the whole colon in all patients and it should be possible to inspect the whole colonic surface. Cross sectional imaging modalities like helical computed tomography or magnetic resonance tomography combined with appropriate image post-processing may represent a perfect screening tool, fulfilling all the above mentioned conditions. Already in 1994 Vining and Gelfand reported their experiences with applying methods of virtual reality to helical computed tomography in order to produce endoluminal views similar to colonoscopy [11, 4]. Today CT colonography refers to a multirow detector CT investigation, where the colon is fully prepared and distended by room air or carbon dioxide. Afterward, offline post-processing of the axial computed tomography slices is done by generating 2D and 3D reconstructions including virtual endoluminal views of the colonic surface. Although the underlying CT examination captures the full

colonic surface, this virtual endoscopic view visualizes only a small part of the inner surface of the colon at once. Due to the complex structure of the colon the navigation is a time consuming and labor intensive work.

In this paper, we describe our initial experience on virtual dissection of the colon, a promising display technique for detecting colonic polyps. This method displays the entire inner surface of the colon without the need for navigation. The method is to digitally straighten and then flatten the colon using multirow detector CT images. Experimental data sets that contained simulated polyps in artificial and cadaveric phantoms were used to test the method.

2 Method

In order to obtain the image of the virtually dissected colon, a number of steps have to be executed. First, the images of the region of interest (i.e., the patient's abdomen or the part of the phantom with the polyps) are acquired by a multirow detector CT and reconstructed with standard reconstruction software. Then the colon is segmented with a method based on fuzzy connectedness. Its central path is extracted using a skeletonization algorithm, and orthogonal cross sections are calculated. Finally, the colon is flattened and the inner surface is visualized via surface rendering.

2.1 Segmentation Using Fuzzy Connectedness

After some quick experiments with "simple" techniques (a 3D generalization of region growing and a method using splines) we found that, although the object of our study (i.e., the colon) is clearly defined in the CT images for the human eye, the results produced by these techniques were not well reproducible and the processes were inefficient (regarding the amount of the necessary human interaction and the way human input is required). Since this is the very first step in the post-processing operation line in our project, everything depends on the segmentation results. Therefore, a validated, accurate and highly reproducible method should be used to avoid problems later in the process. This is what made us decide on utilizing the segmentation method using fuzzy connectedness [9, 10].

Images are inherently fuzzy. Their inhomogeneity is partially due to the graded composition of the material being imaged and to the imaging process (e.g., blurring, partial volume effects). Fuzzy connectedness captures this fuzziness as well as the spatial hanging-togetherness of the voxels in a well defined manner. It has a mathematically formulated theory proving its robustness. This framework and the algorithms have now been successfully applied extensively on 1000's of 3D volumes in several routine clinical applications.

In our case (i.e., segmenting the colon from a 3D CT image volume), air has a well defined range of Hounsfield units so the parameters for fuzzy affinities required by the method may be set once and used for all studies without a per-study training. The user specifies the rectangular volume of interest (VOI)

using a GUI, as well as one (or a few, if necessary) “seed points” within the colon for seeding the fuzzy connected objects. The VOI operation is to reduce the data volume to be processed by the fuzzy connectedness method, as well as to exclude unwanted regions from the processing volume that would otherwise make several additional post-processing steps necessary. Seed points are specified by clicking on voxels “deeply” inside the colon to avoid the uncertain regions near the boundary.

Fuzzy connectedness is a very robust segmentation technique that very effectively handles inhomogeneity (coming from both material properties and the imaging procedure). It is also robust against the selection of seed points (if boundaries are avoided). Since the only user input (apart from the VOI selection) is the seed specification, and all other parameters are fixed, our segmentation inherits the reproducibility and efficiency of the fuzzy connectedness method.

Since we use absolute fuzzy connectedness and a single object is segmented, the uncertain boundary regions (due to partial volume effects) are not captured by the fuzzy objects, therefore, the segmentation was too tight (uniformly along the entire object) according to the physicians opinion. A 3D dilation using a $3 \times 3 \times 3$ structuring element is applied to the segmented fuzzy connected object, thus making it better fit the observers expectations. Finally, the segmented binary 3D volume is interpolated into cubic voxels, because some of the successive steps (e.g., skeletonization) require isotropic data.

2.2 Skeletonization

The notion of *skeleton* was introduced by Blum [2] as a region-based shape descriptor which summarizes the general form of objects/shapes. A very illustrative definition of the skeleton is given using the prairie-fire analogy: the object boundary is set on fire and the skeleton is formed by the loci where the fire fronts meet and quench each other.

During the last two decades skeletonization (i.e., skeleton extraction from digital binary images) has become a challenging research field. There are two major requirements to be complied with [8]. The *geometrical* requirement means that the skeleton must be in the “middle” of the object and invariant under some relevant geometrical transformations including translation, rotation, and uniform scaling. The second requirement is *topological* requiring that the produced skeleton must be topologically equivalent to the original object.

There are three major skeletonization methods for discrete objects: the distance transform method, the method based on Voronoi diagrams, and thinning.

The first method is to find local maxima or ridges in the distance map calculated from the border points of an object. The distance map is a (non-binary) picture in which each point has a value that approximates the distance to the nearest border point [3]. The result of the distance transformation depends on the selected distance and the ridge extraction is a rather difficult task. The distance map based method fulfills the geometrical requirement if a good approximation to the Euclidean distance is applied, but the topological correctness is not guaranteed.

The Voronoi diagram of a discrete set of points (called generating points) is the partition of the given space into cells so that each cell contains exactly one generating point and the locus of all points which are nearer to this generating point than to other generating points. It has been shown that the skeleton of an object which is described by a set of border points can be approximated by a subgraph of the Voronoi diagram of that generating points. Both the geometrical and the topological requirements can be fulfilled by the skeletonization based on Voronoi diagrams but it is regarded as an expensive process, especially for large and complex objects [6].

The thinning is a frequently used method for producing an approximation to the skeleton in a topology-preserving way [5]. Border points of a binary object that satisfy certain topological and geometric constraints are deleted in iteration steps. The entire process is repeated until only the "skeleton" is left.

The topologically oriented thinning pays less attention to the metric properties of the object, therefore, the invariance under rotation (object orientation) and scale change is not guaranteed.

The skeleton represents some kinds of *local object symmetries* [8]. The skeleton of a 3D object can contain some surface patches (representing mirror symmetry and/or rotational symmetry) and some line segments (representing axial symmetry).

In case of "near tubular" 3D objects (e.g., airway, blood vessel, and gastrointestinal tract), it is generally desired to determine their central paths. It is rather difficult to extract "surface-free" skeletons by applying the distance transform based method or even the Voronoi-based one. Thinning has a major advantage over the other two methods since two kinds of approximations to the skeleton can be produced: *curve-thinning* produces the *medial lines* while *surface-thinning* produces the *medial surfaces* of objects. The results of the surface thinning algorithms are closer to the skeleton than the "skeletons" produced by curve thinning algorithms. Axial symmetry is emphasized by curve thinning and other kinds of symmetries are suppressed.

We use an effective sequential 3D curve-thinning algorithm for extracting medial lines from elongated binary objects [7].

2.3 Flattening the Colon

This step is to digitally straighten and then flatten the colon obtaining a new three dimensional volume. First cross sections orthogonal to the central path were calculated at a distance of 1 mm. Because of the complex structure of the colon nearby cross sections may conflict, and as a result, polyps may be missed or counted multiple times. To resolve this conflict between two intersecting cross sections we have interpolated them iteratively until no intersection occurred. This resulted that some of the cross sections were not orthogonal to the central path any more, which may cause a stretch or compression of the internal and external colon surface. At this moment we obtained the three dimensional cross sectional profile, which is the digitally straightened colon.

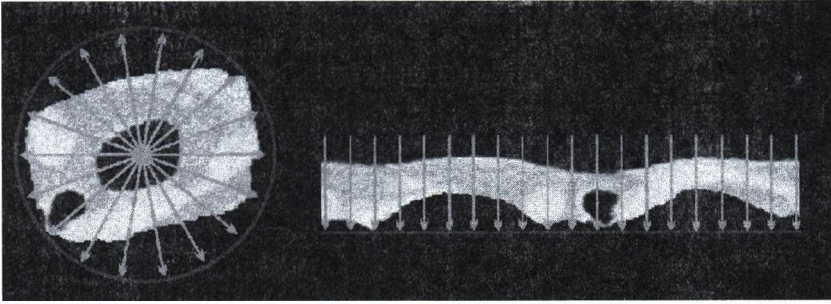


Fig. 1. An orthogonal cross section before (a) and after virtual dissection (b).

Then this volume is remapped into a new three dimensional gray level volume that contains the virtually stretched and flattened colon. This is done with constant angle sampling, with a step of one degree. The remapping process is illustrated in Figure 1. The image in the left side of Figure 1 represents a cross section of the colon before flattening, and the right image in Figure 1 is the same cross section after applying the flattening process.

The last step is to visualize the inner surface of the digitally flattened colon. This is done with the well-known surface rendering technique. This way the entire inner surface of the colon can be inspected from above and the physician can identify the polyps. Figure 5b shows the image of a part of a cadaveric colon after virtual dissection.

3 Results

Two types of phantoms were created to test the method, an artificial and a cadaveric phantom. In both cases the polyps were simulated. The phantoms were scanned by a multirow detector CT (GE Lightspeed QXI, GE Medical Systems) using a collimation of 2.5 mm and a high quality pitch. Axial source images were reconstructed with the CT machine reconstruction program with a slice thickness of 1.25 mm, an increment of 0.5 mm and a matrix of 512×512 . Thus the resulting in-plane pixel size was 0.585 mm. Then we applied virtual dissection on the phantoms and compared positions of the polyps in the resulting images with those on real dissection.

In our first experiment we used an artificial colonic phantom. The phantom was a 100 cm long plastic tube, the polyps were simulated by some small plasticine beads inserted in the ends of the tube. Six polyps were created. In this case we omitted the simulation of the folds of the colon. The photograph of the parts of the phantoms where the polyps were inserted is shown in Figures 2a and 3a. The transparent model of the segmented volume with the central path is presented in Figures 2b and 3b, while the image obtained after virtual dissection is shown in Figures 2c and 3c. All 6 polyps appearing as bumps on the inner surface can be clearly depicted.

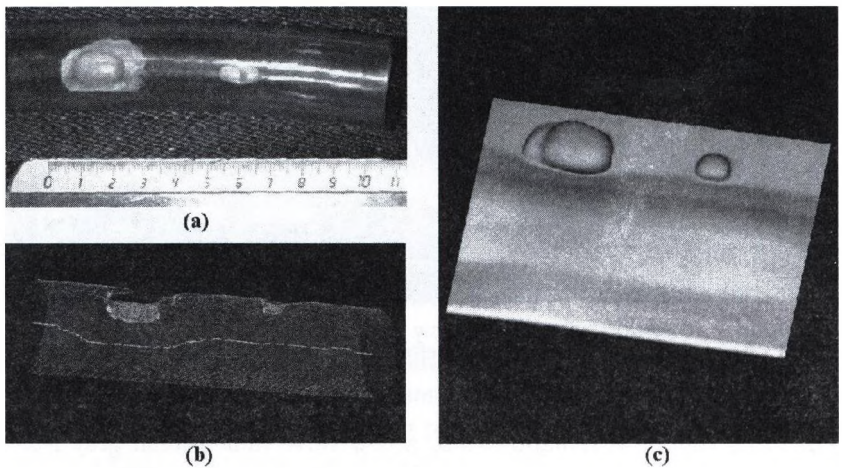


Fig. 2. Photograph of one end of the artificial phantom (a), rendered VRML model of the segmented volume with its center line (b), and the inner surface of the phantom after virtual dissection (c).

After obtaining the permission of the of the ethical comity, for the next experiment we prepared a 50 cm long cadaveric colon. The patient's colon was healthy, therefore the bowel was cleansed and polyps of different sizes were simulated using fat tissues. A total of 13 artificial polyps were created having sizes between $4.1 \times 5.2 \times 2.1 \text{ mm}^3$ and $12.9 \times 15 \times 11.4 \text{ mm}^3$. After air insufflation, the specimen was placed in a 5 l bath containing 5 ml Gastrografin solution imitating the environment of the colon in the human abdomen. A photograph of the phantom can be seen in Figure 4a, and the rendered VRML model of the phantom with its center line can be seen in Figure 4b. After virtual dissection we performed the real dissection of the phantom to compare visually the positions of the polyps on the inner surface of the colon. The photograph taken after the real dissection is presented in Figure 5a and Figure 5b shows the image obtained after virtual dissection. The polyps are marked in order to compare their positions on the photograph taken after real dissection and in the image generated with the real dissection method. This phantom is clearly more complex than the artificial phantom because of the presence of the colon folds. The polyps appeared as bumps or asymmetric broadening of the folds. All simulated polyps could be detected by visual inspection.

4 Conclusion

All polyps can be recognized on both phantoms. While polyps appeared as bumps on the artificial phantom, they appear as bumps or asymmetric broadening on the colon folds on the cadaveric phantom. Both techniques visualize only a part of the inner surface of the colon. Navigation represents another difficulty, which

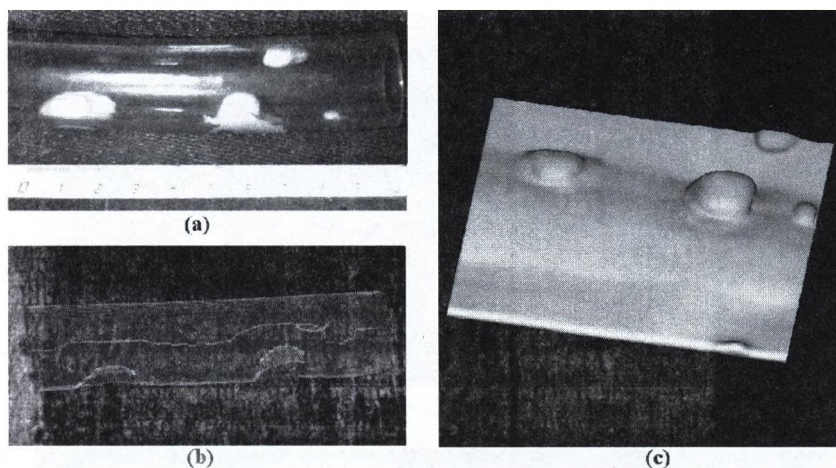


Fig. 3. Photograph of one end of the artificial phantom (a), rendered VRML model of the segmented volume with its center line (b), and the inner surface of the phantom after virtual dissection (c).

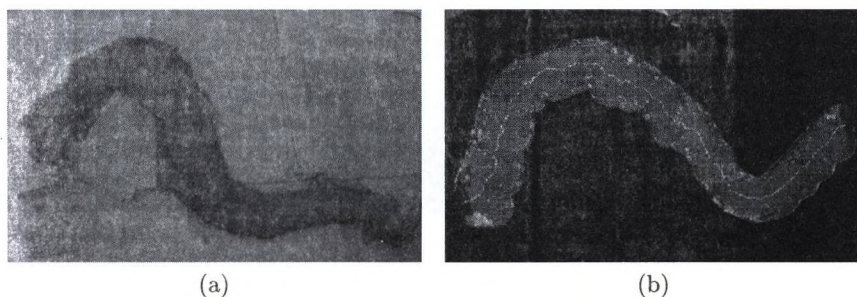
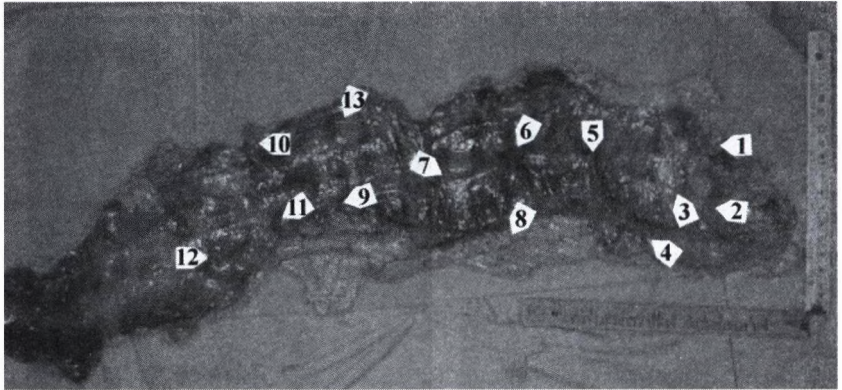


Fig. 4. The cadaveric colon before dissection (a) and its rendered VRML model with its central path (b).

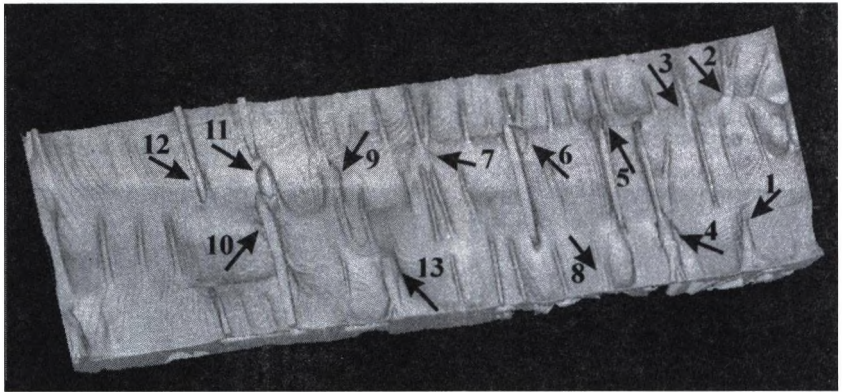
can be laborious and time consuming. The presented method of virtual dissection makes it possible to stretch the colon virtually and to cut it along its central path similarly to what happens on the pathologist's table. The entire inner surface can be inspected from above without the need for navigation. This technique for virtual dissection requires only a minimum of operator interaction. This display technique together with virtual endoscopy has the potential of becoming an alternative to the fiberoptic colonoscopy.

Acknowledgements

This work was supported by the CEEPUS A-34 Grant.



(a)



(b)

Fig. 5. The cadaveric colon after real dissection (a) and after virtual dissection (b). On both images the corresponding polyps are labeled.

References

1. Beaulieu, C.F., Jeffrey, R.B.Jr, Karadi, C., Paik, D.S., and Napel, S.: Display modes for CT Colonography, Part II. Blinded comparison of axial CT and virtual endoscopic and panoramic endoscopic volume-rendered studies, *Radiology* (1999), vol. 212, 203–212.
2. Blum, H.: A transformation for extracting new descriptors of shape. *Models for the Perception of Speech and Visual Form*, MIT Press, (1967) 362–380.
3. Borgefors, G.: Distance transformations in arbitrary dimensions, *Computer Vision, Graphics, and Image Processing* **27** (1984) 321–345
4. Johnson, C.D.: CT Colonography: The Next Colon Screening Examination?, *Radiology* (2000), Vol. 216, 331–341.
5. Kong, T.Y., Rosenfeld, A.: Digital topology: Introduction and survey. *Computer Vision, Graphics, and Image Processing* **48** (1989) 357–393.

6. Náf, M., Székely, G., Kikinis, R., Shenton, M.E., Kübler, G.: 3D Voronoi skeletons and their usage for the characterization and recognition of 3D organ shape. *Computer Vision, Graphics, and Image Processing* **66** (1997) 147–161.
7. Palágyi, K., Sorantin, E., Balogh, E., Kuba, A., Halmaj, Cs., Erdöhelyi, B., Hausegger, K.: A sequential 3D thinning algorithm and its medical applications. In: *Information Processing in medical Imaging, IPMI 2001, Lecture Notes in Computer Science*, Vol. 2082. Springer (2001) 409–415.
8. Székely, G.: Shape characterization by local symmetries. Habilitationsschrift, Institute for Communication Technology, Image Science Division, ETH Zürich, 1996.
9. Udupa, J. K., and Samarasekera, S.: Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation, *Graphical Models and Image Processing* (1996), vol. 58, 246–261.
10. Udupa, J. K., Odhner, D., Samarasekera, S., Goncalves, R. J., Iyer, K. , Venugopal, K., and Furuie, S.: 3DVIEWNIX: An Open, Transportable, Multidimensional, Multimodality, Multiparametric Imaging Software System, In *Proceedings SPIE* (1994), vol. 2164, 58–73.
11. Vining, D.J. and Gelfand, D.W.: Noninvasive colonoscopy using helical CT scanning, 3D reconstruction, and virtual reality, 23rd Annual Meeting and Postgraduate Course of the Society of Gastrointestinal Radiologists, Maui, Hawaii (1994).
12. Wang, G., McFarland, E.G., Brown, B.P., and Vanie, M.W.: GI tract unraveling with curved cross-sections, "IEEE Transactions on Medical Imaging (1998), vol 17, 318–322.
13. Wang, G., Dave, S.B., Brown, B.P., Zhang, Z., McFarland, E.G., and Haller, J.W.: Colon unraveling based on electrical field: Recent progress and further work, In *Proceedings SPIE* (1999), vol. 3660, 125–132.

Színes képosztályozás alkalmazása szívátültetés során fellépő károsodást csökkentő anyagok hatásának vizsgálata során

Csomain István¹, Szabó Gábor², Szabó József¹

¹ PICTRON Kft, 1114 Budapest, Bartók Béla út 3
pictron@axelero.hu

² Chirurgische Universitätsklinik Abteilung Herzchirurgie
69120 Heidelberg Im Neuenheimer Feld 110
Gabor.Szabo@urz.uni-heidelberg.de

Absztrakt. Szívátültetés során a beültetett szív vérkeringésbe való újbóli bekapcsolása (reperfúzió) jelentős kockázati tényezőt jelent a szívizom károsodás tekintetében. A károsodást megelőző illetve csökkentő anyagok vizsgálatára a Heidelbergi Szívsebészeti Klinikán — nemzetközi kutatási projekt részeként — a károsodott szívizom vizsgálatára színes képosztályozó programot alkalmaztak. Az immunhisztológiai eljárással megfestett szövetmintákról mikroszkópra szerelt digitális kamera készítette a felvételeket (40x és 400x nagyítással). A háttér és a pozitív módon megfestődött szövetrészek szétválasztása színminták szerinti osztályozási eljárással történt. A megfestődött részeken belül az elszíneződés mértékének megfelelően 4 fokozatot különítettünk el. Az értékelés alapja az így elhatárolt részek területének meghatározása volt. Előadásunkban a projekt ismertetését követően a képelemző programmal végzett vizsgálat módszerét és eredményét mutatjuk be.

1. Bevezetés

A szív vérellátásának megszakadását (pl. koszorúsér elzáródás, szív-tüdő motorral végzett szívműtét) ischémiának, a szív vérkeringésbe való újbóli bekapcsolását reperfúciónak nevezzük. Az utóbbi évtized kutatásai megmutatták, hogy nemcsak önmagában a szív keringésének a megszakadása, hanem az azt követő ún. reperfúziós fázis is jelentősen hozzájárul a szívizom károsodásához, szívinfarktus kialakulásához. A szívtranszplantáció esetén ez a probléma különös jelentőséggel bír, hisz ebben az esetben a szív vérellátása órákra megszakad, ugyanakkor az újonnan beültetett szívnek rögtön, még a reperfúziós időszak alatt meg kell kezdenie a teljes működését. A nemrégiben felfedezett "poly-ADP-ribose polymerase (PARP)" út központi szerepet játszik a reperfúziós károsodás kialakulásában [1]. Egy új tanulmányunkban megmutattuk, hogy a PARP enzim gátlása jelentősen javítja a transzplantált szív funkcióját reperfúzió alatt patkány modellben [2]. A pontos hatásmechanizmusok azonban még nem ismertek. Korábbi tanulmányok [1], [2], [3] megmutatták, hogy a PARP enzim gátlása jelentősen növeli a sejtek energetikai tartalékait. Újabb adatok [3]

arra utalnak, hogy a PARP-gátlók egyéb kedvező hatásokat is kifejtnek. A reperfüziós károsodás egyik központi eseménye a vérben keringő fehér vérszettek és az érfal közötti kapcsolat kialakulása, amely végül a fehér vérszetteknek a szövetbe való kivándorlásához és sejtkárosító anyagok felszabadulásához vezet. A fehér vérszettek kitapadásához az érfalakon aktiválódó ún. adhéziós molekulák szükségesek. A jelen immunhisztológiai tanulmányban a PARP enzim szerepét vizsgáltuk a P-Selectin és az ICAM-1 adhéziós molekulák aktiválódásában patkány szívtranszplantációs modellben.

2. Módszerek

A kísérletekben összesen 48 Lewis patkányt használtunk. Altatott patkányokból távolítottuk el a donor szíveket és a fogadó állatok hasi verő- és gyűjtőereivel kapcsoltuk össze mikrosebészeti varratokkal. Az ischémiás idő (a vérkeringés megszakadása) egy óra volt. A transzplantált szív vérkeringésének újraindítása előtt az állatokat négy csoportba osztottuk. Két csoportban az állatok placebo, két csoportban pedig PARP-gátlót (PJ34, 3 mg/kg, Inotek Corp. Beverly, MA) kaptak a reperfüzió alatt. Egy-egy kezelt ill. kezeletlen csoportban a szíveket egy óra reperfüzió után, a másik két csoportban 24 óra múlva explantáltuk és szövettanilag feldolgoztuk.

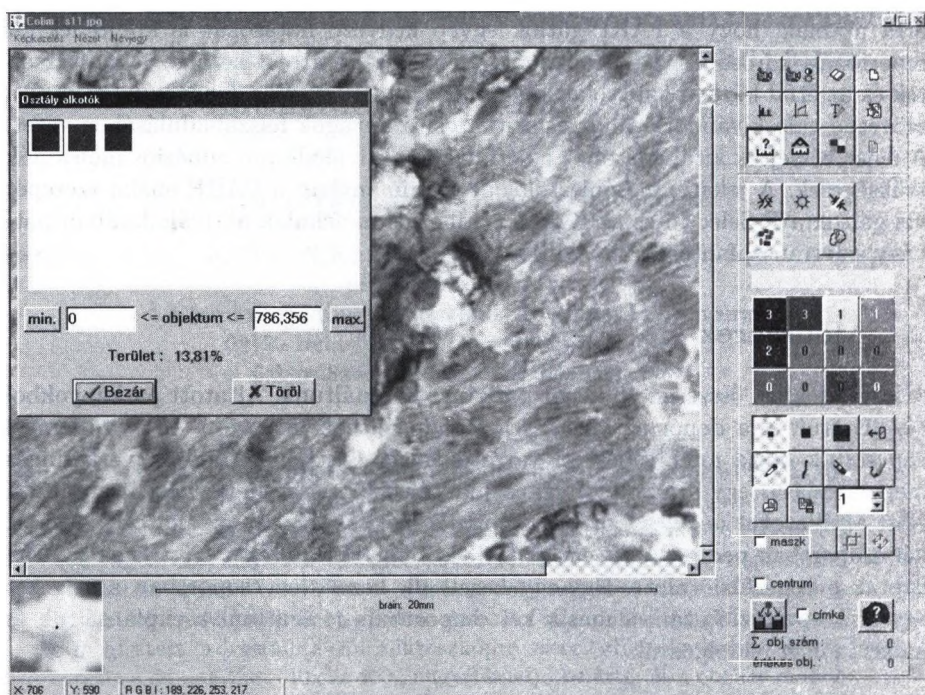
3. Immunhisztologia

PARP kimutatása. Aceton fixált metszeteket elsődleges egér monoklonális anti poly (ADP-ribose) antitesttel (Alexis, San Diego, CA) festettük meg a PARP enzim termékének kimutatására

P-selectin and ICAM-1 immunohistochemistry. Az immunhisztológiai vizsgálatokat az APAAP technika segítségével végeztük [4]. Aceton fixált metszeteket poliklonális nyúl anti human P-selectin (PharMingen Int., San Diego, CA, 1:100 oldás) és monoklonális anti ICAM-1 (Seikagaku America, Falmouth, MA, 1:100 oldás) festettük meg. Mivel a fenti két adhéziós molekula folytonos típusú festést ad az érfal mentén, quantitative morphometriát végeztünk a COLIM program-csomag (Pictron Kft, Budapest, HU) segítségével.

A COLIM programot, és annak egyéb alkalmazási lehetőségeit már korábbi KÉPAF konferenciákon ismertettük [5], [6], ezért itt eltekintünk a program ismertetésétől. Újdonságként annyi emelhető ki, hogy a program a korábbi BMP mellett már TIFF és JPEG formátumot is tud kezelni, és a nemzetközi kooperációban történt felhasználás érdekében elkészült az angol változata. A kezelés további egyszerűsítése érdekében kis mértékben megváltozott a főmenü képernyő is.

Digitális kamerát használtunk a mikroszkópos képek kis (16-40x) és nagy (400x) nagyítású felvételeinek beviteléhez. Öt szín osztályba soroltuk a képpontokat. Az egyik szín osztály a nem festődött terület, a háttér volt, a pozitívan festődő területeket, festődés mértéke alapján négy színosztályba soroltunk. A mintaterületek kijelölését követően a program automatikusan mérte a különböző



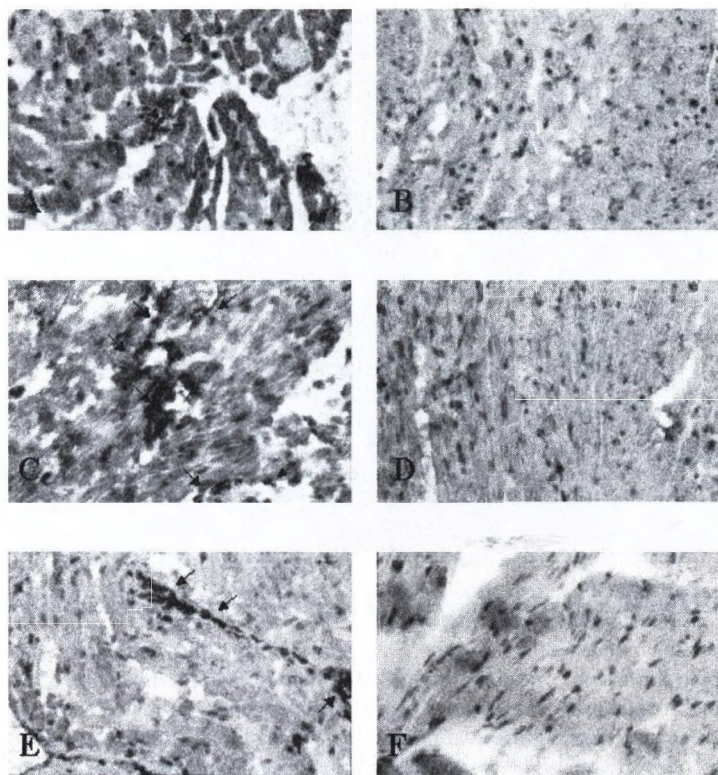
1. ábra: A megváltozott képernyőjű COLIM programmal mintaterületek kijelölés

színosztályokba tartozó objektumok területét. A színosztályok eloszlása alapján a metszeteket egy score szisztéma segítségével hasonlítottuk össze a következő módon: 0 negatív festődés, 1-3 egyre növekedő intenzitású festődés, 4 extenzív festődés.

4. Eredmények

A PARP gátló alkalmazása teljesen megelőzte a PARP enzim aktiválódását. A különböző adhéziós molekulák immunhisztológiai festése a 2. ábrán látható. 1 óra reperfüzió után mind a P-Selectin mind az ICAM-1 expressziója szignifikánsan alacsonyabb volt a PARP-kezelt állatokban (P-Selectin: 1.48 ± 0.03 vs. 0.20 ± 0.01 $p < 0.05$ és ICAM-1: 2.76 ± 0.15 vs. 0.95 ± 0.10 ($p < 0.05$). 24 óra múlva P-Selectin egyik csoportban sem mutatott festődést. Ezzel szemben ICAM-1 továbbra is szignifikánsan magasabb volt a kezeletlen csoportban ($2,08 \pm 0.22$ vs. 0.73 ± 0.17 , $p < 0.05$).

A 3. ábrán egymás mellett láthatjuk az eredeti festésnek megfelelő színekkel, és a szín szerinti osztályozás után, az egyes fokozatoknak megfelelően átszínezett képet. Osztályozáskor a háttérhez sorolt részek eredeti színezése nem változik meg. Az egyes színosztályoknak megfelelő területarányok a képernyőn közvetlenül megjeleníthetők, a terület értékeket a program dBase formátumú fájlban

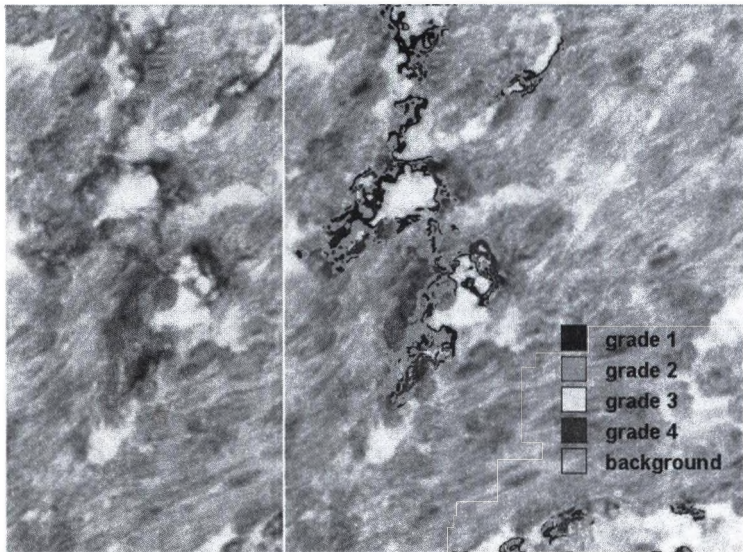


2. ábra: A festés eredménye vizuálisan. **A, B:** P-Selectin 1 óra múlva, **C, D:** ICAM-1 1 óra múlva, **E, F:** ICAM-1 24 óra múlva. **A, C, E:** kezeletlen minták, **B, D, F:** kezelt minták

tárolja. A tárolt eredményeket EXCEL táblázatkezelővel dolgoztuk fel a továbbiakban. Jelen alkalmazásnál a terület arányok kiértékelésére volt szükség, ezért a terület meghatározás képpontokban történt, de a rendszert kalibrálva a mérési eredményeket a kívánt mértékegységben is megkaphatjuk.

5. Diszkusszió

Először mutattuk meg, hogy a PARP gátló PJ34 hatásosan gátolja a PARP enzim aktiválódását szívtranszplantáció után. Ezen kívül bizonyítottuk, hogy a PARP enzim gátlása jelentősen csökkenti a P-Selectin és az ICAM-1 adhéziós molekulák aktiválódását. Ezen eredmények alapján, a PARP-gátlás az adhéziós molekulák szabályozásán keresztül is csökkenti a reperfüziós károsodást.



3. ábra: Képrészlet eredeti színezéssel és osztályozás utáni álszínes megjelenítéssel

Irodalom

1. Szabó, C (Ed): Cell death: the role of PARP. Boca Raton, Florida: CRC Press; 2000
2. Szabó, G., Bährle, S., Stumpf, N., Sonnenberg, K., Szabó, E., Pacher, P., Csont, T., Schulz, R., Dengler, T.J., Liaudet, L., Southan, G.J., Vahl, C.F., Hagl, S., Szabó, C.: Poly (ADP-ribose) Polymerase Inhibition Reduces Reperfusion Injury after Heart Transplantation. *Circ Res* 2002 (in press)
3. Zingarelli B., Salzman AL., Szabó C.: Genetic disruption of poly (ADP-ribose) synthetase inhibits the expression of P-selectin and intracellular adhesion molecule-1 in myocardial ischemia/reperfusion injury. *Circ Res.* 1998;83:85-97.
4. Mason DY.: Immunocytochemical labelling of monoclonal antibodies by APAAP immunoalkaline phosphatase technique. *Techniques in Immunochemistry.* Bullock GR, Petrusz P, eds. New York: Academic Press; 1988;3:25-42
5. Urbanics R., Csomai I.: Fokális agyi ischémia morfológiai analízise interaktív képanalizáló rendszerrel. KEPAF Konferencia, Keszthely 1997. 171-176
6. Csomai I., Szabó J.: Színes kép osztályozó program és alkalmazásai. KEPAF Konferencia, Noszvaj 2000. 121-124

A szívizom vérátfolyásának vizsgálatát támogató összehasonlító adatok előállítására matematikai módszerrel

Máté Eörs¹, Dudásné Nagy Marianna¹, Kári Béla²

¹ SZTE TTK, Alkalmazott Informatikai Tanszék
mate@inf.u-szeged.hu marcsi@inf.u-szeged.hu

² SE ÁOK Radiológiai és Onkoterápiás Klinika
kari@radi.sote.hu

Absztrakt. Képkalkotó eljárással végzett orvosi vizsgálatok értékelésénél nagyban segít, ha lehetőség van egészséges személyeken végzett vizsgálatok eredményével - normális összehasonlító anyaggal - történő összevetésre. Enélkül a képek orvosi megítélése erősen szubjektív és nehezen reprodukálható. Normális összehasonlító anyag előállításának legelterjedtebb módja egészséges személyekről készült vizsgálatok összegyűjtése. Ez a módszer etikailag elfogadhatatlan, ha a vizsgáló eljárásnak egészséget károsító kockázata van. Előadásunkban olyan eljárásra mutatunk példát, amely lehetőséget nyújt arra, hogy korábban elvégzett vizsgálatok előzetes szelekció nélküli felhasználásával normális összehasonlító anyagot állítsunk elő. Módszerünket a szívizom vérátfolyásának radioaktív izotópos vizsgálatára alkalmazva mutatjuk be. 84 beteg adatait használtuk fel, a vizsgálatok fele képezte a tanuló anyagot, a tanuló anyag alapján előállított normál anyag segítségével kiértékeljük az összes vizsgálatot. Eredményeinket hat egymástól függetlenül értékelő orvos eredményeivel vetettük össze.

1. Bevezetés

Leletezéskor az orvos a vizsgálati eredményeket mindenkor összeveti az egészségesnek vélt adatokkal. Pl. vérnyomás mérés esetén az egészségesnek vélt adatok egyszerűen hozzáférhetők, más esetekben sokkal nehezebb az előállításuk. A legtöbb képkalkotó eljárás esetén ezek az adatok csak a leletező orvos tudatában léteznek. Képkalkotó eljárások esetén normális összehasonlító adatok előállításának legelterjedtebb módja, hogy önként jelentkező egészséges személyeken elvégezzük a vizsgálatot, az elkészült képeken lényeg kiemelését végzünk, a kiemelt lényeg standardizáljuk, a standardizált eredményeket statisztikailag kiértékeljük.

A statisztikai értékelés eredményét normál anyagnak tekintjük.

A normál anyagot oly módon tudjuk felhasználni, hogy a vizsgálandó személy esetén is végrehajtjuk a fenti lényeg kiemelését és standardizálást, majd a kapott eredményeket összehasonlítjuk a normál anyaggal.

A fenti eljárás alkalmazhatóságának komoly gátat szab, hogy önként jelentkező egészséges személyeken kell végezzük a vizsgálatot, ami egészséget károsító kockázattal járó vizsgálatok esetén etikai problémákat vet fel. Az etikai problémák elkerülésének leggyakoribb módja, hogy az elvégzett vizsgálatokból utólag válogatják ki azokat, amelyek egészségesnek bizonyultak - akár később elvégzett vizsgálatok, akár a klinikai kép alakulása alapján. Ennek az eljárásnak a korrektségét az kérdőjelezi meg, hogy egyrészt ezek a személyek sem lehettek teljesen egészségesek, ha a vizsgálat elvégzése indokolt volt, másrészt helyes indikáció esetén nehezen gyűlik össze megfelelő számú egészséges személyről készült vizsgálat.

Összehasonlító adatok előállítására szolgáló módszerünkről korábban beszámoltunk [1]. Ebben a cikkben a módszer elméleti megalapozását részletezzük, és a generált normál anyag segítségével végzett automatikus értékelés eredményeit vetjük össze hat függetlenül értékelő orvos véleményével.

2. Normál anyag generálásának módszere

A szívizom vérátfolyásának vizsgálatát ma már legtöbbször $^{99m}\text{Tc-MIBI}$ [2–4] alkalmazásával végzik, de az ilyen vizsgálatokhoz nincs hozzáférhető normál anyag.

Normál anyagot előállító módszerünk alapötlet, hogy a korábban elvégzett vizsgálatokhoz tartozó cirkumferenciális profil görbéknek (CPG) [5, 6] csak az egészségesnek bizonyuló részét használjuk fel. A CPG a szívizom vérátfolyását, perfúzióját jellemző görbe. Ha a szívizom egy adott területén vérátfolyási zavar van, akkor a CPG a megfelelő szakaszon alacsonyabban fut a vártnál, az F normál átlag függvényénél. A feladatunk éppen ennek az F -nek a meghatározása, és annak eldöntése, hogy a CPG alacsonyabb volta kóros mértékű perfúziós zavarra utal-e.

Két hipotézist állítottunk fel:

1. hipotézis. *Kevés olyan CPG van, amely több, mint felén perfúziós zavart mutat.*

2. hipotézis. *A perfúziós kiesések eloszlása viszonylag egyenletes a $[0-360^\circ]$ intervallumon.*

Az 1. hipotézis azt biztosítja, hogy minden CPG elég nagy részét használhatjuk, a 2. hipotézis pedig azt, hogy minden szög esetén elegendő adatunk lesz a statisztikai értékeléshez.

Az 1. hipotézis teljesülését a program tudja ellenőrizni (lásd a későbbiekben: α becslése). Válogatás nélküli beteg anyag esetén a 2. hipotézis automatikusan teljesül.

Jelölések: A CPG-k n pontban definiált diszkrét függvények (esetünkben $n = 60$). Jelölje $f_j(x_i)$ a j -edik beteg CPG értékét az i -edik pontban. Osszuk fel a $[0, 360^\circ]$ intervallumot 10 egyenlő részre, ezeket a részeket nevezzük dekádoknak.

CPG illesztés

Tételezzük fel, hogy F ismert, és illesszük f_j -t F -hez. Képezzük az

$$r_{jk} = \frac{S_k}{s_{jk}}, \quad j = 1, \dots, m, \quad k = 1, \dots, 10 \quad (1)$$

hányadosokat, ahol S_k illetve s_{jk} az $F(x_i)$ illetve $f_j(x_i)$ értékek összege a k -adik dekádon:

$$S_k = \sum_{i=1}^6 F(x_{6(k-1)+i}), \quad k = 1, \dots, 10, \quad (2)$$

$$s_{jk} = \sum_{i=1}^6 f_j(x_{6(k-1)+i}), \quad j = 1, \dots, m, \quad k = 1, \dots, 10. \quad (3)$$

Rögzített j esetén legyen ρ_j az öt legkisebb r_{jk} átlaga. Nyilvánvaló, hogy ez az öt legkisebb hányados a legjobb vérátfolyású - tehát az 1. Hipotézis alapján egészséges - dekádokhoz tartozik. A j -edik beteg k -adik dekádját "normális"-nak nevezzük, ha

$$r_{jk} < \alpha \cdot \rho_j. \quad (4)$$

Statisztikai megfontolások alapján α -t 1.28 -nak választottuk.

 α becslése

Tegyük fel, hogy van N teljesen egészséges személyről készült cirkumferenciális profil görbénk. Ezek a görbék alkalmasak lennének normál anyag előállítására. Ha az eljárásunkban $\alpha = 1$ -t alkalmaznánk, akkor mindegyik görbének csak egy kis részét, a legkisebb r_{jk} értékekhez tartozó dekádokat - a legjobb perfundált régiók adatait - használnánk fel, tehát túl magas átlag görbékét állítanánk elő. Ezen probléma elkerülésére vezettük be α -t az alábbi megfontolások alapján.

Tegyük fel, hogy r_{jk} m várható értékű és σ szórású normális eloszlású valószínűségi változóból vett minta. Minden görbe esetén a legkisebb öt r_{jk} kiválasztása megközelítőleg úgy értelmezhető, mint egy

$$\frac{2}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad -\infty < x < m \quad (5)$$

eloszlású valószínűségi változóból vett minta. Határozzuk meg ennek az m' várható értékét és σ' szórását! Felhasználjuk az ismert

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}, \quad (6)$$

$$\int_0^{\infty} x e^{-x^2} dx = \frac{1}{2}, \quad (7)$$

$$\int_0^{\infty} x^2 e^{-x^2} dx = \frac{\sqrt{\pi}}{4} \quad (8)$$

összefüggéseket.

$$m' = \frac{2}{\sqrt{2\pi}\sigma} \int_{-\infty}^m x e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Az $u = \frac{x-m}{\sqrt{2}\sigma}$ helyettesítés alkalmazásával

$$m' = \frac{2m}{\sqrt{\pi}} \int_{-\infty}^0 e^{-u^2} du + \frac{2\sqrt{2}\sigma}{\sqrt{\pi}} \int_{-\infty}^0 u e^{-u^2} du.$$

Felhasználva (6)-ot és (7)-et

$$m' = m - \sigma \sqrt{\frac{2}{\pi}}. \quad (9)$$

Hasonlóan

$$\sigma'^2 = \frac{2}{\sqrt{2\pi}\sigma} \int_{-\infty}^m (x - m')^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx =$$

$$\frac{4\sigma^2}{\sqrt{\pi}} \int_{-\infty}^0 u^2 e^{-u^2} du + \frac{4\sqrt{2}\sigma(m-m')}{\sqrt{\pi}} \int_{-\infty}^0 u e^{-u^2} du + \frac{2(m-m')^2}{\sqrt{\pi}} \int_{-\infty}^0 e^{-u^2} du,$$

$$\sigma'^2 = \sigma^2 - \frac{2\sqrt{2}\sigma(m-m')}{\sqrt{\pi}} + (m-m')^2. \quad (10)$$

$(m-m')$ -t (9)-ből kifejezve és beírva (10)-be

$$\sigma'^2 = \sigma^2 \frac{\pi-2}{\pi}. \quad (11)$$

A CPG-k vizsgálatánál az átlag $\pm 2\sigma$ szórás között futó részt szokás normálisnak tekinteni. (9) és (11) felhasználásával kapjuk, hogy

$$m = m' + (m - m') = m' + \sigma \sqrt{\frac{2}{\pi}} = m' + \sigma' \sqrt{\frac{\pi}{\pi-2}} \left(\sqrt{\frac{2}{\pi}} \right), \quad (12)$$

$$m \pm 2\sigma = m' + \sigma' \sqrt{\frac{\pi}{\pi - 2}} \left(\sqrt{\frac{2}{\pi}} \pm 2 \right). \quad (13)$$

Méréseink során $\sigma' = 0.0606 * m'$ adódott, ezt behelyettesítve (13)-be azt kapjuk, hogy a normálisnak tekinthető dekádokon r_{jk} értéke a $[0.88m', 1.28m']$ intervallumban helyezkedik el. Ez indokolja az $\alpha = 1.28$ választást.

Megjegyezzük, hogy ha egy f_j görbe esetén a legkisebb öt r_{jk} érték nem esik a $[0.88m', 1.28m']$ intervallumba, akkor a görbe nagy valószínűséggel nem teljesíti az 1. Hipotézist, ezért az ilyen görbét a normál anyag generálásánál nem vesszük figyelembe, a normál anyag felhasználásánál - a leletezésnél - pedig figyelmeztetjük a felhasználót, hogy a görbe illesztés nem megbízható.

A normál anyagot generáló algoritmusunk főbb lépéseit a 1. ábrán szemléltetjük.

1. algoritmus. Normál anyag generálása

1. lépés: Inicializálás

Legyen első közelítésben az F normál átlag-görbe az eredeti görbék átlaga:

$$F(x_i) = \frac{1}{m} \sum_{j=1}^m f_j(x_i), \quad i = 1, \dots, n. \quad (14)$$

2. lépés: CPG illesztés

Ezt a lépést minden f_j -n végrehajtjuk. Legyen $r f_j$ az a függvény, amely csak a (4)-et kielégítő dekádokon van értelmezve, és ott legyen $r f_j = n f_j$, ahol $n f_j$ az f_j függvény F -hez illesztésének az eredménye.

3. lépés: Új átlag görbe előállítása

Legyen rF az $r f_j$ függvények átlaga. Az átlag képzésben minden x_i pont esetén természetesen csak azok a függvények vesznek részt, amelyek az adott pontban értelmezve vannak. A 2. Hipotézis biztosítja, hogy minden $rF(x_i)$ érték elegendően sok $r f_j(x_i)$ érték átlagaként adódik.

4. lépés: Ellenőrzés

Ha

$$|F(x_i) - rF(x_i)| > \varepsilon, \quad \varepsilon > 0 \quad (15)$$

valamely x_i -re, akkor legyen rF az új F , és ugrás a 2. lépéshez.

5. lépés: Alsó határ görbe előállítása

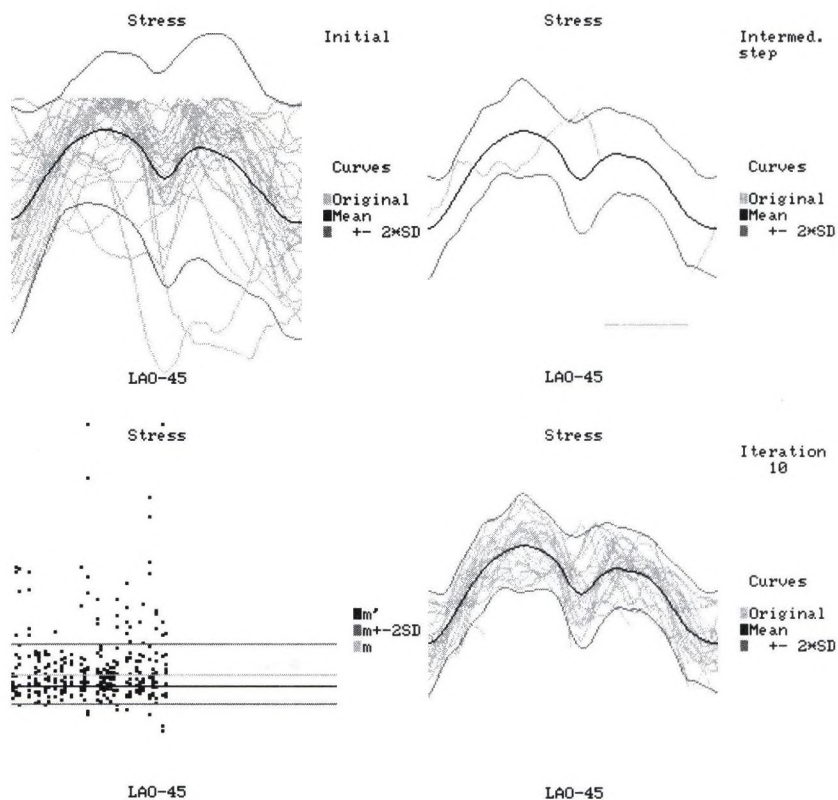
Legyen

$$l(x_i) = F(x_i) - 1.5 \cdot SD(x_i) \quad (16)$$

illetve

$$L(x_i) = F(x_i) - 2 \cdot SD(x_i) \quad (17)$$

az enyhe illetve súlyos alsó határ görbe, ahol $SD(x_i)$ az x_i -ben definiált $r f_j$ függvények x_i -beli értékének szórását jelöli.



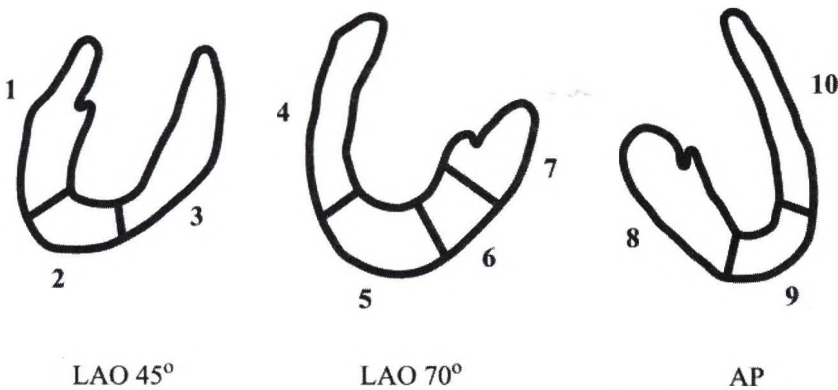
NIBI029

1. ábra: Az algoritmus lépései. Fent balra: Inicializálás. Az eredeti, az átlag és átlag $\pm 2 \cdot SD$ görbék. **Lent balra:** oszloponként az r_{jk} -k úgy skálázva, hogy a legkisebb 5 átlaga m' legyen, a kiszámított m és $m \pm 2 \cdot SD$ érték. **Fent jobbra:** CPG illesztés, a vízszintes vonalnak megfelelő rész nem vesz részt az illesztésben. **Lent jobbra:** Eredmény görbék a 10. iteráció után.

3. Klinikai validálás

A legnagyobb problémát az jelentette, hogy a validáláshoz nem állt rendelkezésünkre arany standard, etalon.

A validálást 84 beteg adatain végeztük. Egy orvos hagyományos módon kiértékelte a vizsgálatokat, további öt orvos mondott véleményt a korábbi értékelés eredmény képei és CPG-i alapján. A 2. ábrán bejelölt 10 régiót mindnyájan az alábbi kategóriákba (score) sorolták: egészséges (0), enyhe (1), súlyos (2) vérátfolyási zavar. Minden beteg mindegyik régiója esetén képeztük a hat orvos által adott score egészre kerekített átlagát, és ezt tekintettük "arany standard"-nak.



2. ábra: A szívizom három vetületének sematikus rajza a 10 vizsgált régióval.

A vizsgálatok véletlenszerűen kiválasztott felét tekintettük tanulóanyagnak, a másik felét tesztanyagnak. A tanulóanyag alapján generáltuk a normál anyagot. A normál anyag segítségével automatikusan kiértékeljük az összes CPG-t. A találati valószínűségek alig térnek el a tanuló- és tesztanyagban, ezért a hat orvosnak és az algoritmusunknak csak a teljes anyagra vonatkozó eredményét mutatjuk be az 1. táblázatban az "arany standard"-hoz viszonyítva. T a True - igaz, F a False - hamis eredményre utal, az első index az értékelő, a második index az "arany standard" score értékét mutatja. Meghatároztuk az orvosi közleményekben szokásos jellemzőket is:

Ezek a jellemzők a táblázat alsó soraiban láthatók.

4. Következtetések

Az orvosok erősen eltérő véleménye az összehasonlító adatok szükségességét húzza alá, és egyben kérdésessé teszi, hogy a hat orvos véleményének átlaga

$$\begin{aligned} \text{Szenzitivitás} &= \text{True}^+ / (\text{True}^+ + \text{False}^-), \\ \text{Specifitás} &= \text{True}^- / (\text{True}^- + \text{False}^+), \\ \text{Pontosság} &= (\text{True}^+ + \text{True}^-) / (\text{True}^+ + \text{False}^+ + \text{True}^- + \text{False}^-), \\ \text{Pozitív prediktív érték} &= \text{True}^+ / (\text{True}^+ + \text{False}^+), \\ \text{Negatív prediktív érték} &= \text{True}^- / (\text{True}^- + \text{False}^-). \end{aligned}$$

valóban arany standardnak tekinthető. Mindenesetre az 1. táblázat alapján megfigyelhető, hogy a módszerünk találati, tévedései az egyetlen F₂₀ kivételével az orvosok találati és tévedési számai között vannak, ami arra utal, hogy az automatikus értékelésünk minősége összemérhető az értékelést végző orvosok értékelésének minőségével.

5. További tervek

A módszerünket szeretnénk valódi arany standard ismeretében is kipróbálni, és a szívizom 3D-s vizsgálatára is kiterjeszteni.

6. Köszönetnyilvánítás

A szerzők köszönetet mondanak Dr. Pusztay Györgyinek, Dr. Farkas Máriának, Dr. Pártos Oszkárnak, Dr. Zámbo Katalinnak, Dr. Schmidt Erzsébetnek és Dr. Udvaros Eszternek a beteg vizsgálatok rendelkezésünkre bocsátásáért és értékeléséért.

