

BUDAPESTI MŰSZAKI EGYETEM
MÉRNÖKTOVÁBBKÉPZŐ INTÉZET

Álló Géza — Főglein János
Hegedűs Gy. Csaba — Szabó József

BEVEZETÉS A SZÁMÍTÓGÉPES KÉPFELDOLGOZÁSBA



BUDAPEST
1993



BUDAPESTI MŰSZAKI EGYETEM
MÉRNÖKTOVÁBBKÉPZŐ INTÉZET

Álló Géza — Főglein János
Hegedűs Gy. Csaba — Szabó József

BEVEZETÉS A SZÁMÍTÓGÉPES KÉPFELDOLGOZÁSBA

KÉZIRAT
Javított kiadása

BUDAPEST
1993

A kötet szerzői:

dr. Álló Géza
okl.villamosmérnök, okl.matematika-fizika tanár
11. és 23. fejezet; 13.2. és 21.1. alfejezet

dr. Főglein János
okl.matematikus
12.1.2. alpont; 13.3 alfejezet; 31. és 32. fejezet

dr. Hegedűs Gy. Csaba
okl.fizikus
13.1. és 13.4. alfejezet; 22. fejezet; 4. rész

Szabó József
okl.matematika-fizika tanár
12. és 21. fejezet
(kivéve: 12.1.2. pont és 21.1. alfejezet)
A 31. fejezetet átdolgozta:
Kelemen Dezső okl. matematikus

Szerkesztette és lektorálta:
dr. Álló Géza

A kéziratot ellenőrizte:
Dr. Loványi István
Dr. Vajta László

ISSN 865-3313
ISBN 963 431 777 4

Megjelent
a Budapesti Műszaki Egyetem Mérnöktovábbképző Intézete
fennállásának 50 éves jubileumi tanévében

A kiadásért felelős:
a Budapesti Műszaki Egyetem Mérnöktovábbképző Intézetének igazgatója
Megrendelve: 1993. július. Megjelent: 1993. november. Példányszám: 618
Készült: kisofszet eljárással, az MSZ 5601-59
és az MSZ 5602-55 szabvány szerint

Készült a Dabas-Jegyzet Kft-ben
Felelős vezető: Marosi György ügyvezető igazgató
Munkaszám: 93-0525

TARTALOMJEGYZÉK

1. ALAPISMERETEK	11
11. BEVEZETÉS.....	11
11.1. A vizuális információ gépi feldolgozása...	11
11.1.1. Történeti áttekintés.....	11
11.1.2. Képfeldolgozási területek.....	12
11.2. A látáselmélet alapjai.....	14
11.2.1. Az emberi látórendszer.....	14
11.2.2. A fekete-fehér látás.....	15
11.2.3. A színes látás.....	17
11.2.4. Alak és textúra.....	19
12. A KÉPFELDOLGOZÁS ESZKÖZEI.....	21
12.1. Hardware eszközök.....	22
12.1.1. Képfeldolgozó rendszerek.....	22
12.1.1.1. A képmű.....	23
12.1.1.2. Képkezelő és multimédia kártyák.....	25
12.1.1.3. Speciális háttértárak...	28
12.1.2. Párhuzamos architektúrájú számítógépek.....	29
12.1.2.1. A futószalag-szervezésű számítógép.....	30
12.1.2.2. Homogén, hierarchikus rendszerek.....	30
12.1.2.3. Inhomogén rendszerek....	31
12.1.3. Képfelvető eszközök.....	31
12.1.3.1. Kamerák.....	32
12.1.3.2. A fotódigitalizáló.....	34
12.1.3.3. Űrfelvételek készítése..	34
12.1.4. Megjelenítő és képrögzítő eszközök	35
12.1.4.1. Elektronikus képmegjelenítők.....	36
12.1.4.2. Képrögzítők.....	37
12.1.5. Interaktív eszközök.....	39
12.2. Software eszközök.....	42
13. MATEMATIKAI HÁTTÉR.....	45
13.1. Mátrixalgebra.....	45
13.1.1. A mátrix fogalma és főbb típusai..	45
13.1.2. Mátrixok összeadása, kivonása és szorzása.....	46
13.1.3. Osztás négyzetes mátrixszal.....	48

13.2.	Függvénykalkulus.....	49
13.2.1.	A Dirac-delta.....	50
13.2.2.	Konvolúció.....	51
13.2.3.	Differenciál-operátorok.....	54
13.2.4.	Függvénytranszformációk.....	56
13.2.4.1.	Ortogonalis függvény- rendszerek.....	56
13.2.4.2.	Fourier-transzformáció..	57
13.2.4.3.	Hadamard-transzformáció.	60
13.3.	Valószínűségszámítás	62
13.3.1.	Események és eseménytér	62
13.3.2.	Valószínűségek és alaptételek	65
13.3.3.	Valószínűségi változók	68
13.3.4.	Statisztikai jellemzők	72
13.4.	Diszkrét geometria.....	74
13.4.1.	Alapfogalmak és definíciók.....	74
13.4.2.	Topológiai jellemzők.....	74
13.4.3.	Geometriai jellemzők.....	76
13.4.4.	Diszkrét egyenesek.....	79
2.	<u>ELŐFELDOLGOZÁS</u>	80
21.	KÉPKORREKCIÓS ELJÁRÁSOK.....	80
21.1.	Digitalizálás és visszaállítás.....	81
21.1.1.	Mintavételezés.....	82
21.1.2.	Kvantálás.....	85
21.1.3.	Visszaállítás (rekonstrukció).....	87
21.2.	Képjavítás	90
21.2.1.	Világosságkód-transzformációk.....	91
21.2.1.1.	A hisztogram.....	91
21.2.1.2.	Skálázások.....	92
21.2.1.3.	Képvágások.....	96
21.2.1.4.	Hisztogram-transzformá- ciók.....	98
21.2.2.	Élkiemelés (élesítés).....	102
21.2.3.	Zajelnyomás (simítás).....	103
21.2.4.	Többsávós képek feldolgozása.....	105
21.2.4.1.	Színkorrekciók.....	106
21.2.4.2.	Pontonkénti műveletek... ..	107
21.2.5.	Képjavítások frekvenciatartományban... ..	108
21.2.5.1.	Zajszűrés.....	109
21.2.5.2.	Élkiemelés.....	110
21.2.5.3.	Kontrasztjavítás.....	112
21.3.	Helyreállítás (restauráció).....	112
21.3.1.	Elvi és gyakorlati nehézségek.....	112
21.3.2.	Inverz szűrés.....	113
22.	GEOMETRIAI KORREKCIÓK.....	115
22.1.	Lineáris geometriai korrekciók.....	115
22.1.1.	Elemi koordinátatranszformációk... ..	116
22.1.2.	Homogén koordináták.....	118
22.1.3.	Eltolás és perspektív transzfor- máció.....	119
22.1.4.	Gyakorlati megvalósítás.....	121

22.2.	Nemlineáris transzformációk.....	122
22.2.1.	Felületelemes közelítés.....	123
22.2.2.	Felületelem transzformációja.....	124
22.3.	A geometriai korrekció definiálása.....	126
22.3.1.	Elméleti módszerek.....	126
22.3.2.	Kísérleti módszerek.....	126
23.	SZEGMENTÁLÁS.....	130
23.1.	Foltelemzés.....	132
23.1.1.	Foltok kijelölése.....	132
23.1.1.1.	Globális módszerek.....	132
23.1.1.2.	Lokális módszerek.....	134
23.1.2.	Foltvizsgálat.....	135
23.1.2.1.	Hámozás (erózió).....	136
23.1.2.2.	Mintaillesztés.....	136
23.2.	Élkitűzés.....	138
23.2.1.	Él(gyanús) pontok kijelölése.....	139
23.2.2.	Élkorrekció.....	140
3.	<u>KÉPOSZTÁLYOZÁS</u>	144
31.	STATISZTIKUS ALAKFELISMERÉS.....	144
31.1.	A döntési folyamat előkészítése.....	145
31.1.1.	A sajátságvektorok előállítása....	146
31.1.1.1.	Mérhető jellemzők keresése.....	146
31.1.1.2.	Többszörös felvételek jellemzése.....	148
31.1.1.3.	Számítástechnikai megfontolások.....	148
31.1.2.	Lényegtömörítés.....	149
31.1.2.1.	Karhunen-Loève transzformáció.....	150
31.1.2.2.	Tömörítés célfüggvény alapján.....	152
31.1.3.	A döntési feladat megfogalmazása..	153
31.2.	Bayes-döntésemélet.....	156
31.2.1.	A Bayes-féle döntési eljárás.....	157
31.2.2.	Speciális esetek.....	159
31.2.3.	Becslési módszerek.....	166
31.3.	Távolságmérésen alapuló módszerek.....	168
31.3.1.	Legközelebbi szomszéd módszerek...	169
31.3.1.1.	Gyors LKS-kereső algoritmusok.....	170
31.3.2.	A téglatest-módszer.....	172
31.3.3.	Lineáris diszkriminánsfüggvények..	173
31.3.3.1.	Döntés diszkriminánsfüggvények alapján.....	173
31.3.3.2.	Lineáris szétválaszthatóság.....	174
31.3.3.3.	Lineáris diszkriminánsfüggvények előállítása..	175
31.3.4.	Egyéb módszerek.....	177

31.4.	Klaszterezés.....	178
31.4.1.	Agglomerációs (hierarchikus) mód- szerek.....	178
31.4.2.	Klaszterezés célfüggvény alapján..	179
31.4.3.	Iteratív módszerek	181
31.4.3.1.	A négyzetes hiba mini- malizálása.....	181
31.4.3.2.	Az ISODATA módszer.....	182
31.5.	Környezetfüggő módszerek.....	185
32.	SZINTAKTIKUS ALAKFELISMERÉS.....	187
4.	<u>KÉPTÖMÖRÍTÉS</u>	191
41.	ALAPFOGALMAK.....	191
41.1.	Redundanciák.....	192
41.1.1.	Kódolási redundancia.....	193
41.1.2.	Képi redundancia.....	193
41.1.3.	Pszichovizuális redundancia.....	193
42.	KÉPTÖMÖRÍTÉSI ELJÁRÁSOK.....	194
42.1.	Veszteségmentes tömörítés.....	194
42.1.1.	Változó hosszúságú kódok.....	195
42.1.1.1.	Huffman kódolás.....	195
42.1.1.2.	Aritmetikai kódolás.....	196
42.1.2.	Bitsík-kódolás.....	196
42.1.3.	Homogén foltok kódolása.....	197
42.1.4.	Határvonal-kódolás.....	197
42.1.5.	Kódolás előrebecsléssel.....	198
42.2.	Veszteséges tömörítés.....	199
42.2.1.	Tömörítés előrebecsléssel.....	200
42.2.2.	Transzformációs kódolás.....	200
42.2.3.	A JPEG szabvány.....	200
42.2.4.	Fraktál-tömörítés.....	204
	Zárszó.....	205
	Irodalomjegyzék.....	206

ELŐSZÓ

A számítógépes képfeldolgozás gyakorlati jelentősége ugrásszerűen megnőtt az utóbbi 15 évben, és néhány éve már hazánkban is megjelentek az első alkalmazások. Ezek egyre fontosabb szerepet játszanak a tudományos, úrkutatási, orvosi-biológiai, ipari, mezőgazdasági, államigazgatási, oktatási stb. feladatok megoldásában. Ilyen alkalmazások pl. a földfelszín automatikus kiértékelése és erőforrások kutatása az űrfelvételek alapján; (elektron)mikroszkópi, termovíziós, radioaktív sugárzási felvételek elemzése; automatikus térképkészítés légifelvételekről; ipari anyagvizsgálat röntgenfelvételek alapján; automatikus közlekedésirányítás; művészeti alkotások és játékok létrehozása stb. Nincs messze az az idő, amikor a képfeldolgozó rendszerek aktív részt kapnak a betegségek felismerésében és a diagnózis elkészítésében; "látó" robotok veszik át a munkát a nagy megterhelést jelentő, veszélyes (pl. vízalatti) és rutinszerű munkát igénylő területeken; - a felsorolásnak csak a fantázia szab határt.