Irodalom

1. Dudásné Nagy M., Máté E., Farkas M., Pusztay G., Kári B.: Normál adatok előállításának új módszere a nukleáris medicinában szívizom perfúziós vizsgálatok kiértékeléséhez. KÉPAF Konferencia, Keszthely (1997), Kiadvány 25–30
2. Gopal B. Sa: Fundamentals of Nuclear Pharmacy, Third edition, Springer-Verlag, Chapter 7 (1992) 121
3. Wackers FJTh, Berman DS, Maddahi J et al.: Technetium 99m hexakis 2 Methoxy-IsoButyl Isonitrile (MIBI) human biodistribution, dosimetry, safety and preliminary comparison to Thallium 201 for myocardial perfusion imaging. J Nucl. Med. **30** (1989) 201–311
4. Berman DS, Kiat H, Van Train K, Germano G, Maddahi J, Friedmann J: Myocardial perfusion imaging with 99mTc sestaMIBI. Comparative analysis of available imaging protocols. J. Nucl. Med. **35** (1994) 681–688
5. Burow RD, Pond M, Schafer AW, Becker L: "Circumferential profiles": A New Method for Computer Analysis of Thallium-201 Myocardial Perfusion Images. J. Nucl. Med. **20** (1979) 771–777
6. Garcia E, Maddahi J, Berman D, Waxman A: Space/Time Quantitation of Thallium-201 Myocardial Scintigraphy. J. Nucl. Med. **22** (1981) 309–317

1. táblázat: Az értékelést végző hat orvos és a generált alsó határ görbék alapján végzett automatikus értékelés összehasonlítása az átlagos orvosi véleménnyel.

eset	orvosok						automatikus értékelés
	1	2	3	4	5	6	
True- T ₀₀	1039	991	912	1044	971	1005	969
True+ T ₁₁	14	158	108	66	291	34	115
	T ₂₂	119	73	119	56	36	116
False- F ₀₁	246	206	34	309	86	133	152
	F ₀₂	1	1	0	8	2	5
	F ₁₂	1	47	2	57	83	0
False+ F ₁₀	1	55	127	3	76	13	28
	F ₂₀	7	1	8	0	0	29
	F ₂₁	122	18	240	7	5	215
Szenzitivitás	35%	48%	86%	25%	66%	52%	57%
Specificitás	89%	93%	71%	99%	92%	80%	83%
Pontosság	76%	79%	73%	75%	84%	75%	77%
Pozitív prediktív érték	51%	76%	38%	92%	80%	37%	54%
Negatív prediktív érték	81%	80%	96%	74%	85%	88%	85%

Digitális képtároló és képtovábbító rendszer (PACS) a Szegedi Tudományegyetemen

Almási László¹, Nagy Antal², Alexin Zoltán², Nyúl László², Kuba Attila²,
Csernay László¹

¹SZTE Orvosi Informatikai Intézet

²SZTE Alkalmazott Informatikai Tanszék

Összefoglalás. 1995 óta, 2 CT-t, MR-t, 3 röntgenátvilágító berendezést, röntgenfilm szkennert, ultrahang és izotópdiagnosztikai készülékeket kötötték a szerzők PACS hálózatba a Szegedi Tudományegyetemen. Silicon Graphics központi képszervert, valamint egy HP Kayak (PC) szerverhez telepített HP SureStore DAT egység állnak rendelkezésre a képek archiválására. A képtárolás és továbbítás a nemzetközileg elfogadott DICOM szabványnak megfelelően történik. A képek visszakeresését ORACLE adatbázis kezelő segíti. Kettő, egyenként 3 db 20" képernyős Silicon Graphics Indy munkaállomással ellátott oktatási kabinet lett kialakítva radiológiai gyakorlatok oktatására. A rendszer 24 órás folyamatos működtetése során a PACS képalkotó modalitásain készült valamennyi vizsgálat képei tárolásra kerülnek és a különböző klinikákon elhelyezett képmegjelenítő munkaállomásokon (jelenleg 43 db) - az elérési jogoknak megfelelően - a képek megtekinthetőek.

1 Bevezetés

A PACS az angol „Picture Archiving and Communication System” kezdőbetűiből alkotott betűszó. A PACS a digitális képalkotó készülékek, képmegjelenítő és képfeldolgozó munkaállomások valamint képtároló eszközök számítógépes hálózata. A PACS magában foglalja a felsorolt eszközök közötti kapcsolatot biztosító kommunikációs struktúrát valamint a képek és hozzájuk tartozó alfanumerikus adatok kezelését is. A PACS elsősorban orvosi digitális képek (Mágneses Rezonanciás Tomográfia (MR), Computer Tomográfia (CT), digitális radiológia (CR), digitális ultrahang, Egyfotonos Emissziós Computer Tomográfia (SPECT), stb.) tárolására, elérésére és megjelenítésére szolgál. Hosszútávon a PACS célja a „filmnélküli” radiológia megvalósítása, amikor minden radiológiai kép digitálisan keletkezik, tárolódik, digitálisan elérhető és megjeleníthető. Az ehhez szükséges digitális röntgenkészülékek, nagyfelbontású displayek (min. 2000x2000x10 bit), gyors elérésű többszáz terrabájtnyi tárolókapacitás és különlegesen gyors hálózatok drágasága miatt csak kevés intézményben működik a világon teljeskörű PACS [1].

1994-ben, elsősorban FEFA III. és IV. pályázatok anyagi támogatásával egy szűkített, oktatásközpontú PACS fejlesztésének kezdtünk neki, amely segítségével az Egyetem területén lévő digitális orvosi képalkotó készülékek képeiből, valamint röntgenfilm-szkennelrel digitalizált felvételekből folyamatosan bővíthető és

fejleszthető oktatási célú képadatbázis készült és amely lehetővé tette az orvosi képalkotó eljárások komparatív tanítását és vizsgáztatását erre a célra létrehozott, 2 egyenként 3 db 20"-os monitorral ellátott számítógépekkel felszerelt oktatási egységben. (1997-óta ezekben a kabinetekben folyik a radiológia gyakorlatok oktatása az orvostanhallgatók IV. évfolyamának.) [2],[3],[4],[5]

2 Célkitűzés

Az oktatásközpontú PACS fejlesztésének fő céljai a következők voltak:

1. A különböző képalkotó modalitásokon keletkezett képek a rendszerbe kapcsolt bármely képmegjelenítő munkaállomáson elérhetőek és megjeleníthetőek legyenek.
2. A radiológiai képek konzultációja az egyes klinikákon
3. A radiológiai képek processzálása, szegmentálása, fűziója
4. A radiológiai képek szeparált gyűjtése tudományos célból és az oktatás fejlesztéséhez
5. Lehetővé kell tenni a hálózatra kapcsolt digitális orvosi képalkotó készülékek képeinek rövidtávú (15 nap) teljeskörű archiválását, egyrészt az oktatási képadatbázis karbantartása miatt, másrészt klinikai konzultációkra történő felhasználása céljából.
6. Alapvető képfeldolgozási és 3-D rekonstrukciós technikák használatát hozzáférhetővé kell tenni a képmegjelenítő munkaállomásokon.
7. Az oktatás korszerűsítésén túl, eredeti céljaink között szerepelt a rendszer 24 órás folyamatos működtetése, hogy ily módon a PACS képalkotó modalitásain készült valamennyi vizsgálat képei tárolásra kerüljenek és a különböző klinikákon elhelyezett képmegjelenítő munkaállomásokon - az elérési jogoknak megfelelően - a képek megtekinthetőek legyenek.
8. Az előző pont megvalósításához a képek hosszútávú (örökös?) tárolása.

3 Standard képformátum és kommunikációs protokoll

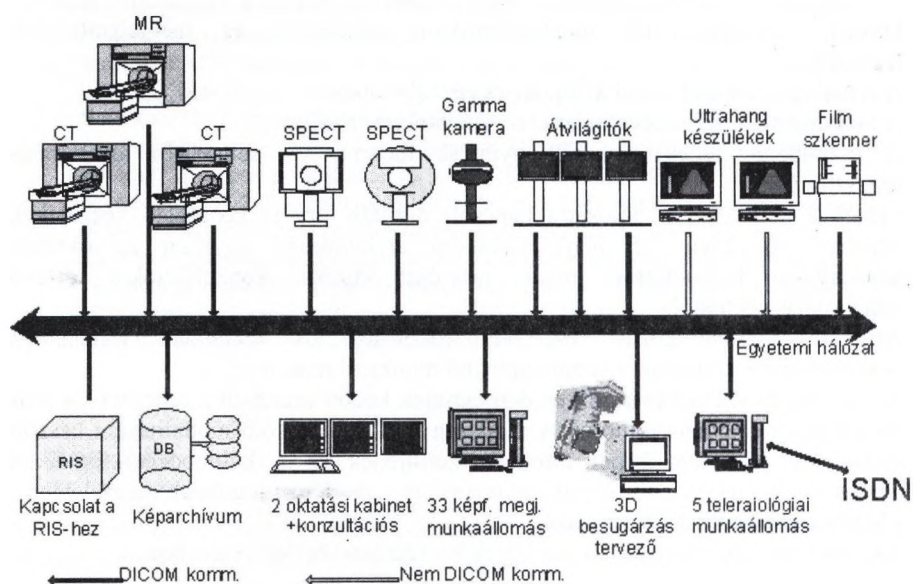
A digitális képek tárolására, a képalkotó és képmegjelenítő állomások összekapcsolására a DICOM 3.0 (Digital Imaging and Communication in Medicine) standardot választottuk.

A DICOM 3.0 az American College of Radiology (ACR) és a National Electrical Manufacturers Association (NEMA) által különböző gyártók készülékeinek összekapcsolására szolgáló ACR/NEMA 1.0 (1985) és 2.0 (1988) standard továbbfejlesztése. Alkalmazásával elérhető, hogy egy gyártó által előállított készülék vagy szoftver kommunikálni tudjon egy másik gyártóéval. A DICOM egyik lényeges tulajdonsága, hogy nem csupán az adatformátumot írja le, hanem megadja azt is, hogy két számítógép hogyan cseréljen információt, hogyan kommunikáljanak (azaz a DICOM egy ún. kommunikációs protokoll is). Ennek következtében a felhasználó számára a DICOM nem csupán a viszonylag könnyen megvalósítható adatátvitel lehetőségét jelenti, hanem a tényleges információcserét is a DICOM szabványnak

megfelelő gépek között. Kapcsolódási protokollt szolgáltat a radiológiai információs rendszerekhez (RIS) és kórház információs rendszerekhez (HIS) is. Ezt a különböző DICOM objektumokhoz kapcsolódó szolgáltatások segítségével nyújtja. [5],[6],[7].

4 Az SZTE PACS felépítése

4.1 Hardver



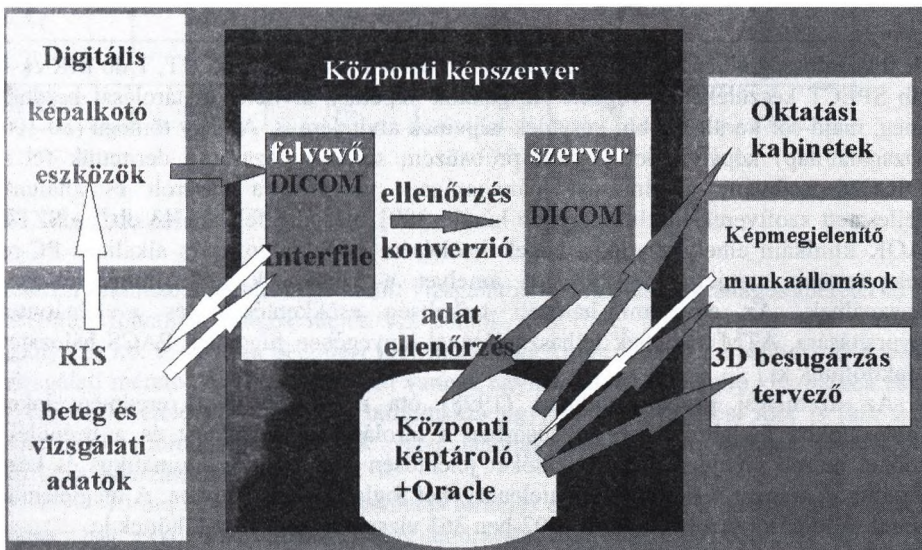
1. ábra. Az SZTE PACS hardver összetevői

Az 1. ábra felső része a PACS input oldalát tartalmazza, a radiológiai és nukleáris medicinai modalitásokat DICOM szolgáltatással ellátva és DICOM szolgáltatások nélkül. Ez esetben Gateway eszközök oldják meg a DICOM kapcsolódást, vagy eseti, kézi beavatkozással érjük el és konvertáljuk az adott eszköz képeit (ultrahang készülékek). A központi rész, a képtároló és képszolgáltató PACS szervert egy SGI Challenge DM számítógép 27 GB merevlemezzel ellátva, ehhez kapcsolódik egy PC alapú HP SureStore DAT egységgel rendelkező szervert. A központi gépek vezérlését SGI Indy munkaállomás látja el. Az output oldal főleg képfeldolgozó és megjelenítő munkaállomásokból (Unix munkaállomások, PC-k) áll, de fontos elem az Onkoterápiás Klinika 3D besugárzástervezője is. Az összeköttetést biztosító hálózati rész az SZTE hálózat Token Ring, ATM gerincéből valamint Ethernet helyi szegmenseiből épül fel.

4.2 Szoftver

A 12 képképző modalitás közül 2 UH és az egyik SPECT készülék nem rendelkezik gyári DICOM opcióval, az utóbbi külön (Passport) egységen keresztül csatlakozik a PACS-hez. A UH egységek képesek képet tárolni és küldeni, de ezek esetleges DICOM konverzióját esetleg kell elvégezni. A felvevő és szervert program a Mallinkrodt CTN (ingyenes) DICOM programcsomagján [8] alapuló saját fejlesztés, TCL/TK-ban és C-ben megírva (Alkalmazott Informatikai Tanszék). PNG képtömörítést alkalmazunk. A szoftver SGI Irix 5.3 operációs rendszer alatt fut. A képek visszakeresését ORACLE adatbázis kezelő segíti. A megjelenítő munkaállomások kereső, letöltő és kitömörítő szoftvere szintén TCL/TK-ban és C-ben készült saját fejlesztésű program, a letöltött képeket OSIRIS [9] programmal lehet megjeleníteni. Ezek a programok Win 95 és Win 98 operációs rendszer alatt futnak.

4.3 Adatáramlás



2. ábra. Adatáramlás

A 2. ábra mutatja az adatok és képek keletkezésének, módosításának, ellenőrzésének és áramlásának folyamatát. Képi adatok a képképző eszközöknél, beteg és vizsgálati adatok a Radiológiai Információs Rendszerben és a képképző eszközöknél keletkeznek. (Mivel a képképző modalitások az átvilágító készülékek kivételével nincsenek a RIS-sel kapcsolatban, a beteg és vizsgálati adatok többnyire kétszer kerülnek rögzítésre.) A képek és vizsgálati adatok a képszerverre, a szerverről vissza az erre alkalmas modalitásokra, 3D besugárzástervezőre és a megjelenítő munkaállomásokra kerülnek. Az ábra középső részén az archivumba történő beépítési folyamat is látható, amely két lépcsőben történik: a képek először ideiglenesen a

modalitásnak kiosztott, felvevő munkaterületen tárolódnak, innen a felvevő program a sikeresen konvertált és ellenőrzött vizsgálatokat a szerver munkaterületre helyezi át, majd a szerver program beépíti azokat az archívumba és egyúttal az ORACLE-be. 60 nap elteltével a képek hosszútávú archívumba, DAT-ra kerülnek.

A felvevő és szerver program kétféleképpen, automata módon és kézi vezérléssel is működhet. Az ellenőrzés során hibásnak talált vizsgálatok a munkaterületen maradnak, a hiba oka felderíthető és ha lehetséges a vizsgálatok kézi beavatkozással kijavíthatóak. Mindkét program biztosítja ugyanis a DICOM formájú vizsgálatok fejlődésének módosításának lehetőségét.

A felvevő program tartja a Radiológiai Információs Rendszerrel a kapcsolatot. A háttérben futó automatikus beépítés megvalósítása elkerülhetetlenné teszi a kapcsolat felépítését, hiszen egyes modalitások alapvető információkat sem vesznek fel a betegről (pl. Helix-SPECT szül. dátum), amit a PACS-ben feltétlenül tárolni szeretnénk.

5 Eredmények

A fokozatosság elvét betartva, üzemelési próbaként elsőként 2 db CT, 1 db MR és 1 db SPECT készülékeken végzett vizsgálatok képeinek átvitelét és tárolását kezdtük meg, majd sor került a többi készülék képeinek átvitelére is. A nagy tömegű (80-100 vizsgálat/nap) képátvitellel végzett próbaüzem során fokozatosan derítettük fel a DICOM szabvány alkalmazási hiányosságait, valamint a vásárolt és általunk fejlesztett szoftverek kisebb-nagyobb hibáit. 2001 második felében 43 db, az SZTE ÁOK klinikáin elhelyezkedő, a képek letöltésére és feldolgozására alkalmas PC-re telepítettünk megjelenítő szoftvert, amelyet a klinikusok több, mint fél éve használnak. Az egyetemi hálózati terheltség csökkentésére és a képátvitel gyorsítására, ATM switchek felhasználásával lényegében független PACS hálózatot alakítottunk ki.

Az üzemelési próba elindítása (1998) óta eltelt idő alatti eredményeinket táblázatokba foglaltuk össze. Elsősorban a tárolási helykapacitást és a beépülési időket mértük, majd a beépítési időket jelentősen befolyásoló automatikus és kézi üzemmód arányát, valamint a hibajelenségeket foglaltuk táblázatokba. A megjelenítő munkaállomásokra a klinikákon 2002-ben 361 vizsgálati sorozatot töltöttek le.

1. táblázat. Vizsgálatok tárolóhely foglalása és beépítési ideje

	Átlagos méret MB	beépítési idő perc
MR+CT ₁ +CT ₂	11,9	8,9
SPECT	6,5	1,0
Átvilágítók	4,5	2,4*
Szkenner	0,6	4,3**
* sok kézi beavatkozást igényel		
** szkennelési idővel (~2,5 perc/film) együtt		

2. táblázat. Képtárolási statisztika az üzemi próba (1998,1999) és a képmegjelenítő program klinikákra történő bevezetése (2002 második fele) egy-egy szegmensében

	Vizsgálat	Autom. beépült	Kézzileg beépült	Nem épült be
1998 márc.	1559	1406 (90%)	138 (9%)	15 (0,9%)
1990 márc.	1914	1627 (85%)	286 (15%)	1 (0,05%)
2001 nov.	1999	1986 (99%)	13 (1%)	0

3. táblázat. Beépítési hibák megoszlása

Modalitás	Össz.	„Normál”	Humán	Program	Humán
		1990 március			2001 nov.
MR	86	65	19	2	5
CT ₁	71	56	15	0	4
CT ₂	63	52	10	1	0
SPECT ₁	66	64	2	0	3

6 Megbeszélés

Először tekintsük át a tárolásra került vizsgálatok tárfoglalásának átlagos méretét és a beépítési folyamat átlagos idejét. Az adatok tömörítve, átlag napi 1 GB tárolási igényt, és kb. 9 - 10 óra beépítési időt mutatnak. Amint látható a beépítési idők és a vizsgálati méretek egyenes arányban vannak egymással, az átvilágító készülékek és a szkennerek kivételével. Az utóbbi tartalmazza ugyanis a szkennelési időt is (kb. 2.5 perc/film) és mindhárom készülék vizsgálatainak beépítése az átlagosnál jóval több kézi beavatkozást igényel.

A 9 - 10 óra beépítési idő egyenletes terhelés esetén éppen kihasználja a szerver kapacitását, ekkor a várakozási lista (ellenőrzésre, tömörítésre és beépítésre várakozó vizsgálatok) kb. ½ óra, egyenetlen terheléskor a várakozási idő több órára is nőhet, sőt időnként segédszámítógép igénybevétele is szükséges elsősorban a képtömörítés elvégzésére.

Érdekes eredményt ad az 1998, 1999 márciusi és a 2001 novemberi archiválási statisztikák összehasonlítása. A 2. táblázat 2 CT, 1 MR és egy SPECT összesített adatai mutatják, hogy hány vizsgálatot sikerült automatikusan vagy kézzileg beépíteni illetve hányat nem sikerült archiválni. A 3. táblázat modalitásonként és hibafajtaként is tartalmazza az automatikusan be nem épült adatokat. Automatikusan a vizsgálatok 90% - 85%-a, kézi módosítás után a vizsgálatok 9% - 15%-a került archiválásra 1999 márciusában. Nem épült be a vizsgálatok 0,9% - 0.05%-a. 2001 végére a kézi beavatkozást igénylő programhibák megszűntek (ezért nem szerepelnek a táblázatban), a "humán" hibák negyedére csökkentek. Itt kell megjegyezni, hogy 2

DICOM modalitás csatlakoztatásának beüzemelése jelenleg is folyik és ezek adatai nem szerepelnek a táblázatban.

A szükséges kézi beavatkozásnak 3 fő oka volt:

1. Az ábrán "normál"-lal jelölt ok. Ilyenkor az adatbázisban már szerepelt az adott beteg a beépítendő vizsgálathoz képest eltérő adatokkal, a program döntést várt a régi adatok megtartására vagy felülírására vonatkozóan. Ez a probléma következhet operátori figyelmen kívül hagyásból (pl. születési dátum elírása, név elírása), de adódhat a konkrét gyári modalitáson futó szoftver hiányosságából (pl. nem lehet bevinni a születési dátumot) illetve egyik modalitás ékezetes betűkkel a másik az angol abc betűivel került beépítésre.
2. Nem megfelelő figyelemmel dolgoztak a modalitásoknál az operátorok (rontott TAJ szám, többször elküldött vizsgálat). Ilyenkor a szerkesztő programmal ki kell javítani a TAJ számot.
3. A képpalkotó modalitás vagy a képtovábbító gateway hozott létre bosszantó, de kézíleg kijavítható (DICOM) hibákat (pl. felesleges szóközpont beírása).

Be nem építhető vizsgálatok 3 fő okból keletkeztek:

1. Saját DICOM adatbázis hiányosság (a fejlesztés befejezése óta új DICOM definíciós táblák megjelenése)
2. Saját programhiba (automatikus beépítésnél időzítési hiba) - kiküszöbölhető
3. Modalitás illetve képtovábbító gateway programhiba

A modalitásokon, vagy a modalitást a hálózatra kötő számítógépen futó szoftvereknél fordult elő a legtöbb hiba. A Fájllista Meta Információs fejléc (információt ad a további adatok kódolásának formájáról) vagy teljesen hiányzott, vagy rossz szintaktikában volt megadva. A modalitások illetve gateway eszközöket gyártók hibájából rosszul volt megadva számos ún. elem-multipllicitás, hiányoztak vagy üresek voltak kötelező mezők, szóközzel voltak feltöltve vagy kiegészítve mezők stb..

Fontosnak tartjuk megemlíteni, hogy a fent említett hibák nagy része nem jelentkezett a kezdeti, lokális teszt stádiumban, csak a nagy mennyiségű képátvitelnél fordultak elő.

A hibákat elsősorban az okozta, hogy a modalitások konvertálóprogramjai a fejlesztés indításakor még kezdeti stádiumban voltak a DICOM szabvány lényegében egyidejű megjelenése miatt. A hibákat a gyártók (Elscent, Siemens, GE, Orel) felé jeleztük, ezek egy részét elhárították, másik részét a konverziós és ellenőrző programok átírásával nekünk kellett leküzdenünk és egy része még megoldatlan maradt. Az utólagos belenyúlás természetesen újabb hibák előfordulásához vezetett a saját programjainkban is.

7 Konklúzió

A fejlesztés indulása óta eltelt 6 év fő tapasztalatai a következők:

1. Különböző gyártók különböző radiológiai és izotópdiaosztikai képpalkotó készülékei PACS-be integrálásához időben hozott, jó döntés volt a DICOM

szabványra történő építés. A döntés helyességét a DICOM rohamos terjedése igazolja.

2. Az előzőhöz kapcsolódik, de külön kiemelő, hogy új modalitás beszerzésénél csak DICOM kompatibilis képképző készülék vásárlása ajánlatos, még ha pillanatnyilag nem is szándékoznak hálózatba kötni.
3. A központi képtároló szoftvernek minimum meg kell valósítani a DICOM „storage, send, retrieve” szolgáltatásokat és lehetőleg ismernie kell a „print” szolgáltatást is.
4. A DICOM szabvány nem megfelelő implementálása miatt a központi képtároló szoftvernek rendelkeznie kell kézi és automatikus formai és tartalmi javítás lehetőségével.
5. A DICOM szempontjából „új típusú” modalitások bekötéséhez a DICOM definíciós tábláknak könnyen bővíthetőnek kell lennie.
6. Automatikus, háttérben folyó archiválás növeli a rendszer hatékonyságát, csökkenti de nem teszi teljesen elkerülhetővé az operátori beavatkozást.
7. Nagy kiterjedésű, több intézetet érintő PACS nagy sávszélességű hálózatra van szükség.
8. A nem PC alapú hardverek javítási költsége enormis, a PC-k jelenlegi fejlettségi szintje lehetővé teszi minden DICOM eszköz PC alapú megvalósítását.

Irodalom

1. Dietric Meyer-Ebrecht: Digital image communication. European Journal of Radiology 1993; 17: 47-55
2. A. Kuba, L. Csernay, L. Kardos, Z. Alexin, A. Nagy, L. Nyúl, L. Almási: Educational PACS at the Medical University of Szeged. CAR'96 (Ed.: H.U. Lemke), Elsevier, p. 1027 (1996).
3. L. Csernay, L. Almási, A. Kuba, L. Kardos: PACS at the Medical University in Szeged for Educational Purposes. Radiology, Supplementum, page 562 (1996).
4. Almási L., Sóti Zs., Kuba A., Alexin Z., Nagy A., Nyúl., Csernay L.: Experience with the SZOTE PACS starting operations. EuroPACS '98, Barcelona, pp. 43-44 (October 1998)
5. Csernay L., Sóti Zs., Almási L., Kuba A.*: Picture archiving with the SZOTE PACS for scientific purposes. EuroPACS '98, Barcelona, pp. 45-46 (October 1998)
6. American College of Radiologists, National Electrical Manufacturers Association, "Digital Imaging and Communications in Medicine (DICOM): Version 3.0". NEMA Standards Publication PS3, 1-12, Washington, DC, 1995.
7. Csernay L, Kuba A., Nyúl L., Nagy A., Almási L., Kardos L.: Micsoda és mire jó a DICOM 3.0? Magyar Radiológia Suppl. 1. kötet, 67-72 (1996).
8. <http://www.erl.wustl.edu/DICOM/ctn.html>
9. <http://www.expasy.ch/www/UIN/html1/projects/osiris/osiris.html>

Orvosi szoftver keretrendszer műtéti tervezéshez

Hajdu András¹, Zörgő Zoltán²

¹ Debreceni Egyetem

Matematikai és Informatikai Intézet

hajdua@math.klte.hu

² Debreceni Egyetem

Matematikai és Informatikai Intézet

zorgoz@delfin.klte.hu

Absztrakt. Az információtechnológia és a számítástechnika rohamos fejlődése természetes módon magával hozta, vagy éppen kialakította az igényt arra, hogy a mindennapi élet de legfőképpen a különböző szakmai tevékenységek minél szélesebb területén ezek a fejlesztések alkalmazásra kerüljenek, megkönnyítve így a szakemberek munkáját. Az orvostudomány különböző területein is egyre nagyobb teret hódít a számítógép. A mesterséges intelligenciát már a 1976-tól hatékonyan alkalmazzák az orvosi diagnosztikában [7]. A számítógépes képfeldolgozás és a grafikus megjelenítést támogató orvosi adatgyűjtő eszközök fejlődése lehetővé tette, hogy olyan rendszerek kerüljenek fejlesztésbe, melyek a szakorvosokat segíthetik nemcsak a diagnózis felállításában, hanem az esetleges beavatkozások kimenetelének és hatékonyságának megbecslésében. Leírásunkban egy olyan orvosi szoftver keretrendszert kívánunk bemutatni, amely a végeselem rendszerre támaszkodva műtétek tervezését is lehetővé teszi.

1. Az orvosi képfeldolgozás általános problematikája

Az orvosi képfeldolgozással foglalkozó kutatók és a klinikusok számára az irodalomból megismert korszerű módszerek és szoftverek rendkívül nehezen, vagy egyáltalán nem érhetőek el, mivel az új fejlesztéseket elsősorban tudományos céllal, kutatólaboratóriumokban, illetve komoly informatikai háttérrel rendelkező klinikákon dolgozzák ki. Amennyiben ezek a szoftverek elérhetőek, klinikai alkalmazásuk nem tekinthető optimálisnak, mert elsősorban fejlesztőik igényei szerint készültek, így integrálásuk a helyi diagnosztikai rendszerbe körülményes, gyakran lehetetlen.

Az orvosi képfeldolgozással foglalkozó cégek csak nehezen tudják implementálni saját szoftvereikbe a legújabb algoritmusokat, mert a két szoftverfejlesztési stratégia, azaz a saját kutatási célra, illetve kereskedelmi célra történő fejlesztés között lényeges különbségek vannak. Az intézetekben a cél a gyors fejlesztés: elegendő az ismert fájlformátumok használata, megengedett bármilyen helyben meglévő vagy implementált szoftver, illetve szoftverkönyvtár alkalmazása és a

kész programok egyszerű az adott célnak megfelelő felhasználói felülettel rendelkeznek. A szoftverfejlesztő cégek programjainak viszont ismernie kell az összes szabványos fájlformátum mellett a piaci szempontok miatt jelentős formátumokat is, a programoknak korszerű és sok szolgáltatást nyújtó felhasználói felülettel kell rendelkezniük. Ezek a szoftverek üzleti megfontolásból zártak, vagy nehezen fejleszthetők.

Ezt az ellentmondást információs technológiával és a multimodalitású képfeldolgozással foglalkozó egyetemi kutatók, valamint az orvosi képfeldolgozás területén hardver- és szoftverfejlesztéseket végző cég mérnökei által közösen szervezett hosszú távú fejlesztői projekt elindításával fel lehet oldani. Ezt az elképzelésünket egy tudományos pályázatunk keretében kívánjuk megvalósítani, melynek célja egy platformfüggetlen multimodalitású informatikai fejlesztői környezet kialakítása, amely lehetővé teszi a kutatási eredmények termékekben történő megjelenését, valamint megfelelő informatikai háttérrel biztosít a képfeldolgozás körében végzett alap- és alkalmazott kutatásokban történő fejlesztésekhez. Ezáltal növeli az intézetekben folytatott kutatómunka hatékonyságát, illetve a fejlesztő cég részére biztosítja piacképes termékek előállítását.

A kidolgozandó fejlesztői rendszer integrálja az alapszintű képfeldolgozási algoritmusokat és a korszerű szegmentációs, regisztrációs és képfúziós technikákat, a felületmodellezési algoritmusokat, valamint a térinformatikai szoftverekben alkalmazott interaktív komputergrafikai megoldásokat.

1.1. A projekt alapvető tudományos és műszaki célkitűzései

A cél egy, a kereskedelemben beszerezhető szoftverkönyvtárakból kialakított multiplatformú, háromrétegű fejlesztői környezet kialakítása, tesztelése és hatékonyságának, illetve használhatóságának bemutatása demonstrációs programok segítségével. A demonstrációs programok specifikálása, tesztelése három klinikai partner bevonásával történik. Az egyes rétegek az orvosi képfeldolgozás különböző témaköréhez kifejlesztendő programok sajátosságainak megfelelően a *funkcionális (vagy matematikai) feladatokhoz, a megjelenítéshez és az interaktív grafikához* valamint a felhasználói felület kezeléséhez kidolgozott szoftverkönyvtárakat jelentik.

- A multidimenzionális (2-, 3- vagy több dimenziós), parametrizált típusú adatstruktúrák modellezése és kezelésükhöz szükséges algoritmusok absztrakt módon megfogalmazhatók. Így az egyes tomográfias leképező eszközök (PET, SPECT, MRI, DSA, CT) képi információi általánosan kezelhetők: a tomográfias mérések a 3D látómező valamely (folyamatos eloszlású) fizikai paraméterének meghatározott szerkezetű 3D képmátrix /térács/ mentén vett leképezését jelentik. Ez az absztrakció biztosítja, hogy az adattárolással, illetve az elemi képfeldolgozással kapcsolatos algoritmusok - képalgebra, mintavételezés, térbeli lineáris- és nemlineáris transzformációk, térbeli szűrések és konvolúciók - ebben az alapkönyvtárban megoldhatók. A képmátrix elemei hordozzák a speciális (radioaktivitás koncentráció, relatív abszorpciók együttható, proton relaxációs-ido, stb.) információkat, amelyekre általában csak a képfeldolgozási folyamat legvégén van szükség.

Hasonló absztrakcióval ki lehet dolgozni a fájlformátum független input/output osztályokat, amelyek felhasználásával a rendszert tetszőleges formátum kezelésével ki lehet bővíteni az alapkönyvtárak változtatása nélkül. A képfeldolgozás további fontos objektumai a profil és kinetikai görbék valamint a VOI (volume of interest) struktúrák. A görbék kezelésére egyszerű tervezési módszerekkel el lehet készíteni az alaposztályokat. A VOI kezelés szervesen kapcsolódik a projekt egyik kutatási területéhez: a műtéti tervezés elemzése végeelem módszer segítségével. E módszerben a 3D felületek térháló segítségével modellezhetők a csomópontok, illetve az adott vizsgálat szempontjából fontos tartományok kijelölésekor a digitális topológia és a matematikai morfológia eszközeivel. Ennek megfelelően a VOI-k automatikus vagy interaktív, esetleg ezek kombinációjával történő kijelöléséhez speciális algoritmusok kidolgozására van szükség.

- A második önálló programozás-technikai feladat a megjelenítési és a 3D interaktív komputergrafikai algoritmusok kidolgozása. Erre a szabványos OpenGL szoftverkönyvtárak segítségével, ezt a megjelenítés-technikát támogató hardver és szoftverkörnyezet specifikálása után megfelelő megoldást lehet találni. E fejlesztés során a funkcionális rétegben specifikált tárolási osztályok és az OpenGL szoftverkönyvtárak közötti kapcsolatot kell megoldani.

Jelentős erőfeszítést igénylő feladat az interaktív VOI és a referenciapontokat kijelölő grafikus felület kialakítása a térinformatikai szoftverekben alkalmazott módszerek mintájára. Ennek a szoftvermodulnak azért van komoly jelentősége, mert a VOI analízis, illetve a referenciapontok kijelölése az orvosi képfeldolgozás leggyakoribb, szakértelmet és komoly időráfordítást igénylő feladatai közé tartoznak. A piacon elérhető professzionális szoftverek a hagyományos kijelölési algoritmusokat teszik elérhetővé a felhasználók számára: a rajzolás 2D unimodalitású metszeteken történik, így abban az esetben, amikor a cél a 3D VOI kijelölése (3D tervezés, térbeli anatómiai struktúrák mentén számított kinetika vagy statisztika) a VOI definiálása a szomszédos metszeteken berajzolt kontúrvonalak meghatározásával lehetséges. A kidolgozandó szoftverkönyvtár erre a feladatra egy olyan alternatív megoldást biztosít, amely a kijelölésre és módosításra fordítandó időt jelentős mértékben csökkenti.

Hasonló problémát vet fel a végeelem analízis segítségével végzett műtéti tervezés elemzése, amelyben a csomópontok, illetve az adott vizsgálat szempontjából fontos tartományok kijelölésekor a digitális topológia és a matematikai morfológia eszközei alkalmazhatók. A 3D-s csontfelszín modellhez illesztett térháló és a térháló egyes elemeihez rendelt anyagparaméterek segítségével kiszámítható a modell viselkedése különböző fizikai ráhatások esetén. Ezzel az eljárással a modelleken végrehajtott módosítások mechanikai vizsgálata is elvégezhető. Módszereink kidolgozásánál fontos szempont, hogy az eredmények könnyen alkalmazhatók legyenek az emberi szervezet különböző részeiről készített rétegfelvételekből előállított modellek hasonló jellegű elemzésére. Távlati célként pedig a protézis-beültetés vizsgálatához, illetve személyre szabott protézisek kifejlesztéséhez, majd CAM rendszerrel végrehajtott gyártásához szükséges grafikai és modellezési funkciók kialakítását lehet megfogalmazni.

- A célprogramok elkészítése szükségszerűen megköveteli egy felhasználói felület használatát. Ehhez célszerű olyan rendszert kiválasztani, amely több platformú fejlesztést is lehetővé tesz. A kiválasztott szoftvercsomag és a kidolgozott OpenGL alapú megjelenítési és interaktív szoftverkönyvtárak illesztésével ez a problémakör kezelhető.

A Debreceni Egyetem TTK, Információ Technológia Tanszék részt vesz a projektben, melynek címe *MEDIP - Platformfüggetlen szoftver keretrendszer orvosi képfeldolgozáshoz*. A következőkben az itt folytatott, projekthez kötődő tevékenységet szeretnénk ismertetni, az elért eredményeket bemutatni. A tanszék keretein belül végzett tevékenységünk túlnyomórészt arra irányult, hogy a Debreceni Egyetem Radiológiai Klinikájáról származó CT felvételeket olyan formára alakítsuk, hogy azok alkalmasak legyenek orvosok által végrehajtható virtuális csontműtétek elvégzésére. Klinikai szempontból jelentkező igény miatt elsősorban a combcsont, illetve a medence és a combcsont ízületeinek vizsgálatakor felmerülő informatikai feladatokra koncentrálnak.

2. Az orvosi képállományok mérnöki formátumba alakításának lépései

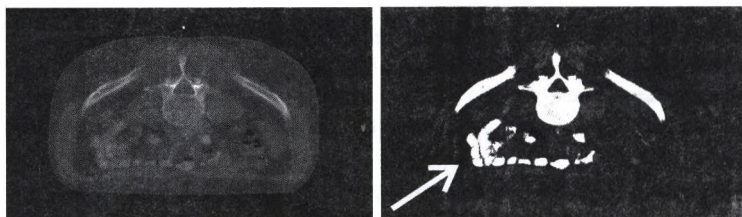
2.1. A bemenet

A Radiológiai Klinika által használt CT berendezések egy, az orvosi képformátum szabványaként elterjedt DICOM (*Digital Imaging and Communications in Medicine*) formátum ([3]) különböző implementációit használják a térszeletek exportálására (például ACR/NEMA 2.0, SPI). Ezek a formátumok sem a DICOM leírással, legtöbbször pedig még egymással sem kompatibilisek. Általában elmondható, hogy egy-egy képsorozat tagjai 512x512x12 bitesek, ahol a szín-mélység a vizsgált személy megfelelő *voxel*-jének (térpont) röntgen-abszorpciós képességét jelenti, amely eléggé korrelál a sűrűséggel.

Kutatásaink elsősorban ortopédiai beavatkozások tervezésére irányulnak, így számunkra a metszeteken található, csontsűrűségnek megfelelő képpontok jelentenek hasznos információt. Tudni kell azonban, hogy az emberi csontok korántsem képeznek homogén mezőket. Még a két lényegesen különböző sűrűségű részen (*szivacsos és tömör állomány*) belül is vannak eltérések. Ezenkívül a csontszövet tengelymenti bordázottsággal rendelkezik, mely a terheléstől függően alakul.

A csontszövet behatárolásának kiindulópontja az úgynevezett csontablak, amely a 12 bites skálán ¹ egy 1500 értékű sáv. Ez jól szűkíthető egy 100 értéket átfedő sávra, de még így is sok olyan részlet marad meg, ami nem csontszövet. Ez jól látható az 1. ábrán is, a jobb oldali képen a szürke nyílal jelölve.

¹ A radiológiában -1024 és 3071 közötti értékeket használják, a hivatkozási csontablak -450 és 1050 között van.



1. ábra: A hivatkozási és a szűkített csontablak.

Az átlagos csípőszélesség körülbelül 300 mm, mely távolság a CT felvételeken 300 pixel körüli, továbbá mivel a sugárterhelés miatt a metszetek távolsága nem lehet 3 mm-nél kisebb, látható, hogy egy voxel körülbelül 1x1x3 mm-es, ami igen nagy². Munkánk kezdeti fázisában medencecsont felvételeken dolgoztunk. Elég nagy nehézségeket okozott, és továbbra is megoldatlan probléma a felbontás miatt "eltűnő" csípőcsont-combcsontfej illeszkedés, ahol a valóságban egy folyadékkal kitöltött vékony rés van. Ennek hiányában nem tudjuk elkülöníteni a combcsontot a csípőtől. A metszettávolság miatt továbbá, az egyes metszetek közötti átmenet durva, a csontrészek elágazásai és összefutásai elég hirtelen következnek be. Bizonyos esetekben, mint például a combcsont esetében, ezek a hiányosságok nem jelentenek problémát. Így kutatási tevékenységünk nagyon szerteágazóvá vált, és sok esetben kellett egy-egy irányt felfüggesztünk vagy elvetnünk.

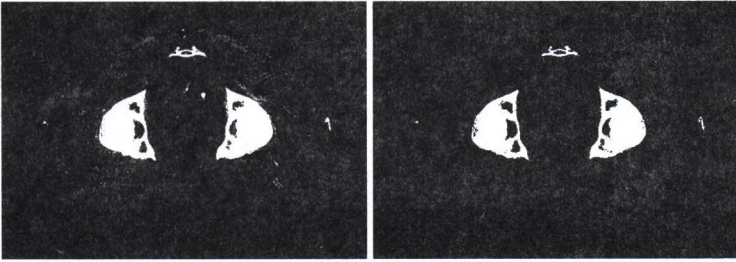
Mindazonáltal kialakult az az elméleti útvonal, melyen metszeti képekből kiindulva felépíthető egy általános geometriai majd testmodell, melyen végelem analízis végezhető. A virtuális beavatkozás többféleképpen is történhet, ennek optimalitása a beavatkozási területtől függ.

2.2. Transzformációk a képtérben és a téglamodell

Az előző szakaszban bemutatott csontablak képezi a kiindulópontját a szintrevágásnak. A későbbiekben a pontosabb becslés érdekében több szövettípust is el szeretnénk különíteni, melyek természetesen különböző statikai jellemzőkkel rendelkeznek. Kutatásaink aktuális fázisában csak egy típusú csontszövetet kezelünk, ezért a képeket csak két szintre vágjuk. Az egyszerű szintrevágási algoritmusok ([4]) mellett egy gradiens-elvű algoritmust ([4]) is alkalmazunk, amely a nagy intenzitás-változásoknál áll meg, így a kritikus illeszkedési terület is jobban kiemelkedik, amint az a 2. ábrán is látható. Ez a módszer sem adta a kívánt eredményt, ezért további módszereken dolgozunk.

A továbblépéshez szükség volt a manuális beavatkozásra, a képeket képszerkesztő program segítségével szintrevágás után kézzel kijavítottuk, eltüntettük a nem-csont elemeket, kisimítottuk a szivacsos-tömör állomány határán fellépő durva egyenlőtlenégeket, megjelöltük a kritikus illeszkedési tartományt minden egyes metszeten.

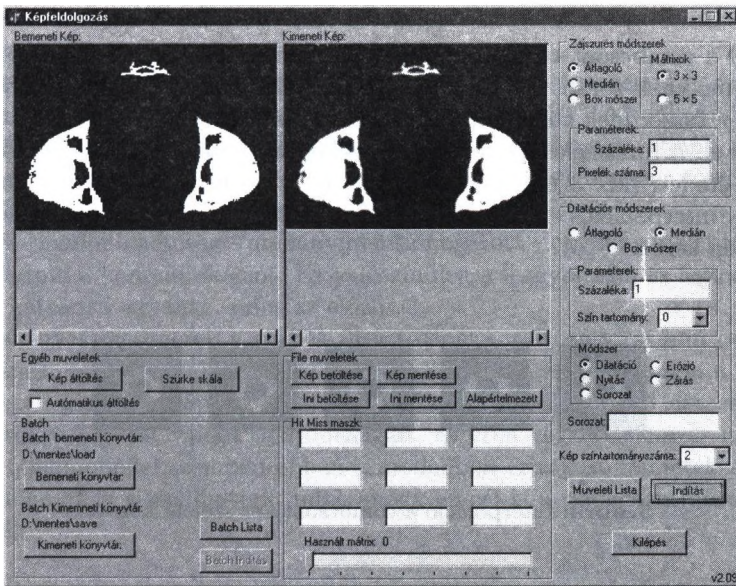
² A kalibrációt egy biliárdgolyóval végeztük, melynek mérete ismert.



2. ábra: Az egyszerű és a gradiens alapú algoritmussal két szitre vágott kép

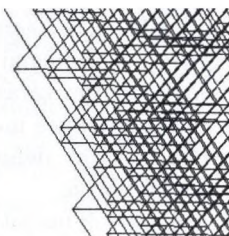
Ezen lépések egy részét már automatikusan is el tudjuk végezni. Ennek a folyamatnak a további automatizálásához készült egy tesztprogram (3. ábra), melynek segítségével az ismert ([4]) morfológiai eljárásokat (zajsztűrés átlagoló, medián, box módszerrel; morfológiai nyitás, zárás, dilatació, erózió, ezek sorozata, általános hit-miss transzformáció a szinteknek megfelelő egyedi maszkolási lehetőséggel, stb.) paraméterezhetjük, végrehajtási sort definiálhatunk belőlük, kötegelt feloldozást végezhetünk velük egy képsoron.

A programban implementált eljárások beépíthetők a végleges képjavító modulba az ugyancsak ebben a programban kikísérletezett paraméterekkel.



3. ábra: A morfológiai műveletek tesztprogramja

A 3D-s testmodell elkészítéséhez választhatjuk a hasonló kutatásokban sokszor használt téglapító eljárást. A megjeleníthetőség szempontjából ez a megközelítés teljesen megfelelő. A szintrevágás során előtérpontnak megfelelő képpontoknak egy-egy téglát feleltetünk meg, melynek mérete megegyezik a korábban említett voxel-méretnek. Az így kapott test felülete nem sima. Ez nem is jelent problémát egészen addig, amíg nem olyan vizsgálatokat végzünk a testmodellünkön, ahol az élek, a meredek törésvonalak potenciális repedések kiindulópontjait jelenthetik. A 4. ábrán a combcsont téglamodelljéből látható egy részlet. A téglamodell optimalizálása céljából egy algoritmus került kidolgozásra, amely összevonást alkalmazva csökkenti a téglák számát. A combcsont esetében ez 30 százalékos fölötti nyereséget jelentett.



4. ábra: A combcsont téglamodell egy részlete

Mivel a combcsont téglamodellje olyan ortopédiai beavatkozás vizsgálatára készült, melynek során a sebész a csontba befűrészelt, de azt keresztül nem vágja, így ésszerűnek tűnt, hogy ezt a virtuális beavatkozást még a képtérben végezzük el. A fűrész éle azonban sokkal vékonyabb, mint a metszetek távolsága, ezért az eredeti (kb. 3 mm magas) téglákon ezt nem lenne indokolt megtenni. Ezért egy interpolációs eljárás segítségével minden két metszeti kép közé további három kép kerül.



5. ábra: Interpoláció a combcsont két szelete között

Így a téglák száma megnő ugyan az első fázisban, de mivel a virtuális beavatkozás végrehajtása után a modellt a fentebb említett módon optimalizáljuk, ez nem jelent gondot. A 5. ábrán egy ilyen interpoláció látható. A csontmetszést szintén automatikusan hajtjuk végre az erre a célra készült tesztprogrammal. Kétféle vágási módot implementáltunk: hossz tengelyre merőleges és spirális vá-

gást. A vágás mélysége és a lépésköz illetve az emelkedés és szögelfordulás szabadon meghatározható.

2.3. Vektorizálás

Mivel a téglamodell nem mindig felel meg a követelményeknek, ezért egy másik megközelítést is használunk. A bitképekből előállítunk egy geometriailag viszonylag egyszerűen leírható sima testet, mely az elérhető végeelem programba betölthető és abban felparaméterezhető a csont anyagparamétereivel [1].

Az első lépése a raszteresről a vektoros formátumra való áttérésnek a metaszetek ponthalmazzá alakítása egyszerű kontúrkijelölő eljárással. Ezután a kontúrpontra görbéket illesztünk. Az egymás fölötti görbék közé felületeket feszítünk, a felületek által határolt térfogatok egyesítéséből pedig létrejön a testmodell. Mivel a tulajdonképpeni vizsgálatot, azaz a virtuális mütét előtti és utáni statikai állapotot egy végeelem rendszerben tudjuk csak elvégezni, ezért a modellünket egy ilyen rendszerben építjük fel.

Mivel az egyes végeelem rendszerek CAD felületének más és más szerkesztési lehetősége van, három lehetőség kínálkozik:

- Saját magunk állítjuk elő és egy általánosan elterjedt fájlformátumba exportáljuk a testmodellt, majd ezt importáljuk az adott rendszerbe. Ilyen formátum például az IGES [5]. Ez a formátum meglehetősen általános és kötött, ezért az output nehezen állítható elő és nagy méretű lesz (a 6. ábrán látható 45000 pont-entitást tartalmazó első modellünk nagyobb, mint 10MB ebben a formátumban).

- A második lehetőségünk, hogy amennyiben a végeelem rendszer rendelkezik script vagy session nyelvvel (mint például a Patran/NASTRAN [6], ANSYS), ezt megismerve, a rendszer lehetőségeihez igazított outputot generáljunk. Ennek előnye, hogy könnyű létrehozni és kis méretű, de a háttérrendszer cseréjekor ezt a modult át kell írni. Mi jelenleg ezt a megoldást részesítjük előnyben.

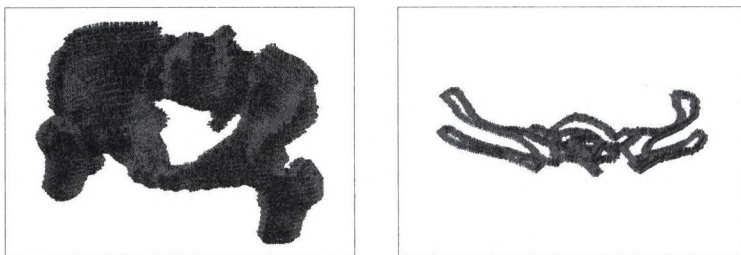
- A harmadik lehetőség – amennyiben támogatott – egy magas szintű nyelven megvalósítani a kommunikációt. Ez valószínűleg a leggyorsabb és legdinamikusabb, de hátránya ugyanaz, mint az előzőnek.

Mivel egyes rendszerek a görbék illesztésénél korlátozott³ számú pontot engedélyeznek, egy olyan algoritmust implementáltunk, ami redukálja a pontokat a környezeti görbület alapján. A görbült felületekre épülő modell simasága miatt jobban tükrözi a test valódi tulajdonságait, az ezen végzett elemzés megbízhatóbb, a téglamoddellel szemben ebben a modellben érintkező felületek statikáját és dinamikáját is vizsgálhatjuk, mint ahogy ez cél is a csípőcsontmütét [2] tervezésénél.

2.4. Végeelem analízis

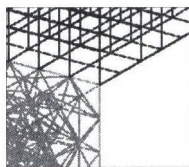
A geometriai testmodell elkészítése után a testet alkotó térfogatókat véges elemekre osztjuk egy kiválasztott topológia szerint kifeszített térháló segítségével

³ Csípő esetén nem több mint 20



6. ábra: Az első modellünk teljes képe és két különálló metszet

(*meshing*). A rendszerek a test geometriájától függően általában kétféle hálót ismernek: kockahálót és tetraéderhálót, lásd 7. ábra.



7. ábra: Kockaháló (fent) és tetraéderháló (balra)

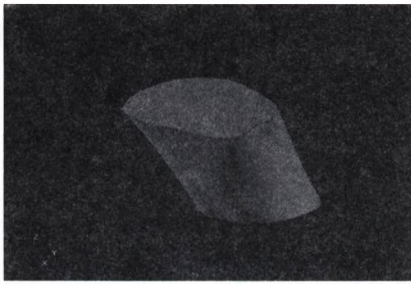
Az így előállt térháló csomópontjaira és lapjaira lehet statikai vizsgálati paramétereket megadni, így kényszererőket (a szabadsági fokok rögzítését), illetve terheléseket (erőhatást, nyomást). A csomópontoknak statikai szempontból hat szabadsági fokuk van: három eltolási és három forgatási.

A vizsgálatok elvégzéséhez feltétlenül szükséges az anyagparaméterek megadása. A rendszerek általában izotropikus, orthotropikus, anizotropikus és kompozit objektumokat ismernek. Lineárisan rugalmas, nem-lineárisan rugalmas, elasztó-plasztikus és más anyagtípusokkal egyaránt tudnak dolgozni. A csont felparaméterezésénél lineárisan rugalmas anyagot adunk meg. A szakirodalom [1] szerint a csontszövet rugalmassági (Young) modulusa 20GPa, Poisson száma 0.3. Ez a két érték elegendő a jelenlegi célunk, a létrejövő deformáció és az ébredő feszültség számításához.

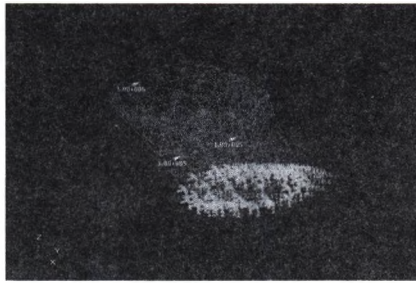
Miután az anyagi jellemzőket hozzárendeljük a térháló elemeihez, lefuttatható az analízis. Szemléltetésül a combcsont modellünk két metszetének pontjaiból alkotott testen végeztünk elemzéseket, ennek eredményét és a paraméterezés folyamatát a 8. ábra mutatja. Megjegyzendő, hogy ebben a példában a csont nem üreges.

2.5. A virtuális műtét elvi menete

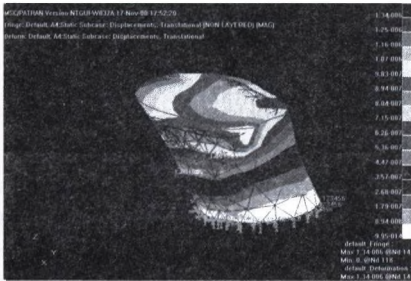
Mivel a végelem rendszerek CAD felületének kezelése nem szakképzett személy számára nagyon nehézkes, ezért nem várható el az ortopédusoktól, hogy ezen a



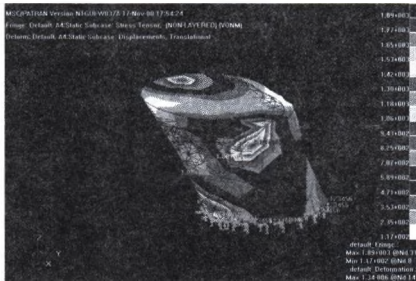
(a). Két metszeti szeletből felépített test



(b). Tetraéderháló, kényszererők és erőhatások



(c). Az elmozdulás szemléltetése

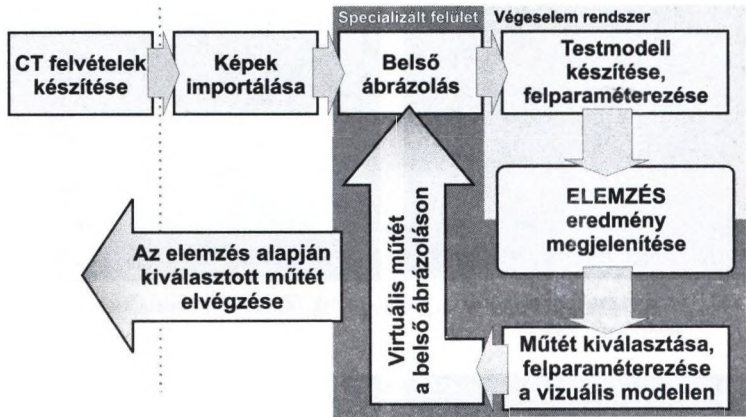


(d). A megjelenő feszültségek

8. ábra: Végeselem analízis egy combcsont darabon

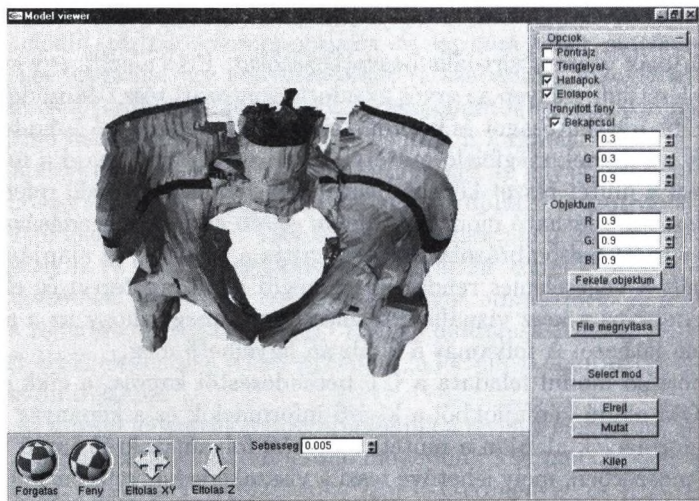
felületen hajtsák végre a virtuális beavatkozásokat. Erre a célra egy specializált felületet fejlesztünk, melyen az orvos az adott csontozati rész testmodelljéhez hasonló – inkább a láthatóságot és körüljárhatóságot hangsúlyozó – modellt lát. Az orvos adott területnek megfelelő műtétek közül választhat, majd a műtétet felparaméterezi a műtét részét képező beavatkozásoknak megfelelő referenciapontok megadásával a látható modellen, illetve egyéb adatok megadásával. Ezután a program a saját belső ábrázolásán végrehajtja a műtétet, ez alapján generálja a testmodellt a végeselemes rendszerben, végül felparaméterezi és elindítja az elemzést. Az orvos a kész vizuális értékelést kapja meg, ahogy az a 8. ábra (c) és (d) részén látható. A folyamat a 9. ábrán figyelhető meg.