Nem kétséges, hogy az alkalmazások fokozatosan elterjednek majd a hétköznapi életben is, hiszen az 5. generációs számítógép-rendszerek egyik legfontosabb célkitűzése a vizuális ember-gép kapcsolat emberközeli megvalósítása, az ember látóképességének minél tökéletesebb gépi realizálása.

A jegyzetet elsősorban azoknak a szakembereknek szántuk, akik már felfigyeltek erre az új alkalmazási területre. Az olvasóról feltételezzük, hogy rendelkezik bizonyos felsőbb matematikai, programtervezési és/vagy programozási ismeretekkel; de a képfeldolgozással mélyebben nem foglalkozott.

A rendszerező munkák és a magyar nyelvű szakirodalom szinte teljes hiánya miatt az 1. részben áttekintjük a képfeldolgozó rendszerek alapelveit és a számítógépes eszközöket, valamint röviden a látáselméletet és a szükséges fontosabb matematikai ismereteket.

A 2. részben foglalkozunk a digitális számítógépes képfeldolgozás témaköreivel; ezek között szerepel a digitális kép létrehozása és javítása, transzformálása, szegmentálása és elemzése.

A 3. részben statisztikus és szintaktikus alakfelismerési eljárásokat mutatunk be, amelyek célja a képi információ tömör leírása, a kép felismerésének, "megértésének" előkészítése.

A jegyzet bevezető jellege és erősen korlátozott terjedelme miatt nem tudjuk részletesebben tárgyalni az egyes témaköröket; az érdeklődők számára irodalomjegyzéket állítottunk össze.

Megjegyezzük még, hogy bár kifejezetten törekedtünk a magyar nyelvű szakkifejezések használatára, megadjuk az angol nyelvűeket is, hogy segítsük a tájékozódást az irodalomban. Kifejezéseinket javaslatnak tekintjük, amelyek helyett a gyakorlatban esetleg jobbak vagy találóbbak fognak elterjedni.

Budapest, 1984. október

A szerzők

ELŐSZÓ A 2. KIADÁSHOZ

Amikor közel 10 évvel ezelőtt leadtuk az 1. kiadás kéziratát, azt jósltuk, hogy a számítógépes - mai szóhasználattal inkább: digitális - képfeldolgozás széleskörű elterjedése várható minden olyan területen, ahol képi információkat kell feldolgozni. Ezt a várakozásunkat még felül is múlták a néhány éve megjelent multimédia eszközök nyújtotta lehetőségek. A korábbi, több kártyából álló, bonyolult, önálló rendszerként megjelenő, drága képművek élesen elváltak az analóg, illetve a digitális képek kezelését és egyidejűleg 1-2 fekete-fehér képet (8 bit/képpont) tudtak tárolni. Az integrált áramköri technika nem is sejtett fejlődése révén az egy kártyára zsugorodott mai képmű legalább egy - sokszor 2-8 - valódiszines képet tárol (15-24 bit/képpont), miközben néhány elemi képműveletet (részletkivágás, ráközelítés (zoom), ablakmozgatás, átszínezés stb.) hardverben végez, továbbá együtt kezeli az analóg és a digitális képjelet a grafikus kártya képjelével (sőt hangjelet is tud generálni, ettől multimédia). Így jöhettek létre az egy (multiszinkronos) monitorral működő, személyi számítógépes képfeldolgozó rendszerek, amelyek akár többféle felbontásban is meg tudják jeleníteni a digitális képet és árban is elérhetőek minden érdeklődő felhasználó számára.

Jóslatunk beteljesedésén érzett büszkeségünket azonban nagymértékben lelohasztja az a tény, hogy a digitális képfeldolgozás elmélete azóta sem alakult ki. Bár Hubel és Wiesel alapvető, Nobel-díjjal is jutalmazott munkássága (ld. pl. [1-9]) jelentős lökést adott a kutatásoknak, a tudomány még ma sem ismeri pontosan a látás folyamatait. Az mindenesetre nyilvánvalóvá vált, hogy csak párhuzamos architektúrájú rendszerekkel lehet kielégítő eredményeket elérni. Noha a transzputerek és a sejtprocesszorok megjelenése ezen a téren áttörést hozott és az utóbbi években kialakuló neurális (az emberi idegrendszer analógiájára működő) számítórendszerek is ígéretes eredményeket produkáltak, a mai eszközeink még mindig 3-5 nagyságrenddel elmaradnak az emberi agy teljesítményétől mind tárolókapacitásban, mind működési sebességben, mind pedig - és talán ez a leglényegesebb - az aktív elemeik közötti kapcsolataik ("szinapszisaik") számában. Így aztán az univerzális látórobot egyelőre még mindig csak a tudományos-fantasztikus (sci-fi) irodalomban létezik.

Maradt hát kutatási terület bőven a most felnövekvő kutatónemzedékeknek is; ezért döntött úgy a kiadó, hogy újból megjelenteti ezt a jegyzetet. Az anyag ugyan lényegében változatlan, de számos helyen pontosítottuk és a lehetőségek szerint ki is egészítettük. Alaposabban a 12., a 13., a 21., a 23. és a 31. fejezetet kellett átdolgozni; az újonnan beiktatott 4. részben pedig képtömörítéssel foglalkozunk. Ez az utóbbi témakör ugyan szorosan véve nem képfeldolgozás, a sok (valódiszínes) képpel dolgozó és/vagy képátvitelt is megvalósító rendszerekben azonban nélkülözhetetlen technika.

Reméljük, hogy munkánk továbbra is hasznos segédeszköz lesz a témával ismerkedőknek.

Budapest, 1993. június

A szerzők

1. ALAPISMERETEK

Ebben a részben röviden - inkább csak utalásszerűen - összefoglaljuk azokat az ismereteket, amelyek - tapasztalataink szerint - elengedhetetlenül szükségesek ahhoz, hogy a képfeldolgozási feladatokat számítógép segítségével meg tudjuk oldani.

A 11. fejezetben néhány elvi kérdést érintünk. A fő gondot az okozza, hogy ez idő szerint a számítógépes képfeldolgozásnak nincs egységes elmélete. A felmerülő problémákra az esetek túlnyomó többségében eseti megoldásokat dolgoztak ki, és csak a legutóbbi időkben kezdtek az általánosításra törekedni. Emiatt egy-egy feladatra sokszor több megoldás is létezik, amelyek különböző szempontokból előnyösek; a nehézség gyakran abban van, hogy kiválasszuk a számunkra legjobbat. Mivel az eredményeket végső soron legtöbbször az ember értékeli ki, itt foglalkozunk röviden az emberi látási folyamattal is.

A 12. fejezetben áttekintjük a képfeldolgozás speciális hardware eszközeit.

A 13. fejezetben vázoljuk a legfontosabb matematikai ismereteket. A gyakorlati munkához szükséges további részletek megtalálhatók az igen jó, magyar nyelvű matematikai szakirodalomban.

11. Bevezetés

11.1. A VIZUÁLIS INFORMÁCIÓ GÉPI FELDOLGOZÁSA

11.1.1. Történeti áttekintés

Az 50-es évek második felében jelentek meg az irodalomban az első cikkek, amelyek a képi információk feldolgozásával foglalkoztak. Azóta a számítógépes képfeldolgozás témaköreinek skálája jelentősen kibővült és a 80-as évek elejére önálló alkalmazási területté fejlődött. Egyes becslések szerint az idevágó cikkek száma eléri, esetleg meg is haladja az évi ezret.

A fejlődési folyamat természetéből következően először az optikai és az analóg jelfeldolgozás (főleg az optikai képfeldolgozás) módszereit adaptálták. Tömegesen születtek az

egyes konkrét feladatokra kidolgozott heurisztikus módszerek, amelyeknek érvényességi feltételeit és konvergenciáját sokszor nem is vizsgálták. A 60-as évek végére már elég sok tapasztalat gyűlt össze, és elég nagy volt a káosz ahhoz, hogy megindulhasson a matematikai módszerek intenzív kutatása.

A 70-es évektől megjelenő cikkekre a matematikai meg-alapozottság igénye a jellemző. A "legdivatosabb" módszer a Fourier-transzformáció (majd ezt követően egy sereg más, egyszerűbben programozható és/vagy "olcsóbban" futtatható transzformációs eljárás) és a klaszterezés lett. Ahogyan azonban nőttek a feldolgozással szemben támasztott követelmények, nyilvánvalóvá vált, hogy az egzakt algoritmusokat nem lehet elfogadható költségszinten és elviselhető válasz-idők mellett realizálni a rendelkezésre álló hardware eszközökkel.

Az évtized második felétől hatalmas erővel megindult a speciális célú hardware eszközök fejlesztése (nagyfelbontású képfelvető, tároló- és megjelenítő eszközök; nagy sebességű (párhuzamos) feldolgozó egységek; a képfeldolgozás alap-feladatait megvalósító LSI elemek stb.). Ezzel egyidejűleg megjelentek az első univerzális célú rendszerek, amelyek az emberi látórendszert kívánták többé-kevésbé megközelíteni. Pillanatok alatt kiderült azonban, hogy ez olyan bonyolult feladat, amelyet "próbálgatással" nem lehet megoldani.

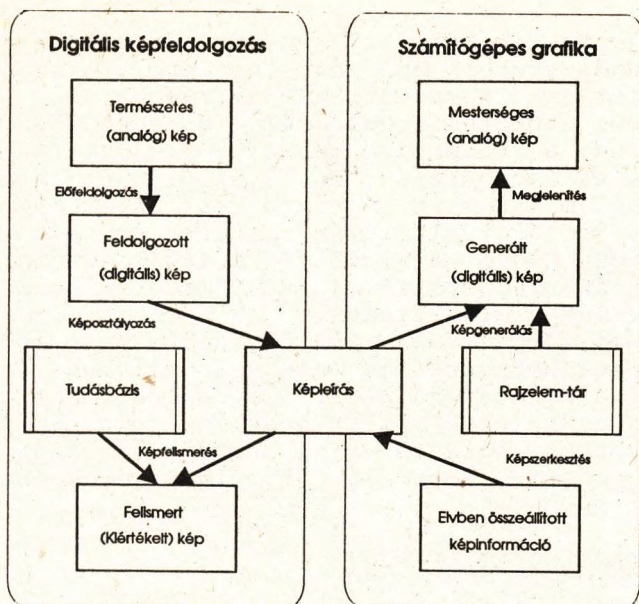
Ez a felismerés indította el az elméleti alapok kidolgozására irányuló nagyszabású kutatómunkát, amelyet az idő-közben társadalmi szükségletté lett számítógépes képfeldolgozás iránti egyre magasabb igények is ösztönöznek. A munka, amelynek az 5. generációs távlati terv további lendületet adott, napjainkban is folyik; ez idő szerint azonban az egy-séges elmélet még nem született meg.