Az importáló modul feladata a CT berendezéstől kapott, a cikk elején tárgyalt DICOM alapú képfájlokból a kísérő információk és a képanyag kinyerése. A kísérő információk alapján a műtéti tervezési folyamatokról napló készül egy adatbázisrendszerben, mely lehetővé teszi a visszakeresést, referenciálul szolgálva későbbi műtétekhez is. A képanyag kinyerése után a modul végrehajtja a szintrevágást és a morfológiai transzformációkat. Az így nyert előtérpontok három dimenziós ponthalmazként kerülnek belső ábrázolásra. Egy kiinduló statikai elemzés készül, mely a beteg aktuális állapotáról nyújt plusz információt az orvosnak. Ezután adható meg a virtuális műtét, amelyből újabb testmodell és elemzés készül. Az orvos ezek után az eredeti modellre visszatérve vagy az előző műtét eredményéből kiindulva újabb virtuális műtéteket hajthat végre és értékelheti



9. ábra: A műtéti tervezés elméleti folyamatábrája

azok eredményeit. A keretrendszerre két tesztprogram is készült, az egyik az importálást és a belső formátumra hozást hajtja végre – aktuális állapotában a téglamodell generálásához is használjuk végeselemes session fájl formájában –, a másik (10. ábra) az OpenGL-es megjelenítést, a körüljárást, majd a műtét kiválasztását, illetve az ahhoz tartozó referenciapontok megadását teszi lehetővé.



10. ábra: A keretprogram megjelenítő tesztváltozata

3. Távlati terveink

- Az OpenGL-es megjelenítést úgy szeretnénk megvalósítani, hogy a már elterjedt, 3D szemüveget vezérlő hardvert is kezelje, ezáltal még realiztikusabb megjelenítést valósítva meg.
- Terveink között szerepel többféle csontszövet megkülönböztetése, amelyek eltérő mechanikai paraméterekkel rendelkeznének, így a végeelemes modellezés precízebben lefolytatható lenne.
- A CT képekből készült statikai modell lehetővé teszi CAM rendszeren személyre szabott protézisek készítését is.
- A finomabb, több csontot tartalmazó ízületek (például bokaízület) feldolgozásához szükséges egy szegmentáló eljárást is beépíteni az egyes csontok megkülönböztetéséhez.
- Érdekesnek (bár egyúttal igen problematikusnak is) ígérkezik a csontmodell lágyszövetekkel (például ínnyalagokkal, izmokkal) való kiegészítése, a bekövetkező változások minél valóságosabb elemzéséhez.

Köszönetnyilvánítás

Ezúton szeretnénk köszönetet mondani a Debreceni Egyetem Ortopédiai Klinikája, Információ Technológia Tanszéke és Műszaki Főiskolai Kara azon dolgozóinak, akiktől jó ötleteket és sok segítséget kaptunk eddigi kutatásainkhoz.

Irodalom

1. Abé H., Hayashi K., Sato, M.: Data Book on Mechanical Properties of Living Cells, Tissues, and Organs. Springer-Verlag Tokyo (1996)
2. Bombelli, R.: Structure and Function in Normal and Abnormal Hips Springer-Verlag Berlin (1996)
3. DICOM online documentation, National Electrical Manufacturers Association Rosslyn, Virginia, USA
4. Gonzalez, R.C., Wintz, P.: Digital Image Processing Addison-Wesley Publishing Company USA (1987)
5. IGES online documentation, Officers and Committees of the IGES/PDES Organization (1991)
6. MSC Patran/Nastran v9.0 online help (2000)
7. Shortliffe, E. H.: Computer based medical consultation: MYCIN American elsevier (1976)

MedEdit: Műtéti tervezést segítő orvosi képfeldolgozó rendszer

Ollé Krisztián¹, Kuba Attila¹, Varga Endre²

¹ Alkalmazott Informatikai Tanszék, Szegedi Tudományegyetem
{h531987,kuba}@inf.u-szeged.hu

² Traumatológiai Klinika, Szegedi Tudományegyetem
varga@orto.szote.u-szeged.hu

A csontrendszer sérülésekor végzett műtétek gyakran nagy bonyolultságúak és nagy pontosságot igényelnek (pl. implantátumok behelyezése). Éppen ezért a műtéti beavatkozás előtt célszerű az adott helyzetet geometriai és mechanikai szempontból is leíró modell készítése, amelyen a különféle műtéti megoldások szimulálhatók, hatásuk pedig lemérhető. Egy ilyen számítógépes rendszer - amit MedEdit-nek neveztünk el - kifejlesztésében elért eredményeinkről szeretnénk röviden beszámolni.

A rendszer funkcionálisan 3 részből áll:

- Az első rész a vizsgált tájék csontjainak geometriai modelljét készíti el. Ehhez a betegről készült CT-felvételeket használjuk fel, amelyeken a csontnak megfelelő részeket szegmentáljuk ki a keresztmetszeti képekből, majd a metszetekből építünk fel egy 3D struktúrát.
- Általában egy ilyen geometriai modell még túl bonyolult, ezért különféle felület-egyszerűsítési technikákkal igyekszünk csökkenteni a modellben használt felületi elemek számát a MedEdit második részében.
- A harmadik rész feladata a különféle műtéti megoldások kipróbálása az előállított modellen. Ehhez 3D szerkesztési funkciókat (pl. vágás, fúrás, implantátum behelyezése) valósítottunk meg.

Betegvizsgálati képeken mutatjuk be az előállított geometriai modellt, az orvos számára kialakított felhasználói felületet és az eddig megvalósított funkciókat.

Gravitáció és kristályosodás

Póliska Csaba¹, Gácsi Zoltán¹, Réger Mihály³

¹ Miskolci Egyetem Fémtechnológiai Tanszék
femtangz@uni-miskolc.hu

² Bánki Donát Műszaki Főiskola Gépészmérnöki Kar

Absztrakt. Dolgozatunk célja, hogy bemutassuk a gravitációs mező hatását a kristályosodás során kialakuló dendrites szerkezet jellemzőire. A kristályosítási kísérleteket három különböző irányban hajtottuk végre. Először vízszintesen, majd kétféle függőleges elrendezésben történt a dendritágak növesztése, úgy, hogy először fentről lefelé, azután pedig lentől felfelé mozgott a dermedési front. A mérések alkalmával a létrejött dendrites szerkezet kristályosítási paramétereit, a primer dendritág távolság, a dendritcsúcs sugár és a szekunder dendritágakra jellemző paraméterértékeket hasonlítottuk össze a kristályosítási irányok függvényében. Ennek meghatározására Quantimet 570C és a Quantimet 500 Image Workstation képelemző berendezéseket és számítógépes programokat használtunk. Vizsgáltuk az olvadékokban létrejövő folyamatokat is. Megpróbáltuk láthatóvá tenni az áramlásokat és a hőmérséklet- ill. koncentráció-gradienst is.

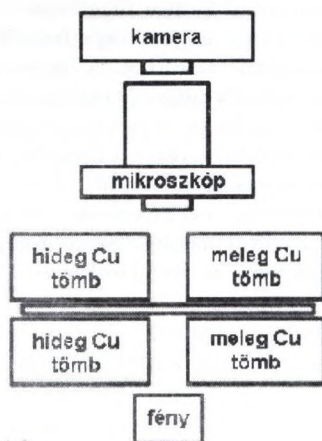
1. Bevezetés

A kísérletekhez használt modellanyag a succinonitril (továbbiakban SCN), amely több előnyös tulajdonsággal rendelkezik, így például alacsony az olvadáspontja (57-58°C), a folyékony SCN átlátszó, megdermedve pedig áttetsző. Ezen tulajdonságok miatt a mérés könnyen kivitelezhető, hiszen egyszerű kísérleti berendezést igényel, illetve az áttetszősége miatt a kristályosodás folyamatai könnyen láthatóvá tehetőek. A primer és szekunder dendritág távolságokat és a dendritcsúcs sugarakat állandósult állapotban mértük, ami azt jelentette, hogy a szilárd-olvadék határfelület mozgási sebessége állandó volt. Kétféle sebességet alkalmaztunk: 0,0015 mm/s, ami 0,005 K/s hűlési sebességnek felel meg, a szekunder dendritágakkal kapcsolatos mérésekhez, 0,013 mm/s, ami 0,045 K/s hűlési sebességnek felel meg, a többi paraméter méréséhez. A hőmérsékleti gradiens átlagos értéke mindkét esetben 3,23 K/mm volt.

2. Kristályosító berendezés

A mintaanyag dermedése egy hő átadó és egy hő elvonó egység között történik úgy, hogy a modellanyagot tartalmazó, üveglapokból összeállított mintatartót a két egység között áthúzzuk [Réger, 1997]. A kristályosodási folyamat a két tömb közötti résben mikroszkóppal megfigyelhető. Mindkét oldalon nagy hőkapacitású,

rézbetétes tömbök találhatóak, melyek állandó hőmérsékletéről folyamatos vízke-ringtetés gondoskodik. Mindkét oldalon termosztatok szolgáltatják a kiválasztott hőmérsékletű vizet. A hőmérsékleti gradiens a tömbök hőmérsékletével és egymástól mért távolságával, a hűlési sebesség pedig elsősorban a húzási sebességgel szabályozható. A kristályosító tömbök között, a húzási irányra merőlegesen helyezkednek el a kristályosodás optikai megfigyeléséhez szükséges eszközök. Ez a rendszer a kristályosító alatt megvilágító egységet, fölötte pedig mikroszkópot és egy hozzá csatlakozó kamerát tartalmaz. A kamera közvetlenül egy videomagnóhoz kapcsolódik. Az egész optikai rendszer elmozdítható a kristályosítóhoz képest, így a tömbök közötti tartomány bármely része megfigyelhető, a kívánt nagyításban. A kristályosító berendezés elvi felépítését mutatja az 1. ábra.

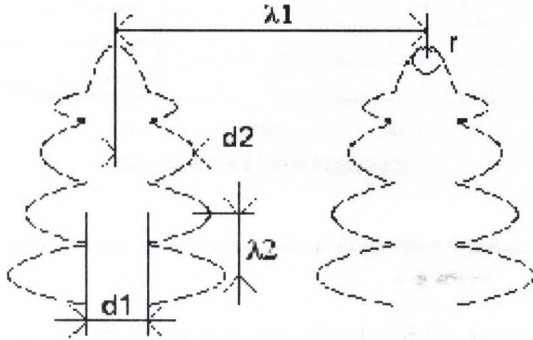


1. ábra: Kristályosító berendezés

3. Mérési módszer

Első lépésként mérésre alkalmas képeket kellett készíteni. A képek átalakítása úgy történt, hogy mindig a mérendő jellemző szempontjából fontos képrészleteket emeltük ki. Ennek megfelelően a különböző paraméterek méréséhez a képeket különbözőképpen alakítottuk át. A képek felvétele a már fentebb említett kristályosító rendszerhez kapcsolt optikai eszközök segítségével történt. A kristályosodás menetét folyamatosan rögzítettük videoszalagra majd az elkészített videofelvételből képeket vágtunk ki, hogy a méréseket elvégezhessük. Több száz képet mentettük el, azonban a képek alakja és mérete nem volt megfelelő a képelemzővel történő elemzéshez, ezért szükség volt a képek négyzetesítésére és nagyítására. A képek végső mérete 512*512 pixel lett, 8 bites felbontásban,

ami azt jelenti, hogy 256 szürkességi szintből állt egy kép. Ez a képméret pontosan megfelelő volt a képelemző berendezés számára. A képeken a mérendő paramétereknek megfelelően különböző átalakításokat végeztünk, amelyek segítségével a dendritágakat, dendritcsúcsokat jobban ki lehetett emelni a környezetükből és így a képeket alkalmasabbá tettük a mérés céljaira. A mért paramétereket a 2. ábrán mutatjuk be.

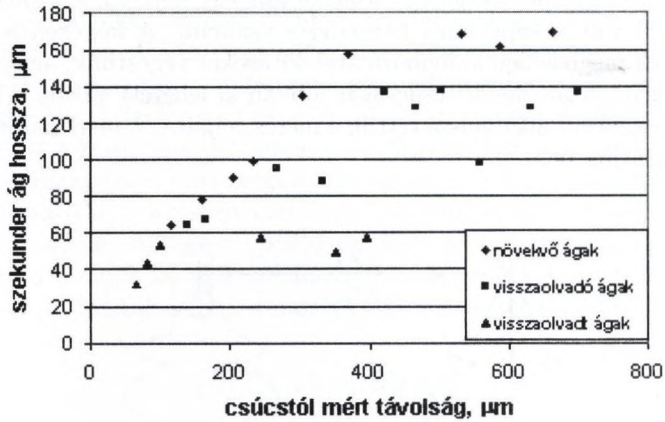


2. ábra: A mért paraméterek

A 2. ábrán található jelölések magyarázata: $\lambda 1$: primer dendritág távolság, $\lambda 2$: szekunder dendritág távolság, $d1$: törzsátmérő, $d2$: a szekunder dendritág hossza, r : dendritcsúcs sugár.

4. Dendrites szerkezet jellemzése

A méréseket - mindhárom kristályosítási irányban - 10-10 egymást követő képen végeztük el. A képeken manuálisan mértük meg a szekunder dendritágak hosszát a törzsátmérővel együtt. A kivágott képek között eltelt idő 2 másodperc volt. A mérés első lépéseként a primer dendritágot megfeleztük a "törzse" mentén. Ezután az egyedi szekunder dendritágak csúcsaitól kiindulva ehhez a vonalhoz képest mértük meg a hosszukat [Dahlström, Fredriksson, 1996]. Minden képen az első 25 szekunder dendritág hosszát mértük meg. Mint ahogy a 3. ábrán is látható, a távolsáértékek három részre oszthatóak. Az első csoportba tartoznak azok az ágak, melyek növekedését semmi sem akadályozza. Ezen ágak száma a csúcstól való távolság növekedésével egyre jobban csökken, ezáltal egyre nagyobb lesz a szekunder dendritág távolság. A második csoportba azok az ágak tartoznak, amelyek már elkezdtek visszaolvadni, de a szomszédaikhoz képest még jelentős hosszal bírnak. Ezen ágak visszaolvadása — tapasztalataink alapján — egyre jobban felgyorsul. A harmadik csoport szekunder dendritágai már majdnem visszaolvadtak, ezért hosszuk nem sokkal haladja meg a törzsátmérő felét.

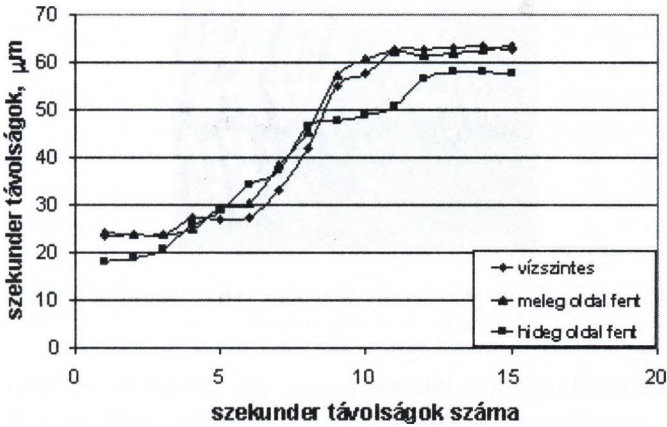


3. ábra: A szekunder dendritágak hossza a csúcstól mért távolság függvényében

Ahogy várható volt, a különböző irányokban kapott mérési eredmények nem egyformák. Tendenciájuk azonban megegyezik a primer dendritágak mérésénél tapasztaltakkal, mint ahogy erről már a korábbiakban [Póliska, Kovács, Gácsi, 2000] beszámoltunk. A legkisebb szekunder dendritág hosszértékeket akkor kaptuk, amikor a dendritek fentről lefelé nőttek, míg a legnagyobb értékek a vízszintes kristályosítás esetén jelentkeztek. Következő lépésként megpróbáltuk meghatározni a kialakuló valódi szekunder dendritág távolságokat. Ennél a mérési folyamatnál csak azon szekunder ágak távolságát vettük figyelembe, melyek hossza az azt megelőzőénél nagyobb volt, vagy ha nem is volt nagyobb, valószínűnek tűnt, hogy az ág nem fog eltűnni. Miután az előzőleg megmért tíz kép mérési eredményeit átalakítottuk, összesítettük az eredményeket. Készítettünk egy diagramot, melyen az első 15, folyamatosan növekvő szekunder dendritág távolságát ábrázoltuk. Mint ahogy a 4. ábrán látszik, a szekunder ágak távolsága a növekedés után egy állandó értékre áll be.

5. Olvadékokban lejátszódó folyamatok

A gravitáció a folyadékok és gázok mozgását alapvetően befolyásolja. A hőmérséklet növekedésének a hatására a folyadék kitágul, és ennek következtében sűrűsége lecsökken. [Bárczy, 1999]. Hűléskor ennek az ellenkezője történik, a sűrűség megnő. Sűrűségkülönbséget okoz minden koncentrációváltozás is, ami a kristályosodási folyamatokat kíséri. Gravitációs térben a nagyobb sűrűségű rész lefelé, a kisebb sűrűségű rész pedig felfelé mozdul el. Ezt az áramlást nevezzük konvekciónak. A folyadékok és gázok esetében a konvekció állandóan működik gravitációs térben. A létrejövő áramlások szabályszerűségeinek a megállapítása nagyon bonyolult, az esetek túlnyomó részében megoldhatatlan probléma. Még a legegyszerűbb áramlási feladat sem oldható meg teljes mértékben. Va-



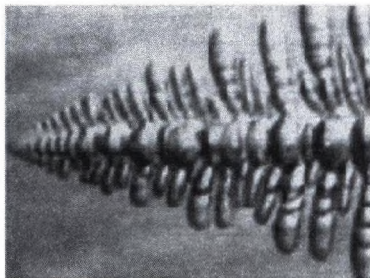
4. ábra: A szekunder dendritág távolságok változása

lóságos körülmények között pedig az igen élénk kölcsönhatások miatt a viszonyok még bonyolultabbá válnak. A konvekció jelentősége annál nagyobb, minél nagyobbak az anyagban a hőmérséklet- és koncentráció-gradiensek. Ilyen eset a kristályosodás is, ahol ráadásul a kristályosodási front stabilitását éppen a megfelelően nagy hőmérséklet-gradienssel lehet beállítani. Mivel az SCN transzparens anyag, megfelelően kicsiny szilárd részecskék olvadékba juttatásával lehetségesnek tűnik az áramlások kimutatása.

6. Koncentráció- és hőmérséklet-gradiens

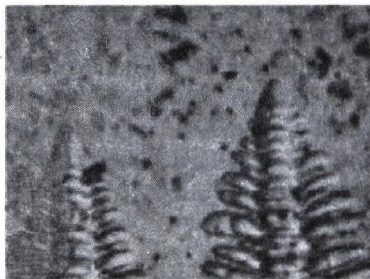
Elsőként a szilárd-olvadék határon fellépő koncentráció-gradiens kimutatásával próbálkoztunk. Egy adott hőmérsékleten a szilárd és az olvadék fázis koncentrációja nem egyezik meg. Bár az SCN-aceton rendszer esetén nem olyan hatalmas az eltérés, de reméltük, hogy ennek ellenére sikerül kimutatnunk a különbséget. Ahhoz azonban, hogy a koncentrációkülönbséget kimutathassuk, valamilyen olyan anyagot kellett találnunk, amely az SCN - acetone rendszernek csak az egyik összetevőjében oldódik, és oldódási koncentrációjától függően változtatja a színét. A szilárd halmazállapotú kobalt (III) - klorid + 6 H₂O vegyület megfelelőnek tűnt (5. ábra).

Annak ellenére azonban, hogy makroszkopikus méretben nagyon jól működött a színátmenet, felnagyítva — valószínűleg a mintatartó kis vastagsága miatt — nem volt tapasztalható jelentős színkülönbség a szilárd és olvadt részek között. Bár a szekunder dendritágak tövéénél látható némi színhatás, nem jelenthető ki teljes bizonyossággal, hogy az a színezőanyag hatásának tudható be. Megpróbáltunk olyan anyagot találni, amely csak a keverékünk egyik alkotójában oldódik és a hőmérséklet változásának hatására, változtatja a színét. Erre a célra a réz-klorid tűnt megfelelőnek. Az SCN - acetone keverékhez réz(I)-, ill.



5. ábra: A modellanyag (SCN+aceton) és kobalt (III) - klorid + 6 H₂O rendszer

réz(II) - kloridot adtunk. A mikroszkóp alatt azt tapasztaltuk, hogy a réz(II) — kloridnak nem változott a színe, a réz(I) — klorid pedig nem oldódott fel maradéktalanul a rendszerben (6. ábra). Maga a kristályosodási folyamat nem változott, de az oldatlan részek annyira zavaróak voltak, hogy mérhetetlenné vált bármilyen kristályosodási paraméter.



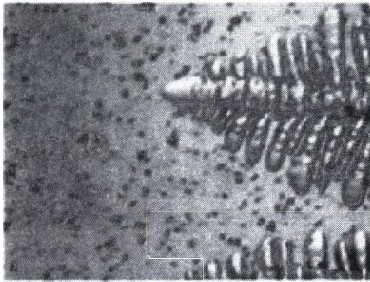
6. ábra: A modellanyag és réz(I) - klorid rendszer

Sajnos a színátmenettel is voltak problémák, mivel a mintatartó nagyon vékony — 0,15 mm — a makroszkopikusan nagyon élesen változó színek mikroszkopikus méretekben itt is jelentősen gyengültek. Ezen valószínűleg csak úgy lehetne változtatni, ha a mintatartó vastagságát növelnénk. A vastagítás hatására viszont több rétegben nőnének a dendritek, és így mérhetetlenné válna a mikroszkopikus kép.

7. Áramlások

Az SCN - aceton olvadékban kialakuló áramlások kimutatását úgy akartuk elérni, hogy az olvadékba olyan apró, szilárd részecskéket juttattunk be, melyek reményeink szerint az olvadékban kialakuló áramlásnak megfelelően mozogtak volna. Erre a célra négyféle anyagot próbáltunk ki: a ként, a talkumot, a grafitot és a polietilént-t (PE). Az elő kísérletek alapján a kén nem felelt meg a

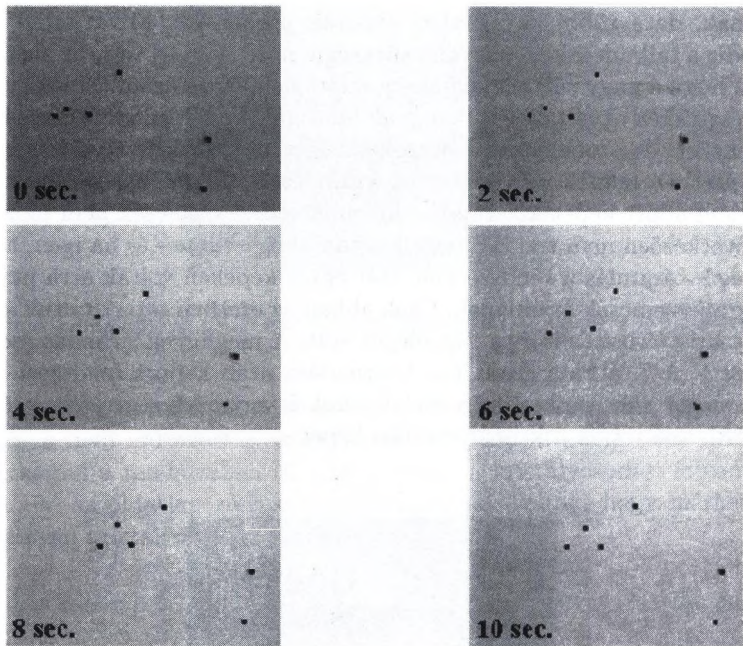
céljainknak, de a többi anyaggal is akadtak problémák. Mivel mind a grafit, mind pedig a talkum sokkal nagyobb sűrűségű, mint az SCN vagy az acetone, ezért féltünk, hogy a nagy sűrűségkülönbség miatt nem valós áramlási képet kapunk, vagy a részecskék egyáltalán nem fognak mozogni. Ennek ellenére összeállítottuk a keverékeket, és próbakristályosításokat végeztünk. A kristályosítás sebessége minden esetben 0,0015 mm/s volt. A grafit és a talkum bár bekerült a rendszerbe, nem adott képet az áramlási folyamatról. A szemcsék nem mozogtak és ebből következően nem tudtuk megállapítani, hogy van-e - és ha igen, akkor milyen jellegű - áramlás a rendszerben. Bár egyes képeken voltak arra utaló jelek, hogy a grafit szemcsék áramlanak. Csak abban az esetben tapasztaltuk áramlást, amikor a kristályosítás iránya függőleges volt. A megfigyelt áramlás esetlegesen jelentkezett. A 7. ábra a grafit por hozzáadása után kapott rendszert mutatja. Látható, hogy sok esetben a grafit részecskék egymáshoz tapadtak és eléggé átláthatatlanná tették a kristályosodási képet.



7. ábra: A modellanyag és grafit rendszer

A PE sűrűsége 0,91-0,93 g/cm^3 , szemnagysága pedig 0,09 és 0,06 mm között volt. Ezen tulajdonságai miatt sokkal kedvezőbbnek tűnt az előző két anyagnál, de már a keverék elkészítésekor adódtak gondok. A rendszer megolvasztása után a PE felúszott az olvadék tetejére. Bár a sűrűségkülönbség csak 0,05 g/cm^3 volt, a PE mégis elég élesen elvált a keverék másik két alkotójától. A szétválás megnehezítette ugyan az anyag mintatartóba való betöltését, de reméltük, hogy ez nem akadályozza meg az áramlások kimutatását. Először a vízszintes irányú kristályosítást hajtottuk végre, de abban az esetben nem tapasztaltunk semmilyen áramlást. Az apró PE részecskék bár láthatóak voltak az olvadékban, de nem mozogtak. Ezután következett a függőleges irányú kristályosítás és itt a részecskék már mozogtak. A függőleges irányú kristályosításnál azt tapasztaltuk, hogy a PE részecskék nagy része a mintatartóban lévő olvadék tetejére úszott fel, a látótérben maradt 10-15 db PE részecske viszont az olvadékáramlás hatására változtatta a helyzetét (8. ábra). Valószínűsíthető, hogy a részecskék mozgása az olvadékban lejátszódó konvekciós áramlás eredménye.

Mindkettő függőleges irányú kristályosításnál tapasztaltunk olvadékmozgást, az 8. és 9. ábra képeinél a kristályosítási irány a gravitáció irányával megegyező - a

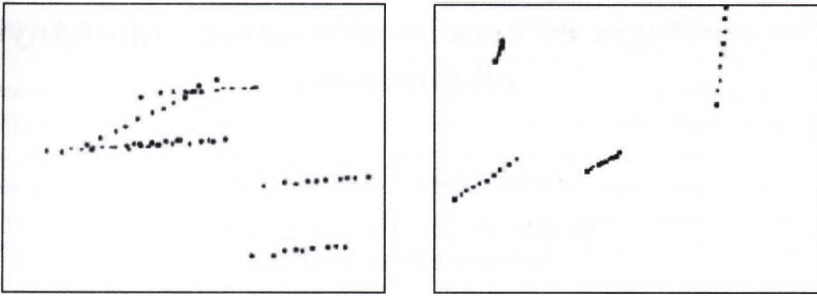


8. ábra: A PE részecskék áramlása

képeken jobbról balra mutat. A 9. ábra képeit úgy készítettük, hogy a mintatartó két különböző helyéről, a szilárd olvadék front közelében, 2 másodperces időkülönbséggel képeket készítettünk, melyeket egymásra másoltunk. Az így kapott ábrák az egyedi részecskék mozgásának az útját mutatják az olvadékban.

8. Összefoglalás

Mérési eredményeink alátámasztották azt a feltételezést, hogy a legdurvább dendrites szerkezet a vízszintes irányú kristályosítás során tapasztalható. A kristályosító berendezés sajátos elrendezése miatt, a dendritek mindig a hossz tengely irányába nőnek, és ebben az irányban alakul ki a koncentráció- és hőmérsékletkülönbségből adódó sűrűségkülönbség. Tekintettel arra, hogy vízszintes irányú kristályosításnál a gravitációs vektor merőleges erre a sűrűségkülönbségre, nem jön létre jelentős mértékű áramlás. Ezzel ellentétben függőleges irányú kristályosításnál a gravitáció hatására a "nehezebb" és "könnyebb" olvadékrészek konvekciós áramlást indítanak el, amelyet sikerült láthatóvá tennünk. Eredményeinket igazolták más szerzők [Weiss, Zimmermann, Rex, 1999] mérései is. Az olvadékban jelenlévő folyamatok láthatóvá tétele sikerrel járt, az áramlásokat kimutattuk függőleges irányban, és biztató kísérleteket folytattunk a koncentráció- és hőmérséklet-gradiens kimutatásával kapcsolatban is.



9. ábra: Az áramló PE részecskék útja

Irodalom

1. M. E. Glicksman: Met. Trans. A, Vol. 7A, pp. 1747-1759, November, 1976.
2. R. A. Brown: Convection and Bulk Transport, A Contribution to the Scientific Basis of Space Processing, Materials Sciences in Space, 1994.
3. S. H. Han, R. Trivedi: Acta Metall. Mater., Vol. 42, pp. 25-41, 1994.
4. J. Dahlström, H. Fredriksson: Experimental Study of the Development of Sidearms in Dendrites, Materials Science Forum Vols. 215-216 pp. 235-242, 1996.
5. Réger M.: Tranziens kristályosodási folyamatok kísérleti vizsgálata, Budapest, 1997.
6. Bárczy P.: Mikrogravitációs kutatások a Miskolci Egyetemen, Hungary, Kohászat, pp. 269-274, 1999.
7. Cs. Póliska, K. Tomolya, J. Kovács, Z. Gácsi, M. Réger: Effect of the Solidification Direction to the Dendritic Structure, Third International Conference on Solidification and Gravity, Hungary, Miskolc-Lillafüred, pp.III/180, 1999.
8. Cs. Póliska, K. Tomolya, J. Kovács, Z. Gácsi, M. Réger: Effect of Direction of Solidification on the Dendritic Structure, Materials Science Forum Vols. 329-330, pp. 291-296, 2000.
9. Póliska Csaba, Kovács Jenő, Gácsi Zoltán: Dendrites szerkezet jellemzése, Képfeldolgozók és alakfelismerők 2. Konferenciája. Noszvaj, 163-168 old., 2000.
10. A. Csontos, P. Bárczy, Gy. Czél: Investigation of Liquid Heat Diffusivity by Direct Temperature Distribution Measurements under Microgravity, Space Forum, Vol. 6, pp. 153-157, 2000.
11. Póliska Csaba: A kristályosodás "in situ" vizsgálata, Diplomamunka, Miskolci Egyetem, 2001.

Sztereológia és képelemzés anyagtudományi alkalmazása

Gácsi Zoltán¹, Sárközi Gábor

Miskolci Egyetem Fémteni Tanszék
femtangz@uni-miskolc.hu

Absztrakt. Az összetett anyagok térbeli szerkezetének modellezésével és a mikroszkópos szövetképek számítógépes feldolgozásával kapcsolatos ismeretek kutatása mindenütt jelen van, ahol az anyagok szerkezetét mikroszkóppal vizsgálják. Viszonylag új, interdiszciplináris tudományterületről van szó, amelynek jelentősége az 1980-as évektől kezdve fokozatosan növekszik, s napjainkra az anyagok szerkezetével kapcsolatos kutatások egyik kulcsterületévé vált. A dolgozat röviden áttekinti a sztereometrikus mikroszkópia matematikailag megalapozott elméletét. Tárgyalja az anyagtudományi szempontból kiemelkedő jelentőséggel bíró mikro-szerkezeti modellek között a viszonylag új keletű fraktál modelleket. A digitális képfeldolgozás és a számítógépes képelemzés tárgyalása az anyagtudományi problémák köré épül. A képi transzformációkat, a bináris képek létrehozását és a bináris műveleteket az anyagtudomány gyakorlati feladatainak keresztül illusztrálja. Bemutatja a karbonszálakkal erősített fémmátrixú kompozitok mikro-szerkezetét; transzparens anyagok kristályosodásakor kialakuló dendrites szerkezet jellemzési lehetőségeit; töretfelületek jellemzését; valamint kerámia szemcsék eloszlását alumínium alapú kompozitban.

1. Sztereometrikus mikroszkópia fogalma

Az anyagtudomány művelőinek legfontosabb célja az anyagok térbeli szerkezete és fizikai tulajdonságai közötti kapcsolat megismerése és annak tudatos használata. Ilyen típusú összefüggések alkalmazásával ugyanis lehetővé válik az adott célra legjobban használható anyag kiválasztása és előállítás. Nem nélkülözhető ez a módszer az anyagtulajdonságok modellezésekor, vagy éppen a számítógéppel segített tervezés, gyártás illetve az anyagminősítés területén sem. A sztereometrikus mikroszkópia ('stereology') feladata az anyagok mikroszkópon megfigyelhető szerkezetének geometriai leírása és annak térbeli interpretációja. Tulajdonképpen segédeszköz, amely egzakt formában fogalmazza meg a vizsgálandó paraméterek és a számított mennyiségek közötti összefüggéseket. A sztereometrikus mikroszkópia elméleti alapjaival a matematikai morfológia foglalkozik. A sztereológia ugyanakkor a gyakorlati alkalmazások felé fordul: művelői sikerrel alkalmazzák a differenciálgeometria elemeit az anyagtudományi szerkezetek jellemzésére. A sztereológiai problémákat különböző módon közelíthetjük meg. Egyik lehetőség a statisztikai-geometriai módszer, amely nagyszámú két-dimenziós mérést, s az adatok feldolgozását jelenti. Ez akkor használható, ha

az anyagot alkotó fázisok geometriai-alakzatai véletlen eloszlásúak a térben. Ebben az esetben egy metszet vagy vetített kép elegendő, feltéve, ha a metszet statisztikailag elegendő számú alakzatot tartalmaz. Ha az objektumok hasonló alakúak és a térben szabályos elrendeződésűek (ezek az anizotróp szerkezetek) akkor sok különböző irányú metszet vagy vetített kép szükséges. A metszési síkoknak a térben véletlenszerűen kell elhelyezkednie, vagy pedig szisztematikus, párhuzamosan elhelyezkedő síkokra (ún. sorozat metszetekre) van szükség, az alakzatok orientációjától függően. Abban az esetben, ha az objektumok bonyolult alakúak, a rekonstrukciójukhoz szintén sorozatmetszetet használunk. Hasonló technikát alkalmazunk akkor is, ha csak egy, vagy néhány térbeli alakzatunk van. Szttereometrikus mikroszkópia alapösszefüggései A szttereometrikus mikroszkópia dilemmája, hogy amennyiben a térbeli mennyiségeket kétdimenziós metszeten vizsgáljuk, milyen egzakt összefüggések léteznek a vizsgálandó fázisra jellemző pontok száma (P), a jellemző vonalak hosszúsága (L), a felületek nagysága (S) és az alakzatok térfogata (V) között. Az leggyakrabban használt mért és számított mennyiségeket az 1. ábrán hasonlítjuk össze. A táblázatban valójában olyan arányok szerepelnek, amelyeknek a számlálójában a vizsgálandó geometriai alakzat különböző mennyiségei (jellemző pontjainak darab száma: P, vonalhosszúsága: L, területe: A, térfogata: V) a nevezőjében pedig a vizsgált próba adatai vannak. Így például $S_V = S/V$, ahol az S jelenti az alakzatok felszínét, míg V a vizsgált próba térfogatát, ennek a mennyiségnek a neve fajlagos felület.

Az alakzatok típusa	A geometriai mennyiségek mértékegysége			
	mm ⁰	mm ⁻¹	mm ⁻²	mm ⁻³
Pontok	(P _P)	(P _L)	(P _A)	{P _V }
Vonalak	(L _L)	(L _A)	{L _V }	
Felületek	(A _A)	{S _V }		
Térfogatok	{V _V }			

1. ábra: A mért () és a számított { } geometriai mennyiségek összehasonlítása

Hasonlóan a P_P mennyiséget (pontarányt) úgy származtathatjuk, hogy $P_P = P_\alpha/P$, ahol a számlálóban lévő P jelenti a mikro-szerkezeti alakzatba eső pontok számát, míg a nevezőben lévő P reprezentálja a vizsgált próba összes pontját. A táblázatban nyíllal jelöltük a leglényegesebb összefüggéseket, amelyek korrekt

matematikai bizonyítása meghaladja e közlemény kereteit (Underwood Ervin E.: Quantitative Stereology. Menlo Park, California. 1970. Addison-Wesley Publishing Company. 23 p.). A táblázat első oszlopában, valamint a második, a harmadik és a negyedik oszlopban levő mennyiségek közötti kapcsolatot a következő egyenletek mutatják:

$$V_V = A_A = L_L = P_P \quad (1)$$

$$S_V = \frac{4}{\pi} L_A = 2P_L \quad (2)$$

$$L_V = 2P_A \quad (3)$$

$$P_V = 1/2 L_V S_V = 2P_A P_L \quad (4)$$

A gyakorlati alkalmazhatóság szempontjából nagy jelentősége van annak, hogy ezek az összefüggések nem tartalmaznak semmilyen feltételezést vagy közelítést az alakzatok méretére, alakjára, távolságára vonatkozóan. Egyetlen feltétel, hogy a mérésnek véletlenszerűnek kell lennie, és statisztikailag reprezentálnia kell a teljes próbatestet.

2. Mikro-szerkezeti modellek

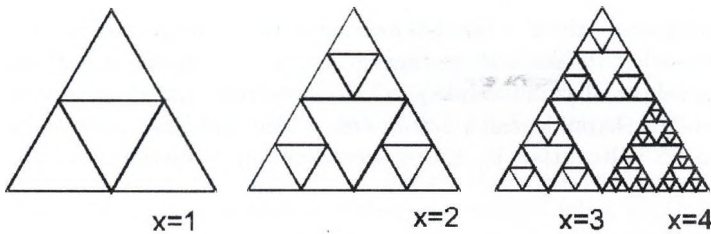
Az anyagok mikro-szerkezetének geometriai alakzatai a rendezettség tekintetében igen sokfélék. Szerkezetük a szinte tökéletesen rendezettől a majdnem teljesen rendezetlenig változhat. A fraktálokkal jellemezhető struktúrák a rend alacsony fokát képviselik, ha nem is teljesen rendezetlenek. Tipikus tulajdonságuk az önhasnoltság, ami azt jelenti, hogy bizonyos jellemvonásaik függetlenek a nagyítástól. Bár a fraktál viszonylag új keletű elnevezés (Mandelbrot B. B., Passoja D. E., Paullay A. J.: Fractal Characterization of Fracture Surfaces of Metals. Nature, 308 (1984) 721. p.), a fraktál - geometria alap gondolata nem új, ugyanis a részhalmazok geometriai elméletének kidolgozása a század elejére nyúlik vissza. Mandelbrot volt ennek a matematikai elméletnek az első gyakorlati alkalmazója: geográfiai, metallográfiai, áramlástani problémák megoldására. A fraktálok két csoportját különböztette meg: a szabályos (vagy geometriai), valamint a véletlen (vagy 'random') fraktálokat. A geometriai fraktálok előállítására meghatározott növekedési szabály segítségével történhet. Az így létrejött geometriai alakzat valamely tulajdonságának (pl. területének) logaritmususa a lépések száma logaritmusának függvényében lineáris. A véletlen fraktálok olyan rendszerek, amelyek bár nem szabályos geometriai alakzatok, de ha tulajdonságaik statisztikus átlagát képezzük, akkor hasonló linearitást kapunk. A síkgörbe és a felület fraktál - dimenziója (D) az alábbi összefüggésből adódik:

$$\ln L_x = \ln L_0 + (1 - D) \ln \epsilon_x \quad (5)$$

és

$$\ln S_x = \ln S_0 + (2 - D)\ln E_x \tag{6}$$

ahol: L_0, S_0 a fraktál hosszúsága illetve felülete az $x=0$ rekurziós lépésben, L_x, S_x a fraktál hosszúsága illetve felülete, ϵ_x, E_x a szegmens hossza. Egyenes esetén a fraktáldimenzió egy ($D=1$), míg a görbe összetettségének fokozódásával a kettőhöz tart. Felületek fraktál - dimenziója, analóg módon 2 és 3 között változik. Gyakorlati vizsgálatok során az ϵ_x, E_x a nagyítással változtatható, míg az L_x, S_x többféle geometriai adattal (kerület, felület, térfogat) kapcsolatba hozható. Az átalakulások jellemzésére is alkalmas a Sierpinsky - féle alakzat (2. ábra), amelyet úgy állítunk elő, hogy egyenlő oldalú háromszögek oldalait két egyenlő részre osztjuk ($n=1/2$), s a kapott szakaszok felhasználásával újabb háromszöget ($N=3$) rajzolunk az előző háromszög belsejébe ($D=\log 3/\log 2 \approx 1,59$).

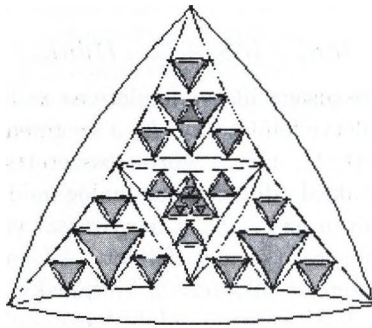


2. ábra: Sierpinsky - féle alakzat

A Sierpinsky-alakzat módosításával jól modellezhető a martenzit-kristályok több lépésben történő kialakulása (3. ábra). A fraktál tulajdonságot leíró paraméter a maradék ausztenit — martenzit határfelületének hossza, az átalakuló ausztenit szemcse átmérő függvényében. Az anyagtudomány területén számos "vonal" elemzése szükséges, ilyenek például a diszlokációk, amelyek "érdeessége" a pontszerű rácshibákkal és egymással történő kölcsönhatásaikból fakad. A vizsgálatok különösképpen nagy vakancia koncentrációjú, neutronsugárzásnak kitétt anyagokban és edzett ötvözetekben érdekesek (Hornbogen E.: Fractals in microstructure of metals. International Materials Reviews, 34. (1989) 6. 277. p.). A fémek nagyszögű szemcsehatárai a koherens (koincidencia-, iker-) határoktól az inkoherensig változnak. Teljesen újrakristályosított anyagban a szemcsehatárok majdnem síkok, görbületük minimális, fraktál - dimenziójuk kettőhöz közeli érték. Az ilyen típusú szemcsehatár sűrűségére vonatkozó sztereológiai összefüggés az alábbi formában érvényes (feltéve, ha ekvixiális szemcsékről van szó) Hornbogen E. (1989):

$$\rho_{hatar} = \frac{\sum A_{hatar}(\epsilon)}{V} = \frac{\pi^2}{4d} \tag{7}$$

ahol: ρ_{hatar} a szemcsehatár sűrűsége, mm^2/mm^3 , A_{hatar} a szemcsehatár felülete, mm^2 , V a szemcsék térfogata, mm^3 , d a szemcse átmérője, mm . A szemcsehatár

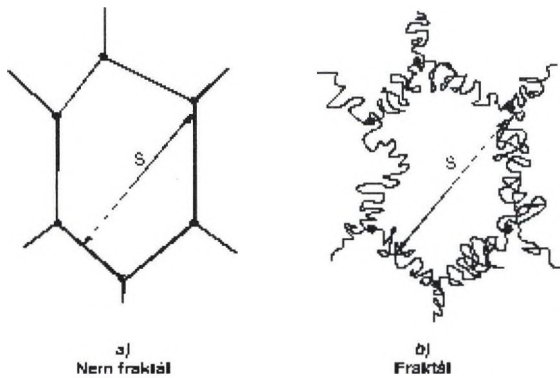


3. ábra: A martenzit (fehér) és a maradék ausztenit (szürke) sematikus ábrázolása

sűrűség felső határa abból a tényből származtatható, hogy a szemcse nem lehet kisebb, mint a két diszlokáció vastagságú határ. Így, amint d a $2b$ -hez tart (b - a Burgers-vektor abszolút értéke), a kristályos rend amorf fázissá alakul. Hasonló eredményt kapunk, ha a határ érdességét növeljük: még ha az átlagos szemcseátmérő változatlan is. Érdes szemcsehatárok esetén a valódi szemcsehatár-sűrűséget a

$$\rho_{hatar} = \frac{\sum A_{hatar}}{V} = \frac{L_{\epsilon}^2}{L_0^3} \quad (8)$$

kifejezés adja. L_0 az átlagos szemcseátmérő függvénye ($L_0 = d\pi/2$), míg L_{ϵ} az adott szegmenshossz ($b < \epsilon < d$), vagyis az adott nagyítás mellett mért határ hosszúság. Ekkor a szemcsehatár fraktál jellegűnek tekinthető.



4. ábra: Különböző típusú szemcsehatárok a) sík lapokkal határolt szemcsehatár, b) fraktál jellegű szemcsehatár kiválásokkal

További érdekes megfigyelésekhez vezethet a fraktáljellegű szemcsehatárok menti kiválások fémtani vizsgálata.

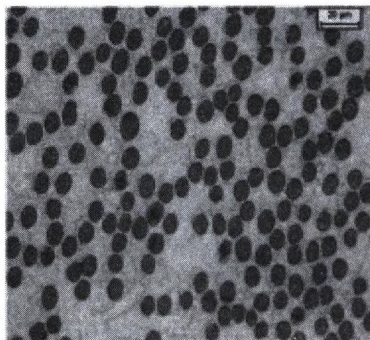
3. Digitális képfeldolgozás

Az anyagok szerkezetére jellemző csiszolatok, vagy metszetek mikroszkópos képeinek feldolgozásakor először a képek olyan típusú átalakítását végezzük el, amelynek segítségével lehetséges a lényeges- és a lényegtelen információk különválasztása. Ilyen átalakítás során lehet kiszűrni az elektronmikroszkópos képeken meglévő "zajt", vagy ezzel a módszerrel lehet a felvételek (vagy csiszolatok) készítése során keletkező "zavarokat" elválasztani az anyag szerkezetét bemutató részletektől. A következő lépés a képelemzés, ami nem más, mint a képek számszerű adatokkal történő jellemzése. Azt is mondhatjuk, hogy ekkor végezzük el a kép információ tartalmának mennyiségi interpretációját. Az így nyert paramétereket az adott környezetbe visszahelyezve, s a sztereológia módszereit felhasználva interpretálni kell, ez a képértelmezés. Tehát egy összetett folyamatról van szó, amelyik a következő főbb lépésekből áll (Gácsi Zoltán (szerk.): Sztereológia és Képelemzés. Miskolc. 2001. WellPress - PHARE):

1. A képek érzékelése és rögzítése. A képek elektronikus formában történő előállítása, valamilyen alkalmas eszközzel, amelyik lehet fénymikroszkópra szerelt videokamera, digitális fényképezőgép, vagy elektronmikroszkóp.
2. A szürkekép előállítása, digitalizálása. Szürkeségi szintek hozzárendelése a kép egyes intenzitás értékeihez vagy egyes színeihez.
3. A szürkeképek digitális feldolgozása. A vizsgálni kívánt jellegzetességek ('features'), vagy objektumok ('object') kihangsúlyozása, vagyis a lényeges és a kevésbé lényeges információk különválasztása.
4. Szegmentálás (detektálás). Az objektumok és a háttér teljes elkülönítése valamilyen tulajdonságuk (pl. szürkeségi szintjük, vagy alakjuk, orientációjuk) alapján, aminek eredményeképpen bináris kép áll elő.
5. Bináris képek átalakítása. Bináris transzformációk végzése, vagyis az objektumok olyan átalakítása, hogy leginkább megközelítsék ideális alakjukat (pl. a képen összetapadt, de a valóságban különálló szemcsék szétválasztása), a mérés előkészítése.
6. A mérés végrehajtása, a mérési eredmények előállítása.
7. A mérési eredmények értelmezése. Az anyagtudományban az eredményeket a különböző mikro-szerkezeti modellek alapján, s a sztereológia eszközrendszerének felhasználásával értelmezzük.

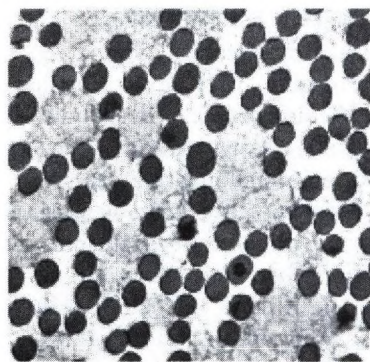
Az eddigiekben vázolt folyamatot gyakorlati példával szemléltetjük. Nevezetesen a nyomásos infiltrációval előállított karbon szálakkal erősített, alumínium mátrixú kompozit vizsgálatának bemutatásával (Szalai Ibolya: Hibridstruktúrák előállítása és szerkezetének vizsgálata. Diplomamunka. Miskolci Egyetem. 2000.).

A kísérletekhez használt különböző minőségű karbonszálakat a ZOLTEK Rt. állította elő, s cég a szálak egy részét felület kezelte, majd poliészterrel, illetve epoxi gyantával vonta be. A szálak felhasználásával a Miskolci Egyetem



5. ábra: Al/C kompozit, fénymikroszkópos felvétel (Szalai Ibolya)

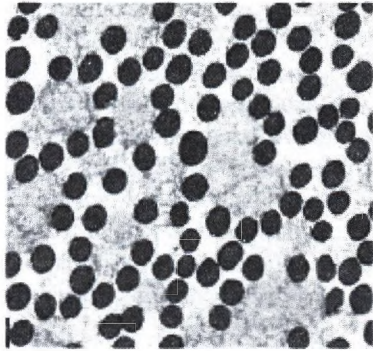
Fémtani Tanszékén, az ott kidolgozott eljárással (Szalai Ibolya, Gácsi Zoltán, Magyar Anita: Karbonszál-erősítésű, alumíniummátrixú kompozit előállítás és szerkezetének vizsgálata. Bányászati és Kohászati Lapok, Kohászat. Budapest. 133. évf. (2000). 279. p.) alumínium olvadékat infiltráltak a karbon szálak közé. Az így nyert kompozitotról készült, fénymikroszkópos felvétel a 4. ábrán látható.



6. ábra: A "White-Sharpener" morfológiai transzformáció

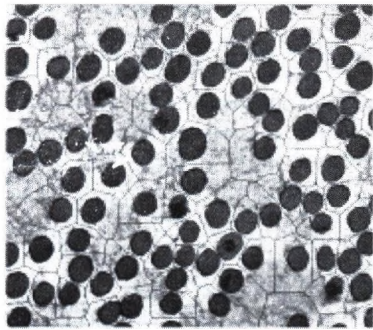
A vizsgálni kívánt karbon szálak környezetét egy komplex morfológiai transzformáció segítségével, a "White-Sharpener" művelettel (Gácsi Zoltán (2001)) sikerült élesebbé tenni (6. ábra).

Ezt követte a detektált kép előállítása. Mivel a detektált képen voltak olyan szálak is, amelyek nem különültek el egymástól, ezért a szálak középpontjainak megkeresésével ('ultimate erosion'), majd a szálak méretének eredetire való növelésével lehetett azokat elválasztani. Ezt követte a karbon szálak közötti területek lecsök-kentése egy képpont szélességűre, vagy másképpen a szálak dilatációja azzal a feltétellel, hogy azok ne érjenek össze. A műveletet a maximális



7. ábra: A bináris átalakítások végeredménye

iterációs számig végezve és eltávolítva a vázszerkezet oldalágait ('inverse pruning'), hatósugár szerinti inverz vázszerkezetet állítottak elő SKIZ ('inverse skeleton by influence zone') eljárással. Ezáltal minden karbon szálhoz hozzá lehetett rendelni egy az alapanyagból elfoglalt tartományt.



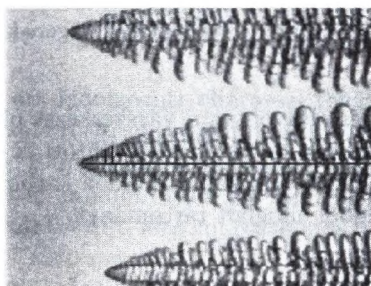
8. ábra: A hatósugár szerinti vázszerkezet

A karbon szálakhoz rendelhető "hatósugár" méretének és szórásának meghatározásával sikerült igazolni, hogy a legegyszerűsebb szál eloszlást, ami az optimális mechanikai tulajdonságú kompozit előállításának az alapja, az epoxi gyantával bevont karbon szálakkal lehet elérni. A szál térfogatarányának meghatározása igazolta, hogy a felületkezelt karbon szálak alkalmazása kedvezőbb tulajdonságú kompozitok kifejlesztéséhez vezet (Szalai (2000)).

4. Anyagtudományi alkalmazások

A kristályosodási folyamatok modellezésekor egyre jobban előtérbe kerülnek az "in situ" vizsgálati eljárások. A legkézenfekvőbb módszer olyan anyag használata,

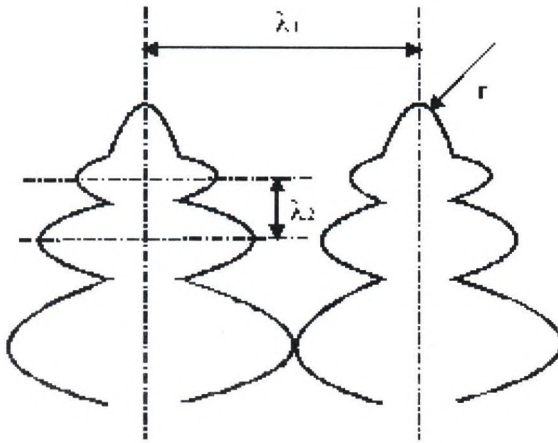
amely transzparens. Ekkor a folyamat hagyományos eszközökkel — kamerával, mikroszkóppal — közvetlenül megfigyelhető. Ilyen típusú kísérletekkel lehet egyszerűen tanulmányozni a gravitációs mező hatását a kristályosodás közben kialakuló olvadékáramlásra. Amennyiben megváltoztatják a szilárd-olvadék határfelület mozgási irányát a gravitációs vektorhoz képest, azt lehet tanulmányozni, hogy kristályosodás közben az olvadékban szükségképpen kialakuló sűrűségkülönbség indukál-e olvadékáramlást, vagy pedig a folyamatot "tisztán" a diffúzió irányítja. A kísérleteknél modellanyagként világszerte succinonitrile (SCN) - aceton bázisú minta anyagot használnak, amely azért kedvelt, mert folyékony és szilárd állapotban egyaránt átlátszó, valamint olvadáspontja igen alacsony.



9. ábra: Növekvő SCN dendrit (Póliska Csaba, Réger Mihály)

A bemutatott kísérletben a berendezésre egy videokamera is fel van szerelve, amelynek segítségével a kristályosodás teljes lefolyását videoszalagra lehet rögzíteni a későbbi feldolgozás céljából (Cs. Póliska, K. Tomolya, J. Kovács, Z. Gácsi, M. Réger: Effect of Direction of Solidification on the Dendritic Structure. 3rd International Conference on Solidification and Gravity 2000, Miskolc-Lillafüred. Materials Science Forum, 329-330 (2000). 291. p.).

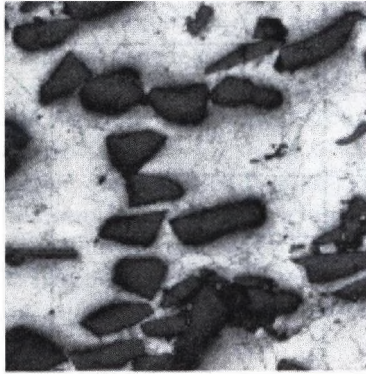
Összehasonlító vizsgálatokat úgy végeztek, hogy a minta anyagot három különböző módon kristályosították, az alkalmazott berendezést Réger Mihály fejlesztette ki (Réger Mihály: Tranzien kristályosodási folyamatok kísérleti vizsgálata. PhD értekezés. 1997.). A szilárd-olvadék front, i) vízszintesen mozgott, ii) fentről lefelé mozgott, iii) lentől felfelé mozgott. Az esetleges olvadék áramlás hatását, a kialakuló dendrites szerkezet (9. ábra) jellemzőire a digitális képfeldolgozás módszerével tanulmányozták. A Miskolci Egyetem Fémteni Tanszékén kifejlesztett számítógépes képelemző eljárással, valamint egy kiegészítő szoftverrel reprodukálható módon határozták meg (10. ábra) a primer dendritágak távolságát (λ_1), a dendritcsúcs sugarát (r), valamint a szekunder dendritágak távolságát (λ_2). Bebizonyították, hogy a különböző irányban elvégzett kristályosítások esetén a dendrites szerkezet különböző. A primer dendritág távolságok mérése során vízszintes irányban tapasztalták a legnagyobb távolságot ($\lambda_1 = 32\mu\text{m}$) és függőleges irányban a legkisebbet, akkor amikor a dendritek fentről lefelé nőttek ($\lambda_1 = 25\mu\text{m}$). A jelenség magyarázata abban rejlik, hogy a függőle-



10. ábra: A dendrites szerkezet jellemző paramétereit

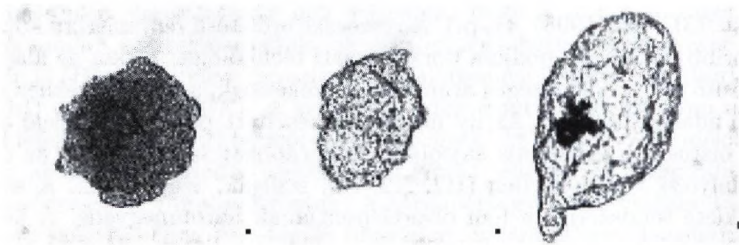
ges kristályosítás során a sűrűség különbség hatására olvadákáramlás alakul ki, ami növeli a szilárd/olvadék front mozgási sebességét és így csökkenti a primer dendritág távolságot. Vízszintesen növekvő dendritok esetén, részben a vékony próba tartó miatt, ez a hatás nem érvényesül. Ezzel a módszerrel tulajdonképpen szimulálni lehet az anyagok mikro-gravitációs viselkedését. Kerámia szemcsékkel erősített fémmátrixú kompozitoknak ('Particle Reinforced Metal Matrix Composites'), olyan fémmátrixú anyagokat nevezünk, amelyek 5-100 μm méretű, homogén eloszlású kerámia szemcséket (SiC, WC, Al₂O₃) tartalmaznak, mintegy 5-40 térfogat-százaléknyi mennyiségben. A szemcsés kompozitokat elsősorban szokatlan tulajdonság kombinációk megvalósítására (pl. kis sűrűség, nagy szilárdság) állítják elő (Csepeli Zsolt, Sólyom Balázs, Gács Zoltán, Buza Gábor, Teleszky Ilona, Kovács Árpád: Részecske- és szálerősítésű fémmátrixú kompozitok előállítási lehetőségei. Bányászati és Kohászati Lapok, Kohászat. Budapest. 131. évf. (1998). 41. p.). A részecske erősítésű fémmátrixú kompozitok leggyakoribb előállítási módja a porkohászati technológia. Ekkor az alapanyagot és az erősítő fázist a szükséges arányban összekeverik, a porkeverékhez sajtoltat könnyítő adalékot adnak. Az ily módon előkészített port a megfelelő alakot és méretet biztosító szerszámba sajtolják. A darabokat sajtolás után az oxidációt megakadályozó védőközegben (H₂, N₂, Ar) izzítják, zsugorítják. A szinterelés hőmérséklete rendszerint a fém olvadáspontjának háromnegyede. A kellő ideig történő izzítás után porózus (11. ábra), de mégis kellően szilárd anyagot nyernek (Szigethy Magdolna: Részecske erősítésű alumínium mátrixú kompozitok előállítási lehetőségei. Diplomamunka. Miskolci Egyetem. (1998)).

A Miskolci Egyetem Fémtani Tanszékén 5%- , 10%- és 15% SiC tartalmú, két különböző szemnagyságú (D^{15} és 7 μm) alumínium mátrixú kompozitot vizsgáltak (Tomolya Kinga: SiC erősítésű kompozit előállítása porkohászati úton. Diplomamunka. Miskolci Egyetem. 2001). A kísérletek során a próbatesteket



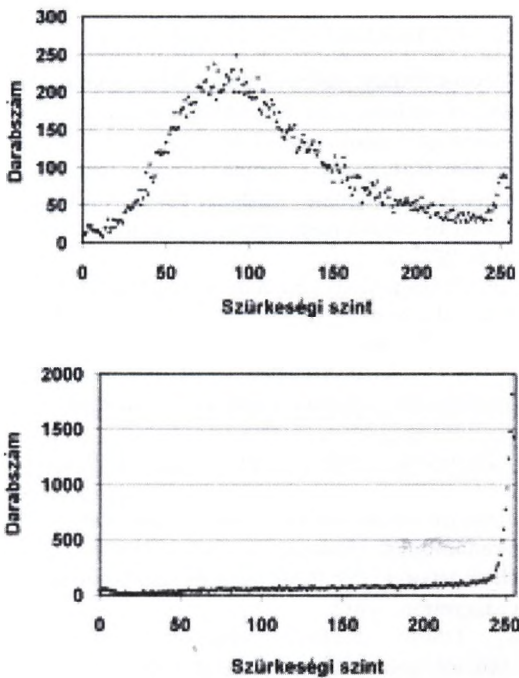
11. ábra: Porózus Al/SiC kompozit (Szigethy Magdolna)

három különböző hőmérsékleten szinterelték: 580 °C, 610 °C, és 640 °C. Megállapították, hogy a SiC erősítésű alumínium mátrixú kompozit egyik legfontosabb, a mechanikai tulajdonságokat meghatározó paramétere a porozitás. A porozitást képelemző segítségével viszonylag egyszerűen meg tudták határozni. A megfelelő nagyítású mikroszkópos képek rögzítését követően a porozitás detektálása, majd a területarány mérése következett. Vizsgálták a SiC mennyiség, a SiC szemcse-nagyság és a szinterelési hőmérséklet porozításra gyakorolt hatását. Azt találták, hogy a hőmérséklet emelése növeli a darabok zsugorodását, s csökkenti a porozitást. Az 5%-os SiC erősítésnél a kisebb szemcse-nagyságú darab volt tömörebb, míg a 10% és ennél nagyobb mértékű SiC erősítés esetén már nincs lényeges különbség a különböző szemcse-nagyságú darabok viselkedése között. Az eljárással a darab keresztmetszetében a porozitás eloszlása is jellemezhető volt. A porozitás képelemzővel történő meghatározása nagy segítséget jelentett az optimális gyártási technológia kialakításához.



12. ábra: Különböző tömörségű PVC szemcsék

Különböző tömörségű PVC szemcsék vizsgálata (Gácsi Zoltán (2001)). Közvetlenül a szürke képeken végzett mérésnek azért van jelentősége, mert néhány esetben az elemezni kívánt fázisokra nem méretük, vagy alakjuk, hanem kép-



13. ábra: Szürkeségi hisztogramok

pontjaiknak a szürkeségi szintek szerinti eloszlása a jellemző. A 12. ábra három különböző típusú PVC szemcsét mutat. Ezek közül az egyik rendkívül porózus (egyenletesen szürke) a), a második rendkívül tömör, ezért az átvilágításos mikroszkópon áttetsző (világos) b), míg a harmadik típusú por szemcséiben nagyméretű üregek találhatók, aminek következtében a szemcsén egy vagy több fekete foltot látunk c). Ezek a jellegzetességek bináris képeken nem jellemezhetők, ellenben nagyon jól megfigyelhetők az objektumok szürkeségi hisztogramjain. Az adott kép szürkeségi hisztogramja nem más, mint a különböző szürkeségi értékek előfordulásának relatív gyakorisága. A 13. ábra felső részén a porózus szemcse (12. a) ábra), míg alsó részén a tömör szemcse (12. b) ábra) hisztogramja látható. Amennyiben a szemcsében üreg van, úgy megjelenik a sötétebb fázisra jellemző csúcs is. Nagyon jellemző az adott szemcsére az átlagos szürkeségi érték, a minimális és a maximális szint, valamint a szórás értéke is. A különböző tömörségű szemcsék ezzel a módszerrel gyorsan és objektíven jellemezhetők.

Megjegyzés

A "III. Országos Anyagtudományi, Anyagvizsgálati és Anyaginformatikai Konferencia és Kiállítás, (Balatonfüred, 2001. Október 14-17.)" előadásának átdolgozott változata.

Irodalom

1. Csepeli Zsolt, Sólyom Balázs, Gácsi Zoltán, Buza Gábor, Teleszky Ilona, Kovács Árpád: Részecske- és szálerősítésű fémmátrixú kompozitok előállítási lehetőségei. *Bányászati és Kohászati Lapok, Kohászat*. Budapest. 131. évf. (1998). 41. p.
2. Gácsi Zoltán, Sárközi Gábor, Réti Tamás, Kovács Jenő, Csepeli Zsolt, Mertinger Valéria: *Sztereológia és Képelemzés*. Miskolc. 2001. WellPress - PHARE.
3. Hornbogen E.: Fractals in Microstructure of Metals. *International Materials Reviews*, 34. (1989) 6. 277. p.
4. Mandelbrot B. B., Passoja D. E., Paullay A. J.: Fractal Characterisation of Fracture Surfaces of Metals. *Nature*, 308 (1984) 721. p.
5. Pólska Cs., Tomolya K., Kovács J., Gácsi Z., Réger M.: Effect of Direction of Solidification on the Dendritic Structure. 3rd International Conference on Solidification and Gravity 2000, Miskolc-Lillafüred. *Materials Science Forum*, 329-330 (2000). 291. p.
6. Réger Mihály: *Tranziens kristályosodási folyamatok kísérleti vizsgálata*. PhD értekezés. 1997.
7. Szigethy Magdolna: *Részecske erősítésű alumínium mátrixú kompozitok előállítási lehetőségei*. Diplomamunka. Miskolci Egyetem. (1998).
8. Szalai Ibolya: *Hibridstruktúrák előállítása és szerkezetének vizsgálata*. Diplomamunka. Miskolci Egyetem. 2000.
9. Szalai Ibolya, Gácsi Zoltán, Magyar Anita: *Karbonszál-erősítésű, alumíniummátrixú kompozit előállítása és szerkezetének vizsgálata*. *Bányászati és Kohászati Lapok, Kohászat*. Budapest. 133. évf. (2000). 279. p.
10. Tomolya Kinga: *SiC erősítésű kompozit előállítása porkohászati úton*. Diplomamunka. Miskolci Egyetem. 2001.
11. Underwood Ervin E.: *Quantitative Stereology*. Menlo Park, California. 1970. Addison-Wesley Publishing Company. 23 p.

Acél széles szalagok szövetszerkezetének morfológiai jellemzése

Csepeli Zsolt¹, Gácsai Zoltán²

¹ Dunaferri Kutatóintézet

² Miskolci Egyetem Fémtani Tanszék
femtangz@uni-miskolc.hu

Absztrakt. Az acélok mechanikai tulajdonságait elsősorban mikro (szövet) szerkezetük határozza meg. A szövetszerkezet kiértékelésével sok fontos információt kaphatunk az anyag előéletéről, várható tulajdonságairól és a jövőbeli alkalmazhatóságáról.

1. Bevezetés

A fémek szemcsenagysága és mechanikai tulajdonságai közötti alapvető kapcsolat régóta ismert. A szemcsék morfológiájának részletes és objektív tanulmányozását azonban csak az elmúlt években tette lehetővé a vizsgálati eszközök és módszerek fejlődése. Célunk a normalizálva és termo-mechanikusan hengerelt acél széles szalagok szövetszerkezetének részletes jellemzése volt, különös tekintettel annak meghatározására, hogy miben hasonlít egymáshoz, illetve miben tér el a kétféle technológiával előállított acélok mikro-szerkezetére. Ebben a cikkben a ferrit szemcsék alak tényezőinek és orientációjának mérettől való függését mutatjuk be. Az eredmények értékelésekor figyelembe kell venni, hogy a ferrit szemcsék mellett gyakran más fázis is jelen van (például perlit, bénit vagy martenzit), melyek lényegesen megváltoztathatják az anyag tulajdonságait, valamint azt is, hogy az acél mechanikai tulajdonságait, mindenek előtt a folyáshatárt a szemcsenagyság és a morfológia mellett az acél kémiai összetétele, a ferrit oldott ötvöző tartalma, diszlokáció sűrűsége, az esetleges kiválások mennyisége, mérete, alakja vagyis a gyártástechnológia is befolyásolja.

2. Vizsgált anyagok és az alkalmazott módszer

A szövetszerkezet vizsgálatot két normalizálva hengerelt próbadarabon (Norm1 és Norm2), valamint két termo-mechanikusan hengerelt próbadarabon (Termo1 és Termo2) végeztük el. A két típusú acél mind összetételében, mind gyártástechnológiájában különbözik. A normalizálva hengerlés az ausztenit statikus újrakristályosodási hőmérséklete felett fejeződik be, majd a széles szalagot gyorsan lehűtik az Ar1 alatti csévézés hőmérsékletre. Termo-mechanikus hengerléskor az utolsó szűrés az ausztenit statikus újrakristályosodási hőmérséklete alatt

történik. A próbadarabok kémiai összetétele az 1. táblázatban látható. A vizsgálatokat 500x nagyítású fénymikroszkópos felvételeken végeztük el. A ferrit szemcsék határait 1% HNO₃ tartalmú alkoholos oldattal tettük láthatóvá. A próbadarabok szövetszerkezete ferrit mellett kisebb mennyiségű perlitet is tartalmaz, azonban detektálással a ferrit és a perlit jól elválasztható volt egymástól, ezért a perlit a későbbi mérést nem befolyásolta. A sötét ferrit határokat "BtopHat" képátalakító művelettel kiemeltük, majd az egyes szemcsék összefüggő határait "Watershed" eljárással megrajzoltuk. Az egymástól így elválasztott szemcséknek megmértük a területét, kerületét, hosszát, szélességét és származtatott orientációját. Próbadarabonként több ezer ferrit szemcse morfológiáját jellemeztük.

1. táblázat: A vizsgált próbadarabok kémiai összetétele

Próbatest	Vastagság (mm)	C	Mn	Si	S	P	Al	Nb
NORM1	2.8	0.127	1.16	0.020	0.012	0.011	0.034	-
NORM2	9.7	0.171	1.42	0.288	0.010	0.015	0.053	-
TERMO1	6.1	0.059	0.78	0.233	0.005	0.013	0.029	0.030
TERMO2	6.1	0.078	0.89	0.016	0.008	0.012	0.040	0.027

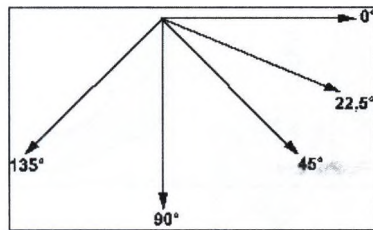
3. Alaktényezők értelmezése

A szemcsék morfológiáját legegyszerűbben célszerűen definiált alaktényezőkkel jellemezhetjük. Alaktényezőnek nevezzük a szemcsék geometriai jellemzőiből származtatott minden olyan számmennyiséget, amely független az alakzat méretétől, valamint síkban, illetve térben elfoglalt helyétől és helyzetétől. Az alaktényezők tehát skalár mennyiségek, amelyek jellegzetes tulajdonságai a következők: mértékegység nélküliek, függetlenek az alakzat síkbeli és térbeli kiterjedésétől, továbbá függetlenek az alakzat helyétől és helyzetétől [1]. Korábbi vizsgálatok alapján a ferrit szemcse metszetek alakjának jellemzésére az alábbi két — geometriai alapadatokból származtatott — alaktényezőt választottuk ki [1,2]:

4. Orientáció definiálása

Az orientáció értelmezéséhez először a Feret átmérőt kell definiálni. A Feret átmérők tulajdonképpen az objektumok maximális méretei valamilyen meghatározott irányban mérve. Azt is szokták mondani, hogy a Feret átmérők tangens átmérők, hiszen tulajdonképpen az adott irányban megrajzolt érintők közül a legmesszebb lévők távolságát jelenti. Mindezekből következik, hogy a Feret átmérők meghatározásakor mindig meg kell adni a mérés irányát is. Leggyakrabban vízszintesen (0°), vagy függőlegesen (90°) szokás mérni ezeket az adatokat. A leggyakrabban használt irányokat a 1. ábra mutatja. A mai korszerű képelemző

programok 64 különböző irányban is képesek mérni a maximális átmérőt. A különböző irányokban mérhető Feret átmérők értelmezését a 2. ábra mutatja. A Feret átmérők közül a legnagyobbat a részecske hosszúságának ('length'), míg a legkisebbet a szélességének ('width') tekintjük. Az orientáció nem más, mint a részecske hosszúságához ('length'), mint vektorhoz rendelt szög (3. ábra). Szögletes és kevésbé nyújtott részecskéknél az orientációs tengely jellemzésére inkább a származtatott orientáció szolgál, amely nem egyetlen maximális átmérőhöz kapcsolódó szög, hanem több átmérőhöz tartozó szög átlaga. A szemcsék orientációját a maximális Feret átmérőkhöz tartozó szögekből számolt úgynevezett származtatott orientációval jellemeztük (4. ábra).

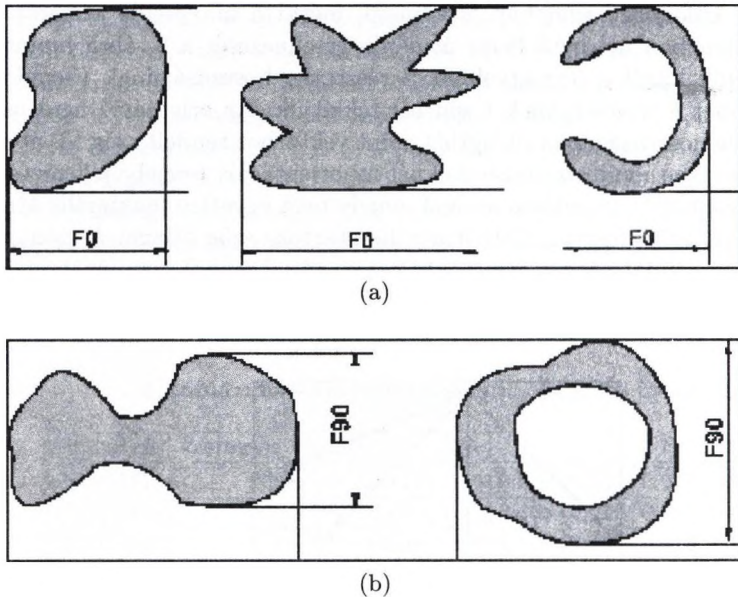


1. ábra: A Feret átmérők meghatározásakor használt irányok

5. Mérési eredmények

Teszt alakzatok jellemzése

Vizsgálataink során részletesen elemeztük a szemcsék alakjának és orientációjának mérettől való függését is. A ferrit szemcsék analóg képeinek digitalizálásából és a mérésből származó hibákat teszt ábrák segítségével becsültük meg. A teszt ábrákhoz Coreldraw programmal nyújtott nyolcszöget rajzoltunk meg, majd ezt több fokozatban kicsinyítettük, illetve elforgattuk. Így a 5. ábrán látható 42 teszt alakzatot kaptunk, melyeken elvégeztük ugyanazokat a méréseket - ugyanazzal a nagyítási értékkel - mint a valódi ferrit szemcsék esetén. A mérési eredményei a 2. táblázatban láthatók. A mért adatokból megállapítható, hogy az alaktényezők mérettől való függése meglepően kicsi. Míg az . ábra első öt sorában a harmincöt különböző méretű és helyzetű szemcsé területa $6.9 \mu\text{m}^2$ és $599 \mu\text{m}^2$ között változik (közel 100 szorosán), addig az alaktényező változása kisebb, mint 10%. Az ábra legalsó sorában levő, pár képpontból álló, kb. 0.5 m^2 területű alakzatoknál azonban az alaktényezők értéke lényegesen megnövekszik, maximuma (P esetén 1.25, (D esetén pedig 1.39 lesz. A próbadarabokon végzett mérések során a kicsi, $5 \mu\text{m}^2$ -nél kisebb területű objektumokat megfigyelve azok inkább tűnnek digitalizálásból, képfeldolgozásból eredő hibáknak, mint az acélt alkotó ferrit szemcsék metszetének, ezért a gyakorlatban csak az ennél nagyobb területű objektumokon mért eredményeket célszerű figyelembe venni.



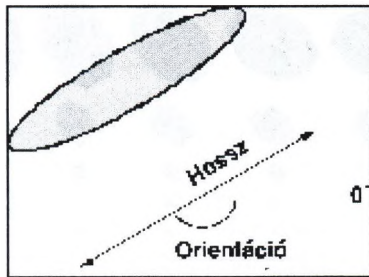
2. ábra: A különböző irányokban mért Feret átmérők értelmezése a) F_0 = a 0° -os szöghöz tartozó Feret átmérő b) F_{90} = a 90° -os szöghöz tartozó Feret átmérő

2. táblázat: Teszt alakzatok jellemzői

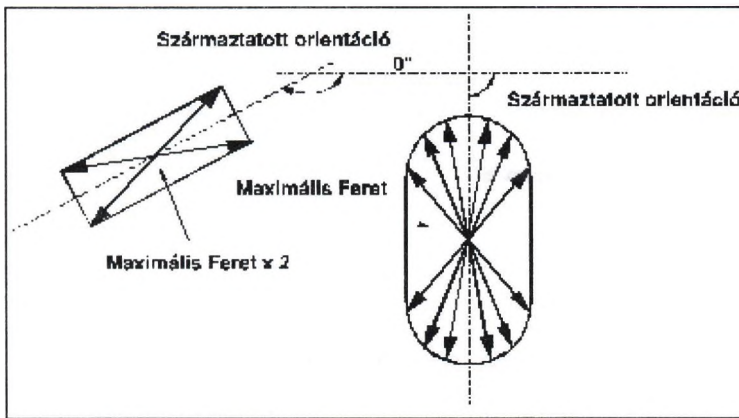
	P	D	Terület
Min	1.06	1.16	0.03
Max	1.17	1.27	0.02
Dev	6.9	599	243

Acél próbadarabok adatai

A szemcsék méret szerinti eloszlása a két normalizálva és két termo-mechanikusan hengerelt acélban a 6. ábrán látható. A szemcséképződéssel és növekedéssel kialakuló szövetszerkezeteknek megfelelően a szemcsé metszet területek növekedésével exponenciálisan csökken az egyes frakciókba tartozó szemcsék mennyisége. A szemcsék kerületéből és területéből számolt (P alaktényező átlagos értéke a nagyobb szemcsé metszetek esetén 40-70%-kal meghaladja a kisebb metszetek esetén mért értékeket (7. ábra). Ez lényegesen nagyobb változás, mint a teszt alakzatoknál mért eltérés, ezért kijelenthető, hogy a szemcsék méretének függvényében mért (P alaktényező változás anyagjellemző, nem a digitalizálási, vagy mérési módszerből adódik). Hasonló, de kisebb mértékű változás figyelhető meg a szemcsék hosszának és területének arányából meghatározott (D alaktényezőnél (8. ábra). A szemcsék mérete csak kis mértékben befolyásolja az orientációs függvény alakját (9-12. ábra). Ha egy próbadarab esetén határozott orientáció figyelhető meg (mint a Termo1 és Termo2 próbáknál), akkor ez minden



3. ábra: Az orientáció értelmezése

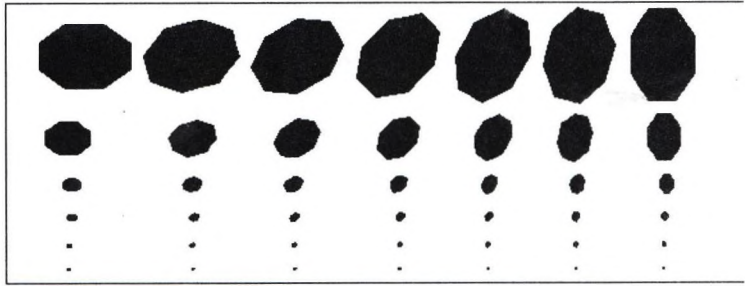


4. ábra: Szármasztatott orientáció meghatározása

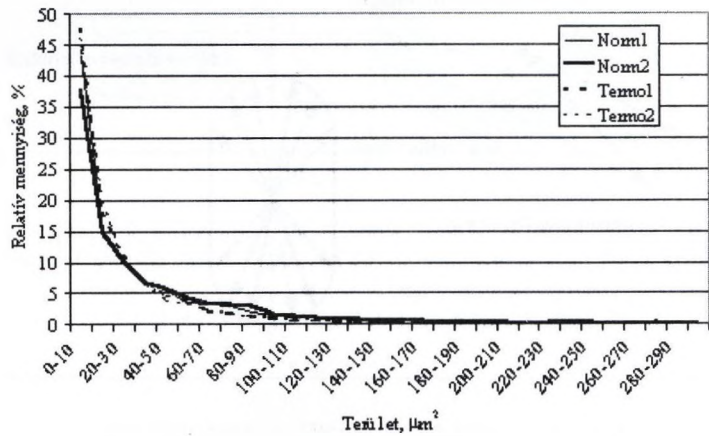
méretű szemcsére jellemző. Ha nincs irányítottság, akkor ezzel ellentétes tendencia semmilyen méretű frakcióban nem látható. A Termo1 és Termo2 próbadaraboknál a legkisebb méretű részecskék orientációja a legkisebb, a legnagyobb méretű részecskénél pedig a legnagyobb.

6. Összefoglalás

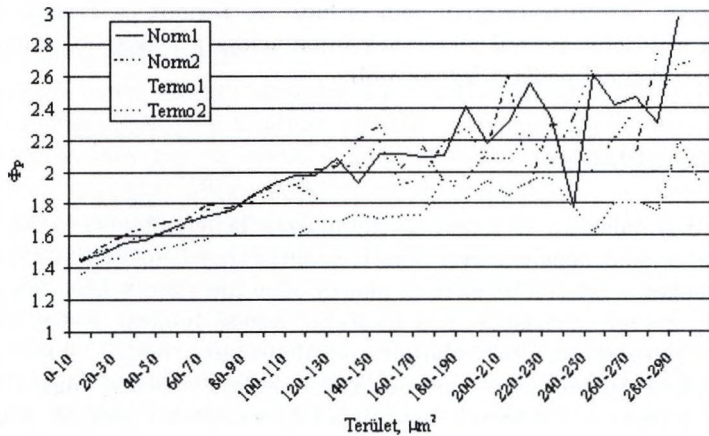
A cikkben normalizálva és termo-mechanikusan hengerelt acél széles szalagok szövetszerkezetének képelemzővel végzett kvantitatív jellemzését mutattuk be. A próbadarabokon mért értékeket teszt alakatokon mért értékekkel hasonlítottuk össze, hogy a digitalizálásból és a mérésből adódó hibákat is figyelembe vehessük. Az elvégzett mérések alapján megállapítható, hogy a ferrit szemcsék morfológiájának jellemezésére használt két alaktényező értéke függ a szemcsék méretétől. A szemcse metszetek kerületéből és területéből számolt alaktényező értéke nagy szemcsék esetén lényegesen magasabb, mint a kisebb szemcséknél meghatározott érték. Az alaktényezők mérettől való függése valószínűleg anyagjellemző, s nem mérési hiba eredménye. A szemcse orientációs függvény alakja



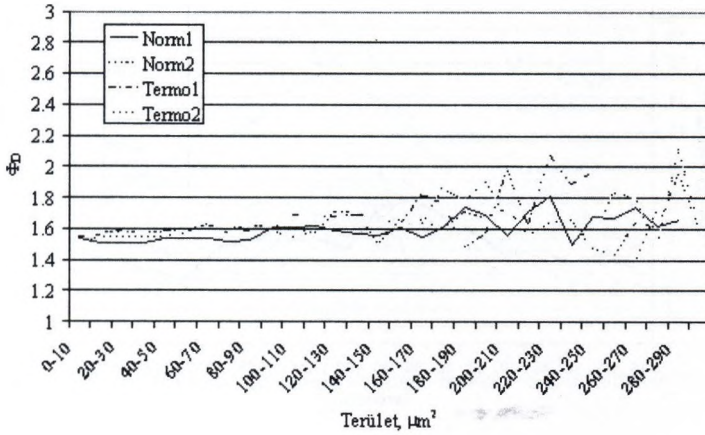
5. ábra: Hét különböző irányban elhelyezkedő, hat eltérő területű (595, 153, 26, 7, 2 és 0,6 μm^2) teszt alakzat



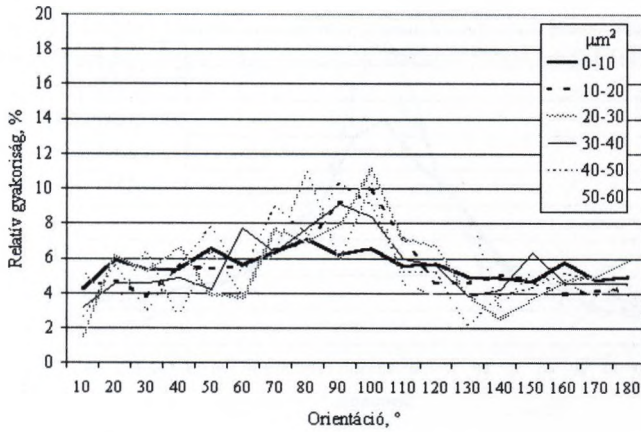
6. ábra: A ferrit szemcsék síkmetszeteinek terület szerinti eloszlása



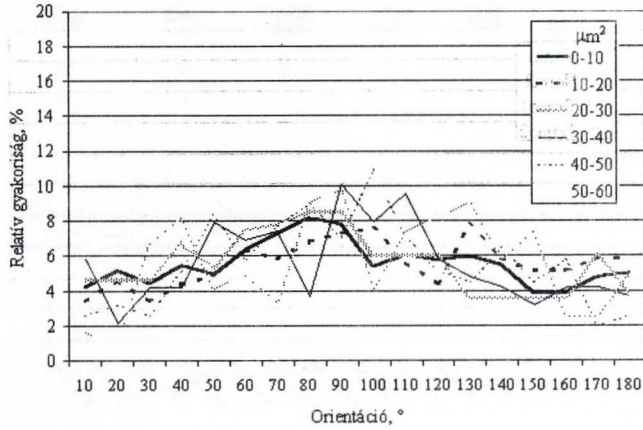
7. ábra: A (P alaktényező értéke különböző méret frakciók esetén



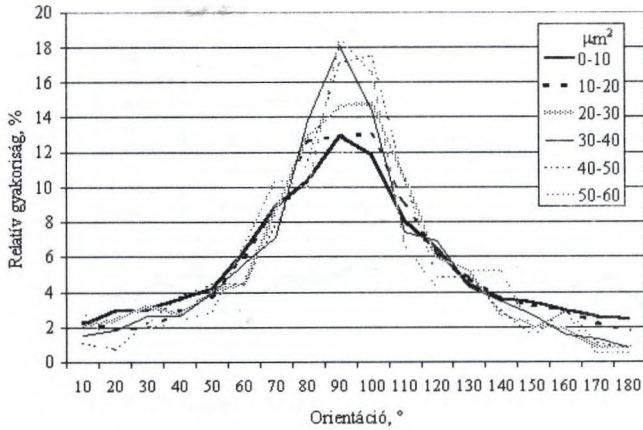
8. ábra: A (D alaktényező értéke különböző méret frakciók esetén



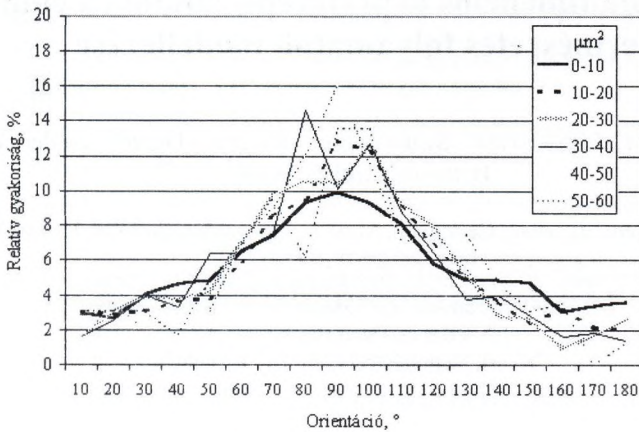
9. ábra: A Norm1 próbadarab ferrit szemcse frakcióinak orientáció eloszlása



10. ábra: A Norm2 próbadarab ferrit szemcse frakcióinak orientáció eloszlása



11. ábra: A Termo1 próbadarab ferrit szemcse frakcióinak orientáció eloszlása



12. ábra: A Termo2 próbadarab ferrit szemcse frakcióinak orientáció eloszlása

csak kisebb mértékben függ a szemcse méretétől. A kisebb területű ferrit szemcsék orientációja valamivel kisebb, mint a nagyobb szemcséké.

Irodalom

1. Gácsi Z., Sárközi G., Réti T., Kovács J., Csepeli Zs., Mertinger J: Sztereológia és képelemzés
2. Underwood E. E.: Quantitative stereology (Addison-Wesley Publishing Company, California 1970)
3. Pratt W. K.: Digital Image Processing (John Wiley and Sons, New York 1987)
4. Leonard E. Samuels: Light Microscopy of Carbon Steels (ASM International 1999)

Háromdimenziós képi (terepi) adatokra épülő természetes folyamatok modellezése

Berke József¹, Sisák István¹, Szabó József², Kelemen Dezső², Szeiler Gábor³,
Berkéné Várbíró Beáta⁴

¹Veszprémi Egyetem, Georgikon, Mezőgazdaságtudományi Kar, Keszthely
berke@georgikon.hu

www.georgikon.hu/digkep/ikta112.htm

²PICTRON Számítás- és Videotechnikai Kft., Budapest

³Szeiler és Társa Bt., Keszthely

⁴Kvark Számítástechnikai Bt., Keszthely

Abstract. A harmadik évezred küszöbén a föld felszínéről gyűjtött információk bősége az információkat előállító rendszerek fejlesztése és marketingje mellett, egyre inkább előtérbe helyezte a hatékony és sokoldalú felhasználást. Felismerve a fentiek hazai jelentőségét, az Oktatási Minisztérium Kutatás-fejlesztési Helyettes Államtitkársága széleskörű hazai kutatási program pályázatát és indítását végezte 2000-ben, IKTA3-KÉPI-2000 „Digitális képi technológiák és alkalmazások” program keretében, a meglévő légi és űrfelvételek korszerű hazai hasznosítása érdekében [17].

Előadásunkban bemutatjuk az IKTA-00112/2000 „Háromdimenziós képi adatokra épülő ökológiai folyamatok modellezése” kutatási program – amely a fenti program része - célkitűzéseit, és az eddig (a projekt 2/3-át bemutató) elért eredményeket:

- természetes megvilágítás szimulációja,
- köd- és páratartalom szimuláció,
- repülés szimulációk,
- ökológiai folyamatok vizuális szimulációja, stb.

1. A projekt célja

Az utóbbi években készített hazai légifelvételek [3, 7, 8] számos nemzetgazdaságot is érintő fejlesztési program megvalósítását teszik lehetővé. Ezek közé tartozik olyan valóság-hű modellek kidolgozása is, amelyek pontosságukkal, gyorsaságukkal, számítógépes szimulációval történő kiegészítésükkel közvetlen alkalmazhatóvá válnak akár a mezőgazdaság, geológia, meteorológia, környezetvédelem, turisztika területein is. Az IKTA3 112/2000 „Háromdimenziós képi adatokra épülő ökológiai folyamatok modellezése” elnevezésű kutatási program, olyan lehetőségek kiaknázását tűzte célul, amelyek háromdimenziós DTM alapú vizuális alaprendszer (vektoros és raszteres is) létrehozására épülnek, a fentiekhez illeszkedő szimulációs vizsgálatokhoz.

Az alábbiakban felsorolás jelleggel ismertetjük a főbb célkitűzéseket [1, 2]:

- Háromdimenziós DTM alapú vizuális alaprendszer (vektoros és raszteres is) létrehozására, EOTR szelvényezés alapján.
- Környezetileg érzékeny területek lehatárolása a vonatkozó EU rendelkezések szerint.
- Kedvezőtlen adottságú teszterületek meghatározása (lejtési, magassági és talajadatok alapján).
- Ökológiai folyamatokhoz 3D szimulációs modellek készítése (EECONET – Egységes Európai Ökológiai Hálózat ajánlásai alapján).
- Megfelelő valósághű, háromdimenziós adatokhoz illeszkedő speciális grafikus adatbeviteli felület kifejlesztését.
- Konkrét szimulációk megvalósítása, azok oktatási és további kutatási alkalmazása.

A fenti program (112/2000) megvalósítását egy konzorcium végzi, melynek tagjai a szerzői felsorolás alatt találhatóak. A KÉPI-2000 elnevezésű kutatási, fejlesztési programok lebonyolítását, koordinálását, támogatását az Oktatási Minisztérium Kutatás-fejlesztési Helyettes Államtitkársága végzi [1].

2. Az alapmodell létrehozása

A program alapját képező terepmodell létrehozásához során elsődleges szempontként a megfelelő referencia adatokra épülő légifelvételeket említhetjük. Esetünkben a légifelvételeket a Magyar Honvédség Térképészeti Hivatala által koordinált éves légi-felmérések (MH TEHI) szolgáltatották [3]. Ezek analóg képek, melyeket *LEICA RC-30* típusú kamerával [4] készítettek. A kamera felépítésében és gyakorlati alkalmazhatóságában jól ismert, nagy pontosságú adatok létrehozását biztosítja. A felvételezés pankromatikus filmre történt, amely a látható spektrum nagy részét (400-670 nm) lefedi, a visszavert fény intenzitását és hullámhosszát olyan szűrkeségi árnyalatokként rögzíti, amely alkalmas a későbbi digitális nyersanyagok előállítására. Ezek a felvételek hazánkban főleg térképezési célra készültek és készülnek napjainkban is. Ezért az alkalmazhatóságuk megítélése is a program feladatai közé került [5].

Az előhívott légifelvételeket napjaink egyik csúcsmínőségű – speciálisan légifilmek digitalizálására kialakított – INTERGRAPH ImageStation PhotoScan szkennerevel kerültek digitalizálásra [3, 6]. A kapott digitális állomány torzításmentes, 7,5 vagy 15 mikron képi felbontású, veszteségmentes raszterkép, GeoTIFF [9] formátumban, 1 mikronos geometriai felbontással és tengelyenkénti 2 mikronos geometriai pontossággal.

A raszteres világban számos képi adatformátum ismeretes [10]. Mi indokolta esetünkben, hogy egy bonyolult, nem helytakarékos formátum képezze az egzakt számításaink alapját. A választásunk indoklása előtt tekintsük át a legjellemzőbb vonásait a GeoTIFF [9] formátumnak. A TIFF formátum azon szakterületeknek lett kifejlesztve, amelyek tetszőleges pixeltartalmú raszteres képeket hardver- és szoftver független módon kívánnak átvinni az egyik rendszerből egy másik számítógépes

(digitális) rendszerbe. GeoTIFF formátummá való továbbfejlesztésére azért volt szükség, mert az eredeti verzió nem támogatta a földrajzi vonatkozásokkal kapcsolatos műveleteket [7]. Ezeknek a műveleteknek a leírásához számos új, különböző típusú paraméterre lett volna szükség, amelyeket a hagyományos módon különálló TIFF címkékkal leírva, új címkék százaira lett volna szükség. Mindez kimerítette volna az új címkék tárolási lehetőségeit. A problémák elkerülése érdekében a GeoTIFF fájlok olyan kulcsokkal (GeoKeys) kerültek kiegészítésre, amelyek támogatják a térinformatikai, távérzékelési rendszerekben alkalmazott eljárásokat, terepi, geológiai szabványokat.

A kutatási területek alapját képező folyamatok modellezéséhez szükséges területek lehatárolása (1. ábra), majd a keletkező adatok becslése után került sor olyan hardver és szoftver rendszerek telepítésére, amelyek alkalmasak a keletkező adatok digitális kezelésére, megfelelő illesztési felületeket tartalmaznak további fejlesztésekhez, szimulációkhoz és természetesen más szakterületekhez is illeszthetőek.

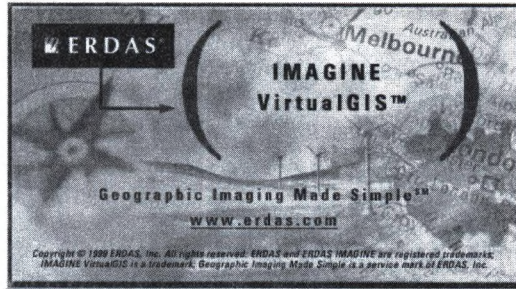


1. ábra Kutatási területek alapját képező folyamatok modellezéséhez szükséges területek egy lehetséges lehatárolása

Ezen főbb szempontok alapján került sor egy SGI 330 grafikus - Windows alapú - munkaállomás [11, 5] és egy hozzá illeszkedő ERDAS IMAGINE Professional szoftverrendszer telepítésére [5, 12, 13, 14, 16] (2. ábra, 3. ábra).



2. ábra SGI 330 grafikus munka-
állomás



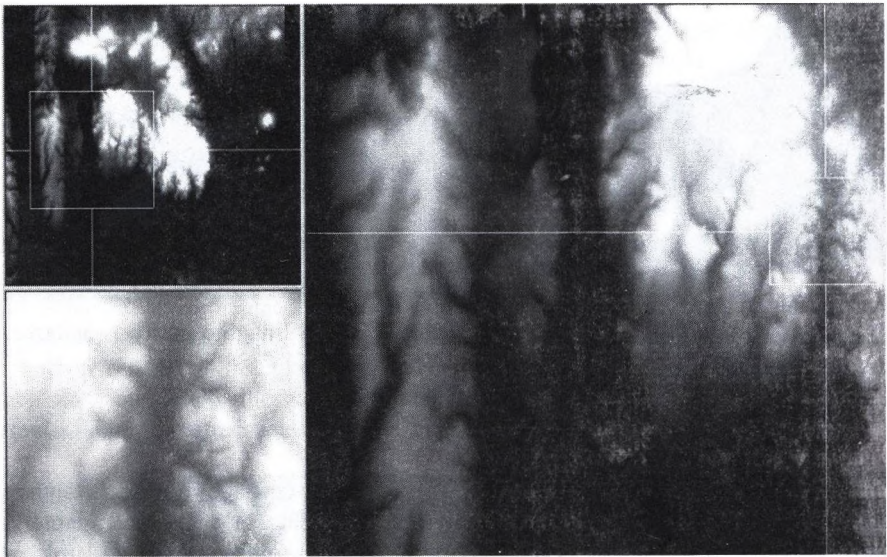
3. ábra ERDAS Imagine szoftver rendszer

A digitálisan előkészített légifelvételek EOV (Egységes Országos Vetületi Rendszer) vetületi rendszerbe kerültek [3, 7], a szükséges alapadatokat a DTA-50 alaptérkép (MSZ K1066 sz. szabvány) [3] tartalmazza. Az így előállított képi adatok már közvetlenül alkalmasak egzakt (térképészeti, térinformatikai szempontokat is kielégítő) vizsgálatokhoz. A magassági alapadatokat a Digitális Domborzati Modell (DDM) adja szelvényenkénti grid állományok formájában, megfelelő illesztéssel.

A 4. ábrán több légifelvételből készített, Keszthely és környékét bemutató, EOV vetületi rendszerbe transzformált kép látható.

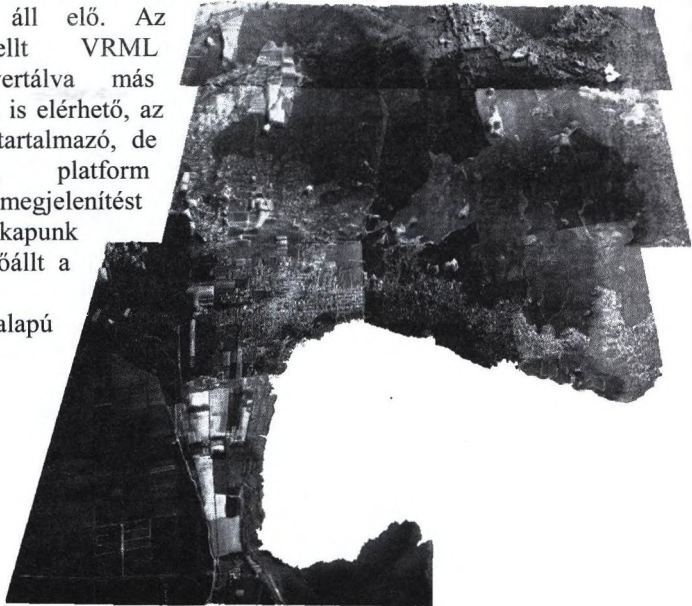


4. ábra Keszthely és környékét bemutató, EOV vetületi rendszerbe transzformált
légifelvétel



5. ábra A magassági adatokat tartalmazó Digitális Domborzati Modell (DDM) egy részlete

A fentiek szerint előállított független alapadatokat (EOV GeoTIFF raszter, DDM, tetszőleges vektoros vagy raszteres rétegek, stb.) egyetlen modellé kapcsolunk, majd 3D alapú megjelenítést alkalmazva közvetlenül megtekinthető és szimulációkra is alkalmas modell áll elő. Az elkészített modellt VRML formátumra konvertálva más rendszerek számára is elérhető, az alapadatokat nem tartalmazó, de OpenGL alapú, platform független megjelenítést támogató modellt kapunk (6. ábra). Ezzel előállt a szimulációkhoz szükséges, 3D alapú alapadatbázis.



6. ábra Digitális alapadatokra épülő, platformfüggetlen megjelenítést biztosító VRML modell

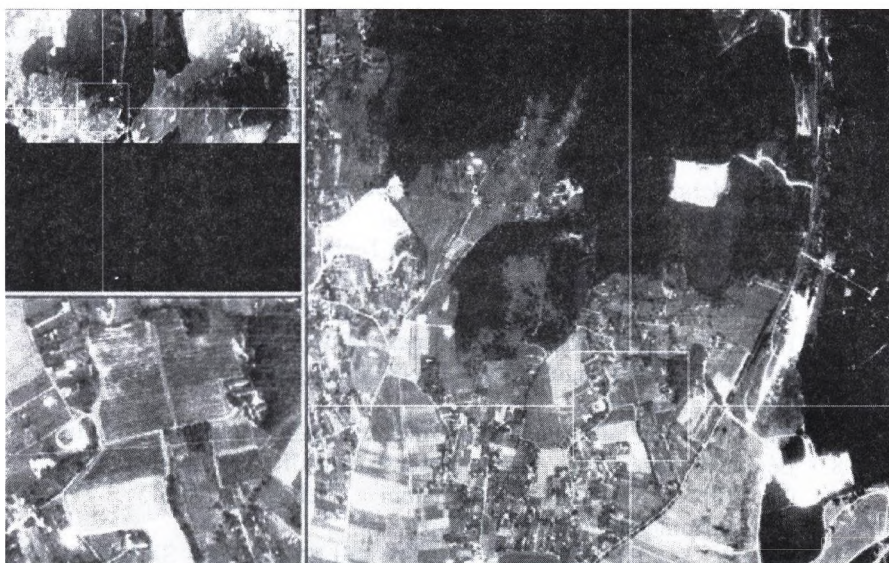
3. Eredmények

A projekt során meghatároztuk a teszterületeket, felmértük és létrehoztuk az ezekre vonatkozó 3D valósághű alapadat rendszert. A teszterületek kiválasztásánál az alábbi szempontokat kívántuk figyelembe venni:

- Légifelvételzés során kapott friss (3 éven belüli) adatok álljanak rendelkezésre.
- A területről nyert adatok illeszthetők legyenek a meglévő információkhoz.
- Megfelelő pontosságú ortofotó létrehozása legyen lehetséges.
- Használható legyen az EOV vetületi rendszer.
- Nem képi jellegű, referenciák legyenek a területről.
- Terepi szintű ismeretsége legyen a kutatóknak a kiválasztott területről.

A fenti szempontok alapján az alábbi területeket választottuk ki egy szakmai workshop keretében:

1. Örvényes és környéke (EOV: 554140, 178670 középponttal)
2. GMK szőlészeti kísérleti telepe (EOV: 500280, 152440 középponttal - 7. ábra)
3. Zalaszentmárton és környéke (EOV: 499700, 152780 középponttal)



7. ábra. Georgikon Mezőgazdaságtudományi Kar szőlészeti kísérleti telepe

A kutatási program keretében elvégeztük [1, 2, 5, 15]:

1. A Keszthelyi-hegység valósághű, digitális modelljének létrehozásához szükséges eljárások, adatok, munkafolyamatok tervezését.
2. A Keszthelyi-hegység valósághű, nagyfelbontású digitális domborzati modelljének létrehozását ERDAS IMAGINE VirtuGIS rendszerben.
3. A terepi adatokkal kiegészítve megoldottuk a modell pontosítását.

4. Panoráma (180 és 360 fokos) felvételeket készítettünk digitális kamerával, majd ezeket geokódoltuk az EOVS vetületi rendszerhez, majd a terepi modellhez illesztettük.
5. Három alap szimulációs modellt hoztunk létre, amelyek a további talajtani-ökológiai szimulációk alapjául szolgálnak:
 - Repülés szimuláció a további légifelvételek készítésének tervezéséhez
 - Természetes megvilágítás (nap) öröknaptár alapú szimulációja
 - Köd és páratartalom egzakt szimulációja
6. Elkészítettünk és gyári úton előállítottunk olyan multimédia alapú interaktív anyagot (*Festmények interaktív kiállítása – NSZMAT v1.0*), ahol a szimulációs modellt egyszerűsített változata felhasználásra került.
7. A kutatási eredményeket hazai és nemzetközi szakmai fórumokon ismertettük [1, 2, 5, 15].

Jelenleg folyik a gyakorlati hasznosítás megvalósítási tervének készítése, valamint a kifejlesztett speciális adatbeviteli felület tesztelése.

Irodalom

1. BERKE J., SZABÓ J., SZEILER G., VÁRBÍRÓ B. (2000): Háromdimenziós képi adatokra épülő ökológiai folyamatok modellezése, IKTA3 képi technológiák projektbemutató, Budapest, 2001.01.24. BERKE, J. (2001): Wireless supporting of agricultural and forestry information systems. 4th Concertation Meeting of Mobile/Wireless/Satellite IST projects, Brussels.
3. Honvédelmi Minisztérium Térképészeti Közhasznú Társaság légifelvételek adatbázisa: <http://www.mhtehi.gov.hu/>.
4. LEICA RC-30 Aerial Camera System Technical Reference Manual.
5. BERKE, J. – SZABÓ, J. – SZEILER, G. – VÁRBÍRÓ, B. (2001): Háromdimenziós képi adatokra épülő ökológiai folyamatok modellezése, IKTA3 képi 2000 workshop, Budapest, 2001.11.14.
6. INTERGRAPH ImageStation – Digital Photogrammetric Solutions.
7. FÖMI légifilmtár:
<http://fish.fomi.hu/fish/adathaz/termekek/Legifelvelekek/legifilm.htm>.
8. Légifelvelekezés: http://www.eurosense.com/UK/serv_photograph_uk.htm.
9. GeoTIFF szabvány formátum leírása /SGI 1992-2001/, <ftp://sgi.com/graphics/tiff/TIFF6.ps>.
10. BERKE, J. - HEGEDŰS, GY. CS. - KELEMEN, D. - SZABÓ, J. (2001): Digitális képfeldolgozás és alkalmazásai. Georgikon – Szeiler – Kvarckeszthely - Pictron Kft., Budapest, ISBN 963 00 5744 1.
11. SGI Magyarország honlapja: <http://www.silicon.hu/>.
12. ERDAS Inc. honlapja: <http://www.erdas.com/>.
13. ERDAS IMAGINE Magyarországi forgalmazójának honlapja: <http://www.bekes.hu/>.
14. ERDAS Inc., White Paper – IMAGINE VirtualGIS, September 1997 - July 2000.

15. BERKE, J. - SZABÓ, J. - SZEILER, G. - BERKÉNÉ VÁRBÍRÓ, B. (2001): Háromdimenziós képi adatokra épülő ökológiai folyamatok modellezése. Média - Informatika - Kommunikáció, Veszprém, 2001. szeptember 13-15.
16. BERKE, J. – PALLÉR, N. (2000): ERDAS IMAGINE gyakorlati alkalmazása a Georgikon Karon. Második Magyar ERDAS Felhasználói Konferencia, Budapest.
17. Oktatási Minisztérium Kutatás-fejlesztési Helyettes Államtitkársága, <http://www.om.hu/>.

Vector-based map interpretation

Endre Katona¹

Department of Applied Informatics
University of Szeged
H-6720 Szeged, Árpád tér 2, Hungary
katona@inf.u-szeged.hu¹

Abstract. Geographic Information Systems are mostly based on vector-coded digital maps. As a consequence, it is a big challenge to solve the automatic conversion of scanned paper maps to vector (CAD) format. This conversion problem involves the recognition of all signs, notations and structural elements of the scanned drawing.

Lots of systems have been published for different kinds of maps, applying a large variety of raster-based and vector-based recognition techniques. In our approach, a totally vector-based processing method is applied. A universal data model, named DG (Drawing Graph), is introduced to handle vector data uniformly during the whole recognition process. To improve algorithmic efficiency, spatial indexing is used.

Our approach has been applied in the Phare Land Consolidation Project to process Hungarian cadastral maps. Further applications on topographic maps and engineering drawings will be investigated in the future.

1 Introduction

The published map interpretation systems (see e.g. [1, 4, 5, 7, 10, 19]) can be classified into three categories:

- *Raster-based systems* perform all recognition on scanned raster data.
- *Vector-based solutions* start with a raw vectorization, and all the recognition steps are performed on vector data.
- *Hybrid solutions* usually perform segmentation to separate text from graphics, and text is recognized with some raster-based OCR technique while graphics is analysed after vectorization.

In the current work we prefer the vector-based approach, the scanned raster image is left only for optional use to support decisions in some crucial situations.

Although a lot of papers have been published presenting map interpretation systems, only a few of them deals with the questions of data storage and handling. Here we give an overview of data models that we have investigated during our work.

It is a natural way to use some graph representation for a vectorized drawing. Lladós et al. [12] define an *attributed graph*, and after extracting minimum closed loops, a *region adjacency graph* is generated. This approach concentrates

on region matching and this fact restricts its applicability. A special approach is applied in [4]: after an initial rungraph vectorization a mixed graph representation is used to ensure interface between raster and vector data.

Some interpretation systems use *relational database tables* to store geometric information of vectorized maps [1, 19]. The advantage of this approach is that commercial database management systems can be used to handle data. Although there are existing techniques to store spatial data in relational tables [16], it is clear that the relational model is not the best choice to store map data.

Object-oriented models are more flexible than relational ones and are widely used in current GIS (Geographical Information System) applications, normally based on the Object Database Standard [3]. Relationships between objects can be coded by pointer lists in such implementations. Object-oriented GIS typically uses a hierarchy of spatial object types, such as defined in the Geometry Object Model of the OpenGIS Consortium [16] supporting interoperability of different systems. The object-oriented concept is excellent for high level map description, but it does not support low level algorithmic efficiency during recognition.

Topological models have been introduced in early GIS [6] and also currently used in Arc/Info, that has been applied for map interpretation in [14]. A topological model can be regarded as a set of cross-referenced drawing elements: each element has a unique identifier (id) to offer the possibility for other objects to reference it. A characteristic example is the node-line-polygon structure of the Arc/Info data model [2] describing region maps with three types of records:

- $Node(id, x, y)$ contains the coordinates of a junction point.
- $Line(id, x_1, y_1, \dots, x_n, y_n, node_1, node_2, lpoly, rpoly)$ describes a linestring between the junction points $node_1$ and $node_2$ with breakpoints $x_1, y_1, \dots, x_n, y_n$ and referencing the left and right neighbouring polygons ($lpoly, rpoly$).
- $Polygon(id, line_1, \dots, line_n)$ describes a region, bounded by the linestrings $line_1, \dots, line_n$.

This model ensures efficient algorithms to compute polygon areas, point-in-polygon decisions, overlay of two polygon coverages, etc. [6].

It is a challenging approach in graphics recognition to use some knowledge representation scheme, such as *semantic networks*. [15] gives detailed description of a general semantic network model, applied both for vector-based [5] and raster-based [18] map interpretation. Such a semantic network model defines *concepts* to describe a priori knowledge, while *modified concepts* and *instances* are built up during interpretation. [8] uses a similar approach to process gray-level map images. Although semantic networks give a rather general approach, applied also in speech understanding and robotics [15], they do not support the low-level efficiency of algorithms like topological models do.

2 The data model

Our DG (Drawing Graph) model combines topological and object-oriented aspects with semantic networks in some sense. A preliminary version of DG has

been used in a signature verification system [9], an advanced version is applied for interpretation of Hungarian cadastral maps [10].

Basic element of the model is the *DG-object*. A set Z of objects, describing the current state of interpretation, is called *DG-document*. Considering the storage capacity of current computers, we suppose that the whole DG-document can be stored in RAM. The object set Z consists of two disjoint subsets:

- Z_n denotes the set of *normal objects*, they correspond to recognized parts of the drawing.
- Z_s denotes the set of *sample objects*, giving an a priori knowledge description (like concepts in a semantic network). For instance, sample objects can describe a symbol library of the map legend or vector fonts of a given language.

Each object has the structure (*id, layer, attributes, references*) where *id* is the object identifier number, and *layer* is a CAD-like attribute to classify objects. Normal objects have 0, 1, 2, etc. layer numbers, while sample objects are kept in a distinguished layer S . Layer 0 is reserved for unrecognized objects. Each object may have references to other objects using their id's (similarly to relationships between objects in [2] and [3]). The set of references form an acyclic directed graph $R \subset Z \times Z$, termed *reference graph*. Two types of references can be distinguished:

- A "contains" reference means that the current object involves the referred object as a component. $R_c \subset Z \times Z$ denotes the set of "contains" references.
- A "defined by" reference means that the current object is a transformed version of a referred sample object. Such references are mainly used to describe recognized instances of sample objects. $R_d \subset Z \times Z_s$ denotes the set of "defined by" references, and $R = R_c \cup R_d$ holds.

Denote $domain(u)$ the set of all objects v that have a reference path from u to v , and denote $scope(u)$ the set of all objects v with a reference path from v to u . Notations $domain_c(u)$ and $scope_c(u)$ mean restrictions to "contains" reference paths. It is required that for any sample object s , $domain(s) \subseteq Z_s$.