11.1.2. Képfeldolgozási területek

A számítógéppel megoldott képfeldolgozási feladatok kö-re túlságosan szerteágazó, együttes tárgyalásuk áttekinthetetlen lenne. Ezért a jegyzetben nem foglalkozunk sem optikai, sem egyéb analóg módszerekkel. Nem tárgyaljuk továbbá a térbeli (pl. sztereo vagy kislátószögű (range)) felvételek feldolgozását, köztük a projektív geometriai módszereket sem.

Számítógépes képfeldolgozáson kizárólag kétdimenziós (síkbeli), digitális képek (ld. 21.1. pontot) gépi feldolgozását értjük. Ennek két, bizonyos mértékig átfedő, de cél-kitűzéseiben és módszereiben lényegesen különböző szakterüle-te van (ld. 1-1. ábra).

1. A számítógépes grafika szűkebb értelemben egy vonal-as kép előállítását (generálását) jelenti a kép leírásából; tágabb értelemben ide soroljuk a feliratok generálását, va-lamint mindazokat az eljárásokat is, amelyek megkönnyítik a



1-1. ábra
A számítógépes képfeldolgozás területei

képleírás elkészítését. (Leírás \rightarrow kép leképezések. A vonalas kép egyenes és/vagy görbe vonalakból és ezekkel határolt idomokból áll, amelyek természetesen színesek is lehetnek.)

A képleírást szintaktikus leírásnak lehet tekinteni; ez tulajdonképpen egy grafikus nyelven írt program a kép felrajzolására. Ebből egy rajzelem-tár felhasználásával készül el a kép gépi (digitális) reprezentációja.

A grafikus rendszerek az ember-gép kapcsolatban alapvetően a kimeneti funkciót valósítják meg; emellett azonban felhasználhatók bizonyos vizuális információk bevitelére is.

Ezzel a területtel a továbbiakban nem foglalkozunk.

2. A digitális képfeldolgozás a természetes képeken (természetes tárgyak optikai vagy elektronoptikai úton készített síkbeli leképezéseiben) foglalt képi információ megértésére irányul. A folyamatot - a rész-célkitűzések szerint - 3 szintre bonthatjuk:

a) Előfeldolgozáson a digitális kép létrehozását, valamint a leképezési hibák kijavítását, a jellemző képtulajdonságok kiemelését és a képnek a további feldolgozásra kedvezőbb, adekvát átalakításait értjük (kép \rightarrow kép leképezések).

b) A képosztályozás során megkeressük a képen előforduló alakzatokat, ill. - általános értelemben - a képet jellemző különféle (nem szükségképpen geometriai) sajátságokat (features). Az ezekhez mindig rendelhető numerikus értékekkel vagy a köztük levő relációkkal leírjuk, majd osztályozzuk a képet, ill. egyes részeit (kép \rightarrow leírás leképezések).

Fontos megjegyezni, hogy ez a leírás lényegesen eltérhet a grafikus rendszerekben említett leírástól, hiszen a saját-ságok sokszor nem is geometriai jellegűek!

c) Végül a képfelismerés célja a képen rögzített valós világ leírása; a (többnyire térbeli, 3 dimenziós) tárgyak felismerése és azonosítása, egy - tanulás útján bővülő - tudásbázisból már ismert alakzatokkal; ennek alapján valamilyen feldolgozási célkitűzés teljesítése (leírás → értelmezés leképezések). A végső cél a látás automatizálása: a vizuális információt közvetlenül "megértő" ezzel vezérelt robotok/automaták előállítása.

A látási folyamatnak ezen egyszerűsített modell szerinti megközelítése sokféle és egymással bonyolult kölcsönhatásban levő módszer és eljárás megvalósítását igényli; ezért az egyes szintek között nem lehet éles határokat vonni. A jegyzetben mégis eszerint csoportosítottuk az anyagot, a könnyebb megértés és a jobb áttekintés érdekében.

11.2. A LÁTÁSELMÉLET ALAPJAI

Az ember és a környezete közötti kapcsolatban a vizuális információ a legfontosabb, minden mást felülmúló tömörsége miatt. Jellemzői a szín, az alak, a mélység (térbeliség) és a mozgás. Ezekkel bővebben nem tudunk foglalkozni, de az alábbiakban röviden áttekintjük az emberi látórendszer azon jellemzőit, amelyeket figyelembe kell venni a képfeldolgozás során.

11.2.1. Az emberi látórendszer

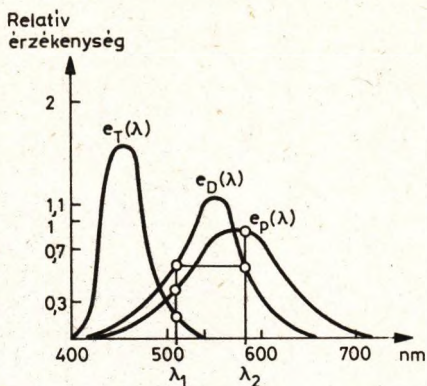
A szem az emberi látórendszer "bemeneti készüléke". Az elektromágneses sugárzásból kicsit több mint 1 oktávnyit: a 380 nm - 780 nm hullámhossztartományt, vagyis a vizuális információt hordozó analóg fényjelet érzékeli. Mint ismeretes, a szemlencse az éles képet a recehártyán (retina) levő látógödörben (fovea) állítja elő, a pupillán keresztül belépő fénysugarakból. (A kép legnagyobb átmérője végtelen távoli tárgy esetén kb. 0,3 mm.) Kétféle érzékelő (receptor) létezik:

- A pálcikák a látógödör középső tartományán kívül a recehártyában helyezkednek el; feladatuk a fényerősség-változások érzékelése. Fontos szerepük van a gyenge megvilágítás melletti látásban és a fekete-fehér képek elemzésében. Számuk mindkét szemben kb. 100-120 millió.

- A csapok a látógödör közepét foglalják el; számuk mindkét szemben kb. 6,5 millió. A színlátásban és az erős megvilágítás melletti látásban játszanak szerepet. P (protonopen), D (deu-

teronopen) és T (trionopen) típusú csapok léteznek, ezek azonos erősségű (világosságú), különböző hullámhosszú (színű) fényhatásokra az 1-2. ábrán látható ingerülettel válaszolnak. A színes látás a 3 különböző csaptípus együttműködésének eredménye.

A fény által a receptorokban kiváltott ingerületek hatására a rezechártyát átitató bíborfolyadékban fotokémiai reakciók mennek végbe. Ezek hatását érzékelik a rezechártyát beborító ideghártyában levő látóidegek, amelyek az így keletkezett ingereket - az oldalsó térdes-testeken keresztül - az agyi látómezőkbe vezetik. Itt dolgozzuk fel a kapott információkat, a fentebb említett 4 jellemző szerint, korábbi vizuális ismereteink ("adat-" és "tudásbázisunk") felhasználásával.



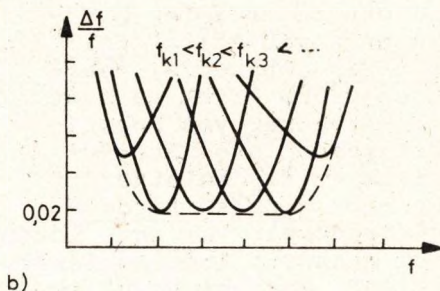
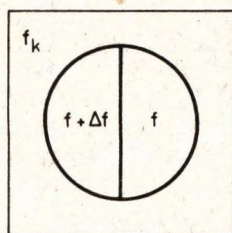
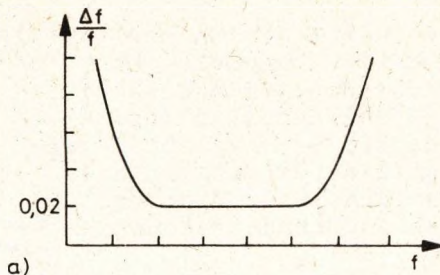
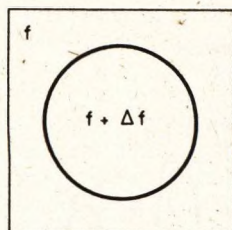
1-2. ábra
A csapok érzékenységi görbéi

11.2.2. A fekete-fehér látás

A fekete-fehér (a továbbiakban: ff) képek legfontosabb jellemző sajátysága a különböző világosságú képelemek fényerőssége közötti különbség, a kontraszt. Szemünk kontraszt-érzékenysége logaritmikus, ezért nem abszolút fényerősségkülönbségeket érzékelünk, hanem a változásoknak az eredeti megvilágításhoz viszonyított arányát. Pl.: az f_1 , f_2 , ill f_3 fényerősségű képelemek között a szem akkor találja azonosnak a kontrasztot, ha teljesül az

$$\frac{f_2 - f_1}{f_1} = \frac{f_3 - f_2}{f_2}$$

feltétel. A legkisebb, még észlelhető Δf fényerősségkülönbséget kontrasztküszöbnek nevezzük. A mondottakból következik, hogy ez függ a fényerősségtől, azonban a $\Delta f/f$ kontrasztküszöb-arány (= Weber-féle tört) elég széles fényerősségtartományban állandó, és értéke 0,02 körül van (ld. 1-3/a. ábra). Lényegesen más a helyzet, ha a különböző fényerősségű képelemek valamilyen más fényerősségű háttérbe vannak ágyazva; erre az 1-3/b. ábrán látunk kvalitatív példát. Ezzel magyarázható az a jelenség, hogy azonos világosságú képrészleteket sötétebb háttérben világosabbnak látunk, mint világosabb háttérben.



b)

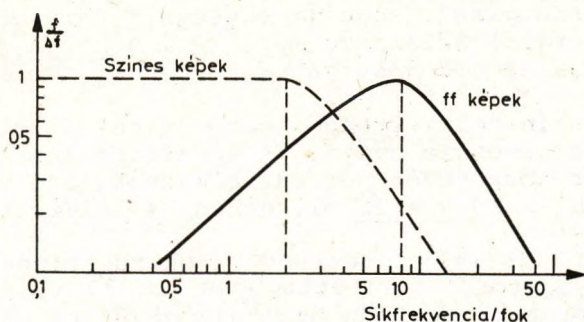
1-3. ábra

A kontraszküszöb és a háttérmegvilágítás összefüggése

A kontrasztérzékelés függ még a síkfrekvenciától (spatial frequency) is. (A síkfrekvencia a kép síkjának egy tetszés szerinti irányában észlelhető periodikus fényerősség-változásokat jellemzi; megkülönböztetendő a frekvenciának az időbeli változásokra vonatkozó hagyományos és általunk is megtartott értelmezésétől.) Ha pl. a fényerősség a kép vízszintes oldala mentén periodikusan változik úgy, hogy a legvilágosabb és a legsötétebb helyek között az átmenet folyamatos, és ezek távolsága pl. balról jobbra haladva csökken (vagyis a síkfrekvencia nő); egyidejűleg pedig függőleges irányban haladva csökken közöttük a fényerősségkülönbség (vagyis csökken a kontraszt); akkor azt tapasztaljuk, hogy mind a kis, mind a nagy síkfrekvenciákon romlik a kontrasztérzékelés, vagyis nő a kontraszküszöb; legkisebb értékét akkor veszi fel, ha egy látószög-fokra 5-10 periódus esik (ld. 1-4. ábra). A szem alacsonyfrekvenciás érzéketlenségén több képjavítási eljárás alapul. Az optimális tartományban - kellő megvilágítás esetén - átlagosan mintegy 500 kontrasztfokozatot tudunk megkülönböztetni.