The DG model contains three basic object types (instances of each may be normal or sample objects as well):

- A *NODE object* represents a point with coordinates, a NODE instance is denoted as $node(x, y)$. A NODE has no references to other objects.
- An *EDGE object* is a straight line section given by "contains" references to the endpoints. An EDGE instance is denoted as $edge(node_1, node_2)$. A "line width" attribute may be attached, if necessary.
- A *PAT (pattern) object* represents a set of arbitrary DG-objects given by "contains" references to its components. A PAT instance is denoted as $pat(obj_1, \dots, obj_n)$.

At a first look, NODE and EDGE objects form a usual graph structure describing the drawing after initial vectorization. A PAT object typically contains

a set of edges identifying a recognized pattern on the drawing, but PATs can be utilized also in very different ways.

The object type *NODE* has an important subtype, termed *TEXT*, representing a transformed instance of a sample object. A *TEXT* instance is denoted as $text(sample, x, y, T, string)$ where *sample* is a "defined by" reference to a sample object, *x* and *y* are coordinates of the insertion point, *T* is a transformation usually given by an enlargement factor and rotation angle, and *string* is an ASCII sequence of characters. In map interpretation, we use two versions of *TEXT*:

- If *string* is omitted, then *sample* refers to the description of a certain symbol. For instance, *sample* may refer to a $pat(edge_1, \dots, edge_n)$ object giving vector description of a map symbol with (0, 0) coordinates as center point. In this case the *TEXT* object describes a recognized instance of the symbol at point (*x*, *y*).
- If *string* is given, then *sample* refers to a vector font $pat(letter_0, \dots, letter_{255})$ where, for any *i*, $letter_i$ is a PAT defining a character shape for ASCII code *i*. In this case *TEXT* describes a recognized inscription on the map.

Since *TEXT* is a subtype of *NODE*, it can be applied as endpoint of an edge. In this way we can represent special map symbols applied for instance at parcel corners on cadastral maps (Fig. 1).

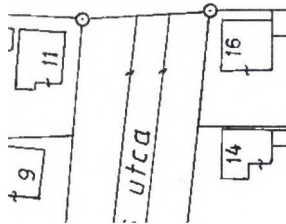


Fig. 1. Small circles denoting geodetical identified points

3 Spatial indexing

The model described above ensures fast data access along references, but spatial searches, for instance to find the nearest node to a given node, may be very slow. The problem can be solved by spatial indexing (for an overview see [17]). There are two main types of spatial indexes:

- *Tree structured indexes* are based on hierarchical tiling of the space, usually quadtrees are applied.
- *Grid structured indexes* use homogeneous grid tiling.

Although quadtrees have nice properties in general, in the case of drawing interpretation a grid index may be a better choice, because of the following reasons. On one hand, drawing density is limited by readability considerations (the number of objects in a grid cell is a priori limited). On the other hand, map interpretation algorithms normally use fixed search window (for instance, when recognizing dashed lines). A grid index of cell size near to the search window size can work efficiently.

To discuss indexing techniques, we define the *minimum bounding box* (MBB) of an object as the minimum enclosing rectangle whose edges are parallel to coordinate axes. Considering the DG model, MBB of an object z can be determined by computing minimum and maximum coordinates of nodes in $domain_c(z)$.

Fig. 2 shows a grid index of 3×3 tiles where a list of object id's is created to each grid cell. An id appears in the i -th list if the MBB of the object overlaps C_i . In this way the same object id may appear in several lists.

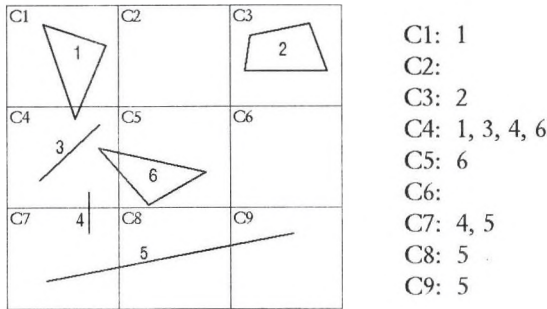


Fig. 2. Example of a grid index

We use a common linked list for grid index implementation, ensuring the insertion of a new id in constant time. As a consequence, a grid index for N objects can be generated in $O(N)$ time, which is better than the usual $O(N \cdot \log N)$ building time for quadtrees. However, object deletion cannot be solved in an efficient way in our index, that is why we recreate the index if necessary, instead of updating it.

4 Interpretation strategy

Initially the DG-document contains only sample PAT objects coding prototypes of symbols and characters to be recognized. Interpretation starts with some raw vectorization process (see [21] for an overview of vectorization methods). As a result of the vectorization, a NODE-EDGE graph description of the drawing is inserted in the DG-document. At this moment all normal objects are in layer 0. Map processing is performed as a sequence of recognition steps, each step consisting of three phases:

1. *Hypothesis generation.* PAT objects are created in the DG-document. For instance, if a set of edges e_1, \dots, e_n is recognized as a map object, then a $pat(e_1, \dots, e_n)$ is created with the layer number associated with the current map object type. Such an operation does not change the underlying data, thus the hypothesis generation is a reversible step ensuring the possibilities of backtracking and ignoring. PAT objects can describe a hierarchy of higher level structures like blocks and entities in [20].

2. *Verification of hypotheses* can be made by the user or by a higher level algorithm: PATs of false hypotheses are marked as "rejected" while correct ones as "accepted".

3. *Finalization.* Rejected hypotheses are dropped and accepted ones are processed, possibly making irreversible changes in the underlying data. In some cases finalization can be omitted or postponed, in this way preserving the possibility of backtracking.

The above procedure will be demonstrated on text recognition (Fig. 3).

1. In the hypothesis generation step, small connected subgraphs are detected as candidates of single characters (see pat_1 , pat_2 and pat_3 in Fig. 3). Next, if an aligned group of character candidates is found, a $pat(pat_1, \dots, pat_n, text_1, text_2)$ object is created where $text_1$ and $text_2$ are not defined yet. Rotation angle α of the string candidate is detected, and characters are recognized with rotation α and $180 - \alpha$, respectively. Recognition results are stored in $text_1$ for α and in $text_2$ for $180 - \alpha$. The recognition itself can be performed with some graph matching technique like in [12] and [13], or using a neural network [10].

2. String hypotheses can be verified by the user or applying some a priori information. For instance, when processing cadastral maps, the set of legal parcel numbers is usually given in an external database.

3. If a hypothesis is rejected, then all PATs and TEXTs, created in the hypothesis generation phase, should be deleted (Fig. 3). If accepted, then only the selected TEXT should be kept and all other objects – including the base vectors of the characters – should be deleted.

5 Application: interpretation of cadastral maps

The DG model has been applied to interpret Hungarian cadastral maps. Main processing phases are sketched below, for algorithmic details see [10, 11].

1. *Vectorization.* A thinning-based vectorization algorithm converts the whole scanned image into a set of vectors. Further on, denote N the number of generated vectors.

2. *Creating topology.* An initial DG-document (a NODE-EDGE graph) is generated from the set of vectors. This process takes $O(N \cdot \log N)$ time (it is based on sorting the nodes according to their coordinates, and unifying nodes of identical coordinates).

3. *Dashed line recognition.* For each dashed line candidate a $pat(edge_1, \dots, edge_n)$ is created. If grid indexing is applied, then recognition can be performed in $O(N)$

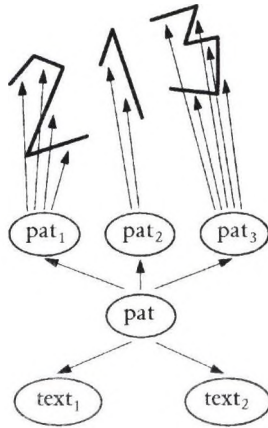


Fig. 3. Example of text recognition using the DG model. Arrows denote references between objects

time. If a dashed line hypothesis is accepted, then component edges are replaced with a single EDGE object having a layer number assigned to dashed lines.

4. *Text recognition.* String candidate PATs are created as shown in Section 4. A 17-element feature vector is generated for each character PAT, and recognition is performed by a feedforward neural network. Although vectorized symbols may have significant distortions as compared to the raster image, the neural network can learn these distortions and can produce successful recognition (Fig. 4).

5. *Recognizing connection-signs.* (A connection sign is applied on the boundary line of two map objects, and expresses that the two objects are logically connected, e.g. a building belongs to a given parcel, see Fig. 4).

6. *Recognizing small circle symbols* (Fig. 1). A hypothesis $pat(edge_1, \dots, edge_n)$ is created to each small closed convex polygon. If accepted, $edge_1, \dots, edge_n$ are deleted and nodes of these edges are replaced with a single TEXT object at the center point of the polygon.

7. *Drawing correction.* The initial raw vectorization has typical anomalies at corners and T-junctions (Fig. 4/b). Each of these anomalies is recognized as a $pat(pat_1, pat_2)$ where pat_1 contains edges to be corrected and pat_2 contains the new (corrected) edges. When accepted, edges in pat_1 are deleted, otherwise edges in pat_2 should be deleted. These corrections are performed only on edges with empty scope, that is, on edges that have not been recognized till now.

8. *Recognizing buildings and parcels.* A complex algorithm is applied, utilizing recognized connection signs, house numbers and parcel numbers [10]. A PAT object is created to each building polygon and parcel polygon. If accepted, the PAT is kept, otherwise rejected.

The above system produces an average recognition rate over 90 % [10], and the processing time of one recognition step takes only a few seconds for a total map sheet.

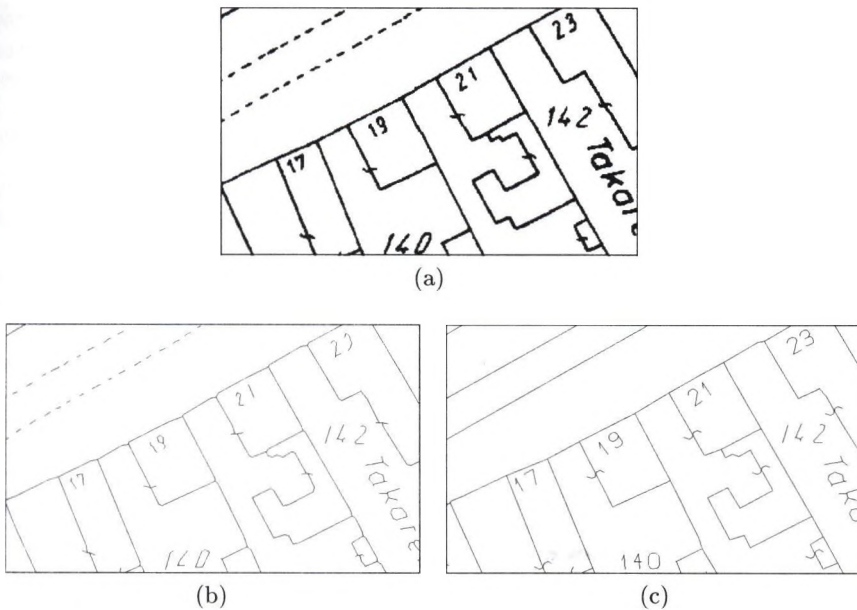


Fig. 4. Automatic interpretation of Hungarian cadastral maps: a) scanned raster image, b) raw vectorization, c) recognition result

6 Conclusions

The presented vector-based interpretation strategy can be applied not only for map processing but also for any kind of graphics recognition. The topological data model and spatial indexing together ensure linear processing time in most cases, this fact supports interactivity and makes it possible to create rather complex algorithms in realistic time.

When applying an interpretation system, it is a usual difficulty that the user is not familiar with inherent algorithms and data structures, and therefore cannot make optimal control of the system. We think that basic ideas of our DG model are simple enough to understand for the user, and this fact supports optimal interactive work.

Current version of the DG model contains only 4 object types, NODE, EDGE, PAT and TEXT (the latter one as subtype of NODE). This model can be extended by defining *ARC* as subtype of EDGE and *CIRCLE* as subtype of PAT to describe circular elements. These objects have less importance in map interpretation, but crucial when processing engineering drawings.

References

1. Chen L.-H., Liao H.-Y., Wang J.-Y., Fan K.-C., Hsieh C.-C.: An Interpretation System for Cadastral Maps. Proc. of 13th Internat. Conf. on Pattern Recognition, IEEE Press, Los Alamitos, California, pp. 711-715, 1996.

2. Arc/Info Data Management. Environmental System Research Institute, Inc., 1994.
3. Atwood T., Duhl J., Ferran G., Loomis M., Wade D.: The Object Database Standard: ODMG-93. Morgan Kaufmann Publishers, San Francisco, California, 1994.
4. Boatto L., Consorti V., Buono M., Zenzo S., Eramo V., Esposito A., Melcarne F., Meucci M., Morelli A., Mosciatti M., Scarci S., Tucci M.: An Interpretation System for Land Register Maps. *Computer*, Vol. 25, No. 7, pp. 25-33, 1992.
5. Ebi N. B.: Image Interpretation of Topographic Maps on a Medium Scale Via Frame-based modelling. *International Conference on Image Processing*, IEEE Press, California, Vol. I., pp. 250-253, 1995.
6. Goodchild M. F., Kemp K. K. (eds): NCGIA Core Curriculum. National Center for Geographic Information and Analysis (NCGIA), University of California, Santa Barbara, 1990.
7. Graphics Recognition (series). Selected papers of GREC workshops, *Lecture Notes in Computer Science*, Springer, No. 1072 (1996), No. 1389 (1998), No. 1941 (2000).
8. Hartog J., Kate T., Gerbrands J.: Knowledge-Based Segmentation for Automatic Map Interpretation. *Graphics Recognition. Lecture Notes in Comp. Sci.* 1072, Springer, pp. 159-178, 1996.
9. Katona E., Palágyi K., Tóth N.: Signature verification using neural nets. *Proceedings of 9th Scandinavian Conference on Image Analysis*, pp. 1115-1122, 1995.
10. Katona E., Hudra Gy.: An Interpretation System for Cadastral Maps. *Proceedings of 10th International Conference on Image Analysis and Processing (ICIAP 99)*, IEEE Press, pp. 792-797, 1999.
11. Katona E.: Automatic map interpretation. Ph.D. Thesis, Department of Physical Geography, University of Szeged (in Hungarian), 2001.
12. Lladós J., Sanchez G., Marti E.: A String-Based Method to Recognize Symbols and Structural Textures in Architectural Plans. *Graphics Recognition. Lecture Notes in Comp. Sci.* 1389, Springer, pp. 91-103, 1998.
13. Messner B. T., Bunke H.: Automatic Learning and Recognition of Graphical Symbols in Engineering Drawings. In R. Kasturi and K. Tombre (eds): *Graphics Recognition - Methods and Applications. Lecture Notes in Comp. Sci.* Vol. 1072, Springer, pp. 123-134, 1996.
14. Nagy G., Samal A., Seth S., Fisher T., Guthmann E., Kalafala K., Sarkar L. L. P., Sivasubramaniam S., Xu Y.: A Prototype for Adaptive Association of Street Names with Streets on Maps. *Graphics Recognition. Lecture Notes in Comp. Sci.* 1389, Springer, pp. 302-313, 1998.
15. Niemann H., Sagerer G. F., Schröder S., Kummert F.: ERNEST: A Semantic Network System for Pattern Understanding. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 9, pp. 883-905, 1990.
16. OpenGIS Consortium, Inc.: Simple Features Specification for SQL - Version 1.1. <http://www.opengis.org>, 1999.
17. Samet H.: *Design and Analysis of Spatial Data Structures*. Addison Wesley, 1989.
18. Schawemaker J. G. M., Reinders M. J. T.: Information Fusion for Conflict Resolution in Map Interpretation. *Graphics Recognition. Lecture Notes in Comp. Sci.* 1389, Springer, pp. 231-242, 1998.
19. Suzuki S., Yamada T.: MARIS: Map Recognition Input System. *Pattern Recognition*, Vol. 23, No. 8, pp. 919-933, 1990.
20. Vaxiviere P., Tombre K.: Celsstin: CAD Conversion of Mechanical Drawings. *Computer*, Vol. 25, No. 7, pp. 46-54, 1992.
21. Wenyin L., Dori D.: From Raster to Vectors: Extracting Visual Information from Line Drawings. *Pattern Analysis and Applications*, Springer, 1999/2, pp. 10-21, 1999.

Számítógépes spermium minősítő rendszer és alkalmazásai

Kelemen Dezső¹, Gábor György², Szabó József¹

¹ PICTRON Kft, Budapest

² Állattenyésztési és Takarmányozásai Kutató Intézet, Herceghalom

Az ember azt gondolná, hogy a laboratóriumi alkalmazásoknál az idő nem igazán okoz problémát. Ezt cáfolandó bemutatjuk a spermiumminősítésre kifejlesztett CASPAR (Computer Aided SPERM AnalyseR) elnevezésű programot. A spermium objektív minősítésének mesterséges megtermékenyítés esetén van jelentősége, hiszen csak egészséges, határozott mozgással haladó spermium esetén van esély a megtermékenyítésre. A mikroszkóp alá helyezett mintát vizsgálva, hihetetlenül gyorsan változó képet látunk. A spermiumok mozgását a Makler kamra alkalmazásával közel egy síkra korlátozhatjuk, és speciális digitalizáló kártya alkalmazásával másodpercenként 25 képet képes a rendszer rögzíteni. Az egészséges spermiumok egy kép-váltás közbeni 1/25 mp alatt annyit mozdu-
nak el, hogy a két pozíció közötti kapcsolat még beazonosítható. Ehhez biztosítani kell, hogy ne túl sűrűen legyenek a spermiumok, ezért hígított mintákat lehet csak alkalmazni. A program több egymást követő képet összead. Az így kapott képen az egészséges spermiumok jellegzetes füzérszerű útvonalat rajzol-
nak fel. Egy-két perces munkával kijelölhetők az álló (elhalt), a sérült és az egészséges (határozott mozgással rendelkező) sejtek és azok útvonalhossza. Ezt követően program kiszámítja az egészséges sejtek haladási sebességét és egyéb fontos jellemzőket (pl. élő/holt arány). Egymás után két-három szekvencia fel-
dolgozása után statisztikailag jól értékelhető, objektív minősítést kapunk. A mérési eredmények, a képek és videó részletek tárolhatók, később előkereshetők és összevethetők az újabb vizsgálat eredményeivel. Korábban szubjektív érték-
ítélet alapján, néhány fokozatú skála szerint sorolták be a mintákat. A néhány perces plusz munkáért, bőven kárpótolnak az objektív és dokumentálható ered-
mények. Előadásunkban a CASAR programot és annak alkalmazási eredményeit mutatjuk be.

Pattern matching by Gabor filters for different views in a multi-camera system

Tamás Szirányi¹, László Tizedes²

¹Analogical Computing Laboratory, Computer. & Automation Research Institute,
Hungarian Academy of Science, H-1111 Budapest, Kende u. 13-17, Hungary
sziranyi@sztaki.hu

²Tateyama Hungary Laboratory, H-1111 Budapest, Zenta utca 1.

¹²also with Univ. Veszprém, Dep. Image Proc. & Neurocomp.,
H-8200 Veszprém, Egyetem u. 10

Abstract In our surveillance multi-camera system different camera views are applied to follow moving objects. The environment might be very confused: shadows, mirrors, obstacles, etc. The cameras are with panoramic lenses, but they have rather low resolution for exact recognition. For tracking, a moving object can be followed if the multi-object scene can be matched for the multiple view by recognizing the same object. We have solved this multiple-object from multi-view problem by using matched Gabor filters.

1. Introduction

Gabor's original paper [1] proposed synthesis of signals using complex sinusoids windowed by a Gaussian function (Gabor functions). His method is thus closely related to the short time (windowed) Fourier transformation with the special Gaussian window. It is also called Gabor transformation. Since Gabor functions are good local spectral descriptors, being very close to the visual response of brain in some pattern analysis tasks, it is natural to use it in many of the pattern recognition problems, like: texture analysis [2], face recognition, iris identification [3], pattern recognition.

Panoramic image acquisition makes multiple or mechanically controlled camera systems needless for many applications. Panoramic Annular Lens's (PAL [4]) main advantage to other omnidirectional monitoring systems (e.g. [6]) is that it is a cheap, small, compact device with no external hyperboloidal surface in a glass sphere as in other panoramic optical devices. Converting the annular image to a rectangular one, we get a low-resolution (cc. 1000x150) image. In [5] we have developed some algorithms, which can analyze this low-resolution image to get motion information for surveillance and smoke detection application. This camera alone is able to solve simple surveillance tasks. In the more complex case, when more cameras are organized together to follow and recognize multiple objects, new problems are arising due to the object-registration and positioning.

Omnidirectional vision is a well-suited tool for general surveillance systems or robotic vehicles [6]. The main problem is with the low resolution, so we should use special methods [7]. [8] proposes an invariant object detection method based on

evidence accumulation and the Gabor transforms feature to detect object parts. This algorithm robustly detects arbitrary shaped objects in cluttered environments with invariance to translation, rotation, scaling and small deformation. Another main problem is the partial occlusion of objects to be tracked. In [9] we see a system for the interpretation of camera images of scenes composed of several known objects with mutual occlusion. A semi-automatic way is shown by at varying background. Objects are recognized by dynamic link matching, where the scenes are analyzed by the recognition of the objects present and by the determination of their occlusion relations through stored model graphs.

In the present work we demonstrate a new searching method to find the same object of different views of different cameras. The (panoramic) images have poor resolution at a wide (full) angle. The only parameter we know is the distance of cameras, and maybe their relative position to the map of the environment.

In our low-resolution system we are not looking for special image parts. Gabor filters scan on the moving blocks of the image, and the highest scores are detected. In the image of another camera view, Gabor scanning also processes the moving blocks, then the scaling of Gabor is varied to get similar series of Gabor coefficients. Different moving blocks of different camera images are paired on this similarity.

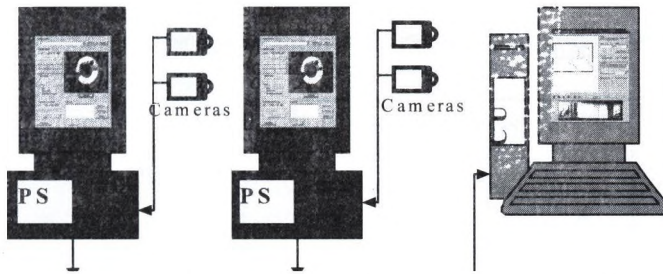


Figure 1 Multi-camera monitoring system and central unit

2. The measuring environment

First, we show in Fig. 1 a simple monitoring system, using two or more cameras (PS) and a central unit for camera and map organization, learning and alarm generation. In Figs. 2 and 3 the camera-controller and the central unit of a multi-camera system is shown.

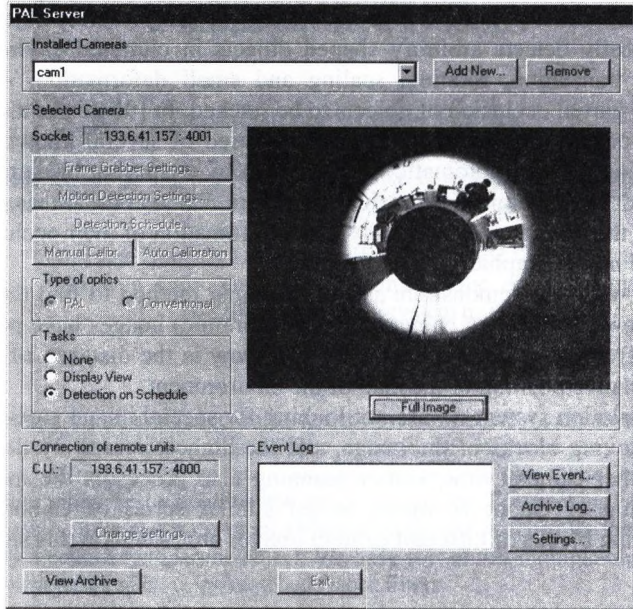


Figure 2 Panoramic Server (PS) module of multi-camera monitoring system

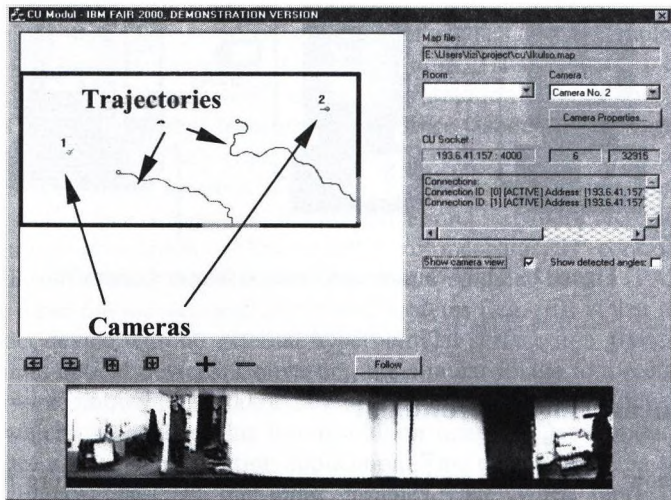


Figure 3 Central Unit (CU) module of multi-camera monitoring system

3. Pattern matching and Gabor filters

Figure 4 shows the same moving object grabbed from two different cameras. We can see that all of the color, rotation and occlusion are different in the two samples, moreover we have low resolution images.

Our pairing task is to find the corresponding objects in the different views. For this registration, we tested three different methods by using

- Color information,
- Projection,
- Gabor filter.

Color information means the simple image statistics (with or without histogram equalization), and the similarity is counted on the differences of the correlated histograms.

Projection means that we produce a projection-histogram into vertical and horizontal directions in the motion window.

Gabor filters are successful in texture analysis and edge registration. Although our objects are under-sampled, resulting in poor edge information, the textural and partial edge information can be exploited by Gabor filters for registration purposes.

Registering by using color histogram statistics and projections is often used in matching and image indexing. In our case there may be problems with the occlusion, scaling and rotation. We might see the same object from the reverse direction, occluded by something in a half-shadow (darkness). However, in the following we will see how to find them for pairing.

The Gabor filter matching algorithm:

The $G(X)$ Gabor filters are different in parameters of scaling and window-width, denoted the parameters by $X \in \mathbb{R}^2$. We have a predefined set of Gabor filters, $G(X_k) \in \mathcal{G} = \{G(X_k), 0 < k < N\}$. Presently this set has $N=10$ elements, defined as the preferred filter set in [10,11]. These Gabor filters are used in the next algorithm:

- Select the $MB(\cdot)$ motion blocks [5] in each camera's view,
- Scan the blocks by the predefined set of different Gabor filters (Gabor transformation),
- Select the highest scores for the Gabor coefficients for each block.

Select an $Im(1)$ view-image and a second $Im(2)$ view for matching trial of objects.

1. Select the i -th motion block in $Im(1)$, signing the block by $Im(1) \Rightarrow MB(i)$. The maximum score of Gabor filters is $G(X_M) \in \mathcal{G}$ at parameter setting X_M .
2. Select the j -th motion block in $Im(2)$, signed by $Im(2) \Rightarrow MB(j)$. Scan the image with $G(X)$ filters, varying X with appropriate steps in a predefined interval.
3. Select the highest score of Gabor filters, denoted by $G(X_{scan})$, in $Im(2) \Rightarrow MB(j)$.
4. Pairing X_M in $Im(1) \Rightarrow MB(i)$ and X_{scan} in $Im(2) \Rightarrow MB(j)$, the scaling parameters of X_M and X_{scan} are related to rescale the set of \mathcal{G} to

characterize $Im(2) \Rightarrow MB(j)$ by the maximum coefficients of the rescaled Gabors, denoted by G_{rs} .

5. The test results of G in $Im(1) \Rightarrow MB(i)$ and G_{rs} in $Im(2) \Rightarrow MB(j)$ are paired and the difference between the corresponding coefficients are counted. We chose the smallest 5 differences of the set of 10 for calculus.
6. The error-summation on the above 5 distances gives the similarity-distance $D_{i,j}$ between $Im(1) \Rightarrow MB(i)$ and $Im(2) \Rightarrow MB(j)$.
7. For more accurate comparison, the second or third highest scores can also be applied as the base of scaling in step 4.
8. If there is no a priori estimation about the position of the compared moving objects, we correlates a $MB(i)$ block of $Im(1)$ with every $MB(j)$ block of $Im(2)$ to find the best matching: change for $j:=j+1$ and go to 2, looking for $j_{min} = \min(D_{i,j}, MB(j) \subset Im(2))$.
9. We step to the next reference block in $Im(1)$, continuing from 1.
10. If the same matching is found for a block in $Im(2)$ referenced to different blocks in $Im(1)$, more coefficients or additional (color etc.) information is considered.

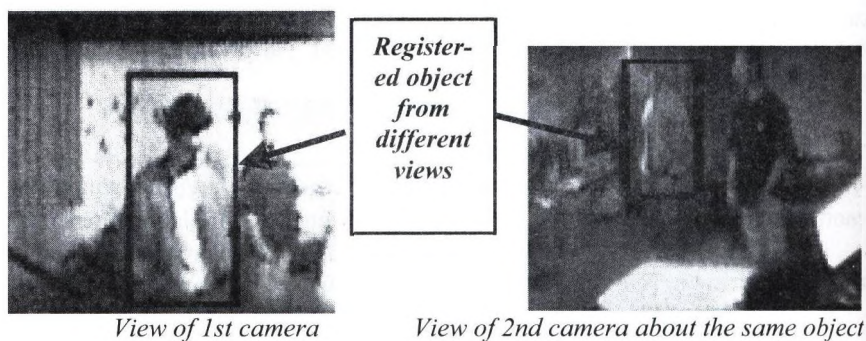


Figure 4

4. Recognition results

In the testing procedure the average size of motion blocks was about 35×55 , and the CPU time was measured for the case of 2 different objects. Processing unit was a 533MHz Celeron.

Results of different methods are compared in Table 1. If we use the data of more than one frame (tracking and evaluating an object for a longer series), then we can achieve a better registration accuracy by using maximum likelihood decision instead of a simple Bayesian-decision. In this way we get a much better recognition result, see

Table 2. The accuracy of Bayesian decision can be studied by the probability curves in Figure 5.

Method	Processing time	Efficiency
Color histograms	8 ms	62 %
Projections	9 ms	69 %
Matched Gabor filters	595 ms	88 %

Table 1 Comparison of different methods for object registration by using only one frame

Method	Processing time	Efficiency
Color histograms	30 ms	68 %
Projections	33 ms	74 %
Matched Gabor filters	1791 ms	92 %

Table 2 Comparison of different methods for object registration by using only 3 consecutive frames

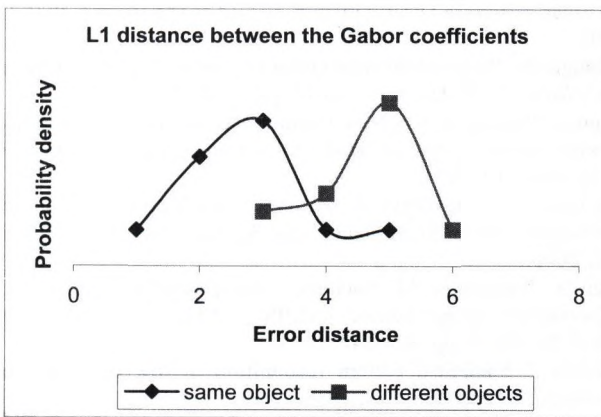


Figure 5

5. Conclusions

Our Gabor-based matching procedure is quite effective in low-resolution multi-camera images. It can help to recognize objects and their relative position when other methods are unsuccessful.

The above exhaustive procedure is necessary when we detect more than one new moving objects in the same time, or color & projection based methods are confused, and we do not have any a prior or additional information. When objects are once discriminated successfully, then the tracking and in-motion occlusion handling can appropriately follow the moving objects [5].

If not only 3 consecutive frames, but a longer series of motion blocks are analyzed, then the most characteristic Gabor filter can be continuously scaled, resulting in a fast search. The statistical data over this matched filter also can be well estimated for a more powerful pattern matching of statistical methods for low-resolution or undersampled patterns [12][13].

Acknowledgements

This work has been supported by The Hungarian research Fund (OTKA) and the Tateyama Hungary Ltd.

References

- [1] D. Gabor, "A Theory of Communication", *Journ. IEE*, V.93, pp.429-457, 1946.
- [2] M.R. Turner, "Texture Discrimination by Gabor Functions", *Biol. Cybernetics*, V.55, pp.71-82, 1986.
- [3] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence", *IEEE Tr. PAMI*, V.15, No.11, pp.1148-1161, 1993
- [4] P. Greguss, "Exoscope – a New Omnidirectional Imaging and Holographic Device for Life Science Studies", *Optical Methods in Biomedical and Environmental Sciences*, Elsevier BV, pp. 309-312, 1994.
- [5] I. Kopilovic, B. Vágvölgyi, T. Szirányi, "Application of Panoramic Annular Lens for Motion Analysis Tasks: Surveillance and Smoke Detection", *15th ICPR*, Barcelona, IAPR & IEEE, 2000.
- [6] Y. Yagi, Y. Nishizawa, M. Yachida, "Map-Based Navigation for a Mobile Robot with Omnidirectional Image Sensor COPIS", *IEEE Transactions on Robotics and Automation*, Vol. 11. No. 5. pp. 624-647, 1995.
- [7] T. Szirányi, "Statistical pattern recognition of low resolution pictures", *Pattern Recognition Letters*, V.8, pp. 221-228, 1988.
- [8] H.J. Park and H.S. Yang, "Invariant object detection based on evidence accumulation and Gabor features", *Pattern Recognition Letters*, Vol. 22 (8), pp. 869-882, 2001.
- [9] L. Wiskott, C. Malsburg, "A Neural System for the Recognition of Partially Occluded Objects in Cluttered Scenes", *Int. J. of Pattern Recognition and Artificial Intelligence*, 7(4):935-948, 1993.

- [10] J. Shao and W. Förstner, "Gabor Wavelets for Texture Edge Extraction", *ISPRS Commission III Symposium on Spatial Information from Digital Photogrammetry and Computer Vision*, 1994, WG/2, Munich, Germany
- [11] T.P. Weldon, W.E. Higgins, "An Algorithm for Designing Multiple Gabor Filters for Segmenting Multi-Textured Images", *IEEE Int. Conf. Image Processing*, Chicago, IL, 4-7 October 1998.
- [12] T. Szirányi, "Subpixel Pattern Recognition by Image Histograms", *Pattern Recognition*, Vol.27, pp.1079-1092, 1994.
- [13] T. Szirányi, A. Hanis, "Sub-pattern texture recognition using intelligent focal-plane imaging sensor of small window-size", *Pattern Recognition Letters*, Vol. 20, pp.1133-1140, 1999.

Creating Video Animations Combining Stochastic Paintbrush Transformation and Motion Detection

Levente Kovács¹, Tamás Szirányi²

¹University of Veszprém, Department of Image Processing and Neurocomputing,
H-8200 Veszprém, Egyetem u. 10, Hungary
levente.kovacs@freemail.hu

²Analogical Comp. Lab., Comp. & Autom. Inst., Hungarian Academy of Sci.,
H-1111 Budapest, Kende u. 13-17, Hungary
sziranyi@sztaki.hu

Abstract In this paper we propose a method for creating animation and animation-like video sequences from an ordinary video recorded by any means available. The method is based on the Paintbrush Transformation earlier patented by our Department [7], and on different motion detection algorithms [3,4,5]. One of the goals of the method is to obtain animations, cartoon-like outputs from a usual camera-recorded image sequence. The method inherits the properties of the paintbrush transformation method like well-defined contours, acceptable distortion, and a painting-like view with no fine details below a limit. The resulting output is a series of video frames stored as brush-strokes and motion data between the frames, compressed for size reduction.

1. Introduction

There are numerous image and video coding and compression methods available nowadays. Images can be interpreted in several ways by decomposition into basic functions: strokes [8,9], fractals [10], etc. Each of these is natural in some sense. The idea behind paintbrush transformation techniques was to transform an image so that it gives the sensation of a painting, the painting process being a simulation of the real painting process by using simplified artificial strokes. The parameters of consecutive strokes can then be used for image description and compression as well.

Taking our stochastic paintbrush transformation as a starting point [1,2] we tried to develop a method of motion picture transformation with the goal of obtaining animation-like outputs from ordinary video inputs, which have the properties of the original paintbrush transformation.

The method is based on the following idea: besides the full-frame paintbrush transformation of the key-frames, the transformation is based on the motion information obtained from motion detection between consecutive frames. By using the optical flow data obtained, the areas where motion occurred are re-transformed and the next transformed frame is obtained from the previous transformed frame and the processed motion areas.

The output of the transformation is an intermediate format, in which the frames of the output video are stored as series of brush-strokes, and the motion data is also stored between the frames, run-length-encoded. The whole intermediate format undergoes a Huffman-encoding [6], obtaining smaller output sizes. By decompression and repainting the frames of the transformed video - using the stored stroke-series and the motion data - can be reconstructed and recoded into any available video format.

2. The Paintbrush Transformation

The development of paintbrush transformation [1] had the goal of achieving an automated method which imitates, simulates the painting process of a real painter, to obtain a picture similar to a real painting, where the purpose of the painter is to portray something which looks like to be a real scenery.

During the transformation process the picture is being constructed randomly with bigger brush-strokes, then it gets refined with smaller and smaller strokes, and the picture increasingly resembles the model.

The transformation starts with bigger strokes then refines the actual picture with smaller strokes. In the original algorithm during the refining process a stroke with given size, orientation and color gets on the picture if it's placement improves it, makes the picture converge more to the model. The coordinates of the next examination were originally picked by exhaustive stochastic search, then an optimized model was developed, in which the exhaustive search was replaced by a decision based on the approximation of the error caused by placing the strokes, where the target density is dynamically fit during the process and the characterizing densities are changing through the iterations.

An important feature of the transformation is that there are sharp edges and well-defined patterns and areas although we have no a priori information about the textures, contours or shapes on the picture.

The generated image can be described by a series of strokes, which can be used for moderate compression, a stroke being described by seven bytes of data (id, position, color-information).

3. Application On Motion Picture

The first step on the way of producing an animation-like output was to paintbrush-transform every frame of an input video. However, because of the stochastic nature of stroke-placement this method generated vibrating image sequences, and the output was quite large because of the large number of strokes necessary to describe the frames.

It seems that the motion detection-based approach gives the answer. The concept of the method is that we only fully transform the key-frames. Between the key-frames we use the optical flow data obtained from motion detection algorithms, and

transform only the motion areas. After eliminating those strokes, which are covered by others, we get much fewer strokes and thus much smaller output – compared to transforming the whole image.

4. Motion Detection

We use basically two motion detection algorithms: gradient-based and block matching motion detections.

The usual starting point for velocity estimation [3,4] is to assume that the intensities are shifted from one frame to the next, and that the shifted intensity values are conserved:

$$f(x, y, t) = f(x + u_1, y + u_2, t + 1) \quad (1)$$

A path $(\mathbf{x}(t), \mathbf{y}(t))^T$ along which intensity is equal to a constant c must satisfy:

$$\frac{d}{dt} f(x(t), y(t), t) = 0 = \frac{\partial f}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial f}{\partial t} \frac{\partial t}{\partial t} = f_x u_1 + f_y u_2 + f_t \quad (2)$$

Since it's impossible to recover velocity given just the gradient constraint over a single position, we combine constraints over a region. To measure the extent to which the gradient constraints are not satisfied we square the constraints and sum them:

$$E(u_1, u_2) = \sum_{x,y} g(x, y) [u_1 f_x(x, y, t) + u_2 f_y(x, y, t) + f_t(x, y, t)]^2 \quad (3)$$

where $g(x, y)$ is window function that is zero outside the neighborhood within we use the constraints.

We solve for u_1 and u_2 :

$$\frac{\partial E(u_1, u_2)}{\partial u_1} = \sum_{x,y} g(x, y) [u_1 f_x^2 + u_2 f_x f_y + f_x f_t] = 0 \quad (4)$$

$$\frac{\partial E(u_1, u_2)}{\partial u_2} = \sum_{x,y} g(x, y) [u_1 f_x f_y + u_2 f_y^2 + f_y f_t] = 0$$

$$M = \begin{bmatrix} \sum g f_x^2 & \sum g f_x f_y \\ \sum g f_x f_y & \sum g f_y^2 \end{bmatrix}$$

$$u \approx -M^{-1} b$$

In the block matching motion detection [3,4,5]; we search for a block with a given radius in an also given radius area around it, along a spiral path. This method is more sensitive to non-translation motion.

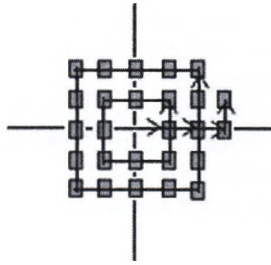


Fig. 1. Spiral path in block-matching

The optical flow field data obtained with these motion detection methods are used when transforming the motion areas of the frames. The motion data calculated are stored as a two-dimensional array, each element containing the position where that element will move when the motion occurred.

5. The Algorithm

The main steps of the algorithm are as follows:

1. Setting of parameters (key-frame frequency, type of motion detection to use, other parameters).
2. Paintbrush transformation of frame $F(i)$ to $F'(i)$ and writing of $F'(i)$ into the intermediate format.
3. Motion detection calculation between $F(i)$ and $F(i+1)$.
4. Transformation of the areas where motion occurred using the optical flow field data calculated at the previous step.
5. Writing the partially transformed frame data onto $F'(i)$, generating $F'(i+1)$ and writing it into the intermediate format along with the motion data.
6. If the next frame will be a key-frame then increase i and jump to step 2., otherwise to step 3.

The steps of the transformation used at step 4 are as follows:

1. Choose stroke-set. If last one, go to step 12.
2. Convolution of motion areas with the strokes of the set (color data estimation).
3. Generation of error image between the actual step of transformation and the original area.
4. Blurring the error image, generation of histogram from absolute values of error image.
5. Picking the coordinates and orientation of the next proposed stroke over the motion area. Color data taken from step 2., orientation data from the generated motion data.
6. If the error on the actual position isn't lower than a threshold value, and the stroke lowers the error value – more then a threshold -, then place the stroke.

7. Calculating the error image between the original and the image got after placing the stroke, over the motion area. If there is improvement over a threshold value, then accept.
8. If we are over a specified number of tries, then check the difference of the actual image from the image got from the previous iteration. If it's under a threshold, we had no real improvement.
9. If there are more strokes in the actual set and there was some improvement at step 8., then go to step 5.
10. Clear those strokes, which are completely covered by other strokes.
11. Jump to step 1.
12. Generate the next transformed frame from the painted motion area and the transformed image obtained from the previous iteration step. The function realizing this algorithm returns with this frame.

After painting we only have to consider those areas where motion has occurred, and place these areas over the previously generated transformed frame. Then we have to write the frame to the intermediate format file, along with the motion data – all required for final video reconstruction.

6. Comparison of Compression methods

For comparison we choose the Cinepak codec, because it has long been the most popular codec for AVI files, and because it provides good replay speed and image quality. It is a middle-class middle-quality codec, that's why we compared out under-development format to it. It must be note that our method is in its infancy; a lot can be done to better compare it with professional codecs.

The two graphs in Figure 2 show the MSE data from frame to frame of a Cinepak-generated output and our output video relative to the original input video.

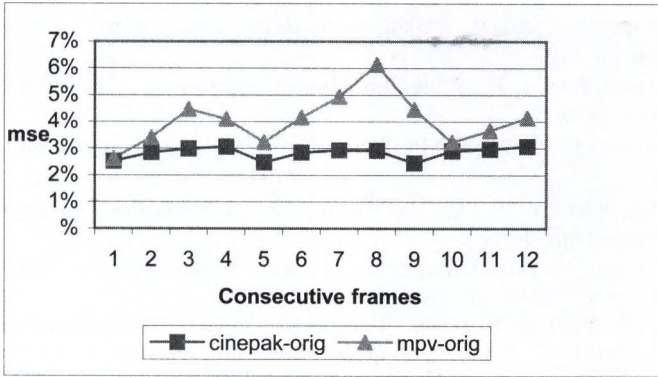
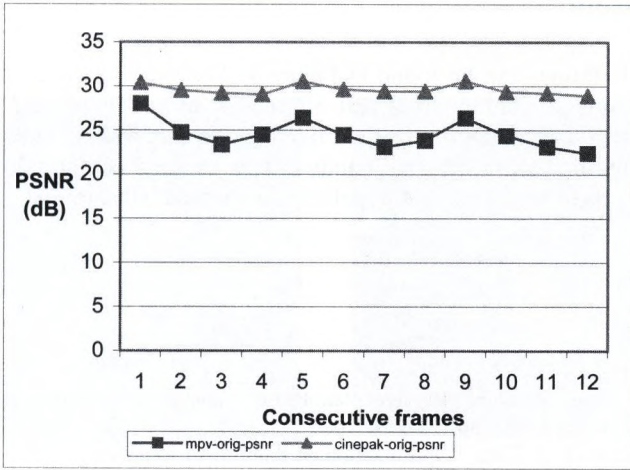


Fig. 2. MSE values of Cinepak and Painting coder

The graph in Figure 3 shows the error between numerous transformed videos and the original input video – and simultaneously the respective Cinepak-coded videos and the original one.

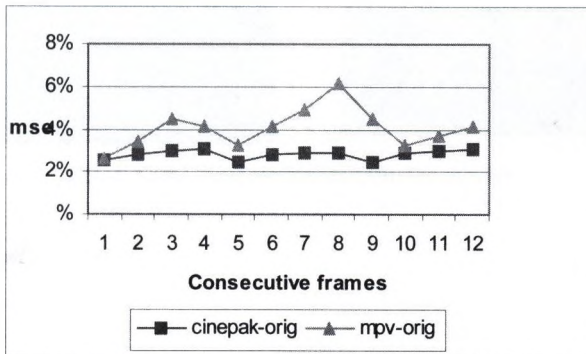


Fig. 3. Overall error rate

7. Conclusion

Some example frames can be found in Figure 4. The method we are developing is a way of generating animations from real-life scenes, in motion picture. The algorithm works like fractal-based algorithms (and like the Cinepak codec): slow encoding, fast decoding, requiring relatively small bandwidth, with good compression ratios. There are still problems to be solved, but hopefully the method will turn out to be usable and useful.

References

- [1] T. Szirányi, Z. Tóth: "Random Paintbrush Transformation", *15th ICPR, Barcelona, IAPR & IEEE, V.3*, pp.155-158, **2000**
- [2] T. Szirányi, Z. Tóth: "Optimization of Paintbrush Rendering of Images by Dynamic MCMC methods", *Lecture Notes Comp. Sci.*, Vol. LNCS 2134, pp.201-215, 2001
- [3] Simoncelli, Eero P., "Distributed Representation and Analysis of Visual Motion", 1993
- [4] Heeger, David. J., "Notes on Motion Estimation", *Psych 267/CS 348D/EE* p. 365, 1998
- [5] J. Barron, D.J. Fleet, "Performance of Optical Flow Techniques", *CVPR*, 1992
- [6] Wells, Richard B., *Applied Coding and Information Theory for Engineers*, Prentice-Hall, 1999
- [7] T.Sziranyi, Z. Toth, I. Kopilovic, "Paintbrush Transformation", *Hungarian Patent Office*, No.: P00 00041, 2000
- [8] H. H.S. Ip, H. T. F. Wong, "Generation of Brush Written Characters with fractal Characteristics", *ICCPOL, 17th Int. Conf. on Computer Processing of Oriental Languages*, pp. 156-161, Hong Kong, April, 1997.
- [9] H. H S Ip, H. T F Wong, "Calligraphic Character Synthesis using Brush Model", *CGI'97, Computer Graphics International conference*, pp. 13-21, 1997
- [10] Y. Fisher. (ed.), *Fractal Image Compression*, Springer Verlag , 1994

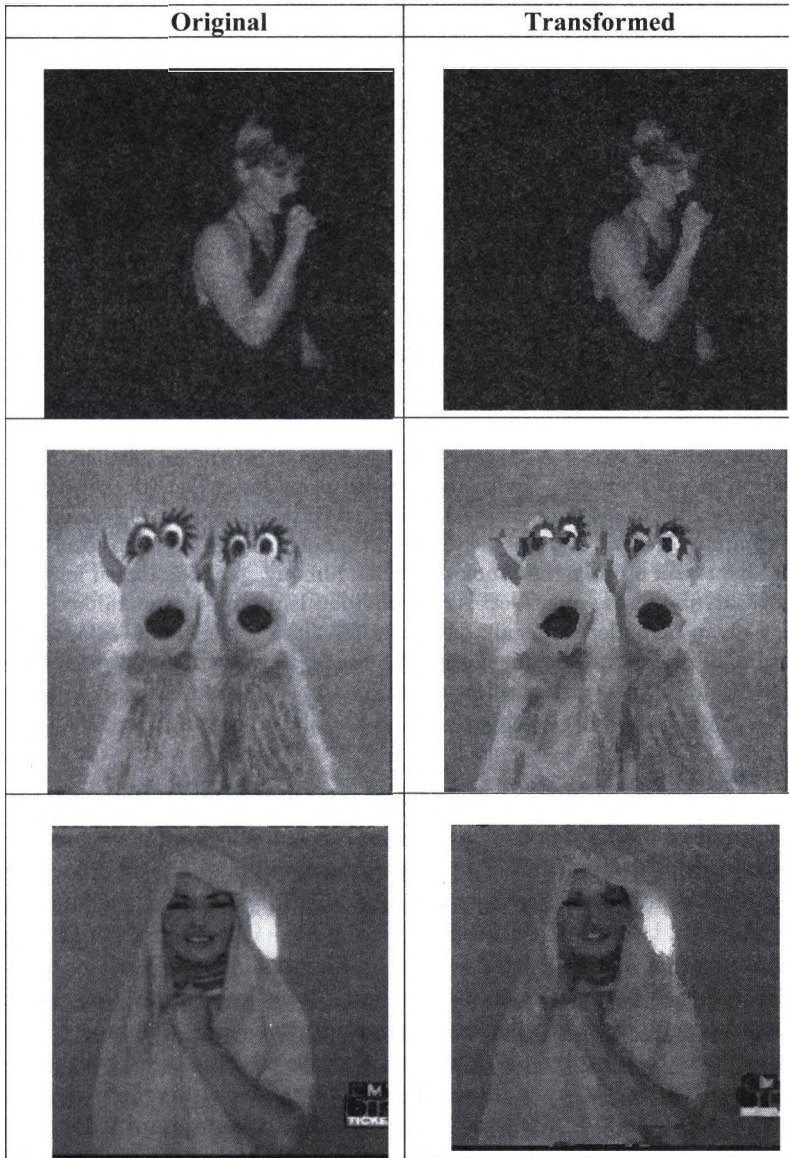


Fig. 4. Some example frames from the movies

Digitális MOZgóképhelyreállító Rendszer Filmarchívumok Számára (DIMORF)

Czúni László¹, Szirányi Tamás², Licsár Attila, Hanis Attila, Schanda János,
Kránicz Balázs, Farkas Péter

¹ Veszprémi Egyetem, Képfeldolgozás és Neuroszámítógépek Tsz.
8200 Veszprém, Egyetem u. 10.
czuni@almos.hu

² MTA SZTAKI, Analogikai és Neurális Számítások Laboratórium,
1111 Budapest, Kende u. 13-17.
sziranyi@sztaki.hu

Absztrakt. Kivonat A Magyar Nemzeti Filmarchívumban jelenleg több ezer film vár mentésre, javításra, a jelenlegi módszerekkel (ill. körülmények között) évente mindössze 1-2 film javítása megoldott. Ugyanakkor napjaink korszerű digitális képbeviteli, feldolgozási és írási eszközei lehetővé teszik a filmarchívumok tömérdek filmanyagának korszerű feldolgozását, javítását, mentését. Az itt röviden ismertetett munkánk célja egy olyan berendezés megalkotása, ami felváltva a hagyományos analóg technikákat, a képfeldolgozás legújabb módszereivel törekszik minél tökéletesebb filmjavítást, mentést elérni.

1. Bevezetés

Az analóg javítási módszerek csak a jellemző hibák kis hányadát képesek kezelni, ugyanakkor gyakran előfordul, hogy egy-egy egyedülálló értéket képviselő film visszafordíthatatlan sérüléseket szenved a javítási folyamatban. Célunk, hogy az általunk alkotott berendezés a hibák nagyobb tárházát legyen képes kezelni a film további sérülésének valószínűségének minimalizálása mellett.

Módszereink: A munka kezdeti fázisában több alternatívát dolgozunk ki a javítás/mentés módjaira, jelenleg ezeknek a feltérképezése, elemzése, prototípus algoritmusok tesztelése folyik. A feldolgozás alapvető menete a következő:

Digitalizálás: A 35mm-es filmek (évjárattól, típustól függően) nagyságrendileg 6000×4000 , 4000×3000 vagy ennél kisebb, de minimum 2000×1000 -es felbontás igényelnek, ami igen megnehezíti a költséghatékony feldolgozást. Mivel a digitalizálás a film mozgatásával jár (és ez számos veszélyt rejt magában a film szakadásától kedve egészen a nitrát alapú filmek öngyulladásáig), ezért a digitális adatmozgatáson kívül a mechanikai részek kidolgozása is nehéz feladat.

Feldolgozás: A feldolgozás nehézségei alapvetően a felbontásból adódó adatmennyiségből adódnak. Ezen bár lehet enyhíteni többfelbontásos feldolgozással, de bizonyos hibák (pl. szemcsézettség, hajszál karcok) csak a legnagyobb felbontás mellett orvosolhatók. A másik alapvető probléma valójában nem csak kimondottan technikai jellegű: mivel a filmalapanyag romlásának következtében

információ vesztes fordul elő, ezért gyakran etikai kérdés, esetleg a restaurátor ízlésének a függvénye hogy milyen is lesz a helyreállított mozgókép színvilága, szerkezete, stb. A filmekben előforduló hibák igen széles skálán helyezkednek el, a leggyakoribbak közöttük: különböző karcok, szennyeződés, por, vibráció, villódzás, penész, hiányzó kockák, feliratok. Ezek digitális korrigálását három fő csoportba oszthatjuk:

1. Hibák korrigálása a világosságban, színben.
2. Szennyeződés, por, karcok és egyéb 1 kockán jelentkező hibák detekciója és javítása.
3. Rezgések megszüntetése.

Természetesen a készülő filmrestauráló rendszer szükséges, hogy tartalmazza a film szerkesztésével kapcsolatos funkciókat is, így a vágásdetekció, optikai mozgásanalízis szerepe a hibajavításon kívül a rendszer automatikus (ill. félautomatikus) működéséhez is nélkülözhetetlen.

Filmírás: A filmírás célja a helyreállított film újbóli, időtálló rögzítése. Mivel a filmes szakma (az orvosokhoz hasonlóan) méltóságteljesen tartózkodik a gyorsan változó és még gyorsabban elavuló digitális technikáktól, valamint különösen érzékeny a digitális veszteséges tömörítésből adódó mellékhatásokra, ezért a rögzítés alapvető módja továbbra is a filmnegatív marad, ami ugyanakkor alapjául szolgál a további másolatok készítésének.

Cikkünkben elsősorban a digitális feldolgozás kérdéseivel foglalkozunk, a digitalizálás és filmírás optikai és mechanikai problémáira nem térünk ki (ezzel az MTA SZTAKIban Manno Sándor és Kas Iván csoportja foglalkozik).

2. Az analóg és digitális restauráció folyamata

A filmrestauráció, filmjavítás, filmmentés és filmarchiválás szorosan összekapcsolódó fogalmak [3]. Így amikor egy olyan rendszert tervezünk, aminek elsődleges feladata régi filmek képi hibáinak javítása és a javított film új hordozón való rögzítése, akkor nem csak a digitális kép (ill. hang) javításának kérdései merülnek fel, hanem a restauráció egyéb a restaurátor által igényelt funkcióit (pl. videó szerkesztés, archív adatok rendszerezése) is érdemes figyelembe venni. Ezeket összevetve a digitális restauráció elsősorban rugalmasságával, nagyobb szabadsági fokával múlja felül a hagyományos javítási-archiválási munkálatokat, ugyanakkor számos olyan új elem jelenik meg a feldolgozási folyamatban, amit csak kellő körültekintéssel lehet szolgálatunkba állítani (pl. a képi adatok analóg-digitális-analóg konverziója).

Az analóg esetben általában a következő lépésekkel jellemezhető egy film javításának fázisai (esetünkben a kiindulási film pozitív, a javított filmet negatívra mentjük):

1. Film azonosítása: A filmmel kapcsolatos információk, technikai adatok rendszerezése: film címe, gyártási éve, gyártó adatai, szereplők, tekercsek száma, hossza, film típusa, a hangsáv típusa, perforáció típusa, stb.

2. Felkészítés a javításra, másolásra: A beavatkozások sorrendiségének meghatározása, a megfelelő javítások meghatározása.
3. Fénymegadás: A másológép megvilágítási paramétereinek meghatározása az egyes filmrészekhez. Ez mintaképek (pl. emberi arc) kivetítésével és a film kivetített jeleneteinek összehasonlításával történik.
4. Másolás: A másolás során történik bizonyos képi hibák javítása és a film tartalmának új hordozóra való átmentése.
5. Előhívás: Ahhoz, hogy a film vizuálisan értelmezhető legyen, pozitív másolat szükséges.
6. A minőség ellenőrzése folyamatosan történik, ami a gyakorlatban azt is jelenti hogy a 3-4-5-ös pontokat gyakran többször is el kell végezni, természetesen nem az egész filmre, csupán annak részeire (ekkor keletkezik a "megítélő" és a "mintatekercs").

Ezzel szemben a digitális feldolgozás főbb lépései a következők:

1. Film azonosítása
2. Digitalizálás
3. Kisfelbontású munkaanyag készítése
4. Felkészítés a javításra, mentésre
5. Javítás-szerkesztés-feldolgozás
6. Filmírás
7. Előhívás
8. Minőség ellenőrzése (5-6-7-es lépések esetleges ismétlése a film egyes részletei esetén).

Mint kiolvasható, a digitális módszer esetén a fénymegadás és javítás alapvetően egy lépcsőben történik, és amennyiben a filmírás és előhívás "átviteli függvénye" jól meghatározott, a minőségi kontrol még a digitális változaton jól elvégezhető, az automatikus és manuális hibajavító eljárásokkal kísérletek végezhetőek. Mint azt ahogy a következő fejezetben láthatjuk, nem csak a feldolgozás menete különbözik jelentősen a digitális javítás során, hanem maguk a vizuális hibák javításának módszerei is jelentősen eltérnek.

3. A filmrestauráció során javítandó főbb képi hibák és hagyományos javítási módjuk

A film felkészítése a javítási munkálatokra alapos körültekintést igényel, hiszen egy-egy meggondolatlan felesleges művelet akár a film maradandó károsodásával is járhat, vagy alaposan megnövelheti a költségeket (emellett egy-egy előhívás meglehetősen időigényes is). Alapvető feladat ilyenkor a film szemrevételezése után a javítási munkálatok pontos meghatározása, a sorrendiség definiálása, az esetleges pótlások elvégzése, a hangsáv javítására tett előkészületek.

Az alábbiakban a leggyakoribb előforduló hibákat soroljuk fel és azok analóg javítási módjait is megemlítjük.

A legtöbb régi film egyik leggyakoribb és legfeltűnőbb hibája a **karcok** jelenléte. Ezeknek számtalan fajtája létezik, mint pl. folyamatos, szakaszos, egy soros, többsoros, éles, dörzs, esőkarc, dobnomás effektus, stb. Javításuk jelenleg leggyakrabban folyadékkapukon keresztül történő másolással történik, amikor is a film anyagához hasonló törésmutatójú folyadékkal töltik fel a kaput. Ezen kívül ismert még fényezési, lakkozási és átmosásos módszer is. Fotográfiai eljárással csak olyan karcok tüntethetők el, amik fizikai sérüléssel jellemezhetők. A film alapanyagán lévő karcok könnyebben javíthatók mint az emulzió jelentkezők.

A **piszok** és **por** fő jellemzője, hogy a kockákon egymástól függetlenül, véletlenszerűen keletkezik sérülés: sötét vagy világos folt látható, atmoszával javítható. Mivel a kemény víz könnyen száradási nyomokat hagyhat a filmen, ezért a víz keménységére ügyelni kell. A száradási nyomok újabb mosással szüntethetők csak meg.

Penész: mivel sokszor a helytelenül tárolt tekercs egy bizonyos részén jelentkezik penészesedés, ezért magán a filmen periodikusan jelenhet meg hiba (pl. minden 20., 30. kockán). A penész erős világosságváltozással vagy színváltozással jár. A penészgomba az emulzió alatt vagy felett lehet jelen, s napjainkban csak a felületi élősködőket lehet mosással eltávolítani. Penészes szennyeződés esetén szükséges a filmen lévő vizuális hatások minimalizálása: folyadékkapuk kezelés (másolás), vákum-kamrás kezelés, egyéb kézzel történő vegyszeres kezelés használatos. Fontos a további fertőzés megakadályozása vegyszeres eljárásokkal. A fakulások - speciális vegyszeres kezelés útján javíthatók - különböző mértékben jelentkeznek a fekete-fehér és a színes képek esetén. Az előbbieket általában kevésbé fakulnak, kivéve ha a fixáló anyagok nincsenek rendesen kimosva, míg a színes képek különböző színösszetevői sok esetben jelentős, különböző veszteségen mennek keresztül. (A színezett, stencilezett ill. kézzel festett filmek jóval stabilabbnak mondhatók.)

A **zsugorodást** víz és plaszticizáló anyagok, oldószerek párolgása okozza, a jelentős zsugorodás komoly akadálya lehet a másolási folyamatnak. A film kezelésekor acetont, glycerolt és vizet segítségével alacsony hőmérsékleten, néhány nap alatt visszanyeri eredeti méretét. Mivel az így kapott méret nem végleges, azonnal meg kell kezdeni a másolást. Egyes másolóberendezések képesek kezelni súlyos zsugorodást is (> 2%). A film kiszáradása során törékennyé és gyűrötté is válhat. Utóbbi jelenségnek az az oka, hogy a film széle és közepe különböző mértékben zsugorodik — ez lehet egy rossz karcolás eltüntetési eredménye is.

A kép **rezgésének**, **remegésének**, a **vibrációnak** az oka többféle lehet. Ilyen pl. a rossz kamera rögzítés, rossz filmtovábbítás a kamerában, ill. a másoló gépben. Ez a fajta hiba nem digitális eszközökkel nem, ill. csak speciális másológépekkel korlátozottan korrigálható.

A globális világosságváltozást vagy színváltozást **villódzásnak** nevezzük. Okai: nem konstans expozíciós idő, film fakulása, interferencia külső elektronikus berendezésekkel.

Hiányzó, vagy erősen sérült kockák is gyakoriak a régi filmeknél. Míg sokszor feliratok zavarják a kép tartalmát jelentősen, máskor **szakadások** vagy

azok nyomai szorulnak javításra. Az erősen megviselt filmeknél a **perforáció hibái** is javítandók, illetve hiánya pótolandó.

Felsorolásunk bár nem teljes — hiszen nem szóltunk a **csillogásról**, a **viaszos, vákumos kezeléstről** [3], stb. — mégis a legfontosabb és leggyakoribb hibákról áttekintő képet ad.

4. A digitális javítások algoritmusai röviden

Az analóg eljárások egyik legnagyobb hátránya az, hogy a legtöbb esetben valamilyen vegyszeres, folyadékos kezelési eljárásnak teszik ki a javítandó filmet, ami — figyelembe véve az eredeti film reprodukálhatatlanságát, egyediségét — igen sok veszéllyel jár. Elsősorban attól kell tartani, hogy a film megsérül, újabb szakadások, vagy szennyneveződések kerülhetnek a filmre (pl. a már feloldott szennyneveződés máshol újból lerakódhat), az oldószerek a film emulzióját is oldhatják, ill. az is kétséges lehet, hogy egyáltalán képesek-e az adott hibák javítására. S mivel sok esetben a kezelés irreverzibilis folyamat, ezért nincs lehetőség korrekcióra. Mindemellett sok vegyszer kezelése nehézkes, nem megfelelő védekezés esetén az egészségre ártalmas.

Ezzel szemben a digitális módszerek rugalmasan alakíthatók, reverzibilisek és a felújítandó film digitalizálásán kívül nincs szükség fizikai mozgatásra, teker-cselésre. Ugyanakkor megjelentek olyan filmbeolvasási módszerek és eszközök, amikor infra megvilágítás segítségével azok a karcok, amelyek a film felületén mélyedésként vannak jelen, detektálhatók, majd az így keletkezett maszk segítségével a karcok által okozott hiba a környezet színértékeinek segítségével digitálisan korrigálható. Ilyen javított eredményt mutat az 1. ábra [11]. (Ez a módszer sajnos nagyon sok esetben nem alkalmazható: fekete-fehér filmek és bizonyos típusú színes filmek esetén sem, valamint olyan esetekben, amikor a karc már fizikailag nincs jelen, csupán az általa okozott információhiány látszik a filmen pl. a karc másolása után.)

Tekintsük át a legalapvetőbb digitális képfeldolgozási módszereket, amiket a fenti hibák legtöbbje esetén sikeresen alkalmazhatunk.

4.1. Piszok, por és egyéb 1 kockán belül jelentkező hibák detekciója és javítása

A javítás során a kép struktúráját és a színét, világosságát egyaránt vissza kell állítani. Lépései a következők:

1. kétirányú mozgáskompenzáció (szubpixeles felbontással)
2. a középső kép és a mozgáskompenzált képek különbségének számítása
3. a hiba intenzitásértékének függvényében küszöbölés
4. a küszöbölt különbségképek függvényében hiba-maszk előállítás
5. hiba maszk dilatációja
6. maszk alapján a mozgásvektorok korrekciója a hibás részek felett átlagolással
7. két szomszédos képből két mozgáskompenzált középső kép számítása majd a két kép átlagának képzése
8. ez alapján a hibás részek korrekciója az eredeti képen.



1. ábra: Felületi karcok szűrése a Nikon Coolscan 4000ED [11] szkennner infra szűrőjének segítségével: karcos kép (felül), karcoktól szűrt kép (alul)

4.2. Rezgések korrekciója

A rezgések eltüntetésének a lényege az, hogy a kép mozgásából a nem jelentős, de domináns mozgást szűrni kell:

1. parametrikus mozgásmodell felállítása
2. mozgásparaméterek becslése többfelbontásos modellben
3. lokális mozgásokhoz maszk készítése, a globális (domináns) mozgás paraméterének finomítása
4. globális mozgásparaméterek integrálása révén a szekvencia vibrálást írjuk le
5. a globális mozgást aluláteresztő szűrővel korrigáljuk
6. a mozgásparamétereket korrigáljuk (a globális mozgást kivonjuk), a képsozrotat az új mozgás alapján "warpoljuk"

4.3. Hibák korrigálása a világosságban, színben

Globális (egész kockára jellemző) hibák javításának lépései:

1. referencia, jó minőségű kockák automatikus vagy manuális kiválasztása (két referencia kocka távolsága maximum 5-20 kocka lehet)
2. bemeneti képsorozat hisztogram kiegyenlítése
3. a referencia kockákhoz inverz hisztogram kiegyenlítés függvény képzése
4. köztes kockákhoz az inverz függvények előállítását interpolációval
5. inverz transzformáció számítása a köztes kockákra.

Lokális, kockán belüli hibák javítása:

1. a kockákat négyzetekre bontjuk kb. 30
2. ahol lokálisan mozgást detektálunk, ott ez a négyzetes felbontás nagyobb az alapfelbontásnál
3. a négyzetekre elvégezzük a (fenti) globális eljárás lépéseit
4. az átlapolásoknál lineáris interpolációt végzünk

Az egyes javítási lépések sorrendjét a következő módon érdemes megválasztani:

1. Globális színkorrekció: lehetővé teszi a pontosabb mozgásanalízist
2. Lokális mozgásbecslés
3. Globális mozgásbecslés
4. Lokális színkorrekció
5. Karcok szűrése
6. Piszok, por szűrése
7. Zaj szűrés
8. Stabilizálás

Valójában a pontos végrehajtási sorrend az adott esetekben a hibáktól függően alakul, lehetőleg törekedni kell arra, hogy legyenek párhuzamosan futtatható eljárások, amik egy többprocesszoros gépen jelentősen megrövidíthetik a feldolgozási időt.

4.4. A filmrestauráció színi problémái

Ha a DIMORF projektnek csak színes mozifilmek másolása, sokszorosítása lenne a feladata, akkor a bevezetőben említett három munkafázis színingermetriai szempontból csupán kétféle redukálódna: a digitalizálásra és a filmírásra. Természetesen a megfelelő minőségű másolat létrehozásához elengedhetetlen ezen két munkafázisban használt készülékek; a szkennerek, az esetleges képmegjelenítők és a filmíró készülék pontos színi kalibrációja.

A filmarchívumokban tárolt filmek általában régi kópiák és ezek restaurációjánál, helyreállításánál figyelembe kell vennünk a filmeknek az elkészültük utáni változásait. A legszembetűnőbb változás a régi filmek fakulása. Ez a folyamat ráadásul többnyire különböző mértékben következik be a színes filmek egyes

rétegeiben. A fakulásból eredő színi torzulások helyreállítása, tehát a színi feldolgozás bonyolult feladat, elfogadható eredmény eléréséhez szükség van a fakulás színíngermetriai modellezésére, melynek eredményeképpen olyan színi transzformációkat kell létrehoznunk, amelyek a lehető legnagyobb mértékben kompenzálni tudják a fakulásból eredő színi torzulásokat. Ezen feladat elvégzéséhez számos film-minta spektrális bemérésére és stabil szín-etalonok használatára is szükség lehet.

A *színi feldolgozás* folyamata általában nem automatizálható az emberi tényező teljes kizárhatóságát jelentő mértékben. A fakulást kompenzáló színi transzformációk eredménye azonban jó alapot nyújthat annak a szakembernek (színmegadó, színesztéta), aki a művészi hatásokat is figyelembe véve jelenetenként vagy meghatározott számú filmkockánként ellenőrzi a folyamatot, és szükség esetén beavatkozik a digitális restaurációba.

4.5. A filmrestauráció minőségmérési problémái

A filmjavítás meglehetősen szubjektív dolog, és valójában a régi filmek már eredetileg sem voltak mindig tökéletesek (optikai torzítás, kép-szinkron, színtechnika, kamera remegése, forgatási sebesség). Emiatt a javításuk eleve nem lehet tökéletes, és sokszor annak az eldöntése is nehéz, hogy valójában mi a hiba? Ugyanakkor a filmanyag fizikai és kémia sérülései, öregedése nem mindig választható el egyértelműen a kép tartalmától: élek, textúrák, apró alakzatok.

A javítást nagymértékben segíti a harmadik dimenzió, az idő. Az egymást követő képek azonos alakzatainak összehasonlításával sok véletlen hibát, sérülést, torzulást lehet korrigálni. Sok esetben ez sem segít, mert a csíkok lehetnek folytonosak, a kémiai folyamatok színfüggőek, az optikai torzítás pedig helyfüggő.

A javítási eljárás minősítésére, elfogadására ezért szubjektív módszereket is kell alkalmazni. Ennek során ki kell választanunk a fontos és kevésbé fontos részeket, a javítható és nem javítható hibákat. A folyamat egésze szempontjából optimalizálni kell a képi hibák javítását, hogy elfogadható sebesség mellett a fontos hibák javuljanak, a kevésbé fontosak pedig többségükben kijavítódjanak, de más helyen kárt ne okozunk.

A fontos helyek kiválasztásánál automatikusan meg kell határozni a néző számára legfontosabb részleteket, a kép alanyát és tárgyát. Az alany az, aki mozog és mozgat, a tárgy pedig az, amire a figyelem irányul. A film nézése közben ezekre figyelünk a legjobban. Ezért a javításban kiemelt helye van a személyek és egyes tárgyak egymáshoz, ill. a nézőhöz való viszonyának a meghatározásának. Ez egy bonyolultabb felismerési eljárás eredménye.

A javítási eljárást úgy készítjük el, hogy a megelőző szubjektív tesztek eredményei alapján automatikusan szabályozzuk a helyreállítás folyamatát. A gépi minősítési módszereket a humán mérések alapján állítjuk be, súlyozzuk a paramétereiket.

A videószekvenciák humán minősítése fizikailag és pszichikailag is sokparaméteres eljárás. Erről szól pl. Az ITU-R BT.500-10-es ajánlás, amiben előírják a képméretet, háttérrel, felbontást, a minták számát, a kísérlet menetét, a pontozási metódust és a statisztikai kiértékelést. Vannak már a természetes látáshoz

jobban illeszkedő eljárások a kép életszerűbb minősítésére, de ezek még nincsenek szabványosítva, ill. még sok elvi kérdés nyitott. Célunk olyan automatikus mérési módszer beillesztése, melyet előzőleg szabványosan végzett humán tesztek alapján állítottunk be.

5. Szerkesztési funkciók

Bár a fejlesztésnek nem célja egy teljes archiváló rendszer elkészítése, mindképpen lehetőséget kell adni a későbbi esetleges bővítésekre. Emellett a használhatóságot alapvetően meghatározza a restaurátor által elérhető szerkesztési funkciók halmaza, amik leginkább egy hagyományos videószerkesztő program funkcióira hasonlítanak:

- Képek, képsorozatok betöltése és kimentése bmp, tiff, jpg, cineon, avi formátumokban.
- Képkockák kivágása, vágólapra helyezése, vágólapról új kockák beszúrása és cseréje.
- Szekvenciák összefűzése.
- A feldolgozandó anyag kijelölése: egész szekvencia vagy egy (kezdő és végpontok által) kijelölt szakasz, kocka és másodperc alapú kijelölések.
- Egy kép vagy képszekvencia részlet megjelenítése kicsinyítve, kockák nagyítása, forgatása, egyéb geometriai transzformációja.
- Vágások detekciója, vágások kézzel való kijelölése, könyvjelzők beszúrása.
- Lejátszás tetszőleges pontról oda-vissza, lassítva, gyorsítva, megállás vágásoknál ill. könyvjelzőnél
- Kockákhoz szöveges információ elmentése és megjelenítése.
- Trükk effektusok, vágások, kép- és videósűrítési funkciók.
- Előnézet (gauss piramison).

A képen elvégzendő feladatokat csoportokba érdemes szervezni, ahol egy csoporton belüli eljárások sorrendisége meghatározott. Szükséges lehet a különböző képfeldolgozási analízis funkciók eredményeinek köztes mentése is a későbbi szűrések gyorsításához (mivel pl. a mozgásbecslést több algoritmus is használhatja).

Nagy segítséget jelenthet a filmes szakemberek számára a film automatikus vágásdetekciója és indexelése. A digitális videó (ill. film) indexelés nem más, mint egy digitalizált film különféle "tulajdonságainak" megkeresése és e tulajdonságok összegyűjtése egy adatbázisba. Egy film tulajdonsága lehet globális, ami a teljes filmre vonatkozik (pl.: film címe, rendezője, témája, színvilága, stb.) és lehet lokális, ami a film egyes jeleneteire vagy csak bizonyos rövid részeire vonatkozik (pl. mi történik egy adott jelenetben, kit ábrázol egy filmkocka, hol játszódik a jelenet, stb.). A tartalom alapú videó indexelésnél a videó szekvenciákat felosztjuk hasonló képkockák csoportjaira, ezeket hívjuk jeleneteknek. A jelenetek határainak detekciója valójában vágásdetekciónak felel meg, az indexelés pedig a csoportokat reprezentáló filmkockákon történik. Az egyes jeleneteket a hosszúságuk és szerkezetük figyelembevételével egy vagy több reprezentatív képkockával írhatjuk

le. A visszakeresés az adatbázisból a reprezentatív képkocka és a lekérdezésnél valamilyen módon megadott referenciakép hasonlóságának vizsgálata alapján történik. Az indexelés és visszakeresés lehet szöveg, szín, alakzat vagy textúra alapú, de sok hasznos információt lehet gyűjteni a mozgásinformáció vizsgálatával is.

Az indexálás előfeldolgozó lépése a film jelenetekre való bontása. A jelenetekre bontás technikája azon alapszik, hogy egy jeleneten belül az egymást követő képkockák képelemei csak kicsit változnak, vágásnál viszont nagy a változás. Ennek a változásnak mérésére sokféle módszert javasol a szakirodalom [1], ahol a vizsgálódások tárgyai a következők lehetnek:

- Az egymást követő képkockák közötti képpont értékek változása
- Az egymást követő képkockák szín hisztogramjainak távolsága.
- Twin-comparison technika [4] (elsősorban átmeneti vágásoknál, pl fade-in, fade-out, dissolve, wipe, stb. effektusok)
- MPEG tömörített filmek esetén az un. DC-képek. (ezek 64x kisebbek mint az eredeti kép, de tartalmazzák a kép átlagos intenzitás tulajdonságait.)
- Mozgásinformáció [2]
- Stb.

6. Jelentősebb filmhelyreállítással foglalkozó projektek, szoftverek

A következőkben néhány jelentősebb filmfelújító projektet ill. szoftvert mutatunk be röviden.

Diamant(1993-2001) [7]

Az ausztriai Diamant szoftver (ill. projekt) több korábbi projekt keretében készített szoftver egyesített végterméke, amely komplett, átfogó szolgáltatásokat nyújt a restaurációs folyamatban. Ebben a sorban a Joanneum Research és a HS-ART cég **Limelight** nevű szoftvere volt az első, amelyben alapvető restaurációs algoritmusokat valósítottak meg. A Limelight egy munkaállomáson közel 10 perc alatt dolgozott fel egy nagyfelbontású képet. Ilyen hosszú feldolgozási idő nagyon megnövelte a film helyreállításához szükséges költségeket, így a feldolgozási idő leszorítása lett a fő cél. Ezért fejlesztették ki a **FRAME** [9] szoftvert amely a HPCN (High-Performance Computing and Networking) technológiát alkalmazva egy képkocka átlagos feldolgozási idejét kb. 72 mp-re (videó felbontásban 15 mp-re) csökkentette. Ezt egészítette ki a Nephew project, amely a hálózati erőforráselosztást valósította meg Windows NT környezetben. A Nephew project célul tűzte ki, hogy több PC-n Windows-NT környezetben párhuzamosan futtathassuk eljárásainkat és így jobb ár/teljesítmény viszonyt érjünk el.

A Vinci Restor Product Group Rs2-es modulja [6]

A Vinci Restor Product Group Rs2-es modulja egy komplett filmhelyreállító szoftver. A rendszer SDTV és HDTV felbontásban dolgozik, képes operátor beavatkozása nélkül is detektálni és javítani bizonyos képi hibákat. A rendszer egy SGI munkaállomásból és a hozzá kapcsolódó da Vinci STOR merevlemezről és egy videó felvevő egységből áll.

A Kodak CINEON rendszere (1997) [5]

A Kodak Cineon FX rendszereinek egyes moduljai a vizuális effektusok létrehozása mellett film helyreállításra is alkalmasak. A rendszer SGI munkaállomáson futtatható. A Kodak saját fejlesztésű hardver eszközeivel végzi a film digitalizálását (Cineon Genesis Plus Film Scanner), filmre (Cineon Lighting II.) és videóra írását. A rendszer automatikusan detektálja és kijavítja a karcok, szennyeződések jelentős részét (Dust-busting) és sok más jellegű lokális hibát is. Egyes moduljai eltüntetik a kép egészének vagy annak részeinek villódzását (Deflicker), csökkentik a film anyagára jellemző szemcse zajt (Grain Management), javítják a színhelyességet, vagy módosíthatják egy jelenet hosszát képkockák elvételével vagy hozzáadásával (Cinespeed).

DIGITAL OZU Project – Tokyo University [8]

A Tokiói Egyetemen a DIGITAL OZU project keretében újjátették fel Yasujiro Ozu "Tokyo Story" című filmét. A munka során egyaránt javították a kép és a hang minőségét. A film szkennelését 6000X4000-es felbontásban végezték, a jellemző szkennelési idő 30 mp. volt kockánként. Az előfeldolgozás alatt a szkennelés hibáit javították valamint a filmet jelenetekre bontották. A digitális helyreállítás a következő hibákat érintette: karcok, foltok, színhelyesség, kontraszt, pozíció, zaj, hanghibák.

SARSA (Systeme Automatique de Restauration de Séquences Animées – Automatic Restoration System for Animated Sequences) [7]

Etienne Decenciere Ferrandiere PhD téziseit mozgó kép helyreállítás témakörében írta. Az ott megvalósított eredményeit foglalta össze egy rendszerben. A rendszer kezeli a kép villódzását, remegését, eltünteti a függőleges karcokat és a képen megjelenő kisebb hibákat. A program interpoláció segítségével pótolja a hiányzó képkockákat.

A fentiekén kívül más érdekes megoldásokkal is találkozhatunk, azonban a terjedelemre való tekintettel azoktól most eltekintünk.

7. Összegzés

A filmrestauráció elkerülhetetlenül nem csak technikai, fizikai beavatkozást jelent, hanem esztétikait, esetleg tartalmat is, hiszen a film valójában a vásznon fog megjelenedni. Egy filmnek a visszaállítás szempontjából elméletileg különböző verziói lehetnek:

- a visszaállító kezében lévő
- az első nézők által látott verzió
- későbbi nézők által látott verziók
- ahogy a film készítője elképzelte (elképzelhette)
- ahogy egy modern közönség látná (akarná látni)
- mai művész által átdolgozott verzió
- kereskedelmi szempontból eladható, forgalmazható verzió.

Tehát a restaurálás etikai kérdései fontos momentumok a restaurátor, a filológus, az esztéta számára. Az általunk fejlesztés alatt álló rendszer reméljük

megkönnyíti majd ezt a feladatot. A hagyományos analóg technikával szemben a következő előnyeit és hátrányait tudjuk felsorolni:

1. táblázat: Nagyfelbontású képeken történő képfeldolgozási műveletek időigénye Intel PIII Dual 1GHz processzoros gépen 1GB memóriával az Intel Image Processing Library [10] optimalizált függvényeinek segítségével

Művelet	Mért eredmények átlaga (mp.)		
Színmélység (bit)	24	24	8
Felbontás (pixel)	6000*4000	3000*2000	3000*2000
Méret (MByte)	68.66	17.16	5.72
Betöltés merevlemezről	1	0.4	0.1
Konvolúció (3*3)	0.9	0.22	0.064
Szeparált konvolúció (3*3)	1.95	0.49	0.13
Gauss átlagolás (3*3)	0.35	0.085	0.03
Gauss átlagolás (5*5)	1.75	0.31	0.1
Median (3*3)	0.35	0.088	0.03
Median (5*5)	1.62	0.41	0.14
Blur(3*3)- átlagolás egyenlő súllyal	1.55	0.39	0.18
Blur(5*5)- átlagolás egyenlő súllyal	1.73	0.44	0.196
DCT(8*8)	3.2	0.82	0.35
DCT(32*32)	5.4	1.36	0.42
FFT(32*32)	6	0.98	0.29
Hisztogram kiszámítása	0.45	0.11	0.05
Hisztogram széthúzás	0.6	0.14	0.05
Hisztogram kiegyenlítés	0.6	0.14	0.05

Előnyök:

- Nagyobb szabadsági fok a szerkesztésben (nemlineáris szerkesztés, képi indexálás).
- Több javítási funkció, a javítás eredményei az írás előtt ellenőrizhetők, a beállítások korrigálhatók.
- Nincs szükséges folyadékos, vegyszeres módszerekre, a film további rongálódásának valószínűsége csökken.
- Digitális tárolás tömörítve, különböző formátumokban, publikációs lehetőség.
- Többféle digitális verzió készítése relatív alacsony költségen.
- Integrált funkciók: a filmrestaurációs és archiválási folyamatok egy közös rendszerbe integrálhatók, rendszerezhetők, megkönnyítve a filmes szakemberek (archivátor és filológus) munkáját.
- Reprodukálhatóság: a restaurációs folyamat pontos rögzítésre kerül, ezért később, esetleg kisebb módosítások után megismételhető.

Hátrányok, problémák, feloldásuk:

- Lassúság: a nagyfelbontású digitális képek mozgatása és feldolgozása nagyságrendileg tovább tart a szokásos videó-rátánál. Remélhetőleg ezt a problémát a technika fejlődése orvosolja. Lásd a becsült számítási idők táblázatát (1. táblázat).
- Költséges: az eszköz beruházáson kívül jelentős a kezelő személyzet igénybevétele, de mivel a munka egyes fázisai párhuzamosíthatók, így a költségek jelentősen csökkenthetők.
- A hosszú távú digitális tárolás csak veszteségesen lehetséges a nagy adatmennyiség miatt, ennek elfogadható mértéke nehezen meghatározható, bár a szakirodalomban jelentős az emberi látórendszert vizsgáló és modellező cikkeknek a száma.
- Rövid elavulási idő: míg a filmarchívumok gyakorlatában nem ritka a matuzsálem korú berendezések használata, addig a számítástechnikában néhány év alatt az eszközök nagy része elavul, technikák merülnek feledésbe, inkompatibilitási problémák lépnek fel.

Köszönetnyilvánítás

Ezúton is köszönetünket szeretnénk kifejezni Beke Évának (Magyar Nemzeti Filmarchívum), Kas Ivánnak, Derényi Károlynak, Kovács Györgynek (MTA SZTAKI) és az NKFP-nek segítségükért, közreműködésükért, támogatásukért.

Irodalom

1. Boreczky, S., Rowe, L.A.: Comparison of Video Shot Boundary Detection Techniques. Proceedings of the Conference on Storage and Retrieval for Image and Video Databases IV, San Jose, CA, SPIE Proceedings Series, Vol. 2670. (1996) 170-179
2. Jain, R.C.: Segmentation of Frame Sequence Obtained by a Moving Observer. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 6. No. 5, (1984) 624-629
3. Paul Read (Editor), Mark-Paul Meyer: Restoration of Motion Picture Film. Butterworth-Heinemann, (2000)
4. Zhang, H., Kankanhalli, A. and Somlinal, S.W.: Automatic Partitioning of Full-Motion Video. ACM/Springer Multimedia Systems, Vol. 1. No. 1, (1993) 10-28
5. Cineon - Digital Film System: <http://www.cineon.com/>
6. da Vinci - RS2: <http://www.davsys.com/rs2.htm/>
7. Diamant: <http://diamant.joanneum.ac.at/>
8. Digital Ozu projekt:
http://www.um.u-tokyo.ac.jp/dm2k-umdb/publish_db/books/dm2000/english/01/01-15.html
9. FRAME (Austria):
<http://www.vpcp.univie.ac.at/activities/projects/FRAME/>
10. Intel képfeldolgozó könyvtára:
<http://developer.intel.com/software/products/perflib/ipl/>
11. Nikon Coolscan 4000ED: <http://www.nikontechusa.com/LS4000FAQ.html>

Supervised training based hand gesture recognition system

Attila Licsár¹ and Tamás Szirányi²

¹ University of Veszprém, Department of Image Processing and Neurocomputing,
H-8200 Veszprém, Egyetem u. 10, Hungary

licsara@www.knt.vein.hu

² Analogical & Neural Computing Laboratory, Computer &
Automation Research Institute,

Hungarian Academy of Sciences, H-1111 Budapest, Kende u. 13-17, Hungary

sziranyi@sztaki.hu

Abstract. We have developed a hand gesture recognition system, based on shape analysis of static gestures, for Human Computer Interaction purposes. Our appearance-based recognition uses modified Fourier descriptors for the classification of hand shapes. As always found in literature such recognition systems consist of two phases: training and recognition. In our new practical approach, following the chosen appearance-based model, training and recognition is done in an interactive supervised way: the adaptation for untrained gestures is also solved by hand signals. Our experimental results with three different users are reported. In this paper, besides describing the recognition itself, we demonstrate our interactive training method in a practical application.¹

1 Introduction

In the information society the communication between user and computer has become a very active research area. In this paper we demonstrate a practical method to improve the performance of our gesture recognition system. In the spatial gesture model there are two main approaches: the 3D hand model-based and appearance-based methods. We have chosen appearance-based methods in our system because it is simple and efficient for two-dimensional desktop applications.

Appearance-based systems have two categories: motion-based and posture-based recognition. Posture-based recognition not only handles location of the hand but also recognizes the shape features of the hand. Motion-based systems use dynamic gestures, while posture-based systems use static hand gestures. Our system recognizes static hand postures, because dynamic parameters of the hand e.g. position or movements are used for controlling purpose in the virtual environment.

¹ Hand Recognition Demo can be downloaded from
<http://www.knt.vein.hu/staff/licsara/>

There exist a lot of methods in appearance-based recognition for recognizing static gestures in our system. The majority of methods is using parameters derived from image. In such cases, the model parameters are derived from the description of the shape. Systems can use one or more camera pictures [1]. They include: edge-based contours [2], edges, image moments [3][4], image eigenvectors [5], or geometric moment description of hand shapes. Some other techniques use second order moments (like Zernike methods [6]), which invariant to rotation of the shape. Other methods use orientation histograms [7], which are invariant to lighting conditions and represent summarized information of small patch orientations over the whole image. Geometric moment description is not invariant to rotation, and the invariance of other moment-based methods is restricted. We need a method for contour classification, where the parameters are invariant to translation, rotation, and scaling. The disadvantage of invariant moments is its high computational cost, because features are computed using the entire region. Boundary-based methods, such as Fourier descriptor [8] use only contour-points. However, the Fourier descriptor is sensitive to the starting point of the shape boundary.

2 Overview of our gesture recognition system

In our system we adopt a modified Fourier descriptor method [9] (MFD), which is already used in the field of the character recognition. We apply this method for the gesture-based man-machine interface. The example-based system involves two phases: training and running phase. In the training phase, the user shows to the system one or more examples of hand gestures. The system stores the Fourier coefficients of the hand shape, and in the running phase the computer compares the current hand shape with each of stored shapes by coefficients. The best match gesture is selected by the nearest-neighbor method with the metric of MFD. Our goal is that under the recognition phases the system will be able to correct faulty detected gestures and it can interactively modify and train hand gestures with the help of the user's feedback. With this interaction, the recognition efficiency will be grown.

The organization of this paper is as follows. First, we describe features of MFD method, and then analyze the parameters of the gesture recognition system. Next, we will discuss the interactive learning method and examine the recognition rate of the system with and without interactive learning. Finally, we present our practical application with this interactive training method.

3 Gesture recognition by MFD

The first implementations of Fourier descriptor had some weaknesses, for example, invariance to starting point of the shape contour. Extending this traditional method to the modified Fourier descriptor, it is invariant to transition, reflection, scaling of shapes, as well as the starting point used in defining boundary sequence. The MFD is robust against noise, when shape boundaries often contain

local irregularities due to image noise. Another advantage of the MFD is that feature vectors should be computed efficiently. It has a new distance measure for describing and comparing closed curves.

Some applications, like methods [10] with Fourier descriptor, use other classification algorithms (e.g. neural network), but this simple metric, based on features of Fourier coefficients, is quite fast and reliable: cc. 80% recognition rate for 7-8 gestures. For the calculation of the MFD we need a complex sequence from the x and y coordinates of the n -th boundary points. The MFD is defined as the Discrete Fourier Transform (DFT) of the previous complex sequence. The magnitude and phase of Fourier descriptor coefficients of the compared boundaries are related to each other. From the features of the Fourier transform and orientations of the major axes of the two shapes we can calculate two sequences [9]. It is easy to see that these sequences will be constant if the two compared gestures belong to the same class. The distance measure for magnitude and phase angle is defined as the standard deviation of the previous two calculated sequences. So, the classification method of MFD is fast because it computes a simple standard deviation. The system's processing speed is ~ 12 samples/sec on a Celeron 733MHz computer.

The Fourier descriptors are calculated from the boundary of the palm. The system uses restrictive background to localize the hand efficiently. Hand silhouettes or contours can be easily extracted from the hand image. Since forearm features do not contain important information, the perfect and consequent segmentation of arm and forearm is important. The problem of automatic segmentation has solved by other systems [5], which use the direction of the arm for automatic forearm segmentation. From the image moments [11], we can get the global direction and position of the hand. Since the method is invariant to orientation, position and size of the palm, we may use these parameters to control or manipulate virtual objects. Using these parameters the system can get feedback from the user about the recognition.

We have tested the system with 9 gesture classes (Figure 1). The starting set of training was very small, usually only one, but the continuous supervised/unsupervised training adaptively changed the class-parameters. Testing process was done by a set of 140 samples. The efficiency of the recognition methods has been tested to find the best feature-detection. Methods for selecting the most appropriate Fourier descriptors are as follows:

- (Method 1.) The first 3 coefficients are used in the comparison
- (Method 2.) Coefficients which are greater than a threshold (here 11.0) in case of at least one class
- (Method 3.) Starting from the lowest spatial frequencies where each coefficient is greater than the threshold.

We demonstrate the above methods in Table 1. For each method the distances among the classes are counted. In case of Method 2, the distances are greater than in the other 2 cases. Starting with a one-class/one-sample training, a limited number of classes can be appropriately recognized.

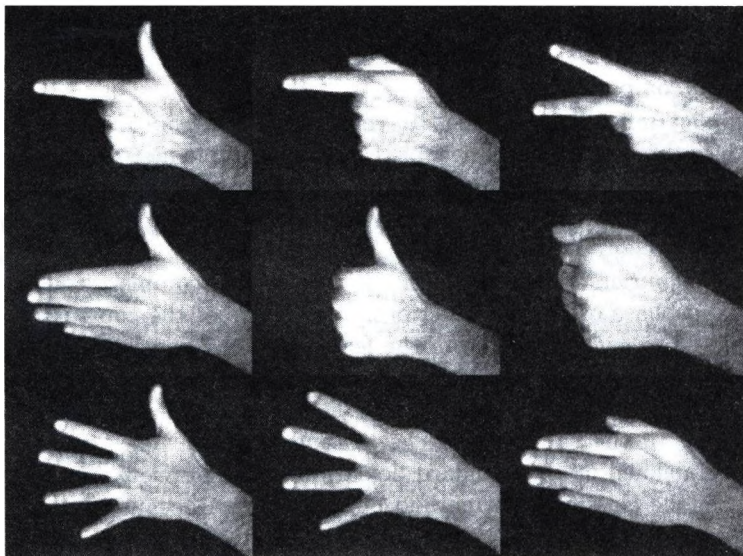


Fig. 1. Some of our hand gestures used in our system

Table 1. Comparing efficiency of selecting methods

Used method	Recognition rate	Average distance
Method 1.	97%	0.73
Method 2.	98%	1.01
Method 3.	97%	0.74

4 Interactive learning system

The conventional learning and recognition phases of the system will be extended by a new interactive training algorithm, which will improve the efficiency of the recognition methods. In the recognition phase the system can correct faulty detected gestures and interactively can modify and teach hand gestures with the user feedback. In this additional phase the user can modify the recognition strategy by modifying his gestures' clustering among classes. The user's feedback signal (crucial shapes to be trained) is a rapid gesture moving or shaking, because under a normal interaction the user doesn't apply rapid moving. After the rapid moving or shaking, the system modifies the decision and it chooses the next most probable gesture. This gesture parameter will be trained by the actual gesture parameter grabbed from the camera. This way of interaction is quite natural considering the human behavior. Under the recognition phase the system refreshes parameters if the decision is right, so it is able to adapt to the gestures of the user. With this interactive training algorithm our system is able to adapt

to gestures of other users. In Figure 2 we see the algorithm of our training method.

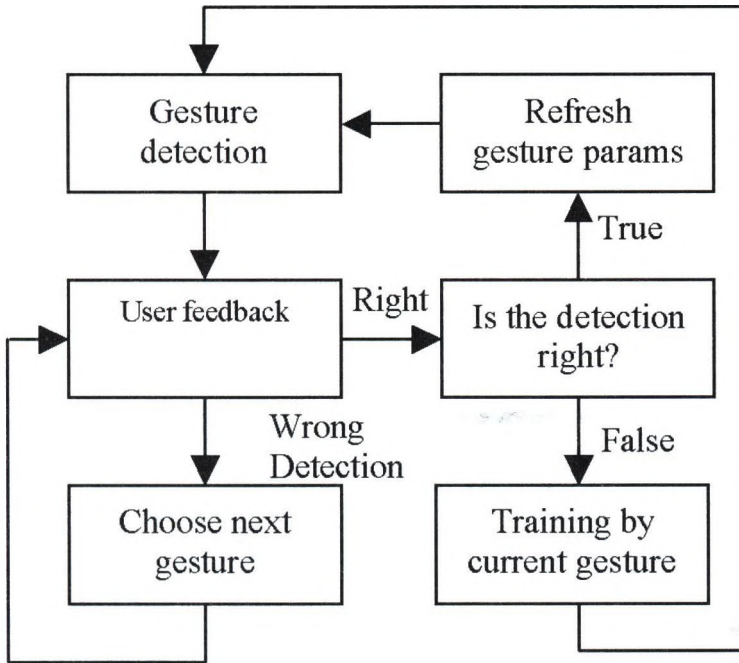


Fig. 2. Interactive training algorithm

In the natural languages, when people show rapid hand shaking, it usually means denial signal. Our method applies the same movement to get user's feedback of negation. The user's feedback can also be hand shaking or rotating. These signals can be extracted from the variations of the palm position and orientation. When the variations of the position or the rotation are greater than a threshold value, the system recognizes a negation signal.

If the detection is true then parameters of the detected gesture continuously will be refreshed by the parameters of the current gesture with a predefined small weight. Continuously refreshing the static gesture, the system is able to adapt to small changes of gestures. For example, when the user is tired and cannot show standard gestures, the system may learn it. If the detection is wrong, the user indicates it by rapid hand shaking. In this situation the system chooses the next most-probable gesture-class and this gesture will appear on the screen. The user can see the result and can generate feedback again until the result is accepted. In this case the system uses supervised training and corrects the false recognition.

$$Z = \frac{w_1 \cdot X + w_2 \cdot Y}{w_1 + w_2} \quad (1)$$

If in the above equation Y is the parameter of the trained gesture and X is the parameter of the trainer's gesture. We can set the training method by w_1 and w_2 . In Figure 3 we can see, the distance of the two gestures in time, with several weights, $w_1 = 30$ and $w_2 = 1$ and so on. In this training the distance decreases slowly since the gradient of the curve is small, while the harder training at the 1. and 3. sec results in a deeper descent of the curve. In unsupervised training the **problem** is that two or more classes may get to close to each other, resulting in **an unstable recognition**. For this reason, the program continuously detects **the distances** among the classes, and it makes alert when the trained pattern is **inconsistent**, or it is close to be unstable.

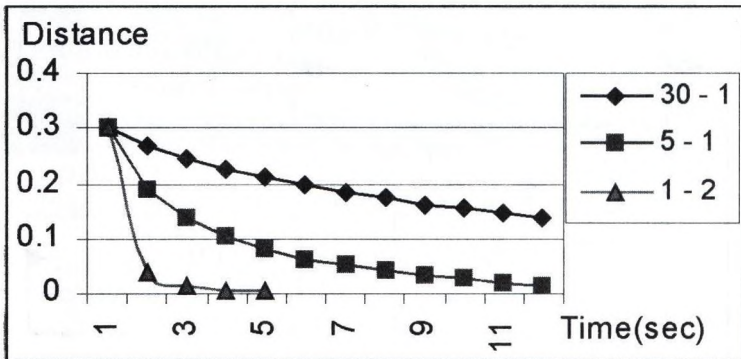


Fig. 3. Training with several weighting value

In Table 2 we summarize the efficiency of our method. First, the trainer tests the method, correcting a little bit to get 99% rate. Then new users are working at 76%-86% rates, correcting the machine-parameters in the automatic interaction, resulting in 92%-95% recognition rates.

Table 2. Testing with a new user on the old training (9 gesture classes, 140 test images)

Users	Recognition rates	
	Unsupervised learning	Supervised learning
Trainer user	98%	99%
New user #1	86%	95%
New user #2	76%	92%
New user #3	82%	94%

5 Practical application

Our gesture-recognizer and supervised interactive trainer system is applied in a practical application. Restoration of old films is an expensive process, and it needs several ways of human-interaction (lighting, contour, color, defects, noises, synchrony). The hand-based interaction can be used in the pre-processing phase: registration of places for hard error corrections or enhancement, notching the key-frames or reference frames. This method supports both the frame-by-frame and frame-sequence operations. Using gestures we are able to track the position of the hand, drawing a continuous line on the screen, browsing among image sequences (start, stop, back), signing defected or key frames (reference frames, cuts).

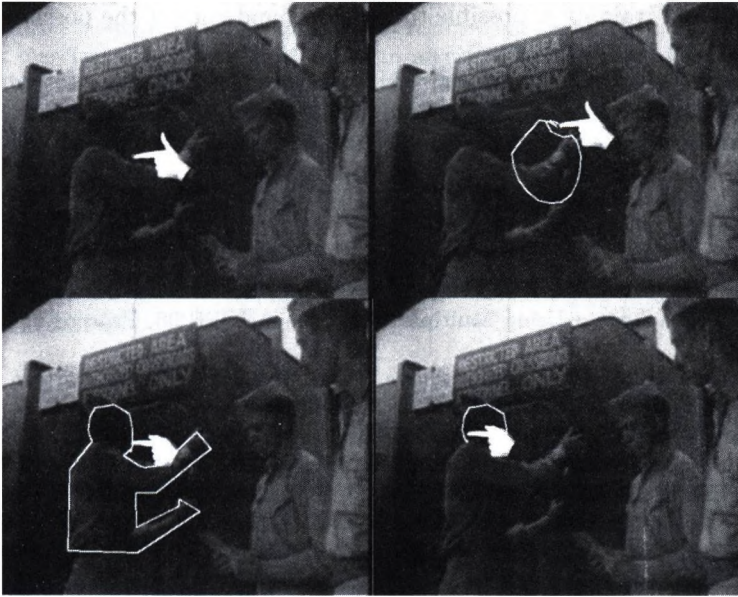


Fig. 4. Signing the feature-areas in sample images

In our software the reference point of the hand gives a virtual cursor (Figure 4), while the recognized gesture generates the command, since the contours of hand gesture are rotation and shift invariant. In the interaction process we see the pictogram of the recognized hand gesture around the cursor, resulting in a continuous feedback about the position and the proposed command. In case of mistake or misunderstanding the rotation/shake of the hand cancels the command and the next probable command (proposed hand gesture) is processed. In this way the method continuously refreshes the gesture-parameters in the unsupervised (accepted cases and fine parameter modifications of slightly different class-features) or supervised (shaking and modifying) training. When there is a

danger of mixing (or getting too close to) the different gesture-classes during the unsupervised training due to a forgetful gesture-series, there is a change in the shape of the plotted cursor to alarm the user for the mistake of posture.

6 Conclusion

The above work has shown that

- Interactive training is user-friendly and user-independent;
- Gestures classes can be trained from a limited number of training sets (it also works with solo training set);
- Unsupervised training can be continuously run to follow the slight changes in the gesture styles;
- Supervised training is possible to recognize and correct the possible overlap among the different classes;
- We have tested the supervised training system with three users and found that the performance of recognition has increased significantly as experimental data shows above.

References

1. Utsumi, A., Miyasato, T., Kishino, F., Nakatsu, R.: Hand Gesture Hand Gesture Recognition System Using Multiple Cameras. In ICPR'96, Proceeding of ICPR, (1996)
2. Cho K., Dunn, S. M.: Learning shape classes. In IEEE Tran. on Pattern Analysis and Machine Intelligence, vol. 16, (1994), 882-888
3. Starner, T., Pentland, A.: Visual Recognition of American Sign Language Using Hidden Markov Models. In Proc. Int'l Workshop on Automatic Face and Gesture Recognition, Zurich, (1995)
4. Alt, F.L.: Digital Pattern Recognition by Moments. In JACM, 9, 2, April (1962), 240-258
5. Imagawa, K., Taniguchi, R., Arita, D., Matsuo, H., Lu, S., Igi, S.: Appearance-based Recognition of Hand Shapes for Sign Language in Low Resolution Image. In Proc. of 4th Asian Conference on Computer Vision, (2000), 943-948
6. Schlenzig, J., Hunter, E., Jain, R.: Vision-Based Hand Gesture Interpretation Using Recursive Estimation. In Proc. 28th Asilomar Conf. Signals, Systems, and Computer, (1994)
7. Freeman, W., Roth, M.: Orientation histograms for hand gesture recognition. In International Workshop on Automatic Face and Gesture Recognition, (1995)
8. Zahn, C.T., Roskies, R.Z.: Fourier descriptors for plane closed curves. In IEEE Trans. on Computers C21, (1972), 269-281
9. Rui, Y., She, A., Huang, T.S.: A Modified Fourier Descriptor for Shape Matching in MARS. Image Databases and Multimedia Search, (1998) 165-180
10. Kohler: Vision Based Hand Gesture Recognition Systems. <http://ls7-www.cs.uni-ortmund.de/research/gesture/vbgr-table.html>
11. Horn, B.K.P.: Robot Vision. MIT Press, Cambridge, Massachusetts, (1986)

Rekonstrukciós algoritmusok hv -konvex 4- és 8-összefüggő diszkrét halmazokra

Balogh Emese¹, Kuba Attila¹, Alberto Del Lungo²

¹ Alkalmazott Informatikai Tanszék, Szegedi Tudományegyetem
{bmse, kuba}@inf.u-szeged.hu

² Department of Mathematics, University of Siena
dellungo@unisi.it

Absztrakt. Három rekonstrukciós algoritmust mutatunk be és hasonlítunk össze, melyek diszkrét halmazokat rekonstruálnak azok két, egymásra merőleges vetületeiből. Mindhárom algoritmusnak két verziója van, egyik a hv -konvex 4-összefüggő, a másik pedig a hv -konvex 8-összefüggő halmazok rekonstrukciójára alkalmas. Mindkét osztályban összehasonlítjuk az algoritmusokat a futási idő és a memóriaszükséglet szempontjából. Az első két algoritmus a szakirodalomban ismert, a harmadik pedig egy új eljárás, mely az előző két módszer kombinációja, örököelve azok jó tulajdonságait.

1. Bevezetés

A *diszkrét tomográfia* egyik legfontosabb kérdése a diszkrét halmazok rekonstrukciója kis számú, gyakran csupán két egymásra merőleges vetületből [9, 10]. Mivel ugyan azokhoz a vetületekhez általában sok különböző megoldás is tartozik, keressük azokat az osztályokat, amelyekben a rekonstrukciós algoritmus polinomiális idejű. Tehát a rekonstruálandó halmazoknak amellet, hogy meg kell feleljenek az adott vetületeknek, bizonyos tulajdonságokkal is rendelkezniük kell. Ezek a tulajdonságok általában a konvexitás és az összefüggőség. Ebben a cikkben a hv -konvex 4- és 8-összefüggő diszkrét halmazok rekonstruálásával foglalkozunk két egymásra merőleges vetületből.

Kuba még 1984-ben publikált egy algoritmust hv -konvex diszkrét halmazok rekonstruálására [12]. Később kiderült, hogy a feladat ebben az osztályban NP-teljes [16]. Barcucii módosította ezt az algoritmust [3, 4], majd egy továbbfejlesztett változatával [5] már hv -konvex 8-összefüggő halmazokat is rekonstruálni lehetett. Ezen - a későbbiekben A algoritmusként említett algoritmus - komplexitása $O(mn \cdot \log(mn) \cdot \min\{m^2, n^2\})$. 1999-ben Chrobak és Dürr jelentetett meg egy jobb komplexitású ($O(mn \cdot \min\{m^2, n^2\})$) algoritmust [6], majd Kuba megmutatta, hogy azon kicsit módosítva, hv -konvex 8-összefüggő halmazokat is lehet vele rekonstruálni. A későbbiekben erre az algoritmusra mint B algoritmus fogunk hivatkozni.

E két algoritmus implementálása és tesztelése során kiderült, hogy bár az A algoritmus komplexitása rosszabb, mint a B algoritmus komplexitása, általában

sokkal gyorsabban megadja a megoldásokat. Hogy erre magyarázatot találjunk, szisztematikus tesztelésnek vetettük alá az algoritmusokat. Ehhez implementálnunk kellett két módszert [11, 2] amelyek egyenletes eloszlással generálnak $h\nu$ -konvex 4-összefüggő illetve $h\nu$ -konvex 8-összefüggő halmazokat, majd rekonstruálva ezeket a különböző algoritmusokkal, mértük a futási időket és a memória igényeket. Az irodalomból ismert két módszer mellett implementáltunk és teszteltünk egy harmadikat is, mely az előző kettő kombinációja, és mely öröklí azok jó tulajdonságait.

A cikk felépítése a következő. A második fejezetben ismertetjük azokat a meghatározásokat és tulajdonságokat, melyek szükségesek a harmadik fejezetben bemutatott algoritmusokhoz. Végül az utolsó fejezet az algoritmusok összehasonlításával, tesztelésével foglalkozik, figyelmet szentelve a különböző futási időkre és memória szükségletekre valamint magyarázatot ad az algoritmusok viselkedésére.

2. Meghatározások

Jelölje \mathbb{Z}^2 a kétdimenziós diszkrét hálót, az (i, j) pont pedig annak egy elemét. Egy diszkrét halmaz \mathbb{Z}^2 egy véges részhalmaza. Legyen $F \neq \emptyset$ egy diszkrét halmaz. Ekkor létezik a $T = \{1, \dots, m\} \times \{1, \dots, n\}$ diszkrét halmaz, amelyik a legkisebb olyan diszkrét téglalap, amelyik tartalmazza F -et, ahol m és n pozitív egészek.

Az F diszkrét halmaz reprezentálható egy $(f_{ij})_{m \times n}$, $f_{ij} \in \{0, 1\}$ bináris mátrixal úgy, hogy

$$f_{ij} = \begin{cases} 1, & \text{ha } (i, j) \in F, \\ 0, & \text{különben.} \end{cases}$$

Legyen N a természetes számok halmaza. Bármely F diszkrét halmaz vektoreit a \mathcal{H} és \mathcal{V} műveletekkel határozhatjuk meg, ahol $\mathcal{H} : \mathcal{F} \rightarrow \mathbb{N}^m$, $\mathcal{H}(F) = H = (h_1, \dots, h_m)$, ahol

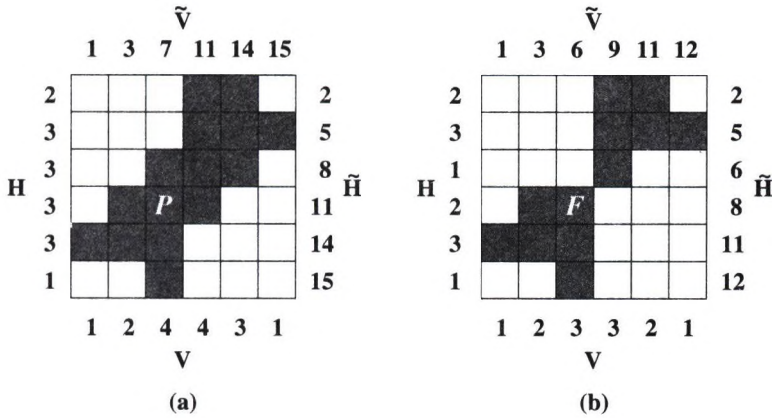
$$h_i = \sum_{j=1}^n f_{ij}, \quad i = 1, \dots, m,$$

és $\mathcal{V} : \mathcal{F} \rightarrow \mathbb{N}^n$, $\mathcal{V}(F) = V = (v_1, \dots, v_n)$, ahol

$$v_j = \sum_{i=1}^m f_{ij}, \quad j = 1, \dots, n.$$

A H és V vektorok az F halmaz sor- és oszlopösszegei (lásd az 1. ábrát).

Azt mondjuk, hogy az F és F' diszkrét halmazok tomográfiailag ekvivalensek, ha $\mathcal{H}(F) = \mathcal{H}(F')$ és $\mathcal{V}(F) = \mathcal{V}(F')$.



1. ábra: (a) P egy hv -convex polyomino. (b) F egy hv -konvex 8-összefüggő de nem 4-összefüggő diszkrét halmaz. A halmazok elemeit sötét négyzetekkel jelöltük. H és V a sor és oszlopösszeg vektorok, \tilde{H} és \tilde{V} pedig a halmpozott összegek, melyek első elemei, $\tilde{h}_0 = 0$ és $\tilde{v}_0 = 0$, nem szerepelnek az ábrán.

A $\tilde{H} = (\tilde{h}_0, \tilde{h}_1, \dots, \tilde{h}_m)$ és $\tilde{V} = (\tilde{v}_0, \tilde{v}_1, \dots, \tilde{v}_n)$ vektorok az F halmaz *halmozott összegvektorai*, ahol $\tilde{h}_0 = 0$, $\tilde{h}_i = \tilde{h}_{i-1} + h_i$, $i = 1, \dots, m$, és $\tilde{v}_0 = 0$, $\tilde{v}_j = \tilde{v}_{j-1} + v_j$, $j = 1, \dots, n$ (lásd az 1. ábrát).

A (H, V) vektorpáros *kompatibilis*, ha létezik m, n , és $A \in \mathbb{N}$ úgy, hogy

- (i) $H \in \mathbb{N}^m$ és $V \in \mathbb{N}^n$;
- (ii) $h_i \leq n$ minden $1 \leq i \leq m$ és $v_j \leq m$ minden $1 \leq j \leq n$;
- (iii) $\sum_{i=1}^m h_i = \sum_{j=1}^n v_j = A$, azaz a két vektor elemeinek összege A .

Az $(i, j) \in \mathbb{Z}^2$ pont *4-szomszédai* az $(i - 1, j)$, $(i, j - 1)$, $(i, j + 1)$, $(i + 1, j)$ pontok és az (i, j) pont. Az $(i, j) \in \mathbb{Z}^2$ pont *8-szomszédai* a 4-szomszédos pontok és az $(i - 1, j - 1)$, $(i - 1, j + 1)$, $(i + 1, j - 1)$ és az $(i + 1, j + 1)$ pontok. Az $(i_0, j_0), \dots, (i_k, j_k)$, $k \in \mathbb{N}$ pontok sorozata egy *4-út/8-út* az (i_0, j_0) és (i_k, j_k) pontok között, ha bármely (i_l, j_l) pont 4-/8-szomszédja az (i_{l-1}, j_{l-1}) pontnak, minden $l = 1, \dots, k$ esetén.