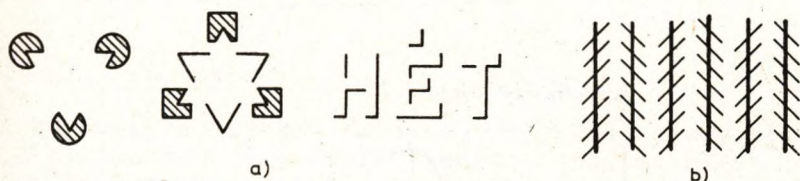
Érdekes jelenség még a hirtelen fényerősség-változásoknál (élek, kontúrvonalak stb. mentén) fellépő "túllövés", vagyis a kontraszküszöb lokális lecsökkenése. Ezzel magyarázhatók pl. az 1-5/a. ábrán látható "kitöltési effektusok".

Ha a síkfrekvencia nő (közeli, különböző fényerősségű helyek), egyszerű vonalas ábrák esetében romlik az irány- és méretérzékelés. (Az 1-5/b. ábrán pl. az azonos vastagságú vonalak párhuzamosak.)



1-4. ábra

A kontrasztküszöb és a síkfrekvencia összefüggése



1-5. ábra

Kontrasztugrások hatásai

11.2.3. A színes látás

Közismert tény, hogy a fehér fény végtelen sok színárnyalat (spektrumszín) keveréke. (Megjegyezzük, hogy a spektrális összetételtől függően különféle "fehér" színek léteznek.) A színes látás azon alapul, hogy a 3-féle csaptípus ingerületeinek eredője minden (megkülönböztethető) színre más. (Egy csaptípust különböző színű fények is gerjeszthetnek azonos módon; pl. az 1-2. ábrán a λ_1 és a λ_2 hullámhosszú fény a D-típusú csapokból azonos ingerületet vált ki; de a másik két csaptípus ingerületei jelentősen eltérnek; így az eredő színérzet is más lesz.)

Az érzékelt színérzetet 3 sajátságára alapján különböztetjük meg egyértelműen:

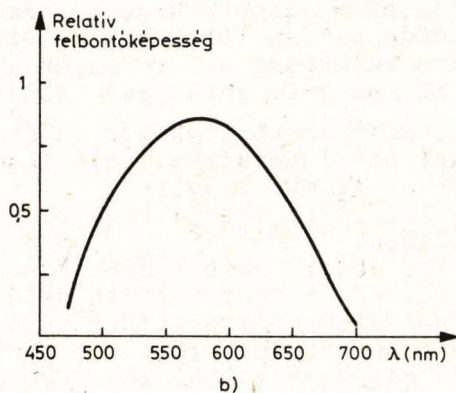
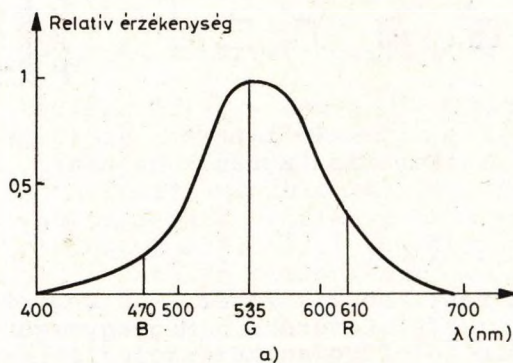
- A színárnyalat (hue), amit a hétköznapi életben egyszerűen színnek neveznek, a fény hullámhosszától függ. Ez azonban önmagában nem elegendő: egyrészt bizonyos színek nem találhatók meg a spektrumban (pl. a bíbor a vörös és a kék szín keveréke); másrészt különböző spektrális eloszlású fényforrások is látszhatnak azonos színűnek.

- A világosságot vagy fényintenzitást (intensity) a fényforrás energiasűrűsége (= egységnyi felületen kisugárzott fényenergia) határozza meg. (Pl. a barna szín spektrális eloszlása azonos a sárgáéval, csak kisebb a világosságértéke.)

- Egy szín telítettsége (saturation) attól függ, hogy mekkora a fehér-összetevője. (A spektrumszínek monokromatikusak, fehér összetevőt nem tartalmaznak, ezért 100%-os telítettségűek. A rózsaszín pl. néhány % telítettségű piros szín.)

A szem a 100%-os telítettségű spektrumszínek közül mintegy 160 árnyalatot, telítettségben kb. 20, világosságban kb. 500 fokozatot tud megkülönböztetni; együttes változásuk esetén azonban összességében mégis csak mintegy 2500-4000 féle színt tudunk felismerni.

A szem spektrális érzékenysége színárnyalatonként és egyénenként is erősen változik; állandó világosságú fényforrások fényességét különbözőnek látjuk az 1-6/a. ábra szerint. Ezenkívül még a szem felbontóképessége is függ a hullám-



1-6. ábra
A szem jelleggörbéi

hossztól az 1-6/b. ábra szerint: a látható tartomány széleihez közeli színekben az észlelt részletgazdagság kisebb. Képjavitási eljárásokban és valódiszínes (színhelyes) képek előállításakor figyelembe kell venni a szem ezen jellegzetességeit.

A mondottakból az is következik, hogy a különböző világosságú spektrumszínek egyesítése más-más "színű" (vagyis spektrális összetételű) fehér fényt eredményez, ennek a színes képek megjelenítésekor (képernyőn, rajzpapíron) van jelentősége: a szem a színárnyalatokat az előállított fehér színhez viszonyítva érzékeli. A fehér fény színhőmérséklete egyenlő annak az abszolút feketetestnek az abszolút hőmérsékletével, amelynek sugárzása azonos spektrális összetételű. A TV-technikában a referenciafehér színhőmérséklete 6500 °K, a 3 alapszín (R = red (vörös), G = green (zöld) és B = blue (kék)) hullámhossza pedig rendre:

$$\lambda_R = 610 \text{ nm}; \quad \lambda_G = 535 \text{ nm}; \quad \lambda_B = 470 \text{ nm}.$$

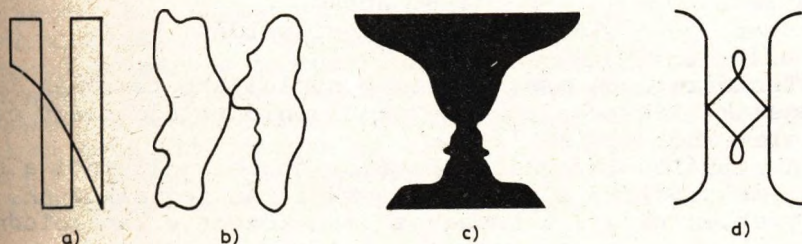
11.2.4. Alak és textúra

A képeket, az ezeken látható tárgyakat az agy nem különböző világosságú vagy színű képpontok halmazaként ismeri fel. A kontraszton, ill. a színen kívül a második alapvetően fontos jellemző a kép makro- és mikroszerkezete.

A makroszerkezetet a képen észlelhető tartományok határozzák meg ("alakfelismerés"). A tartományok körül - legalább részben - éleknek (= hirtelen, ugrásszerű változások a világosságban és/vagy színben) kell lenniük és a belsejüknek valamilyen szempontból (közel) homogénnek és összefüggőnek kell lennie (= foltok).

Anélkül, hogy részletekbe mennénk, megemlítünk néhány jellegzetes alakfelismerési törvényszerűséget:

- elsődleges a "sima" folytonosság; (az 1-7/a. ábrán nem 3 zárt területet, hanem egy négyszöghullámot és egy "sima" görbét látunk);



1-7. ábra
Alakfelismerési példák

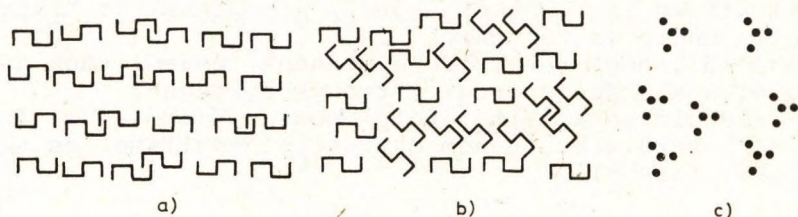
- másodlagos a zárttság (az 1-7/b. ábrán nem egy önma-
gát átmetsző görbét, hanem érintkező zárt területeket lá-
tunk);

- harmadlagos a háttérleválasztás (az 1-7/c. ábrán vagy
egy vázát vagy 2 arcot látunk aszerint, hogy a fehér vagy a
fekete színt tekintjük "háttérnek").

A "szabálysértő" alakzatokat nem, vagy csak nehezen is-
merjük fel. (Pl. az 1-7/d. ábráról el tudjuk képzelni, hogy
egy W és egy M-betűből áll, de nem látjuk annak.)

A kép mikroszerkezetét meghatározó tulajdonságokat tex-
túrának nevezzük. A tartományok nemcsak világosságukat vagy
színüket tekintve lehetnek homogének, hanem textúrájukban is.
A textúraelemek sokszor a szem felbontóképességének határa
közelébe eső méretűek, de a durvább textúrát is felismerjük,
ha az azonos jellemzőkkel bíró textúraelemek száma nagy, és
méretük kicsi ahhoz az alakzathoz képest, amelynek elemei.

A textúraelemek tudatos megfigyelésének szabályai
hasonlók; a nemtudatos megfigyelésben azonban a statiszti-
kai jellemzők játszanak alapvető szerepet: nem különböztet-
jük meg azokat, amelyeknek első és másodrendű statisztiká-
juk azonos. Vannak azonban bizonyos sajátságok - textonok -
amelyek eltérése esetén ez esetben is észleljük a különbsé-
get; ilyen pl. a textúraelemek irányultsága vagy látószöge
(közelsége). Az 1-8/a. és b. ábra egyaránt tükrözött S-ala-
kú textúraelemekből áll, mégis az a. ábrán csak tudatos oda-
figyeléssel lehet észrevenni a textúrákülönbséget, mert a
szárak párhuzamosak. Az 1-8/c. ábrán - közelségük miatt -
inkább a 4, egymáshoz közeli pontot tekintjük textúraelem-
nek, mint magukat a pontokat.



1-8. ábra
Textúrafelismerési példák

Természetesen más, sokkal bonyolultabb textúrák is le-
hetségesek; lényeges, hogy ezek hasonló tulajdonságú és kö-
zeli elemekből épülnek fel.

A vizuális információ további fontos jellemzői a térbe-
liség (mélység) és a mozgás; ezekkel nem foglalkozunk, mivel
a jegyzetben tárgyalt anyaghoz csak közvetve kapcsolódnak.

12. A képfeldolgozás eszközei

Ismeretes, hogy számítógéppel csak digitális képeket lehet feldolgozni. A hétköznapi életben megszokott analóg képekkel (pl. fénykép) szemben a digitális kép képpontokból áll (vagyis csak a sík bizonyos pontjaiban van értelmezve, a közöttük levő helyeken nem). Ezen kívül az egyes képpontok világossága vagy színe sem változik folytonosan, hanem - a feketétől a fehérig - csak bizonyos szürkességi szintek (fekete-fehér (ff) képeken), illetve csak bizonyos színek (színesképeken) fordulhatnak elő.

A jelenleg létező rendszerek ff képek esetén általában 256 (ritkán 1024) féle kvantálási szintet engednek meg. Ezekhez egy-egy számkódot rendelnek hozzá, amelyet a továbbiakban a képpont világosságkódjának nevezünk. Ismeretes, hogy a valódisszines képek 3 monokromatikus (piros, zöld, kék) színösszetevőből állnak, amelyeket külön-külön ff képként digitalizálnak. Ekkor egy képponthoz 3 világosságkód tartozik, amelyeket együttesen színkódnak nevezünk. Ilyen értelemben a világosságkód a ff kép színkódja. A továbbiakban rendszerint az első elnevezést használjuk, mivel a jegyzetben általában ff képek feldolgozásával foglalkozunk. A világosságkód - a kvantálási szintek számától függően - 8 (esetleg 10), a színkód 15-24 biten ábrázolható.

Ílymódon a digitális kép tulajdonképpen egy számmátrix, amelynek annyi sora és soronként annyi eleme van, ahány sorból áll a kép, ill. ahány képpont van egy sorban. A mátrixelemek és a képpontok között kölcsönösen egyértelmű megfeleltetés van; egy adott képpontnak megfelelő mátrixelem értéke az illető pont világosságkódja.

A digitalizálással, ill. a digitális képek megjelenítésével a továbbiakban részletesen foglalkozunk; ehhez a hagyományosaktól eltérő, speciális eszközökre van szükség (ld. a 12.1.2-3-4. pontot.)

Az alábbiakban röviden áttekintjük a fontosabb hardware eszközök működésének, valamint a software kialakításának elveit.

A jegyzet megírása óta közel tíz év telt el, s ezalatt a képfeldolgozási eszközök területén hatalmas fejlődés következett be. Bármennyire elfogultak vagyunk is saját szűkebb szakmánk a képfeldolgozás iránt, el kell ismernünk, hogy az események mozgató rugója a számítástechnika egészét átformáló általános fejlődés volt. Ennek meghatározó eleme a személyi számítógépek (PC) robbanásszerű fejlődése és világméretű elterjedése az elmúlt tíz év során. Egy mai (1993) - már nem is legmodernebb - 386-os PC kapacitása és teljesítménye messze felülmúlja a képfeldolgozás területén

tíz évvel ezelőtt széles körben alkalmazott és tipikusnak mondható PDP 1140-es számítógép lehetőségeit, az ára viszont két nagyságrenddel kisebb. A PC-k megjelenése bővítette az alkalmazási területeket, a megnövekedett sorozatszámok pedig tovább csökkentették az árakat, ami ugyancsak az egyre szélesebb körű felhasználás irányába hatott.

A folyamat a képfeldolgozás vonatkozásában három alkalmazási terület rendkívül gyors fejlődését eredményezte, ezek

- az asztali kiadványszerkesztés (desk top publishing=DTP),
- a karakterfelismerés (optical character recognition=OCR)
- és a vegyes működésű (multimédia) eszközök.

Az első két esetben egy-egy szűkebb alkalmazási terület és az ezeken használt eszközök (szkennerek, színes- és lézernyomtatók, grafikus megjelenítők) fantasztikus méretű elterjedéséről van szó, amit a radikális árcsökkenés tett lehetővé. Várható, hogy a multimédia eszközök teljesen be fognak integrálódni a jövő személyi számítógépeibe úgy, hogy a kép- és a hangkezelés olyan megszokottá válik majd, mint ma az egér használata.

Miután az 1. kiadásban bemutatott eszközök működési elve nem módosult, úgy döntöttünk, hogy a - korábbiakat változatlan formában meghagyva - az egyes alpontok kiegészítésével adunk elvi áttekintést a fejlődésről és az új irányzatokról. A kiegészítéseket a 12.1.1.2., a 12.1.3.1. és a 12.1.4.2. alpont tartalmazza.

12.1. HARDWARE ESZKÖZÖK

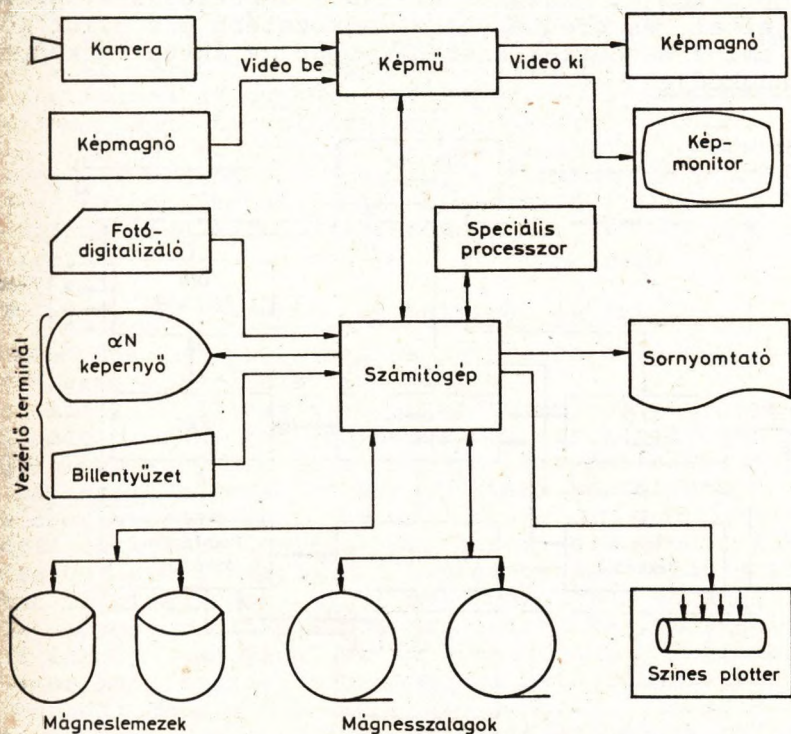
12.1.1. Képfeldolgozó rendszerek

Az 1-9. ábrán általános célú, hagyományos (Neumann-féle) architektúrájú képfeldolgozó rendszer elvi felépítése látható. A felhasznált számítógépek skálája a mikro- (sőt személyi) számítógépektől a mammutméretű szuperszámítógépeken keresztül a speciális architektúrájú sokprocesszoros rendszerekig terjed.

A kiválasztást döntően befolyásolja a felhasználó műszaki-technikai környezete és anyagi ereje. A lehetőségek határain belül figyelembe kell venni a feldolgozandó képek méretét és mennyiségét. Általában előnyösebbek az interaktív beavatkozást lehetővé tevő, többfelhasználós rendszerek. (Ez érthető, hiszen a feldolgozási folyamatban sokszor van szükség szubjektív (vizuális) kiértékelésre; ilyenkor az egyfelhasználós rendszer nincs kihasználva.)

A digitális kép - a felbontás finomságától függően - 0,25 - 1 millió képpontból és képpontonként 8 - 24 bitből áll; így tárolásához 0,25 - 3 Mbyte tárkapacitásra van szükség.

ség. A feldolgozás bonyolultságától függően képpontonként $10-10^3$ műveletet kell elvégezni. Nyilvánvaló, hogy a felmerülő feladatokat - elviselhető futásidők mellett - csak igen nagy tárkapacitású és nagysebességű gépekkel lehet megoldani.



1-9. ábra

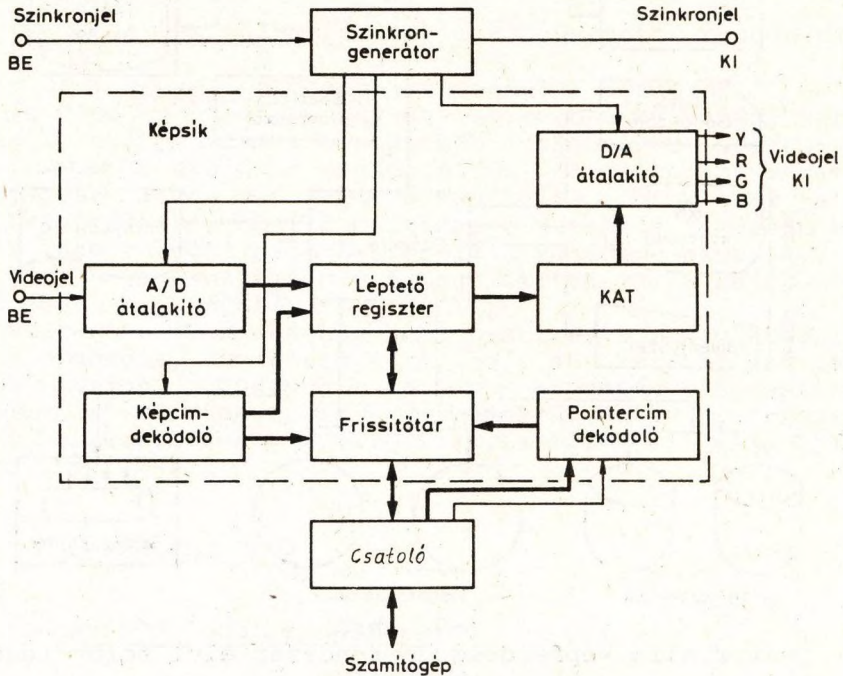
Univerzális képfeldolgozó rendszer elvi felépítése

A Neumann-elvű (soros) számítógépek igazában sohasem optimálisak a képfeldolgozási feladatokra; a hatékonyság a párhuzamos szervezésű gépekkel növelhető (ld. 12.1.2. pont).

12.1.1.1. A képmű

Azoknak a képfeldolgozó rendszereknek, amelyekben analóg képeket dolgoznak fel, ill. az eredményeket vizuálisan is meg kell jeleníteni, alapvető kiegészítő egysége a képmű (display processor). Ez teremt kapcsolatot az analóg működésű (rendszerint TV-technikai) képfelvevő és megjelenítő eszközök, valamint a (digitális működésű) számítógép között.

A képmű elvi felépítését az 1-10. ábrán adjuk meg. A képvétel úgy történik, hogy a bemenetére kapcsolt videojelből (= analóg képjel, ld. 12.1.3.1. alpont) az analóg-digitális (A/D) átalakító az egyes képpontok megjelenésének megfelelő időpillanatokban mintát vesz. A videojel pillanatnyi amplitúdóját összehasonlítja a lehetséges kvantálási szintjeivel és eredményül a legközelebb eső szint kódját adja, ami a mondottak szerint éppen az aktuális képpont világosságkódja.



1-10. ábra
A képmű elvi felépítése

A világosságkódok a frissítőtár megfelelő című bájtjába íródnak, s így létrejön a digitális kép. (A tárcímek és a képpontok geometriai helyzete között kölcsönösen egyértelmű megfeleltetés van.) A mintavételezési és a tárciklus-idő közötti különbséget léptetőregiszter hidalja át. A frissítőtár közvetlen összeköttetésben van a számítógép főtárával, rendszerint cikluslopásos elven működő, gyors csatolókon (DMA = direct memory access) keresztül. A digitális képet így be lehet olvasni a számítógépbe, ill. a feldolgozás eredményét vissza lehet írni a frissítőtárba. Ilyenkor a soron következő képpont címét a pointer-regiszter őrzi.