Az F diszkrét halmaz *4-összefüggő* illetve *8-összefüggő* ha bármely két pontját kiválasztva létezik közöttük egy 4-út illetve 8-út F -en belül, amely összeköti őket. A 4-összefüggő halmazokat szokás *poliominóknak* is nevezni [8].

Az F diszkrét halmaz *h-konvex*, ha sorai 4-összefüggők és *v-konvex*, ha oszlopai 4-összefüggők. F *hv-konvex*, ha h-konvex és v-konvex.

Jelölje \mathcal{F}_4 a hv-konvex 4-összefüggő diszkrét halmazok (hv-konvex poliominók) osztályát, \mathcal{F}_8 pedig a hv-konvex 8-összefüggő diszkrét halmazok osztályát.

3. Rekonstrukciós algoritmusok

Tekintsük a hv-konvex poliominók (\mathcal{F}_4) és a hv-konvex 8-összefüggő diszkrét halmazok (\mathcal{F}_8) osztályán értelmezett rekonstrukciós feladatokat.

REKONSTRUKCIÓ (\mathcal{F}_4).Adott: $m, n \in \mathbb{N}$, $H \in \mathbb{N}^m$ és $V \in \mathbb{N}^n$.Keressük: azon $F \in \mathcal{F}_4$ diszkrét halmazokat, melyekre $\mathcal{H}(F) = H$ és $\mathcal{V}(F) = V$.REKONSTRUKCIÓ (\mathcal{F}_8).Adott: $m, n \in \mathbb{N}$, $H \in \mathbb{N}^m$ és $V \in \mathbb{N}^n$.Keressük: azon $F \in \mathcal{F}_8$ diszkrét halmazokat, melyekre $\mathcal{H}(F) = H$ és $\mathcal{V}(F) = V$.

A szakirodalom több módszert is ismer ezen rekonstrukciós feladatok megoldására [3, 4, 6, 7, 13]. Ebben a cikkben három rekonstrukciós algoritmust mutatunk be és hasonlítunk össze. Ezek közül kettő ismert módszer (nevezzük őket A és B algoritmusoknak.), a harmadik pedig egy új algoritmus, melyre a későbbiekben mint C algoritmus fogunk hivatkozni.

Mindkét algoritmusnak két verziója van, az A_4 , B_4 és C_4 algoritmusok hv -konvex 4-összefüggő halmazokat, az A_8 , B_8 és C_8 algoritmusok pedig hv -konvex 8-összefüggő halmazokat rekonstruálnak.

3.1. Az A algoritmus

Ezt az algoritmust először Barucci közölte hv -konvex poliominókra [3, 4], később továbbfejlesztették és a legújabb verziója [5] már a hv -konvex 8-összefüggő diszkrét halmazokat is rekonstruálni tudja. Az algoritmusok leírásánál az [5] jelöléseit használjuk. Az algoritmus két diszkrét halmazokból álló sorozattal közelíti meg a megoldást. Az egyik sorozat monoton növekvő, elemei a *mag* halmazok, a másik sorozat monoton csökkenő, ennek elemei a *burok* halmazok. A kezdeti magot a vetületekből számolt gerinc elemei alkotják, a kezdeti burkot pedig a diszkrét halmazt magába foglaló legkisebb diszkrét téglalap elemei alkotják. A **mag** tartalmazza az aktuális mag elemeit, a **burok** pedig az aktuális burok elemeit. Az algoritmusban előforduló eljárások (a lábak lehetséges pozíciójának kiszámolása, gerinc számolás, mag-burok eljárás) megtalálható [5]-ben.

 A algoritmus**Input:** Egy (H, V) kompatibilis vektorpár, ahol $H \in \mathbb{N}^m$ és $V \in \mathbb{N}^n$.**Output:** Minden olyan hv -konvex 4-/8-összefüggő diszkrét halmaz, melyeknek vetületei a H és V vektorok.

1. A halmazok összegek (\tilde{H}, \tilde{V}) kiszámolása.
2. A lábak lehetséges pozícióinak kiszámolása.
3. Minden lehetséges lábkonfigurációra
 - 3.1. A gerinc kiszámolása.
 - 3.2. A mag-burok eljárás alkalmazása.
 - 3.3. ha burok \setminus mag $\neq \emptyset$

akkor

ha minden sorban van legalább egy eleme a mag-nak
akkor a 2-SAT algoritmus alkalmazása.

különben {

2 szemközti láb kiválasztása.

A gerinc kiszámolása.

A mag-burok eljárás alkalmazása.

ha $\text{burok} \setminus \text{mag} \neq \emptyset$

akkor a 2-SAT algoritmus alkalmazása.

}

Az A algoritmus komplexitása $O(mn \cdot \log(mn) \cdot \min\{m^2, n^2\})$ [5].

3.2. A B algoritmus

Ezt az algoritmust Chrobak és Dürr publikálta 1999-ben [6], és lényege, hogy a teljes rekonstrukciós feladatot átírja egy 2-SAT feladattá, amely már megoldható polinomiális idő alatt [1]. Később Kuba [13] megmutatta, hogy elhagyva a 4-összefüggőséget biztosító klózokat a 2-SAT kifejezésben, az algoritmus a $h\nu$ -konvex 8-összefüggő halmazokat is rekonstruálja.

B algoritmus

Input: Egy (H, V) kompatibilis vektorpár, ahol $H \in \mathbb{N}^n$ és $V \in \mathbb{N}^m$.

Output: Minden olyan $h\nu$ -konvex 4-/8-összefüggő diszkrét halmaz, melyeknek vetületei a H és V vektorok.

minden $k, l = 1, \dots, m$

{

Írjuk fel azon klózokat, melyek a (k, l) horgonypontokkal rendelkező diszkrét halmazokat definiálják.

Értékeljük ki a felírt 2-SAT kifejezést.

}

Ennek az algoritmusnak a komplexitása $O(mn \cdot \min\{m^2, n^2\})$ [6].

3.3. A C algoritmus

Most egy, az előző két algoritmusra épülő új algoritmust mutatunk be. Az első lépések megegyeznek az A algoritmus első lépéseivel, de később, amikor az A algoritmus két szemközti lábat választ, ez az új algoritmus a B algoritmust hívja meg 2 szemközti láb minden lehetséges pozíciója esetén. Ez azt jelenti, hogy megtartva az A algoritmus gyorsaságát, a C algoritmus komplexitása ugyan olyan jó, mint a B algoritmus komplexitása.

C algoritmus

Input: Egy (H, V) kompatibilis vektorpár, ahol $H \in \mathbb{N}^n$ és $V \in \mathbb{N}^m$.

Output: Minden olyan hv -konvex 4-/8-összefüggő diszkrét halmaz, melyeknek vetületei a H és V vektorok.

1. A halmazok összegek (\tilde{H}, \tilde{V}) kiszámolása.
2. A lábak lehetséges pozícióinak kiszámolása.
3. **Minden** lehetséges lábkonfigurációra
 - 3.1. A gerinc kiszámolása.
 - 3.2. A mag-burok eljárás alkalmazása.
 - 3.3. **ha** burok \setminus mag $\neq \emptyset$

akkor

ha minden sorban van legalább egy eleme a mag-nak
akkor a 2-SAT algoritmus alkalmazása.

különben { **minden** lehetséges lábpozícióra

Legyen k és l a két láb egy-egy sora/oszlopa.

Írjuk fel azon klózokat, melyek a (k, l) horgony-
 pontokkal rendelkező diszkrét halmazokat
 definiálják.

Értékeljük ki a felírt 2-SAT kifejezést.

}

Hasonlóan a B algoritmushoz, ennek az algoritmusnak a komplexitása is $O(mn \cdot \min\{m^2, n^2\})$.

4. Algoritmusok összehasonlítása

Az előző fejezetben bemutatott algoritmusok összehasonlítása érdekében implementáltunk egy-egy algoritmust adott méretű hv -konvex poliominók és hv -konvex 8-összefüggő halmazok egyenletes generálására. A generált halmazokat ezután rekonstruáltuk mindhárom algoritmussal, mérve az első megoldás megtalálásához szükséges időt és minden megoldás megtalálásához szükséges időt, melyeket az 1-es és 2-es táblázatokban foglaltunk össze.

Méret $m \times n$	Algoritmus	Első megoldás (sec)	Minden megoldás (sec)
20 × 20	A_4	0.003	0.003
	B_4	0.068	0.377
	C_4	0.003	0.006
40 × 40	A_4	0.007	0.008
	B_4	0.673	6.045
	C_4	0.007	0.008
60 × 60	A_4	0.015	0.018
	B_4	2.296	28.842
	C_4	0.015	0.018
80 × 80	A_4	0.027	0.031
	B_4	5.152	92.953
	C_4	0.027	0.031
100 × 100	A_4	0.042	0.048
	B_4	7.528	190.810
	C_4	0.042	0.048

1. táblázat: Az A_4 , B_4 és C_4 algoritmusok futási idő eredményei a rekonstruálandó diszkrét halmaz méretének függvényében. A teszthez minden méretben 1000 darab $h\nu$ -konvex polyominót generáltunk egyenletes eloszlással.

Méret $m \times n$	Algoritmus	Első megoldás (sec)	Minden megoldás (sec)
20 × 20	A_8	0.006	0.006
	B_8	0.063	0.359
	C_8	0.007	0.009
40 × 40	A_8	0.009	0.010
	B_8	0.620	5.952
	C_8	0.009	0.010
60 × 60	A_8	0.018	0.021
	B_8	2.328	29.274
	C_8	0.022	0.064
80 × 80	A_8	0.029	0.033
	B_8	5.790	90.548
	C_8	0.029	0.033
100 × 100	A_8	0.044	0.050
	B_8	10.233	198.854
	C_8	0.044	0.050

2. táblázat: Az A_8 , B_8 és C_8 algoritmusok futási idő eredményei a rekonstruálandó diszkrét halmaz méretének függvényében. A teszthez minden méretben 1000 darab $h\nu$ -konvex 8-összefüggő diszkrét halmazt generáltunk egyenletes eloszlással.

Ennél a kísérletnél minden méretből 1000 halmazt generáltunk és rekonstruáltunk. A tesztekéből kiderült, hogy az A és C algoritmusok sokkal gyorsabban megtalálják az első és az összes megoldást, mint a B algoritmus. A futási eredmények a hv -konvex 4- illetve 8-összefüggő halmazok esetén nagyon hasonló, hiszen az algoritmusok verziói között csak nagyon kis eltérés van.

A tesztek azt mutatják, hogy az átlagos futási idők különböznek az algoritmusok komplexitása alapján vártaktól. Hogy megtaláljuk ennek a magyarázatát, tanulmányoztuk, hogy az egyes algoritmusok általában melyik lépés után találják meg az összes megoldást.

Méret $m \times n$	2SAT nélkül	2SAT-tal	Két szemközti lábát választva
20×20	46485 (92.970%)	3345 (6.680%)	170 (0.340%)
40×40	47827 (95.654%)	2147 (4.294%)	26 (0.053%)
60×60	48179 (96.358%)	1814 (3.628%)	7 (0.014%)
80×80	48419 (96.838%)	1572 (3.144%)	9 (0.018%)
100×100	48447 (96.894%)	1551 (3.102%)	2 (0.004%)

3. táblázat: 50000 egyenletes eloszlással generált hv -konvex poliominó rekonstruálásakor azon esetek száma és százalékos aránya, melyeknél 2-SAT nélkül, a 2-SAT algoritmussal, illetve 2 szemközti láb választásával kaptuk meg az összes megoldást.

Méret $m \times n$	2SAT nélkül	2SAT-tal	Két szemközti lábát választva
20×20	46565 (93.130%)	3266 (6.532%)	169 (0.338%)
40×40	47809 (95.618%)	2168 (4.336%)	23 (0.046%)
60×60	48172 (96.344%)	1815 (3.630%)	13 (0.026%)
80×80	48403 (96.806%)	1586 (3.172%)	11 (0.022%)
100×100	48496 (96.992%)	1503 (3.006%)	1 (0.002%)

4. táblázat: 50000 egyenletes eloszlással generált hv -konvex 8-összefüggő halmaz rekonstruálásakor azon esetek száma és százalékos aránya, melyeknél 2-SAT nélkül, a 2-SAT algoritmussal, illetve 2 szemközti láb választásával kaptuk meg az összes megoldást.

Kísérleteink azt mutatták, hogy az A és C algoritmusok az esetek több, mint 90 %-ában rekonstruálnak minden megoldást a gerinc számolása és az azt követő mag-burok eljárás alkalmazása után, és az esetek több, mint 99 %-ában anélkül, hogy két szemközti lábát választanának. E két lépés komplexitása $O(mn \cdot \log(mn))$. Ezen eredményeket méreteként 50000 hv -konvex poliominó

és hv -konvex 8-összefüggő diszkrét halmaz rekonstruálása során kaptuk, lásd a 3.-as és 4-es táblázatokat.

A B algoritmusban a $\min\{m^2, n^2\}$ tag a horgonypontok lehetséges számából adódik, míg az A algoritmusban a két szemközti láb választásából. Amíg a B algoritmus mindig választ horgonypontokat, az A és C algoritmusok kevesebb, mint az esetek 1 %-ában választanak két szemközti lábat illetve hívják meg a B algoritmusból kölcsönzött 2-SAT eljárást, lásd a 3-as és 4-es táblázatokat.

Egy másik szempont, ami szerint az algoritmusok összehasonlíthatók, az a memóriaszükségletük. Minden módszer esetén a ki és bemenő adatok tárolására szükséges egy $m \times n$ méretű tömb (mn bájt) és két m és két n méretű tömb ($4 \cdot (m + n)$ bájt) a vetületek tárolására. Ezen kívül az A és C algoritmusoknak szükségük van még 2 darab m méretű és két darab n éretű tömbre ($4 \cdot mn$ bájt) a halmozott összegek tárolására, és 4 darab m méretű és 4 darab n méretű tömb ($4 \cdot 4 \cdot (m + n)$) a mag és burok tárolására.

Méret $m \times n$	Algoritmus	Futási idő (sec)
20×20	A_4	0.002
	B_4	0.025
	C_4	0.001
40×40	A_4	0.002
	B_4	0.128
	C_4	0.001
60×60	A_4	0.003
	B_4	0.314
	C_4	0.003
80×80	A_4	0.004
	B_4	0.573
	C_4	0.004
100×100	A_4	0.006
	B_4	0.908
	C_4	0.006

5. táblázat: Az A_4 , B_4 és C_4 algoritmusok futási eredményei a rekonstruálandó diszkrét halmaz méretének függvényében. A teszthez minden méretben 50000 darab hv -konvex polyominót generáltunk egyenletes eloszlással.

Méret $m \times n$	Algoritmus	Futási idő (sec)
20×20	A_8	0.001
	B_8	0.026
	C_8	0.001
40×40	A_8	0.001
	B_8	0.128
	C_8	0.001
60×60	A_8	0.003
	B_8	0.314
	C_8	0.003
80×80	A_8	0.004
	B_8	0.573
	C_8	0.004
100×100	A_8	0.006
	B_8	0.908
	C_8	0.006

6. táblázat: Az A_8 , B_8 és C_8 algoritmusok futási eredményei a rekonstruálandó diszkrét halmaz méretének függvényében. A teszthez minden méretben 50000 darab $h\nu$ -konvex polyominót generáltunk egyenletes eloszlással.

A B és C algoritmusok az utolsó lépésben átírják a rekonstrukciós feladatot egy 2-SAT feladattá. Ehhez $20 \cdot mn + 2 \cdot \min\{m, n\} + 8$ klóz szükséges. A 2-SAT feladat megoldásához egy irányított gráfot határozzunk meg. Ehhez $2 \cdot 4 \cdot mn$ csúcspont (4 sarok régió van, ezeket egyenként egy mn méretű tömb, és azon belül a tömb minden eleméhez 2 csúcspont tartozik) és $2 \cdot (20 \cdot mn \cdot 2 \cdot \min\{m, n\} + 8)$ él tartozik. A gráfot a memóriában fix hosszúságon ábrázoljuk. Minden csúcspontból 9 él indul ki, ezek típusa long pointer. Tehát a teljes gráf $2 \cdot 4 \cdot 9 \cdot 4 \cdot mn = 288 \cdot mn$ bájt helyet foglal a memóriában. Tehát átlagosan az A és C algoritmusoknak sokkal kisebb a memóriaigényük.

Egy következő tesztben azt vizsgáltuk, hogy milyen gyorsan állapítják meg az algoritmusok, hogy nincs megoldás az adott osztályon belül. Ehhez először meg kell vizsgálni, hogy a vetületek teljesítik-e Ryser szükséges és elégséges feltételét [15], hogy eldönthessük, hogy van-e olyan megoldás, amelyik kielégíti a vetületeket. A következő lépés azt megvizsgálni, hogy a megoldások között van-e olyan, amelyik kielégíti a kért tulajdonságokat is. Ehhez méretenként 1000 darab általános (tehát nem $h\nu$ -konvex 4-/8-összefüggő) diszkrét halmazt generáltunk egyenletes eloszlással, majd megpróbáltuk őket a három algoritmussal rekonstruálni. A mért átlagos futási eredményeket az 5-ös és 6-os táblázatok tartalmazták.

Méret $m \times n$	Első Megoldás (sec)	Futási idő (sec)
500×500	2.415	2.606
1000×1000	10.658	11.327
1500×1500	25.412	26.701
2000×2000	51.613	54.184

7. táblázat: Az A_4 és C_4 algoritmusok átlagos futási idői a rekonstruálandó diszkrét halmaz méretének függvényében. Minden méretben 100 darab hv -konvex diszkrét halmazt generáltunk egyenletes eloszlással.

Méret $m \times n$	Első Megoldás (sec)	Futási idő (sec)
500×500	2.479	2.638
1000×1000	10.791	11.473
1500×1500	25.972	27.267
2000×2000	53.956	56.244

8. táblázat: Az A_4 és C_4 algoritmusok átlagos futási idői a rekonstruálandó diszkrét halmaz méretének függvényében. Minden méretben 100 darab hv -konvex diszkrét halmazt generáltunk egyenletes eloszlással.

Az A és C algoritmusok erősségét bizonyítandó elvégeztünk még egy kísérletet, növelve a rekonstruálandó halmazok méretét egészen 2000×2000 -ig. A B algoritmust ki kellett hagynunk ebből a tesztből az óriási memóriaigénye miatt. Az átlagos futási eredményeket az 5-ös és 6-os táblázatok tartalmazzák. Ebben a kísérletben minden megoldást rekonstruálni tudtunk a két szemközti láb választása nélkül, illetve a C algoritmus esetén a B algoritmusból ismert 2-SAT eljárás meghívása nélkül.

Az algoritmusok tesztelésére írt C++ programokat egy 533 MHz-es Intel Pentium III processzorral és 192 Mb-al rendelkező PC-t használtunk. A gépen egy Red Hat 6.2-es (Zoot), 2.2.12-5.0-s kernelű Linux futott.

A tesztelés eredményeként kijelenthetjük, hogy az A és C algoritmusok átlagosan jóval gyorsabbak és jóval kevesebb memóriát igényelnek, mint a B algoritmus. Komplexitás szempontjából pedig a B és C algoritmusok jobbák, mint az A algoritmus.

Köszönetnyilvánítás

Ezt a munkát az OTKA T032241 sz. kutatási pályázata támogatta.

Irodalom

1. B. Aspvall, M.F. Plass, and R.E. Tarjan, A linear-time algorithm for testing the truth of certain quantified boolean formulas. *Information Processing Letters* **8(3)** (1979) 121–123.
2. E. Balogh, Generation and reconstruction of *hv*-convex 8-connected discrete sets, *Acta Cybernetica*, to appear.
3. E. Barcucci, A. Del Lungo, M. Nivat, R. Pinzani, Reconstructing convex polyominoes from horizontal and vertical projections, *Theor. Comput. Sci.* **155** (1996) 321–347.
4. E. Barcucci, A. Del Lungo, M. Nivat, R. Pinzani, Medians of polyominoes: A property for the reconstruction, *Int. J. Imaging Systems and Techn.* **9** (1998) 69–77.
5. S. Brunetti, A. Del Lungo, F. Del Ristoro, A. Kuba, M. Nivat, Reconstruction of 8- and 4-connected convex discrete sets from row and column projections, submitted for publication.
6. M. Chrobak, C. Dürr, Reconstructing *hv*-convex polyominoes from orthogonal projections, *Information Processing Letters* **69** (1999) 283–289.
7. A. Del Lungo, M. Nivat, Reconstruction of connected sets from two projections, In [10] (1999) 3–34.
8. S.W. Golomb. *Polyominoes: Puzzles, Patterns, Problems, and Packings*, second edition, Princeton Academic Press, (1996).
9. G.T. Herman, A. Kuba (Eds.), Discrete Tomography, Special Issue. *Int. J. Imaging Systems and Techn.* **9** (1998) No. 2/3.
10. G.T. Herman, A. Kuba (Eds.), *Discrete Tomography: Foundations, Algorithms and Applications* (Birkhäuser, Boston, 1999).
11. W. Hochstättler, M. Loebel, C. Moll, Generating convex polyominoes at random, *Discrete Mathematics* **153** (1996) 165–176.
12. A. Kuba, The reconstruction of two-directionally connected binary patterns from their two orthogonal projections, *Comp. Vision, Graphics, and Image Proc.* **27** (1984) 249–265.
13. A. Kuba, Reconstruction in different classes of 2D discrete sets, In *Lecture Notes on Computer Sciences* **1568** (1999) 153–163.
14. A. Kuba, E. Balogh, Reconstruction of convex 2D discrete sets in polynomial time, *Theor. Comput. Sci.*, to appear.
15. H.J. Ryser, Combinatorial properties of matrices of zeros and ones, *Canad. J. Math.* **9** (1957) 371–377.
16. G.W. Woeginger, The reconstruction of polyominoes from their orthogonal projections, *Information Processing Letters* **77** (2001) 225–229.

Kétirányú konvex bináris képek rekonstrukciója abszorpciós vetületekből

Nagy Antal, Kuba Attila

Szegedi Tudományegyetem, Alkalmazott Informatikai Tanszék

nagya@inf.u-szeged.hu

kuba@inf.u-szeged.hu

Bináris képek (mátrixok) rekonstrukciója sor- és oszlopösszegeikből a diszkrét tomográfia egyik klasszikus problémája. Ismeretes, hogy általában nagyon sok megoldása lehet egy ilyen problémának, ezért a gyakorlat számára a feladatot inkább úgy tűzik ki, hogy bizonyos előírt tulajdonságú bináris mátrixokat rekonstruálnak két vetületből. A tulajdonság kiválasztása nagyon lényeges: lehetőleg olyan célszerű választani, amely a gyakorlati alkalmazáshoz elég általános, de a rekonstrukció még polinomiális időben elvégezhető. Ebből a szempontból az egyik legérdekesebb osztály a vízszintesen és függőlegesen (röviden kétirányú) konvex mátrixok osztálya, amelyben ugyan ez a probléma még NP-teljes, de 4- vagy 8-irányú összefüggőséget hozzávéve a kétirányú konvexitáshoz, már polinomiális időben megoldható problémával állunk szemben.

Előadásunkban azt a kérdést vizsgáljuk meg, hogy mennyiben változik ugyanennek a feladatnak a nehézsége, ha a sor- és oszlopösszegek helyett bizonyos abszorpciós vetületek állnak rendelkezésre. Ez a feladat abból a modellből származik, amikor a rekonstruálandó homogén anyagú tárgy valamilyen ismert abszorpciós közegben van, és ezért a vetületek az abszorpciónak megfelelően torzulnak, ún. abszorpciós vetületek keletkeznek. Megmutatjuk, hogy a kétirányú konvex bináris mátrixok rekonstrukciója két abszorpciós vetületből polinomiális időben elvégezhető (szemben az abszorpciómentes vetületek esetével), ismertetjük a rekonstrukcióra használható algoritmust és néhány képet is bemutatunk az így rekonstruáltak közül.

Dinamikus SPECT vizsgálatok faktorképeinek rekonstrukciója diszkrét tomográfiai eljárással

Dudásné Nagy Marianna, Kuba Attila

Alkalmazott Informatikai Tanszék, Szegedi Tudományegyetem
{marcsi,kuba}@inf.u-szeged.hu

A nukleáris medicina a beteg szervezetében lejátszódó folyamatokat teszi láthatóvá γ -sugárzó radioizotópok segítségével. A felvételek a radioizotóp eloszlásának detektálására alkalmas eszközökkel (gamma kamerával) készülnek. A dinamikus SPECT egy olyan vizsgáló eljárás, mely 4D képeket, azaz 3D képek időbeli sorozatát állítja elő oldalirányú gamma-kamerás felvételekből. Egy-egy irányból egy olyan képsorozatot kapunk, amely az adott irányból nézve adja meg a folyamat időbeli változását. Ha egy ilyen oldalirányú képsorozatban faktoranalízist végzünk, akkor a kapott faktorképek úgy tekinthetők, mint a 3D faktorok megfelelő irányú vetületei. Mivel a 3D faktorok homogén aktivitás-eloszlású objektumok, ezért speciális, diszkrét tomográfiai eljárással rekonstruálhatók. A diszkrét tomográfia olyan függvények rekonstrukcióját végzi néhány vetületből, amelyek csak egy előre meghatározott diszkrét halmazból vehetnek föl diszkrét értékeket.

Matematikailag a megoldandó probléma az, hogy 3D homogén tárgyakat rekonstruáljunk 4 vetületből. A szimulált lehülés elvén alapuló iteratív rekonstrukciót valósítottuk meg. Szimulációs kísérleteket végeztünk: szoftverfantom segítségével generáltunk vetületi adatokat, amelyek tartalmaztak abszorpció és leképezési hibák okozta hatásokat, valamint zajt. A rekonstrukciót minden egyes 3D faktorra metszetenként hajtottuk végre. Vizsgáltuk a különféle paraméterek hatását a rekonstrukció minőségére. Szimulált vesefantomon mutatjuk be az eljárás eredményét.

Körmetszeteket tartalmazó tárgyak rekonstrukciója néhány vetületből

Kiss Zoltán¹, Kuba Attila², Rodek Lajos³

¹ SZTE, Alkalmazott Informatikai Tanszék
h734673@inf.u-szeged.hu

² kuba@inf.u-szeged.hu

³ h734726@inf.u-szeged.hu

Nemroncsoló anyagvizsgálat végzésekor jelentkezett a következő probléma: Rekonstruáljunk egy olyan 3D homogén anyagú tárgyat vetületi képeiből (pl. röntgen vagy neutronos felvételekből), amelyről tudjuk, hogy keresztmetszetei egymást nem átfedő, ismert számú körlapból és körgyűrűből állnak (pl. csapágyak, csövek). A megoldásnál vegyük figyelembe azt, hogy a leképezés – és így a vetületi képek – torzítottak, zajjal terheltek, továbbá szeretnénk a rekonstrukciót minél kevesebb számú vetületből elvégezni.

Megoldásként egy olyan eljárást választottunk, amely a rekonstrukciós problémát optimalizálásként fogalmazza meg. Keressük köröknek azt a konfigurációját, amelynek a vetületei a lehető legkisebb mértékben térnek el a megadott vetületi adatoktól. Az eltérés mértékeként a négyzetes eltérést választottuk. A keresési tér pontjai pedig megfelelnek a konfigurációban szereplő körök sugarai és középpontjai lehetséges értékeinek. A szimulált lehülés elvét használó algoritmust implementáltuk ennek az optimalizálási problémának a megoldására. Beszámolunk a program működésével kapcsolatos tapasztalatainkról és az elért eredményekről.

Diszkrét tomográfiai képek VRML alapú megjelenítése

Ruskó László, Kuba Attila, Balogh Emese

Alkalmazott Informatikai Tanszék, Szegedi Tudományegyetem
{rusko,kuba,bmse}@inf.u-szeged.hu

Diszkrét tomográfia egy képalkotási technika, amely olyan 2D vagy 3D objektumok keresztmetszeti képeit rekonstruálja vetületi képekből, amely objektumok pontjai csak néhány, ismert diszkrét értéket vehetnek fel. Például homogén tárgyak rekonstrukciójánál csak kétféle érték lehetséges: a tárgy anyagára és a környezetére jellemző konstansok. Egy olyan képfeldolgozó rendszer kifejlesztését tűztük ki célul, amely a legfontosabb diszkrét tomográfiai algoritmusok kipróbálását és összehasonlítását teszi lehetővé. Ezt a rendszert a web-en keresztül is elérhetővé kívánjuk tenni. Főbb részei: a tesztadatok generálása, a rekonstrukciós programok futtatása és az eredmények megjelenítése.

A megjelenítés az adatok VRML formátumú dokumentumra történő konvertálását jelenti, amelyet akár az Internet-böngészők meg tudnak jeleníteni. A rekonstrukció eredményeként kapott objektum általában egy 2D vagy 3D bináris kép, mely a vetületekkel együtt egy szöveges állományban található. Ezeket a fájlokat konvertálja a program VRML dokumentummá. Az eszköz interaktív, azaz a felhasználó megadhatja, hogy az adatokat milyen formában kívánja megjeleníteni. Lehetőség van objektumok egyszerű, két különböző objektum egyidejű, valamint objektum és vetületei egyidejű megjelenítésére. A kép- ill. vetületi-pontok stílusa, valamint további információk (pl. vetítési sugarak, egyedileg kijelölt pontok) is megadhatók.

An Algorithm to Register Sets of 3D Points to 3D Lines for Using Arbitrary Frame Devices in Image Guided Percutaneous Therapies

Attila Tanács^{1,2}, Gábor Fichtinger², and Attila Kuba¹

¹ University of Szeged, Department of Applied Informatics, Szeged, Hungary
{tanacs,kuba}@inf.u-szeged.hu

² Center for Computer-Integrated Surgical Systems and Technology,
Johns Hopkins University, Baltimore, MD 21218, USA
{tanacs,gabor}@cs.jhu.edu

Abstract. We present a method for simultaneous registration of sets of 3D points to 3D lines. The algorithm requires a priori knowledge of correspondence between lines and point sets. This registration method is ideally suitable for stereotactic point set registration in image-guided percutaneous procedures and radiotherapies.

1 Introduction

Percutaneous surgical procedures require precise needle placement inside the patient's body. These techniques, include tumor biopsy, radiofrequency ablation, nerve blocking, can greatly benefit from image guidance [1, 2]. Using pre- or intra-operative images of the patient, entry and target points can be selected, while asserting that needle trajectories avoid critical structures, such as blood vessels and bones. In this process, a critical step is to find the correspondence between the physical space and the image space. This step is referred to as registration.

In brain surgery stereotactic head-frames have been widely used for this purpose. The basic idea is to rigidly attach a coordinate frame to the patient's skull. To this frame a localization module can be rigidly attached. The localization modules include several fiducial rods made from materials well visible in the CT or MR volume images. The geometric description and the spatial relation among the fiducial rods is a priori known. When an image plane intersects with fiducial rods, the rods produce distinctive marks in the image (Fig. 2). This is the basis of CRW, BRW [3], Leksell [4], Kelly, and Leibinger-Fischer head-frames (Figs. 1A,1B, 1C,1D), representing most head-frames used in clinical practice.

For procedures that require access to the abdomen or spine, full-body stereotactic frames were developed, however these were designed for fractionated stereotactic radiotherapy and never took ground in interventional procedures. Susil et al. proposed a downsized version of the BRW head-frame [5], which is successfully used at Johns Hopkins University in research of robotic assisted percutaneous therapy of prostate, kidney, liver, and spine (Fig. 1E).

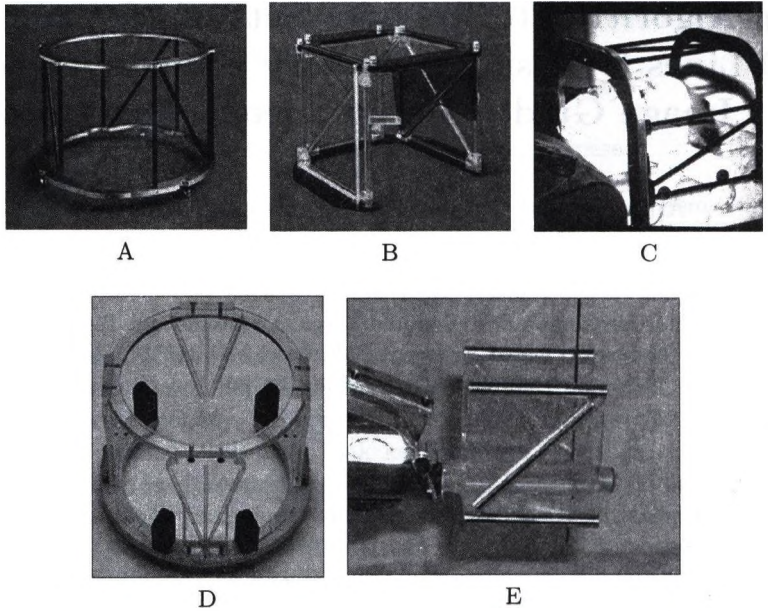


Fig. 1. Different stereotactic frames. From left to right, top to bottom: CRW CT, CRW MR, Kelly CT, Leibinger-Fischer CT, Susil CT.

Considering visualization and trajectory planning, different frames are conceptually identical, only the number and the spatial positions of the rods used for registration are different. Our goal was to develop a unified framework independent from the imaging modality and frame type, where the frame specific details can be read from configuration files. This way new frame types can be easily added.

We developed an iterative algorithm to register sets of 3D points to 3D lines. Point samples of rods are segmented manually or semi-automatically. A single slice containing the cross section of most of the rods (at least 3 or 4, depending on the configuration) is sufficient, using more rods and slices can improve precision. It is necessary to know which set of points represents which rod.

2 Method

Let n denote the number of lines and m_i ($i = 1, \dots, n$) the number of point samples representing the lines. A line L_i is defined by its two arbitrary but different points $k_i = (k_i^x, k_i^y, k_i^z) \in \mathbb{R}^3$ and $l_i = (l_i^x, l_i^y, l_i^z) \in \mathbb{R}^3$. Let $p_{ij} = (p_{ij}^x, p_{ij}^y, p_{ij}^z) \in \mathbb{R}^3$ denote the j th point representing the i th line ($j = 1, \dots, m_i$). Let $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be a rigid-body transformation. To find the optimal T we have to minimize the following cost function:

$$\psi(T) = \sum_{i=1}^n \sum_{j=1}^{m_i} \text{dist}(T(p_{ij}), L_i)^2,$$

where $\text{dist}(q, L_i)$ ($q = (q^x, q^y, q^z) \in \mathbb{R}^3$) is the distance between the point and the line it represents, defined as follows:

$$x_1 = l_i^x - k_i^x, y_1 = l_i^y - k_i^y, z_1 = l_i^z - k_i^z,$$

$$x_2 = q^x - k_i^x, y_2 = q^y - k_i^y, z_2 = q^z - k_i^z,$$

$$\text{dist}(q, L_i) = \frac{\sqrt{(y_1 \cdot z_2 - z_1 \cdot y_2)^2 + (x_1 \cdot z_2 - x_2 \cdot z_1)^2 + (x_1 \cdot y_2 - x_2 \cdot y_1)^2}}{\sqrt{x_1^2 + y_1^2 + z_1^2}}.$$

We use Powell’s method to minimize the cost function [6]. A result of registration of CRW head-frame model to a CT volume is shown in Figs. 2 and 3.

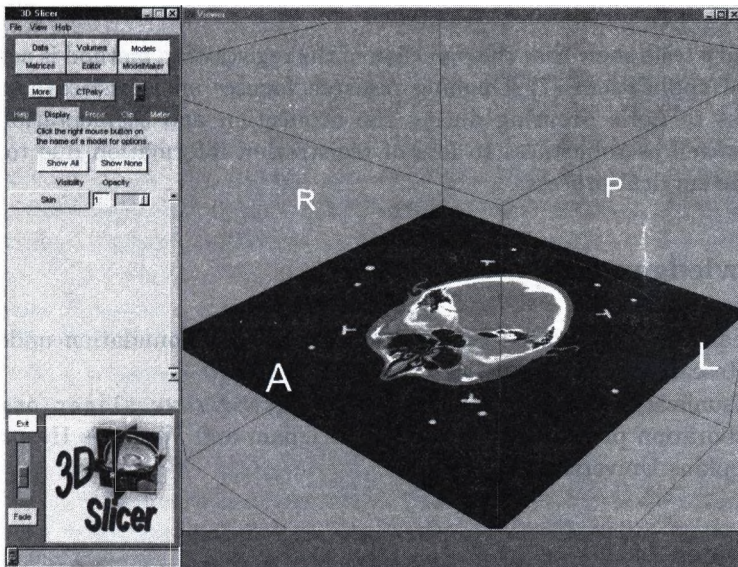


Fig. 2. A slice of a CT head volume. The cross sections of the CRW head-frame fiducial rods are well visible around the head.

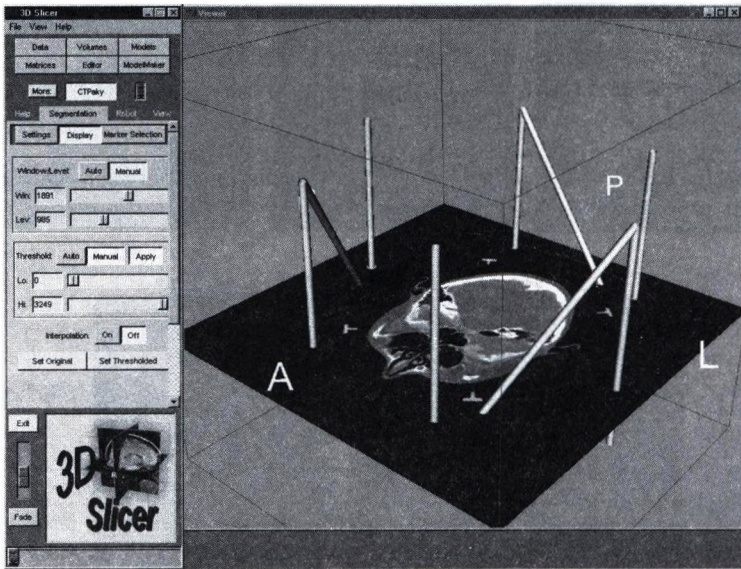


Fig. 3. A slice of a CT head volume and the registered model of the CRW head-frame fiducial rods.

3 Conclusion

Preliminary tests show that the precision of the registration method is acceptable in clinical applications [7]. Ongoing research focuses on statistical analysis of robustness to noise occurring during the acquisition and segmentation of the image, as well as robustness to loss of registration information due to motion during the surgical process.

Acknowledgements

This research was supported by the National Science Foundation under Engineering Research Center grant EEC9731478.

For visualization, we use the Slicer package (<http://www.slicer.org>) which is a collaboration project involving MIT, Brigham and Women's Hospital, and Johns Hopkins University.

References

1. G. Fichtinger, K. Masamune, A. Patriciu, A. Tanács, J.H. Anderson, T.L. DeWeese, R.H. Taylor, D. Stoianovici: Robotically Assisted Percutaneous Local Therapy and Biopsy. Workshop proceedings of the Tenth IEEE International Conference of Advance Robotics, Budapest, 2001, pp. 133–151, ISBN 963–7154–043, 2001

2. G. Fichtinger, T.L. DeWeese, A. Patriciu, A. Tanács, D. Mazilu, J.H. Anderson, K. Masamune, R.H. Taylor, D. Stoianovici: System for Robotically Assisted Prostate Biopsy and Therapy with Intraoperative CT Guidance. *Journal of Academic Radiology*, Vol. 9, No. 1, pp. 60–74, 2002
3. R.A. Brown, T.S. Roberts, A.G. Osborne: Stereotactic Frame and Computer Software for CT-Directed Neurosurgical Localization. *Invest. Radiol.*, Vol. 15, pp. 308–312, 1980
4. L. Leksell, B. Jerenberg: Stereotaxis and Tomography: a Technical Note. *Acta Neurochir*, vol. 52, pp. 1–7, 1980
5. R.C. Susil, J.H. Anderson, R.H. Taylor: A Single Image Registration Method for CT-Guided Interventions. *Lecture Notes in Computer Science, MICCAI99*, Springer-Verlag, Vol. 1679, pp. 798–808, 1999
6. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery: *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed., Cambridge University Press, New York, NY, 1992.
7. S. Lee, G. Fichtinger, S. Chirikjian: Novel Algorithms for Robust Registration of Fiducials in CT and MRI. *Proceedings of the Fourth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, *Lecture Notes in Computer Science 2208*, pp. 717–724, Springer Verlag, 2001

Approximating the Euclidean circle using neighbourhood sequences

András Hajdu¹ and Benedek Nagy²

¹ Institute of Mathematics and Informatics
University of Debrecen

hajdua@math.klte.hu

² Institute of Mathematics and Informatics
University of Debrecen

nbenedek@math.klte.hu

Abstract. Approximating Euclidean object in digital grids is a very important task of digital image processing, and digital geometry. In this paper we show a possible way of approximating the Euclidean circle in the most frequently used digital grids, the square, hexagonal, and triangular grid. In our procedure we observe the approximating rate of digital objects defined by neighbourhood sequences. These digital objects depend on the digital grid, thus we obtain different results, according to these digital grids. To measure the "goodness" of the approximation, we apply a standard geometric measure, the noncompactness ratio. For every investigated digital grid, we construct the best or a well approximating neighbourhood sequences, and give them in closed form.

1 Introduction

The problem of approximating the object of the 2D Euclidean space in digital grids is one of the most important and basic problems in digital geometry and digital image processing. For example, we can think about the several methods for drawing digital circles and lines in many applications. A main problem according to this drawing methods is that they are individual and the digital objects they specify cannot be handled mathematically in a simple way. In this paper we present a way for the approximation of the 2D Euclidean circle in digital grids, which can be described in a nice and simple way.

In our approximation methods we use neighbourhood sequences to construct the approximating objects. Neighbourhood sequences help us to move on these digital grids in a natural and easy-to-follow way, and we can measure distance by the help of them. We involve the 2D square, hexagonal and triangular grids into our analyses that are the most often used grids in digital image processing.

Neighbourhood sequences based on the neighbourhood relations defined on the digital grid. In our analyses we tried to select the most usual neighbourhood relations on these grids. It means that we consider (see Figure 1) two types of neighbours in the square grid, one type of neighbour in the hexagonal grid, and three types of neighbours in the triangular grid. We note here, that other types

of neighbours are considered sometimes on the triangular and hexagonal grid, in these cases the same analyses can be performed. Other planar grids are can also be considered, this case neighbourhood relations should be defined in the grid first.

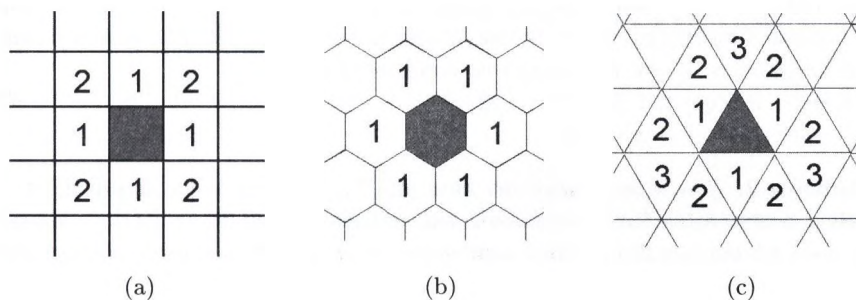


Fig. 1. Types of neighbours in digital grids (a) Square grid (b) Hexagonal grid (c) Triangular grid

The definition of neighbourhood sequences are very similar for these three grids, but they have different behavior in the triangular grid, since this grid has some "non-symmetric" property. Using neighbourhood sequences we can obtain different polygons in these grids, namely octagons in the square grid, hexagons in the hexagonal grid, and dodecagons in the triangular grid. Roughly speaking, we are restricted to use only these planar objects for the approximation of the circle, but we can do so in a clear and elegant way that can be relatively simply analysed.

2 Polygons defined by neighbourhood sequences in different digital grids

The concept of neighbourhood sequences were introduced by Das et al. [1] in 2D by allowing arbitrary mixture of the well-known 4- and 8-neighbourhood. This definition was given for periodic sequences, and a slightly modified definition was given in [3] to generalize it for arbitrary sequences in arbitrary dimension. We use the generalized definition here, since it provides a better approximation tool in our investigations. In this section we recall the definitions and statements we need here, and show some simple example to demonstrate the meaning and different behavior of neighbourhood sequences on the digital grids in which we perform our analyses.

2.1 Neighbourhood sequences in the square grid

In the square grid (Z^2) every grid point has two coordinates, with two types of possible neighbours. We recall the definition of neighbourhood sequences from [3] for the square grid.

Definition 1. Let p and q be two points in Z^2 . The i th coordinate of the point p is indicated by $\text{Pr}_i(p)$. Let k be an integer with $0 \leq k \leq 2$. The points p and q are k -neighbours, if the following two conditions hold:

- $|\text{Pr}_i(p) - \text{Pr}_i(q)| \leq 1$ for $1 \leq i \leq 2$,
- $\sum_{i=1}^2 |\text{Pr}_i(p) - \text{Pr}_i(q)| \leq k$.

Definition 2. The infinite sequence $B = (b_i)_{i=1}^{\infty}$, where $b_i \in \{1, 2\}$ for all $i \in N$, is called a generalized nD -neighbourhood sequence. If $\exists l \forall i : b(i) = b(i+l)$, then we call the neighbourhood sequence periodic, and use the notation $B = (b(1), b(2), \dots, b(l))$.

In Z^2 neighbourhood sequences are actually infinite sequences of 1s and 2s. Value 1 corresponds to 4-neighbours, and value 2 to 8-neighbours. The following definitions from [3] describe how can neighbourhood sequences be used for measuring distance in the square grid.

Definition 3. Let p and q be two points in Z^2 and B a generalized $2D$ -neighbourhood sequence. The point sequence $\Pi(p, q; B)$ – which has the form $p = p_0, p_1, \dots, p_m = q$, where p_{i-1} and p_i are $b(i)$ -neighbours for $1 \leq i \leq m$ – is called a path from p to q determined by B . The length $|\Pi(p, q; B)|$ of the path $\Pi(p, q; B)$ is m .

Definition 4. Let p and q be two points in Z^2 and B a generalized $2D$ -neighbourhood sequence. The shortest path from p to q is denoted by $\Pi^*(p, q; B)$. The distance between p and q is defined as the length of the minimal path, and is written as

$$d(p, q; B) = |\Pi^*(p, q; B)|.$$

We also use the brief notation $d(B)$ for the distance function generated by the neighbourhood sequence B .

In Figure 2 we show some simple examples how neighbourhood sequences spread in the square grid. As it is shown in the figure and as also a known result [2], in the square grid neighbourhood sequences occupy digital octagons after every step. In our analyses we consider the convex hull of these digital octagons, and use parameters of these Euclidean octagons.

Definition 5. Let B be a neighbourhood sequence in the square grid. Let p a point in Z^2 called the starting point of the neighbourhood sequence. For $k \in N$, let

$$\overline{B}_k = \{q \in Z^2 : d(p, q; B) \leq k\}.$$

Let $H(\overline{B}_k)$ be the convex hull of \overline{B}_k considered in R^n , the area a neighbourhood sequence occupies after its k th step.

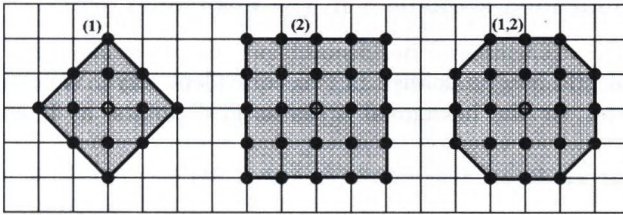


Fig. 2. The area the neighbourhood sequences (1), (2), and (1,2) occupy after 2 steps

It is shown in [2] that in Z^2 the area $H(\overline{B_k})$, an arbitrary neighbourhood sequence occupies after its k -th step, is always an octagon. Moreover, as it is known again from [2], for $\forall k \in N$ the octagon $H(\overline{B_k})$ is symmetric to both of the coordinate axes, and to the axes, obtained by rotating the coordinate axes by 45° .

We show that the length of the sides of the octagons, and hence its perimeter and area can be calculated by the number of the 1 and 2 elements of a neighbourhood sequence.

Definition 6. Let B be a neighbourhood sequence in the square grid, and $k \in N$. Let $k(1)$ and $k(2)$ be the number of the 1 and 2 values among the first k -th elements of B .

For our approximation method we need to determine the perimeter and area of these octagons. With simple calculations, we can prove that the side lengths, perimeters, and areas of these octagons can be given by the functions $k(1)$, and $k(2)$.

Definition 7. Let B be a neighbourhood sequence in the square grid. Let $a(k)$ be the length of the horizontal and $b(k)$ the length of the inclined sides of the octagon $H(\overline{B_k})$. Moreover, let $P(k)$ be the perimeter and $V(k)$ be the area of this octagon.

Remark 1. As a result of these symmetrical properties, for such an octagon the lengths of the horizontal and vertical sides are equal, and the same holds for the lengths of the inclined sides.

Proposition 1. Using the notations introduced above, the following relations hold:

- $a(k) = 2k(2)$,
- $b(k) = \sqrt{2}k(1)$,
- $P(k) = 4\sqrt{2}k(1) + 8k(2)$,
- $V(k) = 2k(1)^2 + 8k(1)k(2) + 4k(2)^2$.

We omit the proof of these statements here, since it is a simple and rather technical task.

2.2 Neighbourhood sequences in the hexagonal grid

The hexagonal grid can be defined in two equivalent ways. We can consider the triangular nodes, or the hexagonal areas as grid points, as it can be seen in Figure 3.

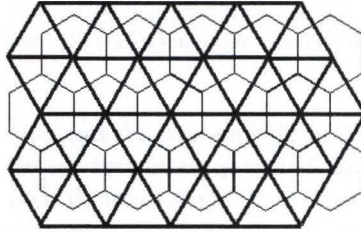


Fig. 3. The connection between the triangular and hexagonal grids

Two objects are neighbours

- in the grid of triangular nodes: if there is a direct connection between these nodes,
- in the grid of hexagonal areas: these hexagons have a common side.

In our analyses we use the first approach, and consider nodes as grid points. As we mentioned in the introduction, we use only one type of neighbourhood relation here, thus every grid point has six 1-type neighbours (see Figure 1). The path and distance (length of the minimal path) can be defined in the same way, as in the case of the square grid.

Since we consider only one type of neighbourhood relation, the concept of neighbourhood sequences is identical to the path defined above. In other words we can use only the neighbourhood sequence (1) here, and introduce only one hexagonal distance, based on the neighbourhood criterion at each step. In this grid, the region the neighbourhood sequence (1) occupies after any steps, is always a regular hexagon, see Figure 4.

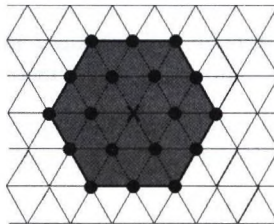


Fig. 4. The area the neighbourhood sequence (1) occupy after 2 steps

So if we apply Definition 5 to the hexagonal grid, we have that $\forall k \in N H(\overline{B_k})$ is a regular hexagon, where the neighbourhood sequence $B = (1)$. To perform the approximation of the circle we calculate the length, perimeter and area of these hexagons.

Proposition 2. *Let $a(k), P(k)$, and $V(k)$ be the side length, perimeter, and area of $H(\overline{B_k})$, respectively. Then*

- $a(k) = k$,
- $P(k) = 6k$,
- $V(k) = 3\sqrt{3}k^2$.

2.3 Neighbourhood sequences in the triangular grid

The triangular grid can be considered as the dual grid of the hexagonal grid. In other words, we can consider the hexagonal nodes, or the triangular areas as grid points, see Figure 3. In our analyses we consider nodes as grid points, instead of areas. In this grid we consider three types of neighbours, as it is shown in Figure 1. Every grid point has three 1-neighbours, nine 2-neighbours, and twelve 3-neighbours. The concept of neighbourhood sequences is a slightly differently can be introduced as it was defined in the square grid.

Definition 8. *The infinite sequence $B = (b_i)_{i=1}^\infty$, where $b_i \in \{1, 2, 3\}$ for all $i \in N$, is called a generalized neighbourhood sequence in the triangular grid. If $\exists l \forall i : b(i) = b(i + l)$, then we call the neighbourhood sequence periodic, and use the notation $B = (b(1), b(2), \dots, b(l))$.*

The concept of the shortest path defined by a neighbourhood sequence can be introduced in the same way as in the square grid. Contrary to the square grid, we can find neighbourhood sequences in the triangular grid that are equivalent with respect to measuring distance. We recall some results from [4] about this equivalence, which we need in our further investigation in this grid.

Definition 9. *Let B and C two neighbourhood sequences in the triangular grid. C is called the minimal equivalent neighbourhood sequence of B if the following conditions hold:*

- $d(p, q; B) = d(p, q; C)$ for all grid points p, q , and
- for each neighbourhood sequence D , if $d(p, q; B) = d(p, q; D)$ for all grid points p, q , then $c(i) \leq d(i)$ for all $i \in N$.

Lemma 1. *The minimal equivalent neighbourhood sequence C of B is uniquely determined, and is given by*

- $c(i) = b(i)$, if $b(i) < 3$,
- $c(i) = 3$, if $b(i) = 3$ and there is no $j < i$ such that $c(j) = 3$,
- $c(i) = 3$, if $b(i) = 3$ and there is some $c(l) = 3$ with $l < i$, and $\sum_{k=j+1}^{i-1} c(k)$ is odd, where $j = \max\{l \mid l < i, c(l) = 3\}$,
- $c(i) = 2$, otherwise.

In the triangular grid, for certain neighbourhood sequences, it can happen that a 3-step is equivalent to a 2-step. Using minimal equivalent neighbourhood sequences we replace these 3 values to 2 values, and obtain a neighbourhood sequence with the same behavior according to the distance measure.

In the triangular grid, just as in the case of the square and hexagonal grids, we consider the convex hulls of the occupied digital regions. In our further analyses we focus on the regions, the neighbourhood sequences occupy, and consider only the finite initial parts of neighbourhood sequences instead of the whole infinite sequences. We call two such parts region-equivalent (or simply equivalent) if they occupy the same region. With respect to this consideration we can classify these initial parts of neighbourhood sequences, and place them in finitely many, namely 6 classes. The classification procedure is based on the results of the following two lemmas. (We omit the proves of these lemmas, since they are simple consequences of basic definitions.)

Lemma 2. *The sequence part 23 is equivalent to the part 32, and 12 is equivalent to 21.*

Lemma 3. *The sequence value 2 is equivalent to 11.*

The first lemma let us to move the 2 values within the sequence, while using the second lemma we can replace every 11 occurrence to a 2 value, thus we have a single 1 value in the modified sequence. Using these lemmas we can construct the following classes.

Definition 10. *An initial part of a neighbourhood sequence is a basic part, if it has one of the following forms:*

(A)	(B)	(C)	(D)	(E)	(F)
2^m	$1(2)^m$	$(13)^n 2^m$	$1(31)^n 2^m$	$3(13)^{n-1} 2^m$	$(31)^n 2^m$

where $(string)^j$ is the repetition of string j times (e.g. $2^3 = 222$, $(13)^2 = 1313$). For (C), (D), (E), (F) we suppose that $n \geq 1$.

The following lemma shows that it is sufficient to deal with the above defined classes of basic parts in our approximation procedure.

Lemma 4. *Let B a neighbourhood sequence on the triangular grid. Every finite initial part of B has an equivalent basic part.*

Proof. Every occupied polygon trivially has a corresponding basic part that produces it, thus it is enough to restrict our attention to basic parts only.

The following table and Figure 5 summarizes the possible polygon types we can produce by using different basic parts.

1	(A), (B)	(C), (D), (E), (F)
triangle	hexagon	enneagon ($m = 0$); dodecagon ($m \geq 1$)

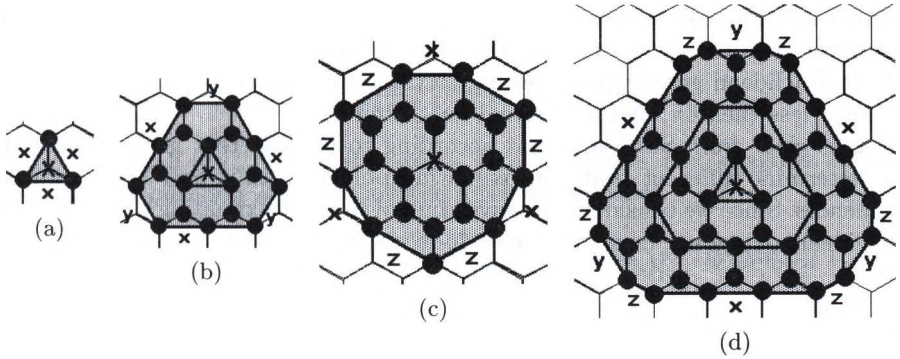


Fig. 5. Possible polygon types defined by basic neighbourhood parts in the triangular grid (a) triangle by part 1 (b) hexagon by part 12 (c) enneagon by part 13 (d) dodecagon by part 132

Since every neighbourhood sequence part can be transformed into a basic part of type (A)–(F), we have to investigate only these basic part types. Let m and n denote the number of the 2 and 3 values in the basic part respectively. Using m and n , in the following table we calculate the length of the sides of the polygons, corresponding to the basic part types (A)–(F). According to Figure 5 we denote the side lengths by x , y and z . Moreover, we also indicate the relation of m , n and the number of the elements of the part, denoted by k .