A képmű digitális-analóg (D/A) átalakítója automatikusan generálja a frissítőtár mindenkori tartalmának megfelelő, a kimeneti átszínező táblával (ld. 12.1.4.1. alpont) átszínezett videojelet, ennek felhasználásával a képmonitoron megjeleníthető pl. az eredménykép. (A D/A átalakító az Y világosságjelen (= ff videojel) kívül a színes (R, G, B) videojeleket is előállíthatja, ha a világosságkódokhoz valamilyen szint rendelünk hozzá. Ekkor - ha a képmonitor erre alkalmas - a ff képet álszínesen is láthatjuk.)

A fentiekben az egyképsíkos képmű elvét ismertettük. Valódiszínes képek feldolgozására (legalább) 3-képsíkos képműre van szükség; ez egyidejűleg digitalizálja a kép 3 színösszetevőjének videojelét (pl. egy színeskamera 3 színjelét). További képsíkokat lehet használni a feldolgozás hatékonyságának növelésére (sebességnövelés, főtárigény csökkentés stb.).

12.1.1.2. Kép(kezelő)- és multimédia kártyák

A képművek felépítésében és működésében az elmúlt évtized két generációváltást is hozott.

Az első 1986-87 táján következett be, amikor az addig külön dobozban, több kártyán megvalósított képműveket sikerült egyetlen, a PC-ben elhelyezhető kártyán felépíteni. Ezek a képkezelő kártyák (frame grabber) közvetlenül a PC sínjére csatlakoznak és a felhasználó a frissítőtárat a PC memóriájának kiterjesztéseként érheti el. (Ipari alkalmazásokra VME sínre csatlakozó képkezelő és -feldolgozó kártyákat is kifejlesztettek.) Bár funkciójuk és a működési elvük azonos az előző alpontban leírtakkal, műszaki paramétereik fokozatosan javultak az évek során. A legismertebbek az amerikai Imaging Technology és Data Translation, valamint a kanadai Matrox cég termékei. A hazai felhasználók körében népszerűek a svájci Leutron Vision cégnek az Szki Pixel Kft. által forgalmazott képkártyái, kedvező áruk és a közvetlen szakmai támogatás miatt.

A képkártyák geometriai felbontása 512x512-től 2048x2048 képpontig terjed. Ez a tárolható kép méretére vonatkozik, a digitalizálás és a megjelenítés nem lépi túl a tv-szabványoknak megfelelő 768x576-os felbontást. A nagy frissítőtár előnye a több kép egyidejű tárolásában és a más forrásból származó nagyméretű képek (pl. úrfelvételek) kényelmes kezelésében jelentkezik.

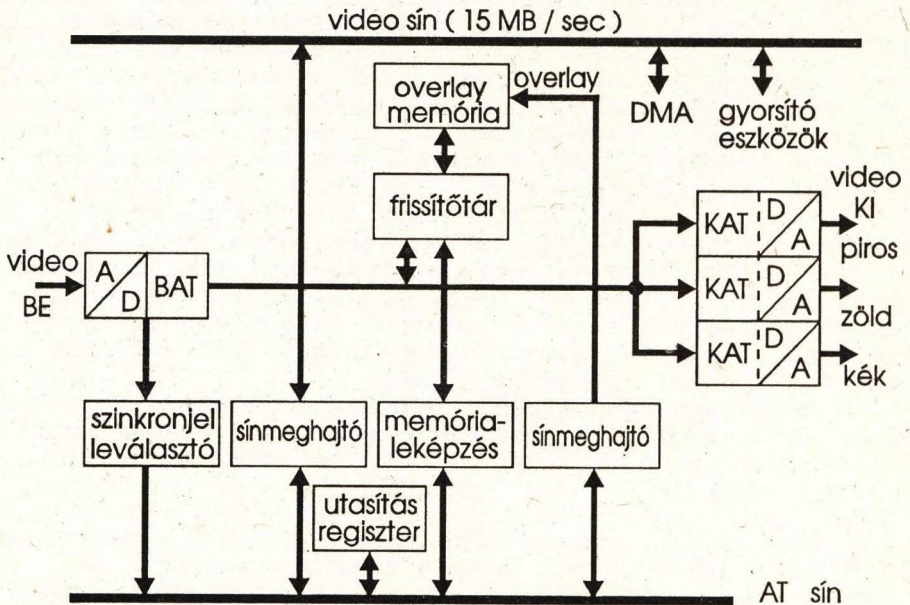
Rendszerint beépített hardver teszi lehetővé a valós idejű nagyítást (zoom), valamint - elég nagy frissítőtár esetén - a (függőleges) görgetést (scroll) és a (vízszintes) úsztatást (pan). Számos esetben néhány (1-4) bites külön grafikus képsíkot is alkalmaznak, miáltal a grafikus információt meg lehet jeleníteni a kép "felett", a képtartalom sérülése nélkül (overlay technika). A gyártók a képkártyák széles skáláját kínálják a 256 szintes monokróm digitalizálóval - és többnyire az álszínes megjelenítéshez szükséges

3 kimenő D/A átalakítóval - valamint be- és kimeneti átszín-
nező táblákkal (BAT, KAT) rendelkező változatoktól a 24 bi-
tes valódiszínű digitalizálást és megjelenítést lehetővé
tevő "luxus" típusokig.

Szemléltető példaként az 1-11. ábrán bemutatjuk a Leu-
ron Vision BFP-AT monokróm képkártyájának elvi rajzát.

A képkártyák után hamarosan megjelentek a képfeldolgo-
zást gyorsító kiegészítő kártyák is; ezeket két nagy csoport-
ba sorolhatjuk:

- az egyikbe a gyakran előforduló képfeldolgozási fel-
adatokra (pl. konvolúció, Fourier-transzformáció, kép-
tömörítés stb.) kifejlesztett gyorsító,
- a másikba a 12.1.2. pontban leírt párhuzamos struk-
túrák valamelyikét korszerű áramkörökkel megvalósító
kártyák tartoznak. Ez utóbbi eszközök előnye, hogy - több-
kevesebb nehézség árán - programozhatók, sőt a legújabb meg-
jelent neurális hálózatok tanulási képességekkel is rendelke-
znek. Az 1-11. ábrán a video sín csatlakozás éppen a különböző
gyorsító eszközökkel való adatcserére szolgál.



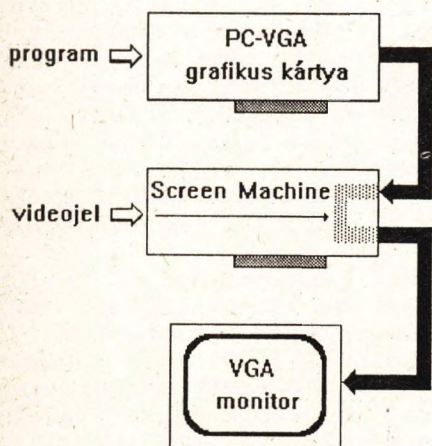
1.11. ábra
A BFP-AT kártya elvi felépítése

A következő váltás a PC-k grafikus kártyáinak paramé-
tereiben bekövetkezett fejlődés eredménye. Amikor a VGA
kártyák felbontása elérte a 640×480 képpontot és a 256 színt,
már várható volt, hogy meg fognak jelenni a hozzájuk csatla-
kozó digitalizáló kártyák is, miáltal megtakarítható lesz a

képmegjelenítéshez szükséges második monitor. Ezt a folyamatot erősítette a multimédia koncepció kialakulása, amelynek egyik sarkalatos pontja a képkezelő eszközök lehetőségeinek - és így az analóg be/kimeneteknek - beépítése a számítógépbe.

Az első - már valóban használható - VGA kártyához kapcsolódó digitalizálót a német FAST cég 1990-ben valósította meg és dobta piacra Screen Machine néven. A kártya lelke egy speciális A/D áramkör, amely a bejövő színes kompozit videojelet digitalizálja és YUV rendszerbe kódolja. A világosságinformációt (Y) 8 bitre bontja, a színinformációt (UV) - gyengébb felbontásban - 4 bitre tömöríti, ezáltal egy színes képpont 12 biten tárolható, az RGB rendszerben szükséges 24 bit helyett. Az információvesztés a színekben levő redundancia miatt a látványképen nem érzékelhető; a képminőség tökéletesen megfelel a multimédia eszközökkel szemben támasztott követelményeknek.

Kimeneti oldalon a kártya RGB jelet állít elő úgy, hogy a kódolt kép megjeleníthető a VGA monitoron. Emellett a VGA csatoló jelét is tudja fogadni, így a két képforrás jele egymásra keverhető. A két kártya kapcsolatát az 1-12. ábra mutatja.



1.12. ábra
Egymonitoros képkártya működési elve

A VGA kártyán előállított grafikus képen bárhol lehet ablakot nyitni, amelyben megjeleníthető az éppen folyamatosan digitalizált, a befagyasztott, vagy a korábban letárolt és visszahívott digitális kép. (A kártyára hanqkezelési lehetőséget is ráépítettek, aminek képfeldolgozási szempontból nincs jelentősége.)

Csak egy évre volt szükség, hogy az amerikai CHIPS cég egy hasonló eszközzel jelenjék meg a piacon; ráadásul - az alapáramkör értékesítésében látva a nagyobb fantáziát - sza-

bad utat nyitott a távolkeleti gyártóknak az erre alapozott olcsó képkártyák előállítására. Ennek köszönhetően a Screen Machine-nál azonos elven működő és egymással tökéletesen azonos felépítésű kártyák elárasztották az egész világot. Magyarországon - az Szki Pixel Kft. jóvoltából - először 1991-ben jelent meg nagyobb számban a tajvani ADDA cég Aver2000 digitálizálója, majd később a szingapúri Creative Technology Ltd. VideoBlaster kártyája.

Bár ezek a kártyák kedvező áruk révén elősegítik a képfeldolgozási alkalmazások elterjedését és számos esetben valóban jól használhatók, fel kell hívnunk a figyelmet arra, hogy a speciális kódolás miatt programozásuk körülményesebb; emiatt megnő a feldolgozási idő és a képdigitalizálás minősége is gyengébb, mint a hagyományos képkártyák esetében.

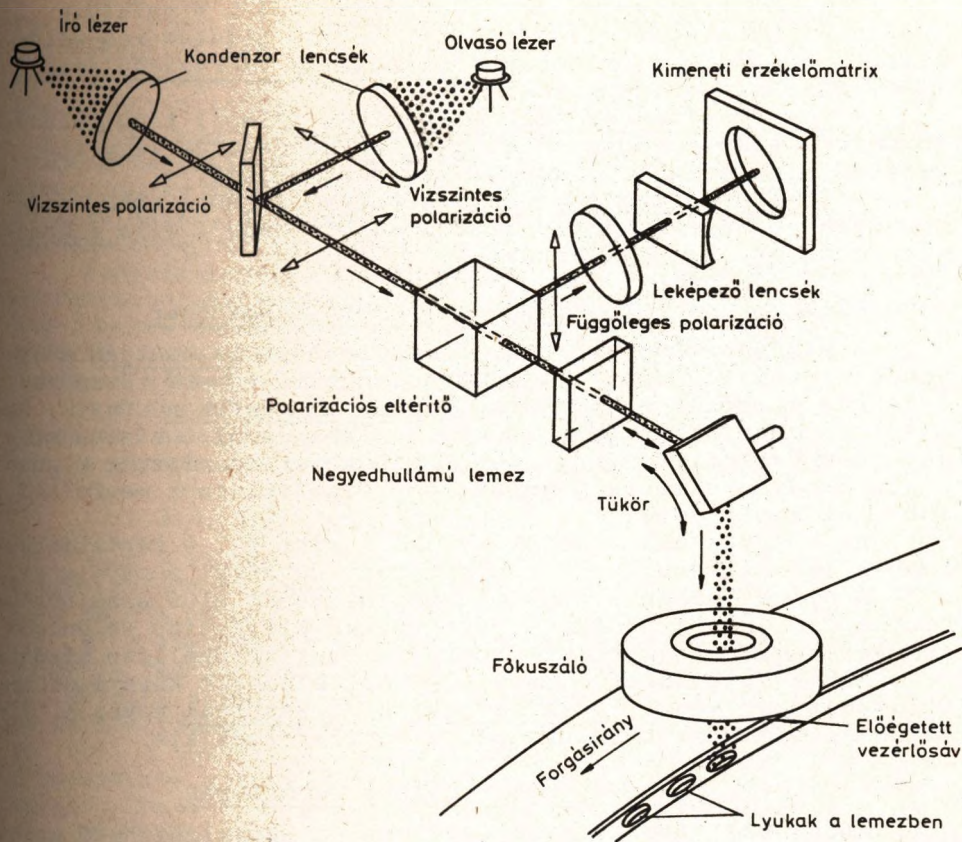
12.1.1.3. Speciális háttértárak

A digitális képek mérete általában nem teszi lehetővé, hogy a teljes képet a főtárban tároljuk. A valósídejű rendszerektől eltekintve - ahol a kiértékelés után már nincs szükség az eredeti képre - rendszerint igény van több egyidejű feldolgozására és a képek archiválására is. Ezért megfelelő háttértár-kapacitásról kell gondoskodni. A feldolgozás alatt álló képet rendszerint mágneslemezen tárolják, archiválás céljaira viszont mágnesszalagot használnak.

Annak illusztrálására, hogy milyen gondokat okozhat a feldolgozandó adattömeg, gondoljuk meg, hogy pl. a LANDSAT MSS rendszerben felvett egyetlen jelenet mintegy 32 Mbyte információ tárolását teszi szükségessé (ld. 12.1.3.3. alpont). Figyelembe véve, hogy az érzékelők felbontóképessége és a sávok száma az elkövetkező rendszerekben növekedni fog, érthető, hogy új tárolási technikák kidolgozására van szükség.

Nagy reményekkel kecsegtet pl. a képlemez, amelynek az elmúlt években megjelent néhány típusa máris lenyűgöző paraméterekkel rendelkezik. (A Philips DOR (Digital Optical Recording) lemeze például 2000 Mbyte kapacitású, 5-10 Mbit/s átviteli sebességgel és 100-500 ms elérési idővel. Az információt lézersugárral beégetett lyukak tárolják, a leolvasás is lézersugárral történik; ld. 1-13. ábra). A képlemez egyik hátránya jelenleg az, hogy nem törölhető. Ezt a nagy tárolási kapacitás némiképp kompenzálja; bizonyos esetekben pedig (pl. archiváláskor) ez a tulajdonság adatvédelmi szempontból előnyként is elkönnyvelhető.

Az ismertetett képlemez korszerűbb változatai, az optikai és magneto-optikai lemez meghajtók (WORM, MO, CD-ROM) ma már a PC világban is megszokott háttértárolók, és ezek egy része többszöri felírást is lehetővé tesz.



1-13. ábra
A képlemez működési elve

12.1.2. Párhuzamos architektúrájú számítógépek

A legtöbb képfeldolgozási algoritmus lokális jellege következtében gyakorlatilag minden képpontban azonos műveleteket kell végezni. Ehhez rendszerint csak az aktuális képpontnak egy kis környezetében kell ismerni a többi képpont világosságcódját, kézenfekvő hát a párhuzamos működésű gépek használata.

Bár ezek jelenleg nálunk nem érhetők el, röviden áttekintjük kialakításuk elveit. Mivel általánosan elfogadott rendszertechnikai osztályozás jelenleg nincs, induljunk ki a géptípusok alábbi, ismert csoportosításából:

- Egy utasítás és adatfolyam (SISD = single instruction, single data stream): ez a hagyományos Neumann-elvű (soros) gépek csoportja; a továbbiakban ezekkel már nem foglalkozunk.

- Több utasítás és egy adatfolyam (MISD = multiple instruction, single data stream): az elv tisztán nem valószínűsíthető meg, a gyakorlatban a futószalag-szervezésű (pipeline) gépeket sorolják ide.

- Egy utasítás és több adatfolyam (SIMD): több feldolgozóműből álló homogén, hierarchikus rendszerek (egy időpontban mindegyik tag azonos műveletet végez).

- Több utasítás és több adatfolyam (MIMD): több, párhuzamosan és esetenként önállóan működő gépből álló inhomogén rendszer, közös adattárral.

12.1.2.1. A futószalag-szervezésű számítógép

A működés elve az, hogy a műveleteket felbontják nagyjából azonos idő alatt, párhuzamosan elvégezhető részműveletekre; s ezeket egyidejűleg különböző adatokon végzik. Az első eredmény akkor jelenik meg, amikor az első adaton minden részművelet végrehajtott; ezt követően viszont minden újabb eredményt már a leghosszabb részműveletnek megfelelő idő után kapunk meg.

Az eljárás akkor hatékony, ha - mint pl. a képfeldolgozásban is - sokszor kell végrehajtani ugyanazt a műveletet.

A gyakorlatban a hagyományos (Neumann) elvű számítógéphez kapcsolják a futószalag-szervezésű kiegészítő feldolgozóművet, amely saját mikroprogramja szerint önállóan működik.

Azt a változatát, amely egy adat bizonyos környezetének feldolgozását végzi, tömb (array) processzornak nevezik. Tipikus alkalmazási területe pl. a Fourier-analízis.

12.1.2.2. Homogén, hierarchikus rendszerek

Népszerű, mivel viszonylag könnyen megvalósítható struktúra. Valójában egy vezérgép (master) dönti el, hogy milyen utasítást kell az összes alárendelt (slave) processzornak végrehajtania. (Ez alól lehet kivétel; pl. az adatoktól függetlenül néhány processzor blokkolt állapotban lehet, így adatvezérelt (data flow) feldolgozásra is mód nyílik.)

A processzorok saját tárral is rendelkezhetnek; ebben tárolják az általuk feldolgozandó adatokat (pl. egy képrészletet), ill. munkaterületként használják fel.

A fejlettebb rendszerekben a processzorok - bizonyos határok között - egymással is tudnak adatot cserélni; az egyszerűbbekben viszont esetleg csak leegyszerűsített végrehajtó egységek vannak.

Néhány példát említünk megvalósított rendszerekre:

a) A londoni University College-ban kifejlesztett CLIP-IV berendezés kifejezetten képfeldolgozási célból készült. 96x96 processzorból áll, elemenként 32 bites saját tárral.

A processzorok egy négyzethálós elemeit alkotják, és mindegyik a 8 szomszédjától kaphat adatokat. Logikai műveletek végrehajtására képes (bár természetesen ezekből természetesen aritmetikai művelet is felépíthető).

b) A NASA megrendelésére készül az MPP (Massively Parallel Processor) berendezés, amely (legalább) 128×128 processzorból fog állni, 1024 bites saját tárrakkal. A működési sebességre jellemző, hogy a tervezők 8 bites összeadás esetén $\sim 6550 \cdot 10^6$ művelet/s sebességet várnak.

c) A londoni Queen Mary College-ben, általános alkalmazásokra fejlesztették ki a DAP (Distributed Array Processor) gépet, amely 64×64 processzorból áll, processzoronként 16 Kbit saját tárral. Bizonyos feladatoknál még a CRAY szuperszámítógépnél is gyorsabb.

Figyelemreméltó megjegyezni, hogy a CLIP és DAP gépek processzorai biteken hajtanak végre műveleteket; a szavas műveleteket szimulálni kell. Ezért, valamint párhuzamos működésük miatt ezeknek a gépeknek a használatát a szokásostól teljesen eltérő gondolkodást igényel.

A fenti műveleti sebességek bizonyos fenntartásokkal értendők; ha ugyanis hozzávesszük az adat – be/kiviteli időket, a rendszer sebessége lényegesen kisebbnek adódik.

12.1.2.3. Inhomogén rendszerek

Ebbe a csoportba igen változatos architektúrájú rendszerek tartoznak; ezeket általában nem elemezzük, csak néhány példát adunk:

a) A linköpingi egyetemen kifejlesztett PICAP-I és PICAP-II rendszer kifejezetten képfeldolgozásra készült. Egy vezérlő számítógéphez csatlakoztatott 16 db 256 Kbyte-os képmemóriából és (legfeljebb) 15 specializált processzorból áll. A vezérlő gép osztja el a feladatokat az alárendelt processzorok között és rendeli hozzájuk a kijelölt képmemóriát. Tehát itt is hierarchikus rendszerről van szó, de az alárendelt processzorok lehetnek eltérők. Így a feldolgozás lokálisan párhuzamosan, de globálisan szekvenciális módon történik.

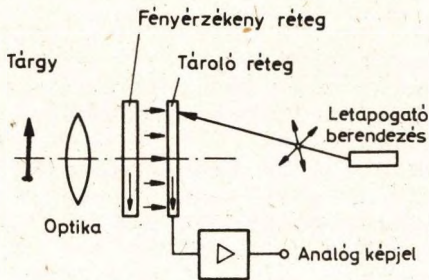
b) A marylandi egyetemen 256 db Z80-ast kapcsoltak össze ZMOB néven. Minden processzornak 64 Kbyte saját adattára és önálló programtára is van. A processzorok közti adatáramlás egyszerű eltolással történik: a láncba felfűzött processzorok addig adják a szomszédjuknak az adatot, míg el nem ér a célhoz.

12.1.3. Képfelvevő eszközök

A számítógépes képfeldolgozásban felvevőeszköznek csak az olyan berendezéseket tekintjük, amelyek a kimenetükön digitalizálható analóg képjelet, ill. közvetlenül digitális képet adnak. (Ilyen értelemben a fényképezőgép nem képfelvevő eszköz.)

12.1.3.1. Kamerák

Az 1960-as években, amikor a képfeldolgozó rendszerek dinamikus fejlődésnek indultak, a televíziós képfelvételi, átviteli és vételi rendszer már széles körben elterjedt, így kézenfekvően adódott a képfeldolgozó rendszerek bemeneti eszközeként az (optoelektronikus) TV-kamera.



1-14. ábra
Az optoelektronikus kamera működési elve

lés hatására a tárolólemezhez csatlakozó ellenálláson a megvilágítással arányos feszültségesség jön létre. Ezt az (analog) elektromos képjelet (egyszerű) videojelnek nevezik. Ehhez legtöbbször hozzákeverik a kép- és a sorszinkron jelet is, így áll elő az összetett videojel. Az OIRT szabvány szerint a képet 625 sorra bontják (ebből azonban csak 576 látható, a többi a képváltás idejére esik); a vízszintes/függőleges képarány 4:3, így egy sor 768 képpontból áll. A képet két félképből (páros, ill. páratlan sorszámú sorok) fésűsen egyesítik (váltottsoros (interlaced) letapogatás). A képváltási frekvencia 25 Hz, így egy kép felvételére 40 ms jut.