Class	x	y	z	k
(A)	m	m	0	m
(B)	$m + 1$	m	0	$m + 1$
(C)	$m + 2$	m	$(3n - 2)/\sqrt{3}$	$2n + m$
(D)	$m + 2$	m	$(3n - 1)/\sqrt{3}$	$2n + m + 1$
(E)	$m + 1$	m	$(3n - 2)/\sqrt{3}$	$2n + m - 1$
(F)	$m + 1$	m	$(3n - 1)/\sqrt{3}$	$2n + m$

Proposition 3. *The perimeter P and area V of the polygon defined by a basic part with elements k is*

- $P(k) = 3(x + y + 2z)$,
- $V(k) = \sqrt{3}(\frac{x^2}{4} + \frac{y^2}{4} + \frac{3z^2}{2} + xy) + 3xz + 3yz$.

Note, that these values depend on the class the basic part belongs to.

3 Finding best approximating neighbourhood sequence

In our approximation approach we calculate the *noncompactness ratio* of the approximating objects, which is equal to $\frac{\text{perimeter}^2}{\text{area}}$. Calculating the noncompactness ratio for approximating the circle is quite a usual technique, see e.g. [2]. In Euclidean geometry, according to the isoperimetric inequality, this ratio is never less than 4π , and is minimal for the circle.

It is clear from Section 2 that according to the different grids we analyse, we determine different best approximating polygons, thus best approximating sequences. Since we consider the convex hull of the approximating digital objects, Euclidean octagons, hexagons, and dodecagons are used to approximate the circle.

3.1 Approximating in the square grid

In the square grid, we can use octagons for approximating the circle. We can calculate the noncompactness ratio of such an octagon in the following way.

Proposition 4. *Let B be a 2D generalized neighbourhood sequence. The noncompactness ratio of $H(\overline{B_k})$ is*

$$NC(H(\overline{B_k})) = \frac{P(k)^2}{V(k)} = 16 \frac{d_{k,2}^2(3 - 2\sqrt{2}) + d_{k,2}(2\sqrt{2} - 2) + 1}{-d_{k,2}^2 + 2d_{k,2} + 1}.$$

where $d_{k,2} = \frac{k(2)}{k}$, the proportional frequency of the 2 values in the neighbourhood sequence.

In the 2D Euclidean space the circle has the minimal noncompactness ratio, namely 4π . So among the approximating octagons we have to look for the octagon, which has minimal noncompactness ratio. It means that as the next step of our approximation method, we have to minimize the function $NC(H(\overline{B_k}))$ for $d_{k,2}$.

Proposition 5. *The function $NC(H(\overline{B_k}))$ takes its minimum at $D(2) = \sqrt{2} - 1$, for every $k \in N$, and this minimum is equal to $32(\sqrt{2} - 1)$.*

Proof. We omit the proof of this proposition, since it requires a simple analytical calculation.

If we observe this minimal noncompactness ratio value $NC(H(\overline{B_k}))$, it clears out that it is the same for the regular octagon. So in our analyses we look for that neighbourhood sequence, which approximating octagons are closest to the regular ones, and the problem of approximating the circle becomes the problem of approximating the regular octagon.

Remark 2. We note, that the similar result can be obtained by looking for that $\frac{k(2)}{k}$ value, for which the horizontal and inclined sides of the octagon are equal. Nevertheless, in our further investigations we use the noncompactness ratio, so we need the function $NC(H(\overline{B_k}))$, as well.

The neighbourhood sequence for approximating the regular octagon can be constructed in such a way that for every k we choose $k(2)$ to make $d_{k,2}$ be as close as possible to $D(2)$. Using this constructive procedure, in the resulting neighbourhood sequence we have

$$\lim_{k \rightarrow \infty} d_{k,2} = \lim_{k \rightarrow \infty} \frac{k(2)}{k} = D(2) = \sqrt{2} - 1$$

for the density of the 2 values, if it exists. In other words, the octagons, defined by this neighbourhood sequence are the most regular ones.

Proposition 6. *Let the neighbourhood sequence $B = (b_i)_{i=1}^\infty$ be defined by:*

$$b(i) = \left\lfloor i(\sqrt{2} - 1) - \frac{1}{2} \right\rfloor - \left\lfloor (i - 1)(\sqrt{2} - 1) - \frac{1}{2} \right\rfloor + 1, \quad i \in N,$$

where $\lfloor . \rfloor$ denotes the floor function. The density of the 2 values in the neighbourhood sequence B is $\sqrt{2} - 1$, and this neighbourhood sequence generates the most regular octagon at every step k .

Proof. The statements of the proposition is a trivial consequence of the way, we construct this neighbourhood sequence.

Remark 3. This proposition generalizes a similar statement in [2], where only periodic sequences were investigated, and it was not possible there to define an infinite sequence with the above density, since $\sqrt{2} - 1$ is irrational. We list the first some elements of the best approximating sequence:

$$(1, 2, 1, 2, 1, 1, 2, 1, 2, 1, 2, 1, 1, 2, 1, \dots).$$

In Figure 6 we can observe, how the noncompactness ratio depends on the density of the 2 values in a neighbourhood sequence. Note, that 0 density represents the neighbourhood sequence (1), while 1 density the sequence (2). In these two cases, we obtain degenerated octagons, namely squares, whose noncompactness ratios are equal to 16.

3.2 Approximating in the hexagonal grid

Since we have only the neighbourhood sequence $B = (1)$ here, we have only the regular hexagon to approximate the Euclidean circle. This time the noncompactness ratio is:

$$NC(H(\overline{B_k})) = \frac{P(k)^2}{V(k)} = 8\sqrt{3}, \quad \forall k \in N.$$

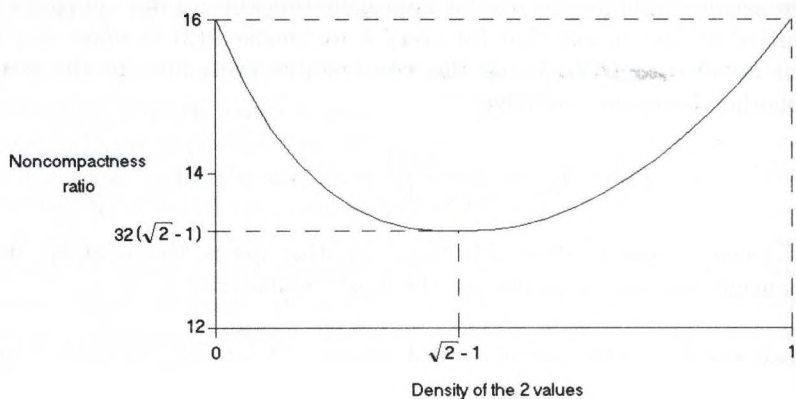


Fig. 6. Noncompactness ratio versus the proportional frequency of the 2 values

3.3 Approximating in the triangular grid

We have a similar result here as in the case of the square grid, namely that the approximation of the circle in the triangular grid is equivalent with approximating regular polygons. Let us observe the noncompactness ratio of the polygons, defined by the members of the classes (A)–(F) in subsection 2.3. We exclude the classes (A) and (B) from our further analyses, since they produce hexagons, so by using them, we cannot have better approximation rate than in the hexagonal grid (see subsection 3.2). In the case of the (C)–(F) classes, we can use dodecagons in the approximation, so we have to find those neighbourhood sequences, which approximate the regular dodecagon.

By minimizing the noncompactness ratio for the proportional frequency of the 2 values m/k , we can have a criteria for a best approximating neighbourhood sequence. This calculation can be performed in the same way for the classes (C)–(F), since in convergence limit we obtain the same minimum value.

Proposition 7. *The noncompactness ratio $\frac{P(k)^2}{V(k)}$ takes its minimum at $2\sqrt{3}-3$, and this minimum is equal to $96-48\sqrt{3}$.*

It means that the density of the 2 values must be $2\sqrt{3}-3$ in the neighbourhood sequence. With a simple calculation we can obtain the density of the 1 and 3 values which is $n = 2 - \sqrt{3}$. To construct a well approximating neighbourhood sequence let us observe the four classes (C)–(F) we can choose from. We can use any of these classes to construct a neighbourhood sequence with the desired density value, so we select a well approximating neighbourhood sequence, which also meets some natural considerations.

If we consider the first element of the neighbourhood sequence, the 3 value gives better approximation than the 1 or 2 value. Moreover, if we compare the

convergence speed of the neighbourhood sequences belonging to different classes, it clears out that class (E) gives us the fastest converging sequence.

If we observe class (E), we see that its sequences have one more 3 values than 1 values. So when we choose a 1 value, we switch to a sequence of (F), thus we have to choose a 3 value to go back to (E). Moreover, a 3 value cannot occur without having a 1 before, since in that case this 3 value could be replaced by a 2 value. By summarizing these conditions, we have that this sequence contains the 1 and 3 values in 13 blocks. We construct a sequence containing 1 and 2 values, where the 1 values stand for the 13 blocks of the neighbourhood sequence. In this sequence the density of the 1 values must be $\frac{\sqrt{3}-1}{2}$, while the density of the 2 values must be $\frac{3-\sqrt{3}}{2}$. We can use the following formula to generate the elements of this sequence:

$$b(i) = \left\lfloor i \frac{3 - \sqrt{3}}{2} - \frac{1}{2} \right\rfloor - \left\lfloor (i - 1) \frac{3 - \sqrt{3}}{2} - \frac{1}{2} \right\rfloor + 1, \quad i \in \mathbb{N}.$$

By inserting a starting 3 value into this sequence, and replacing the 1 symbols with 13 blocks, we obtain the investigated neighbourhood sequence:

$$(3, 2, 2, 1, 3, 2, 2, 1, 3, 2, 1, 3, 2, 2, 1, 3, \dots),$$

and the initial parts of this sequence provide the most regular polygons.

4 Summary

From the digital grids we examined the hexagonal grid provides the poorest result with respect to the approximation of the circle. We can improve this result by considering neighbourhood sequences in the square and triangular grid. In both of these grids we can construct neighbourhood sequences, which can arbitrarily approximate regular polygons. The best result can be achieved in the triangular grid, where this regular polygon is the regular dodecagon.

Acknowledgements

The authors would like to thank Lajos Hajdu for his valuable advice according to this research.

References

1. Das, P.P., Chakrabarti, P.P., Chatterji, B.N.: Distance functions in digital geometry Information Sciences, **42** (1987) 113–136
2. Das, P.P., Chatterji, B.N.: Octagonal distances for digital pictures Information Sciences, **50** (1990) 123–150
3. Fazekas, A., Hajdu, A., Hajdu, L.: Lattice of generalized neighbourhood sequences in nD and ∞D , **submitted**.
4. Nagy B.: Neighbourhood sequences in triangular grids, Technical Reports No. 2001/16, Institute of Mathematics and Informatics, University of Debrecen

Extracting Feature Vectors for Character Recognition by Walsh Transformation

Attila Fazekas¹ and András Hajdu²

¹ Institute of Mathematics and Informatics
University of Debrecen
fattila@math.klte.hu

² Institute of Mathematics and Informatics
University of Debrecen
hajdua@math.klte.hu

Abstract. In this paper we present an effective character recognition algorithm, which is based on Walsh transformation and uses feature vectors to recognize and classify the characters. Our main attempt here is to investigate how the extraction of feature vectors can be improved to make the recognition of the characters more reliable. We compare our algorithm with other popular methods, and make an analyses according to the tolerance against noise corruption, as well.

1 Introduction

Although more and more documents are composed in electronic format, there are machine vision systems and special printed materials that require the use of an optical character recognition method during procession. We can mention the problem of license plate or container number recognition as typical machine vision systems that includes a character recognition step, or we can think about recognizing printed and scanned materials, like recognizing fax or scanned documents. In the case of scanning and recognizing some printed materials, we have to separate machine printed and handwritten text. There are researches in progress to recognize handwritten characters, and there are some OCR packages, which are able to recognize neatly written text, but it seems to take some time to develop a really reliable procedure. To recognize machine printed text, huge number of character recognition methods were developed based on the feature of the character recognition problem. Character recognition methods usually contain a scanning step, preprocessing (binarization, segmentation), feature extraction, the actual recognition and classification, and sometimes some postprocessing and verification. For a comprehensive survey on feature extraction methods see [1], where several algorithms are presented and compared. For classification reason usually a feature vector is composed after the feature extraction step for every character, and the actual recognition is achieved by searching for the closest prototype feature vector. The dimension of the feature vector can vary application by application, and we also have to choose a suitable metrics to measure the difference between two feature vectors. In our analyses we tried to find a

method, which generates easily separable feature vectors, that is the prototype vectors have large distance from one another. We found that the well-known Walsh transformation has this property, so that the feature vectors generated by Walsh transformation can effectively be separated. Using Walsh transformation with underdetermined feature vectors we also obtain a noise filtering effect which is very useful in character recognition processes. We composed a character recognition algorithm, which uses Walsh transformation for generating the feature vectors. We made some experiments to verify the reliability of our character recognition method and some analyses to compare the performance of our feature vector extractor method against other popular character recognition methods (projection histogram, zoning). In the last section of our paper we discuss about some industrial applications that can be improved by involving feature vector optimization to the character recognition procedure.

2 The recognition system

The structure of our recognition system is based on the classical character recognition steps. An important difference from these classical systems is that we skipped feature extraction, since the classification step of our algorithm is based on the Walsh transforms of the image, which can be calculated without any modifications. In our algorithm we focused on classifying the characters in a reliable way, and we did not care much about the technical steps, like binarization or segmentation. We assumed the input image to be a binary image, and chose a simple segmentation algorithm to separate the characters. The detailed description of the segmentation and some segmentation problems are discussed in the following section. During segmentation the original binary image is divided into smaller segments, which are stored in a chained list. These segments are actually rectangles, and beside the image information, the coordinates of the upper left pixel and the size of the rectangle are also stored for every segment. We used the Walsh transformation to assign a feature vector to a segment. After calculating the feature vector of a given segment we judge whether the segment contains text information (letter, number, etc.), or an unrecognizable symbol. Figure 1 below describes how our recognition algorithm works:

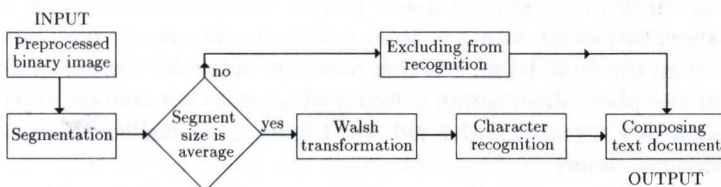


Fig. 1. Recognition system based on Walsh transformation

The main attempt of this paper is to show how the Walsh transformation can be used in finding feature vectors that can be reliably separated. Beside this analyses we also present some results according to noise sensitivity of our algorithm. In our tests we also made some other statistical investigations as well, such as factor analysis.

3 Segmentation

As it was mentioned before, we focused on feature vector extraction in our analyses, so we applied a very simple segmentation algorithm to extract the recognizable segments (characters) from the text. During segmentation we try to determine minimal storing rectangles for those subsets of the image foreground which can be separated by horizontal and vertical lines. These rectangles can be given by calculating their size and the coordinates of their upper left pixel. We mention that it is not always the connected foreground components our segmentation procedure results. This result depends on the font type as well, since in case of OCR fonts we can separate characters comfortably, but e.g. using CMR fonts can cause ligature appearances in the text. To make our algorithm able to recognize ligatures, and other unexpected symbols we integrated a training method into our system. This training procedure is very similar to the ones used in popular OCR packages. The segmentation of the characters achieved by alternating horizontal and vertical segmentation steps. We start from the original binary image as the initial segment, and divide it into smaller segments using the horizontal and vertical steps. The procedure stops if the segmentation steps do not modify the existing segments. This segmentation procedure has a small tolerance according to the rotation of the original image. Horizontal segmentation step: We have a pixel running from left to right, starting from the upper left corner of every segment we already have. If we find an object (foreground) point in the given row then we go down one pixel and start to run a pixel from the beginning of this row. The procedure continues till the running pixel reaches the right side of the segment (we find a row which does not contain object points). In this case we obtain a new segment. The top row of the new segment will be the uppermost row of the original segment which contains an object point. The bottom row of the new segment will be the lowermost row of the original segment which has been already processed and contains an object point. After defining the new segment, we go on with processing the original segment, starting from that row which did not contain object points. Vertical segmentation step: The vertical segmentation procedure is analogous to the horizontal one, but here the pixel runs from top to bottom, starting from the upper left corner of a segment. We go right one pixel till a column is found which does not contain object points. In this case a new segment is defined. On Figure 2 we can see the result of the segmentation procedure.

Moreover, we can find the ligature "fi" in the Hungarian word "figyelő" (observer). We note another problem, namely the segmentation of the accented characters. Accented characters can be recognized in a rather difficult way, since



Fig. 2. Result of the segmentation procedure

the accent is segmented separately. Analyzing the placement of the segments with small size can help to recognize these characters. For example, it can be useful to examine if a vowel takes place below a segment with small size.

4 Character recognition by Walsh transformation

We used the Walsh transformation as a tool to extract the feature vectors, by which the actual character recognition and classification can be performed. Walsh transformation was applied in several cases for several purposes, but never for character recognition. This transformation gives a verbose description of the image. We can check symmetrical relations, or the placement of the foreground and background pixels. During our investigation we found that the feature vectors, composed from some Walsh transforms of the segments, have a large distance with respect to a usual metrics. That was the main reason why we made an analysis to compare the feature vectors of the Walsh transformation and other character recognition methods. The kernel function $W(u, v)$ of the Walsh transformation is a separable and symmetric function. A complete description about the Walsh transformation can be found in [2]. To perform Walsh transformation, the original 2D image has to be an $N \times N = 2n \times 2n$ image for some natural number n . Thus we have to magnify every segment to a size of $2n \times 2n$. In our algorithm we magnified every segment to the size of 32×32 , which is not a considerable requirement, since Walsh transformation is invariant under magnification. One small problem occurs according to magnifying the segments to the same size (32×32). Those characters whose lower and upper cases are similar (e.g. "w" and "W") become identical after magnification. We have to compare the side lengths of the original storing rectangle to restore the proper case of the character. We computed only the following 64 Walsh transforms instead of the $32 \times 32 = 1024$ ones: ($W(0, 0)$, $W(0, 1)$, ..., $W(1, 0)$, ..., $W(7, 7)$). These 64 values are enough to describe the global features, and symmetric relations of the binary image. Restricting the number of the Walsh transforms filter out some noise corruption from the image, since the computation of less Walsh transforms results a blurring effect. We assign a feature vector to every segment to be recognized by composing placing the above described 64 Walsh transforms into a 64D vector. This 64D feature vector is compared with prototype vectors which contain the same 64 Walsh transforms of prototype characters. The prototype characters can be defined in two ways. We can use a prototype alphabet if it is

available, or we can define prototype characters as the average of sample characters. For a given feature vector we find the closest prototype vector by using a suitable distance function (for example the Cartesian distance). Segments with too small or too large size can be excluded from the recognition process. The classification is based on the distance between the feature vector of the analyzed character and the prototype vector.

5 Feature vectors extracted by Walsh transformation

There are two main points according to the distance of the feature vectors of different recognizable characters. We have a reliable recognition if the average distance between different feature vectors is quite large, and the distance of any two feature vectors never goes below a threshold value. It is recommended to find an algorithm that extracts feature vectors according to these two points. In our algorithm the difference (distance) of the 64D feature vectors of different characters is significant, so the recognition of a given character is quite reliable. Figure 3 contains the Cartesian distance values between the feature vectors of prototype digits. We can find strong differences between these values, which fortifies our conception to calculate only a 64 dimensional feature vector for each character.

	0	1	2	3	4	5	6	7	8
1	1221								
2	1494	1607							
3	1113	1532	1727						
4	1353	1338	1723	1566					
5	1208	1253	1466	1271	1417				
6	745	1464	1661	1224	1514	1273			
7	1752	1143	1722	1917	1773	1600	1961		
8	1224	1361	1554	1349	1511	1072	1229	1816	
9	853	1456	1367	1220	1742	1279	1012	1815	1257

Fig. 3. Distance values between prototype digits

According to the above investigation, an interesting problem can be to find a suitable dimension for the feature vectors. Depending on the application, in which we would like to use a character recognition process, the dimension of the feature vectors can be reduced without ruining the accuracy of the recognition. For example, if the recognizable characters only digits then the dimension of the feature vectors extracted by Walsh transformation can be reduced to 48 from 64 without considerable recognition failures. We used factor analysis as a tool to obtain the suitable dimension. Our test indicated that we can have a moderate recognition accuracy (at the level of 90%) if we use only a 16-dimensional feature space.

6 Comparing algorithms according to feature vector extraction

In the followings we present some experimental results about the distance of the feature vectors extracted by our algorithm against the results we got by applying other well-known character recognition procedures. We obtained better results in the case of the Walsh transformation with respect to the feature vectors, and consequently achieved higher recognition accuracy. To make our analysis more realistic, we applied some artificial noise corruption to the original images, and made tests on actually scanned printed material, which is equivalent with a moderate contour noise corruption. We applied both contour and global corruption. Walsh transformation reflected good tolerance against noise, since computing less transforms results a blurring effect. Our algorithm was compared against two classic character recognition method: one of them is based on projection histograms, the other one is based on zoning. The reason why we involved these character recognition algorithms into our analysis is that all of them use feature vectors to classify the recognizable characters. Both of these two methods assign a 64D feature vector to every recognizable segment, similarly to our algorithm. We involved different font types to our analyses (OCR-A, OCR-B, CMR). By fixing a prototype alphabet (letters, digits, punctuation marks, and other symbols), first we computed the average distance for every symbol from the rest of the alphabet. Figure 4 contains our results according to the different character recognition algorithms (we present the first 10 letters here only).

Letter	Walsh transformation		Projection histogram		Zoning	
	CMR	OCR	CMR	OCR	CMR	OCR
a	2375	2795	478	615	316	406
b	2011	2323	335	432	238	310
c	2281	2613	356	470	280	373
d	1974	2396	333	439	244	316
e	2288	2742	390	642	294	417
f	1983	2478	334	477	236	321
g	1930	2377	330	416	233	306
h	1965	2326	338	462	241	314
i	1628	2272	344	487	205	311
j	1655	2328	338	508	194	314

Fig. 4. Average distance values of the first 10 letters from the rest of the prototype alphabet

In our next analysis we corrupted the original binary images with different types of noise at different levels. Figure 5 contains only those noise corruption types and levels, where our algorithm performed perfectly (at 100% accuracy level). We also made this analysis to another character recognition method. Together the results of our method based on Walsh transformation we present

the results of the projection histogram procedure. We applied global and contour noise corruption both separately and together. The experimental analysis indicated that our algorithm tolerates noise corruption quite well. The noise sensitivity of the method can be reduced further by a lexical analysis.

Method	CMR		OCR	
	Global noise (%)	Contour noise (%)	Global noise (%)	Contour noise (%)
Walsh transformation	5	0	30	0
	0	15	0	50
	5	5	20	20
Projection histogram	5	0	5	0
	0	20	0	25
	-	-	5	5

Fig. 5. Noise types and levels with perfect recognition

7 Applications in practice - future trend

As a practical application we built in our character recognition method into an information loss compressive algorithm, see [3]. This compressive algorithm can be used when the main goal is to transmit the compressed document through some telecommunication channel. For example, fax documents usually have these properties. As another application we embedded our method into a license plate recognition system, and achieved quite a reliable accuracy rate. The algorithm can be developed in some places, e.g. parallel processing can be implemented during the segmentation step. Going on with our analysis about the feature vectors extracted by different character recognition algorithms, we plan to improve recognition performance in our applications. It can be achieved e.g. by composing new alphabets or modifying existing alphabets in such a way that the corresponding feature vectors are not so close to one another, thus the recognition becomes more reliable.

References

1. Trier, O.D., Jain, A.K., Taxt, T.: Feature extraction methods for character recognition - A survey *Pattern Recognition Letters*, (1996) 641-662
2. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing* Addison-Wesley, Reading, MA, (1992)
3. Fazekas, A., Hajdu, A.: An algorithm using Walsh transformation for compressing typeset documents *Acta Mathematica Academiae Paedagogicae Nyíregyháziensis*, (1999) 61-68

Support Vector Machines for Face Detection

A. Fazekas, I. Buciu, C. Kotropoulos, and I. Pitas

Department of Informatics, Aristotle University of Thessaloniki
GR-54006 Thessaloniki Box 451, Greece
fattila@math.klte.hu, {nelu, costas, pitas}@zeus.csd.auth.gr

Abstract. In this paper we are going to construct a new kernel function, which is based on Walsh functions, for support vector machines. We prove some theoretical results about the VC-dimension of the support vector machines which are built in the space of the Walsh functions. This information is very important, because the learning capability of the support vector machines depends on the VC-dimension of the kernel function used. The paper also proposes the application of majority voting on the output of several support vector machines in order to select the most suitable learning machine for frontal face detection. The paper also reports the first experimental results related to application of support vector machines with Walsh kernel and majority voting technique to face detection.

1 Introduction

The Bayes likelihood ratio test yields the optimal classifier in the sense that it minimizes the probability of error [4]. However in order to construct the likelihood ratio, the conditional probability density function (pdf) for each class must be known. Although, there are several procedures for estimating a pdf from a finite number of observations [4], the problem of density estimation is ill-posed [10]. An alternative method to solve a two-class pattern recognition problem is to resort to example-based techniques, such as the support vector machines (SVM) [10].

SVM implement the following idea [10]: By mapping the input pattern vectors, which are the elements of the training set, onto a high-dimensional feature space through an a priori suitably chosen mapping, we expect that the elements of the training set will be linearly separable in the feature space. We construct the optimal separating hyperplane in the feature space (as is explained in Section 2) to get a binary decision whether the input vector belongs to a given class or not. For example, in the case of the application studied in the paper, face detection, the input vector comprises gray levels of pixels from a rectangular region of the digital image and the result of the binary decision is the answer whether this region is a face or not.

In general, the determination of the separating hyperplane is not easy, because the dimensionality of the feature space is high. However, in Hilbert spaces, one can estimate the inner product of two vectors in the feature space as a function of the inner product of the two vectors in the input space. These expressions

for inner products are referred as kernel functions. Some kernel functions are well-known, for example the polynomial, the radial, the sigmoid, etc. [1, 2].

In this paper we will construct a new kernel function that is based on the 2-dimensional Walsh system. The 2-dimensional Walsh system is useful in pattern recognition [3]. Because the VC-dimension is the capacity factor of the SVM [11], its knowledge is very important in order to control its behavior. In this paper we prove several propositions about the VC-dimension of the class of the 2-dimensional Walsh functions. Finally, the paper reports experimental results in order to assess the properties of SVMs with Walsh kernels in face detection.

We will give an alternative approach instead of constructing a new kernel function to make a better face detector. We propose to rank an ensemble of SVMs trained on the same set by combining their outputs with majority voting in the decision making process. By doing so, we can define the most efficient SVM, i.e., the one whose outputs appear most frequently in the set of the outputs produced by the ensemble of SVMs. We apply this technique to frontal face detection and report a significant reduction of false acceptance rate.

The structure of the paper is the following: Section 2 and Section 3 are brief overviews of the principles of SVMs and the Walsh system, respectively. Section 4 explains the construction of the 2-dimensional Walsh kernel. Section 5 describes the theoretical results on the VC-dimension of the class of the Walsh functions. Finally, Section 6 presents the promising experimental results obtained with the proposed system of SVMs when applied to the face detection. The application of majority on the output of several support vector machines is explained in Section 7 and its experimental results are reported in Section 8.

The following notation is used throughout the paper. Bold face symbols will indicate vectors or matrices, and normal typeface will be used for vector and matrix elements as well as for scalars. We will denote the set of non-negative integers by \mathbb{N} , the set of positive integers by \mathbb{N}^+ , the set of integers by \mathbb{Z} , and the set of real numbers by \mathbb{R} .

2 Support Vector Machines

SVMs are learning algorithms based on the statistical learning theory [11]. In this section we briefly describe the foundation of SVMs by [11]. Statistical learning from examples aims at selecting from a given set of functions $\{f_\alpha(\mathbf{x}) \mid \alpha \in \Lambda\}$, the one which predicts best the correct response (i.e. the response of a supervisor). This selection is based on the observation of l pairs that build the training set:

$$(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l), \quad \mathbf{x}_i \in \mathbb{R}^m, y_i \in \{+1, -1\} \quad (1)$$

which contains input vectors \mathbf{x}_i and the associated ground "truth" y_i given by an external supervisor.

Let the response of the learning machine $f_\alpha(\mathbf{x})$ belongs to a set of indicator functions $\{f_\alpha(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^m, \alpha \in \Lambda\}$ (see Definition 7 subsequently). If we define the loss-function:

$$L(y, f_\alpha(\mathbf{x})) = \begin{cases} 0, & \text{if } y = f_\alpha(\mathbf{x}), \\ 1, & \text{if } y \neq f_\alpha(\mathbf{x}) \end{cases} \quad (2)$$

that measures the error between the ground truth y to a given input \mathbf{x} and the response $f_\alpha(\mathbf{x})$ provided by the learning machine, the expected value of the loss is given by:

$$R(\alpha) = \int L(y, f_\alpha(\mathbf{x}))p(\mathbf{x}, y)d\mathbf{x}dy \quad (3)$$

where $p(\mathbf{x}, y)$ is the joint probability density function of random variables \mathbf{x} and y . $R(\alpha)$ is called the expected risk. We would like to find the function $f_{\alpha_0}(\mathbf{x})$ which minimizes the risk functional $R(\alpha)$. The selection of the function is based on the training set of l random independent identically distributed observations (1). In order to minimize the risk functional $R(\alpha)$ the empirical risk minimization (ERM) induction principle is usually employed by replacing the expected risk functional $R(\alpha)$ by the empirical risk functional, which is measured on the training set:

$$R_{\text{emp}}(\alpha) = \frac{1}{l} \sum_{i=1}^l L(y_i, f_\alpha(\mathbf{x}_i)). \quad (4)$$

The idea underlying the ERM induction principle is to approximate the function $f_{\alpha_0}(\mathbf{x})$ which minimizes $R(\alpha)$ by the function $f_{\alpha_l}(\mathbf{x})$ which minimizes the empirical risk. This approach may be valid for training sets having large size (ideally infinite). It is known that for some η such that $0 \leq \eta \leq 1$, the expected risk is bounded for arbitrary $\alpha \in A$ with probability $1 - \eta$ [10]:

$$R(\alpha) \leq R_{\text{emp}}(\alpha) + \sqrt{\frac{h(\log(\frac{2l}{h}) + 1) - \log(\frac{\eta}{4})}{l}}, \quad (5)$$

where h is a non-negative integer called the Vapnik-Chervonenkis (VC) dimension (see subsequent Definition 9), and is a measure of capacity of SVMs. In the following, we shall call the right hand side of inequality (5) the risk bound. The second term in the risk bound is called the VC-confidence. Inequality (5) reveals the necessity to minimize both the empirical risk and the VC-confidence. This is the aim of the so-called structural risk minimization (SRM) principle. The SVMs are learning machines that implement the SRM principle in their training.

In order to introduce the basic idea of SVMs, let us consider the construction of the optimal separating hyperplane. Suppose the training data (1) can be separated by a hyperplane, that is $\exists \mathbf{v} \in \mathbb{R}^m$:

$$(\mathbf{v}^T \mathbf{x}_i) + b \geq 1, \quad \text{if } y_i = +1 \quad (6)$$

$$(\mathbf{v}^T \mathbf{x}_i) + b \leq -1, \quad \text{if } y_i = -1, \quad (7)$$

where \mathbf{v} is a normal to the hyperplane, $\frac{|b|}{\|\mathbf{v}\|}$ is the perpendicular distance from the hyperplane to the origin, and $\|\mathbf{v}\|$ is the Euclidean norm of \mathbf{v} . A compact notation for inequalities (6) and (7) is:

$$y_i ((\mathbf{v}^T \mathbf{x}_i) + b) \geq 1, \quad i = 1, \dots, l. \quad (8)$$

Let d_+ (d_-) be the Euclidean distance from the separating hyperplane to the closest positive (negative) example. Define the margin of the separating hyperplane to be $d_+ + d_-$. For the linearly separable case, SVM simply seeks for the separating hyperplane with the largest margin. The optimal hyperplane minimizes

$$\frac{1}{2} \|\mathbf{v}\|^2 \quad \text{subject to the inequalities (8).} \quad (9)$$

After training a support vector machine, one simply determines the side of the decision boundary where a given test pattern \mathbf{x} lies on and assigns the corresponding class label, i.e., $\theta(\mathbf{v}^T \mathbf{x} + b)$.

Let us investigate the generalization to the case where the decision function is not a linear function of the input vector [11]. Now suppose we have first mapped the data to some other Euclidean space \mathcal{H} , using a mapping Φ :

$$\Phi : X \rightarrow \mathcal{H}. \quad (10)$$

Then the training algorithm would only depend on the data through inner products in \mathcal{H} , i.e., on functions of the form $\Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$. If there were a kernel function K such that $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$, we would only need to use $K(\mathbf{x}_i, \mathbf{x}_j)$ in the training algorithm without necessarily explicitly knowing $\Phi(\mathbf{x})$.

How can we use this machine? After all, we need \mathbf{v} that will be in \mathcal{H} as well. The solution of the optimization problem (9) yields a coefficient vector \mathbf{v} that is expressed as a linear combination of a subset of training vectors, whose associated Lagrange multipliers are non-zero. These training vectors are called support vectors. In our case we will denote the images of the support vectors \mathbf{s}_i , $\Phi(\mathbf{s}_i)$. In the test phase, an SVM computes the sign of:

$$f(\mathbf{x}) = \theta \left(\sum_{i=1}^{N_s} \lambda_i y_i \Phi(\mathbf{s}_i)^T \Phi(\mathbf{x}) + b \right) \quad (11)$$

where λ_i are the Lagrange multipliers that are associated to $\Phi(\mathbf{x}_i)$, and N_s is the number of the support vectors. Again, we can avoid computing $\Phi(\mathbf{s}_i)^T \Phi(\mathbf{x})$ explicitly and use $K(\mathbf{s}_i, \mathbf{x}) = \Phi(\mathbf{s}_i)^T \Phi(\mathbf{x})$.

3 The Walsh System

In the literature the term ‘‘Walsh functions’’ refers to one of three orthonormal systems: the Walsh-Paley system, the original Walsh system, or the Walsh-Kaczmarz system [9]. These systems contain the same functions and differ only in enumeration. We will investigate the Walsh-Paley system which will be referred as the Walsh system henceforth. For more details the interested reader may consult [9]. In the following, $n, m \in \mathbb{N}$.

Definition 1. Let $r(x)$ be the function defined on $[0, 1)$ by:

$$r(x) = \begin{cases} 1, & x \in [0, \frac{1}{2}), \\ -1, & x \in [\frac{1}{2}, 1) \end{cases}$$

extended to \mathbb{R} periodically with period 1. The Rademacher system $R = \{r_n(x)\}$ is defined by:

$$r_n(x) = r(2^n x), \quad x \in \mathbb{R}, n \in \mathbb{N}.$$

Definition 2. Given $n \in \mathbb{N}$, it is possible to write n uniquely as:

$$n = \sum_{k=0}^{\infty} n[k]2^k,$$

where either $n[k] = 0$ or 1 for $k \in \mathbb{N}$. This expression will be called the binary expansion of n and the numbers $n[k]$ will be called the binary coefficients of n .

Definition 3. Let x be an arbitrary element of the interval $[0, 1)$. If x has the form $\frac{p}{2^n}$ for some p ($0 \leq p < 2^n$), x will be a dyadic rational in the interval $[0, 1)$.

Definition 4. Any $x \in [0, 1)$ can be written in the form:

$$x = \sum_{k=0}^{\infty} x[k]2^{-(k+1)},$$

where each $x[k]$ is equal to either 0 or 1 . We will call it the dyadic expansion of x . When x is a dyadic rational there are two expressions of this form, one which terminates in 0 's and one which terminates in 1 's. In this paper, we will confine ourselves to the dyadic expansion of x the one that terminates in 0 's.

Definition 5. The Walsh system $W = \{w_n(x)\}$ is product of Rademacher functions in the following way. If $n \in \mathbb{N}$ has binary coefficients $\{n[k], k \in \mathbb{N}\}$ then:

$$w_n(x) = \prod_{k=0}^{\infty} r_k^{n[k]}(x).$$

It is easy to see that this product is always finite, $w_0(x) = 1$ and $w_{2^n}(x) = r_n(x) \forall x \in [0, 1)$. It is worth noticing that each Walsh function is piecewise constant with finitely many jump discontinuities on $[0, 1)$, and takes only the values of either $+1$ or -1 .

Definition 6. The 2-D Walsh system $W^{(2)} = \{w_{(n,m)}(x, y)\}$ is defined in the following way:

$$w_{(n,m)}(x, y) = w_n(x) \cdot w_m(y).$$

4 The Construction of the Walsh Kernel

It is well-known, that the Walsh system is a complete orthonormal system on $[0, 1)$ and the Walsh system is a Schauder basis in L^p for $1 < p < \infty$ [9]. This linear space is a Hilbert space, where the inner product, denoted by $\langle \cdot, \cdot \rangle$ is simply the integral of the product of two functions $\langle p(x), q(x) \rangle = \int_0^1 p(\xi)q(\xi)d\xi$ [9]. Walsh functions have already been used in image analysis [3].

In this section we define a new kernel function for support vector machines. The construction is based on Vapnik's idea [11]. Let us suppose that we would like to analyze an one-dimensional signal in terms of Walsh functions. Let us map the input variable x into the N -dimensional feature space, as follows:

$$\Phi_{(N)}(x) = \left(\frac{1}{\sqrt{2^0}}w_0(x), \frac{1}{\sqrt{2^1}}w_1(x), \dots, \frac{1}{\sqrt{2^{N-1}}}w_{N-1}(x) \right)^T. \quad (12)$$

The inner product of two vectors in this space has the form:

$$K_N(x, x') = \Phi_{(N)}(x)^T \Phi_{(N)}(x') = \sum_{k=0}^{N-1} \frac{1}{2^k} w_k(x) w_k(x'). \quad (13)$$

It is convenient to assume that $N = 2^n$, $n \in \mathbb{N}$. The relations stem from the theoretical characteristics of the Walsh functions [9]. In the case of digital images, the input variable x is the gray level of an arbitrary image pixel. Let us suppose x is normalized to $[0, 1)$. If we were to linearly combine the elements of (12), then the weights of the linear combination would be chosen so that linear separability was maintained in an ideal point-wise SVM. The required inner product computations are then replaced by bit-wise operations that can be executed fast on either dedicated hardware or software.

To construct the SVM for the 2-dimensional vector space $X = \{x \mid x = (x, y)^T\}$, it is sufficient to use the generating kernel that is a product of one-dimensional kernels:

$$K_{(N,M)}^2(\mathbf{x}, \mathbf{x}') = K_N(x, x') \cdot K_M(y, y'). \quad (14)$$

The calculation of the mapping (12) is not a difficult task, because we can use the idea of fast Walsh transformation [5].

5 The VC-dimension of the Class of the 2-D Walsh Functions

The Vapnik-Chervonenkis dimension has a very important role in the statistical learning. The VC-dimension of the support vector machines characterizes the learning capacity of the machine. With control of the VC-dimension one can avoid overfitting of the support vector machines and one can minimize the expected value of the error [11]. So the knowledge of the VC-dimension of the class of the functions employed in the learning algorithm is very important.

At first we quote some definitions from [11], which are important to understand the theoretical results of this section.

Definition 7. An arbitrary $\{+1, -1\}$ -valued function with domain \mathbb{R}^2 is called an 2-D indicator function.

Definition 8. Let f be an arbitrary 2-D indicator function. The sets $\{(x, y) \mid f(x, y) = +1, x, y \in \mathbb{R}\}$ and $\{(x, y) \mid f(x, y) = -1, x, y \in \mathbb{R}\}$ are the separated classes of the domain by using f .

Definition 9. The VC-dimension of a set of 2-D indicator functions is equal to the largest number h of points of the domain of the functions that can be separated into two different classes in all the 2^h possible ways using a function from this set. If for any m there exists a set of m points that can be shattered by the functions of the set, then the VC-dimension is equal to infinity.

In the rest of the paper we assume that the domain of 2-D Walsh functions and of the 2-D indicator functions is the set $[0, 1]^2$.

Theorem 1. The VC-dimension of the class of all the 2-D Walsh functions is equal to ∞ .

The relations $N = 2^n$, $M = 2^m$ and $n, m \geq 2$ are assumed in the following.

Theorem 2. The VC-dimension of the set $W_{(N,M)} = \{w_{(k_1,k_2)}(x, y) \mid k_1 = 0, \dots, N - 1, k_2 = 0, \dots, M - 1\}$ is equal to $\log_2(N \cdot M)$.

Definition 10. Let $G = \{x[n] \mid n \in \mathbb{N}, x[n] = 1 \text{ or } x[n] = 0\}$. Set $I_0(x) = G$ for all $x \in G$. For each $x \in G$ and $n \in \mathbb{N}^+$ define:

$$I_n(x) = \{y \in G \mid y[i] = x[i], 0 \leq i < n\}.$$

We call the sets $I_n(x)$ the dyadic intervals of order n in $[0, 1]$.

Definition 11. By a dyadic step function of order n we mean a finite linear combination of characteristic functions of dyadic intervals of order n in $[0, 1]$.

Definition 12. By a 2-D dyadic step function of order (n, m) we mean a product of two dyadic step functions of orders n and m , respectively.

We use the following notation:

$$\theta(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ -1, & \text{if } x < 0, \end{cases}$$

where $x \in \mathbb{R}$.

Definition 13. Let $f_\alpha(x, y)$ be \mathbb{R} -valued functions. We call the set of indicator functions $\theta(f_\alpha(x, y) - t)$, where $t \in (\inf_{(x,y)} f_\alpha(x, y), \sup_{(x,y)} f_\alpha(x, y))$, the set of indicators for functions $f_\alpha(x, y)$.

Theorem 3. The VC-dimension of the set $\text{lin}(W_{(N,M)}) = \{f(x, y) \mid f(x, y) = \alpha_{(0,0)}w_{(0,0)}(x, y) + \dots + \alpha_{(0,M-1)}w_{(0,M-1)}(x, y) + \dots + \alpha_{(N-1,M-1)}w_{(N-1,M-1)}(x, y), \alpha_{(i,j)} \in \mathbb{R}\}$ is equal to $N \cdot M$.

Corollary 1. *The VC-dimension of the kernel function $K_{(N,M)}$ is equal to $N \cdot M$.*

The theoretical developments presented in this section lead to the construction of an SVM that employs the values admitted by the Walsh basis functions as elements of the training vectors and resorts to coefficients of the linear combination that solve a two-class pattern recognition. The latter coefficients are **not** the ones given by the analysis equation of the Walsh transformation.

6 Experimental Results about Walsh SVMs

The purpose of the designed experiments is to assess the performance of Walsh SVMs in face detection. A training data set of 112 images, 57 images containing a face and another 55 images with non-face patterns, is built. The images containing face patterns have been derived from the face database of IBERMATICA where several sources of degradation are modeled. For a description of this database the interested reader may refer to [8].

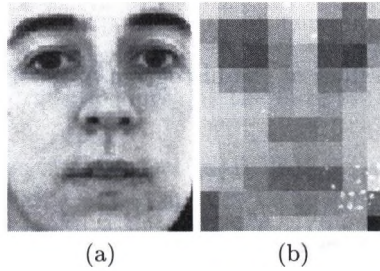
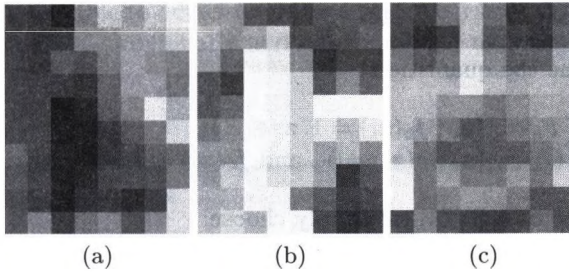
All images in this database are recorded in 256 grey levels and they are of dimensions 320×240 . The procedure for collecting face patterns is as follows. From each image a bounding rectangle of dimensions 128×128 pixels has been manually determined that includes the actual face. This area has been subsampled four times. At each subsampling, non-overlapping regions of 2×2 pixels are replaced by their average. Accordingly, training patterns of dimensions 8×8 are built. The ground truth, that is, the class label +1 has been appended to each pattern. Similarly, 55 non-face patterns have been collected from images in the same way, and labeled by -1.

We have trained the three different SVMs indicated in Table 1. For all experiments the SVMLight toolbox developed by T. Joachims was used [7]. The trained SVMs have been applied to 442 test images (249 faces and 193 non-faces) from the IBERMATICA database that have not been included in the training set. The resolution of each test image has been reduced four times yielding a final image of dimensions 8×8 . The test images are classified as either non-face or face ones.

Table 1 summarizes the results of the test. The first row in Table 1 depicts the cpu time needed for test experiments using each kernel functions for a Pentium III at 543 MHz. As can be seen, the support vector machine that is based on the Walsh function requires more time than the other SVMs. This is attributed to the high dimension of the feature space. The numbers of errors shown in the second row are the misclassification errors, that is either the number of real faces classified as non-faces (number of false rejections) or the number of non-face instances classified as faces (number of false acceptances). From this point of view the Walsh kernel outperforms the other two competitors. Finally, the number of support vectors in the case of Walsh kernel is larger than that of the linear and the polynomial kernels.

Table 1. Experimental Results

	Linear	Walsh	Polynomial with degree 2
Time (sec)	1.78	7.62	5.32
Number of the False Acceptances	18	0	2
Number of the False Rejections	11	0	1
Number of SVs	24	74	14

**Fig. 1.** Correctly detected face in the (a) original, and in the (b) subsampled resolution.**Fig. 2.** False accepted cases (a)-(b), and false rejected case (c) in the subsampled resolution in the case of linear and polynomial SVM.

7 Majority Vote Applied in the case of the Outputs of Several SVMs

Let us consider five SVMs were based on the several kernels as follows: (1) Linear kernel; (2) Polynomial kernel with degree 2; (3) Gaussian Radial Basis Function (GRBF) having σ equal to 10; (4) Sigmoid with $k = 0.5$ and $\theta = 0.2$; (5) Exponential Radial Basis Function (ERBF) having σ equal to 10. The penalty was set to 500. These kernels have the analytical form listed in Table 2, where $\|\cdot\|_p$ denotes the vector p -norm, $p = 1, 2$.

For brevity, we index each SVM by k , $k = 1, \dots, 5$. To distinguish between training and test patterns, the latter ones are denoted by \mathbf{z}_i . Let \mathcal{Z} be the test

Table 2. Kernel functions used in SVMs.

k	SVM type	Kernel function $K(\mathbf{x}, \mathbf{y})$
1	Linear	$\mathbf{x}^T \mathbf{y}$
2	Polynomial with degree q	$(\mathbf{x}^T \mathbf{y} + 1)^q$
3	GRBF	$\exp(-\frac{\ \mathbf{x}-\mathbf{y}\ _1^2}{2\sigma^2})$
4	Sigmoid	$\tanh(k \cdot \mathbf{x}^T \mathbf{y} - \theta)$
5	ERBF	$\exp(-\frac{\ \mathbf{x}-\mathbf{y}\ _1}{2\sigma^2})$

set. We define the two values of the histogram of labels assigned to all $\mathbf{z}_i \in \mathcal{Z}$ as:

$$\begin{aligned} h_1(\mathbf{z}_i) &= |\{k \mid f_k(\mathbf{z}_i) = 1, k = 1, \dots, 5\}|, \\ h_{-1}(\mathbf{z}_i) &= |\{k \mid f_k(\mathbf{z}_i) = -1, k = 1, \dots, 5\}|. \end{aligned} \quad (15)$$

The we may combine the decision taken separately by the SVMs indexed by $k = 1, \dots, 5$ in the following manner:

$$g(\mathbf{z}_i) = \begin{cases} 1, & \text{if } h_1(\mathbf{z}_i) > h_{-1}(\mathbf{z}_i), \\ -1, & \text{otherwise.} \end{cases} \quad (16)$$

Let us now define the quantities:

$$\begin{aligned} F_k &= |\{\mathbf{z}_i \mid f_k(\mathbf{z}_i) = 1, \mathbf{z}_i \in \mathcal{Z}\}|, \\ G_k &= |\{\mathbf{z}_i \mid g(\mathbf{z}_i) = 1, \text{ and } f_k(\mathbf{z}_i) = 1, \mathbf{z}_i \in \mathcal{Z}\}|. \end{aligned} \quad (17)$$

To determine the best SVM, we simply choose:

$$k' = \arg \max_k \frac{G_k}{F_k}. \quad (18)$$

If more than one indices are determined by (18) we count additionally the number of consistent classifications $|\{\mathbf{z}_i \mid f_{k'}(\mathbf{z}_i) = g(\mathbf{z}_i), \mathbf{z}_i \in \mathcal{Z}\}|$, and we resolve the tie by selecting the index that maximizes that number.

8 Experimental Results about Majority Voting Technique

We have trained the five different SVMs indicated in Table 2. The trained SVMs have been applied to six test imaged from the IBERMATICA database that were not included in the training set. Each test image corresponds to a different person. The resolution of images has been decreased four time using a pyramidal algorithm, the final size of the image being of the dimension 15×20 . Scanning row by row the reduced image by a 10×8 rectangular window, test patterns are classified as non-face ones or face pattern. When a face pattern is found by the

Table 3. Ratio G_k/F_k achieved by the various SVMs.

SVM type	Test Images numbers					
	1	2	3	4	5	6
Linear	0.83	0.20	0.57	0.66	1.00	0.74
Polynomial	0.52	0.28	0.57	0.44	1.00	0.71
GRBF	0.67	0.25	0.44	0.44	0.80	0.83
Sigmoid	0.64	0.14	0.15	0.11	0.22	0.13
ERBF	1.00	0.50	0.80	0.80	0.80	1.00

machine, a rectangle is drawn for locating the face. We have tabulated the ratio G_k/F_k in Table 3.

From this table it can be seen that ERBF is found to maximize the ratio in (18) for the five test images. On the contrary the machine built using the sigmoid kernel attains the worst performance with respect to (18). Two quantities measurements have been used for performing the performance of each SVM, namely, the false acceptance rate (FAR) and the false rejection rate (FRR) during the test phase. We have measured FAR and FRR for each individual SVM before and after applying the majority vote procedure. FRR is always zero, meaning that each machine is able to detect faces. Instead FAR varies. The values of FAR attained by each SVM individually and after applying majority vote are shown in Table 4.

Table 4. False acceptance rates (in %) achieved by the various SVMs individually and after applying the majority vote.

SVM type	Test Images numbers					
	1	2	3	4	5	6
Linear	3.9	10.5	6.5	5.2	2.6	6.5
Polynomial	6.5	6.5	6.5	9.2	2.6	6.5
GRBF	5.2	7.8	9.2	9.2	3.9	5.2
Sigmoid	7.8	17.1	31.5	44.7	21.0	47.3
ERBF	2.6	2.6	3.9	3.9	3.9	3.9
Combining	2.6	1.3	2.6	2.6	2.6	3.9

One can see that the application of majority voting reduces the number of false acceptances in all cases and particularly when $F_k \neq G_k$.

Figure 3 and Figure 4 depict face localization determined by the five SVM types along with the face localization using majority vote.

It is seen that majority vote helps to discard many of the candidate face regions returned by single SVMs (Fig.3) yielding the best face localization (Fig.4).

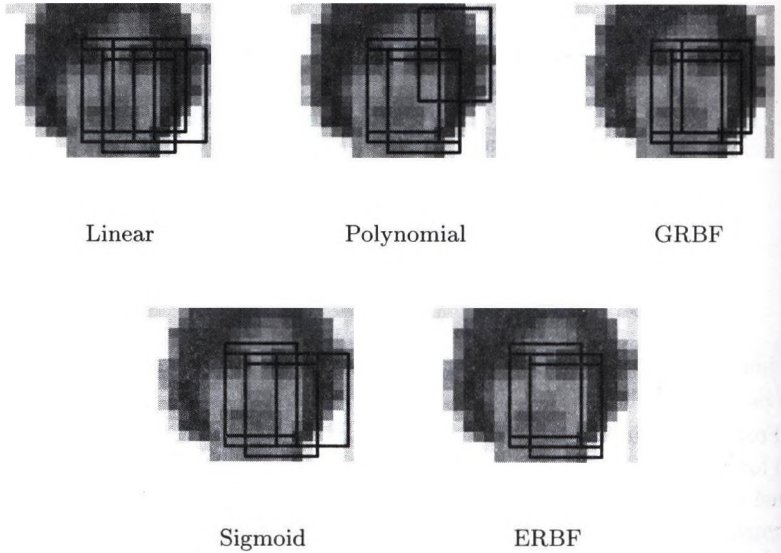


Fig. 3. Face localization in a test image for individual SVMs.

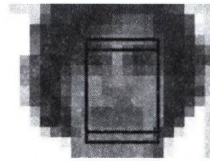


Fig. 4. More accurate face localization after using majority vote procedure.

Acknowledgements

This work was supported by the European Union Research Training Network “Multi-modal Human-Computer Interaction (HPRN-CT-2000-00111)”.

References

1. C.J.C. Burges. A tutorial on support vector machines for pattern recognition, *Knowledge Discovery and Data Mining*, **2**, 1998.
2. L. Devroye, L. Györfi and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New York, 1996.

3. A. Fázekas and A. Hajdu. Recognizing typeset documents using Walsh transformation, *Journal of Computing and Information Technology*, Vol. **9**, pp. 101–112, 2001.
4. K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, San Diego, 1990.
5. R.C. Gonzalez and R.E. Woods. *Digital Image Processing*. Addison-Wesley, New York, 1993.
6. S. Gunn. Support vector machines for classification and regression, *ISIS Technical Report ISIS-1-98*, Image Speech Intelligent System Research Group, University of Southampton, 1998.
7. T. Joachims. Making large-scale SVM learning practical, in *Advances in Kernel Methods - Support Vector Learning*, pp. 169–184, MIT Press, Cambridge, MA, 1998.
8. C. Kotropoulos, A. Tefas, and I. Pitas. Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions, *Pattern Recognition* **33**, 31–43, 2000.
9. F. Schipp, W.R. Wade, P. Simon and J. Pál. *Walsh Series: An Introduction to Dyadic Harmonic Analysis*. Adam Hilger, Budapest, 1990.
10. V.N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
11. V.N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, New York, 1998.



Névmutató

- Alexin Zoltán, 132
Almási László, 132
Aszódi Barnabás, 31
- Balogh Emese, 109, 239, 254
Berke József, 88, 184
Berkéné Várbíró Beáta, 184
Buciu, Ioan, 279
- Chetverikov, Dmitry, 1
Csepeli Zsolt, 175
Csernay László, 132
Csomai István, 118
Czuczor Szabolcs, 31
Czúni László, 218
- Del Lungo, Alberto, 239
Dudásné Nagy Marianna, 123, 252
- Farkas Péter, 218
Fazekas Attila, 272, 279
Fichtinger Gábor, 255
- Gábor György, 201
Gácsi Zoltán, 153, 162, 175
- Hajdu András, 140, 260, 272
Hanis Attila, 218
Hegedűs Károly, 88
- Jankó Zsolt, 10
Jocha Dávid, 98
- Kári Béla, 123
Katona Endre, 192
Kelemen Dezső, 184, 201
Kiss Zoltán, 253
Kolozsár József, 98
Kotropoulos, Constantine, 279
Kovács Levente, 210
Kránicz Balázs, 218
Kuba Attila, 109, 132, 152, 239, 251–255
- Licsár Attila, 218, 231
Máté Eörs, 123
Megyesi Zoltán, 21
- Nagy Antal, 132, 251
Nagy Benedek, 260
Nyúl László, 96, 109, 132
- Ollé Krisztián, 152
Orzó László, 75
- Palágyi Kálmán, 42, 109
Pitas, Ioannis, 279
Póliska Csaba, 153
- Rodek Lajos, 253
Roska Tamás, 75
Ruskó László, 254
Réger Mihály, 153
- Schanda János, 218
Sisák István, 184
Sorantin, Erich, 109
Stepanov, Dmitry, 1
Szabó Gábor, 118
Szabó József, 118, 184, 201
Szeiler Gábor, 184
Szirányi Tamás, 62, 202, 210, 218, 231
Sárközi Gábor, 162
- Tóth Zoltán, 62
Tanács Attila, 255
Tizedes László, 202
Tőkés Szabolcs, 75
- Udupa, Jayaram K., 96
- Varga Endre, 152
Veréb Krisztián, 52
Zörgő Zoltán, 140