A fényérzékeny réteg fizikai tulajdonságainak változtatásával különböző kameratípusokat fejlesztettek ki.

A színeskamerában az optika a képet piros (R), zöld (G) és kék (B) színösszetevőre bontva képezi le a megfelelő szín-sávra érzékeny félvezetőrétegre; ennek megfelelően a tárolórétegen is egyszerre 3 töltéskép keletkezik. A "színes" videojelet 3 együttfutó, de egymástól függetlenül vezérelt elektronsugárral állítják elő.

A hőkamerák az infravörös sugárzási tartományra érzékenyített optikával és félvezetőréteggel rendelkeznek. (Általában a 2-5 μm , ill. a 10-12 μm hullámhossztartományban működnek.) A létező rendszerek egyik TV szabvánnyal sem egyeznek meg; ezért a "hő"-videojel számítógépes feldolgozása eléggé nehézkes.

Az elektronikus képfelbontás elvét az 1-14. ábra alapján érthetjük meg. A képet - megfelelő optikával - egy fényérzékeny félvezető rétegre képezik le. A fotoelektromos hatás következtében ebben az egyes képelemek fényességével arányos elektromos töltés keletkezik, amitől a fényérzékeny réteg mögötti tárolóréteg - töltésmegosztás révén - feltöltődik, a képelemek világosságával arányosan. Ezt a "töltésképet" egy megfelelően vezérelt elektronsugár soronként letapogatja, és a töltéskisü-

A kamerák - viszonylagos olcsóságuk miatt - a mai napig a legelterjedtebb képbeviteli eszközök. További előnyük a nagy felvételi sebesség, valamint az, hogy ipari környezetben is alkalmazhatók. Hátrányuk a kis felbontóképesség és a geometriai pontatlanság.

Az elmúlt évek a képbeviteli eszközök területén is hatalmas változásokat hoztak. A mintegy 10 évvel ezelőtt megjelent töltéscsatolt félvezető eszközök-kel felépülő

(charge coupled devices = CCD) kamerák gyakorlatilag teljesen kiszorították a korábbi képcsöveket. Fő előnyeik:

- nagyfokú linearitás a széleken is;
- nagy mechanikai stabilitás (rázás- és ütésállóság);
- kis fogyasztás.

A CCD eszközök működése szintén azon alapszik, hogy - az optikailag leképezett kép fényeloszlásának megfelelően - a töltés mennyisége megváltozik a félvezetőben. Ebben az esetben azonban az analóg jelet az egyes képelemeknek megfelelő töltéscsomagok elektronikus léptetésével hozzák létre. Egy-egy képpontnak több tranzisztor felel meg, amelyeket sorba vagy mátrixba rendeznek. A tranzisztorok rendre fel-töltés-érzékelés- töltéstovábbítás állapotnak megfelelően működnek.

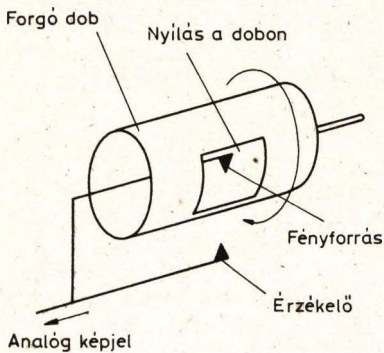
A kamerákba többnyire CCD mátrixot építenek be, ennek elemszáma határozza meg a geometriai felbontást; legelterjedtebbek az 512×512 eleműek. Eléggé szerencsétlen dolog, hogy az ílymódon már mintavételezett képből a kamera elektronikája szabványos analóg jelet generál, amiből a digitalizáló kártya újabb mintavételezéssel (és kvantálással, ld. 2.1.1. pont) állítja elő a digitális képet. Az ebből adódó pontatlanságok elkerülésére a CCD kamerát és a digitalizáló kártyát azonos szinkronjellel vezérlő (képpont-szinkronizált) rendszereket is kifejlesztettek.

Speciális feladatokra CCD vonalkamerákat is alkalmaznak. Ezekben az $1024-5000$ (vagy még több) érzékelő az optika által leképezett kép egyetlen sorát "látja" egyszerre. A kép úgy áll elő, hogy a vizsgált tárgy megfelelő sebességgel mozog a kamera alatt. Vonalkamerát leginkább olyankor használnak, amikor a vizsgálat tárgya - a technológiai folyamatból adódóan - folyamatos mozgásban van (például textilgyártás során a szövési hibák kiszűrésére).

A napjainkban igen elterjedt képbeviteli eszközök, a szkennerek érzékelője szintén CCD vonalkamera. A letapogató során vagy a kamerát vagy a tárgyat (pl. a beolvasandó papírlapot) kiegészítő mechanizmussal kell mozgatni, az érzékelősorra megőleges irányban. A DTP alkalmazások és az irodai archiváló rendszerek elterjedésének köszönhetően a szkennerek igen széles választéka kapható, az egyszerű kézi szkennerektől a lapadagalóval ellátott, nagyfelbontású gyors lapolvasókig; természetesen monokróm és színes változatban egyaránt.

12.1.3.2. A fotódigitalizáló

A mikrodenzitométerek egy tipikus megvalósítása a forgódobos fotódigitalizáló, amely film vagy fénykép bevitelére alkalmas. A filmet a forgódob palástjában kivágott nyílásra erősítik (ld. 1-15. ábra).



1-15. ábra
A fotódigitalizáló működési elve

Az eszköz a felvett képet rögtön digitalizálja is, így a kimenőjele közvetlenül tárolható a számítógépben. A fotódigitalizáló nagy pontosságú, de nehézkes képfeltevő eszköz; a képpont átmérője és a visszaállási pontosság 10-100 μm között van. A képet 512x512, újabban 2048x2048 képpontra bontja. Hátránya viszont, hogy a felvételi idő jelentősen megnő: 5-10 perc, a képmérettől és a digitalizálási pontosságtól függően; emiatt ma már alig használják.

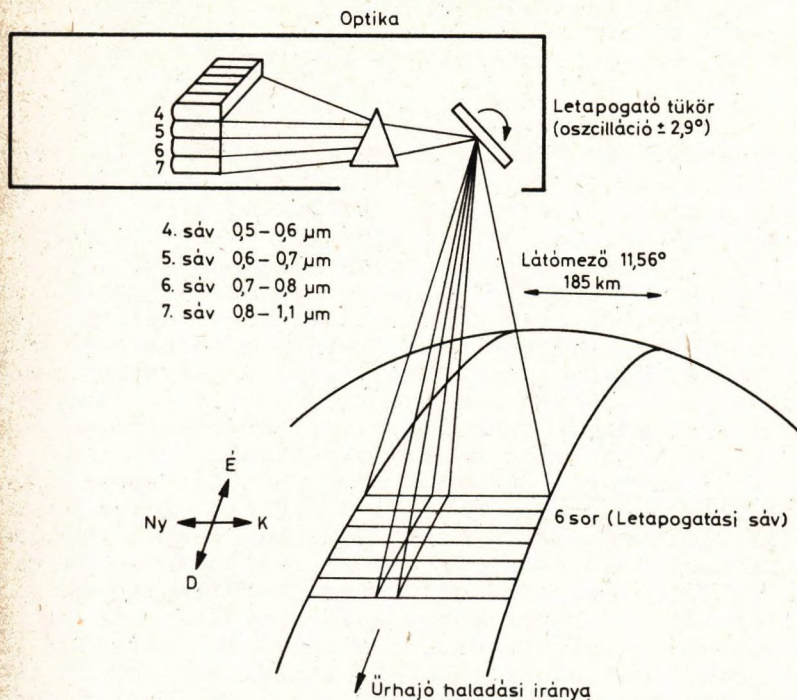
Megfelelő színszűrők alkalmazásával színes film digitalizálására alkalmas berendezéseket is kifejlesztettek.

12.1.3.3. Űrfelvételek készítése

Az űrfelvételek széles körű felhasználása miatt példaképpen vázlatosan ismertetjük a LANDSAT műhold többsávos felvevőrendszerét (MSS = multispectral scanner; ld. 1-16. ábra).

A Föld felszínéről visszavert fénysugár egy oszcilláló mozgást végző tükörrre, majd onnan a fénybontó prizma esik. A prizma a beeső sugarat különböző hullámhosszúságú komponenseire bontja, ezek közül a fényérzékelők négy hullámhossztartományba (két látható és két infravörös sávba) eső sugarakat érzékelnek. Minden sávban hat érzékelő van, azaz egyszerre hat sort pásztáznak végig. A földfelszín letapogatása a műhold haladásának és az oszcilláló tükör mozgásának megfelelően történik. Az érzékelők által keltett elektromos jelet már a műhold fedélzetén digitalizálják.

A felvétel jelenetekből áll. Egy jelenetben a földfelszín egy 185x185 km^2 területű részletét képezik le, sávonként 2340x3380 képpontra. A digitalizálónak 64 kvantálási szintje van, de 8 bites világosságkódokat ad. Egy jelenet



1-16. ábra
 Többsávós űrfelvételi rendszer elve

- a kiegészítő adatokkal - kb. 32 Mbyte-nyi információt tartalmaz. (Megjegyezzük, hogy az újabb rendszerekben mind a sávok száma, mind a felbontás, mind pedig a kvantálási szintek száma nagyobb; így az átviendő információ ennek többszöröse is lehet.)

A jeleneteket rádióan továbbítják a Földre. A vevőközpontokban nagy írássűrűségű mágnesszalagra rögzítik az adást, majd ezekből különböző korrekciók után állítják elő a felhasználók részére a számítógéppel közvetlenül feldolgozható (CCT = Computer Corrected Tape) mágnesszalagot.

Az űr- és légifelvételek készítésekor - az egyes felhasználási területek igényeinek megfelelően - számos más érzékelőt is alkalmaznak, beleértve a mikrohullámú (pl. radar) rendszereket is.

12.1.4. Megjelenítő és képrögzítő eszközök

Bár a digitális számítógépes képfeldolgozás egyik célja számszerűsített eredmények, összefüggések meghatározása, a feldolgozási folyamatban legtöbbször szükség van a vizuális ellenőrzésre vagy elemzésre, sőt a végeredmény sokszor éppen a feldolgozott kép.

A megjelenítők funkciója ellentétes a felvevő eszközzel, azaz a számítógépben tárolt számhalmazt kell az ember számára "élvezhető" formába alakítaniuk.

12.1.4.1. Elektronikus képmegjelenítők

Kézenfekvőnek látszik, hogy a képek megjelenítésére is TV-technikai eszközöket alkalmazzunk. A nehézséget az okozza, hogy a monitor képernyőjén megjelenő képet állandóan (legalább másodpercenként 25-ször) újra kell rajzolni ("fel-frissítés") ahhoz, hogy villogásmentesen lássuk. Ez a feladat túlságosan leterhelné a számítógépet, hiszen pl. egy 512x512 képpontból álló kép és 25 Hz-es ismétlési frekvencia esetén 6,25 Mbyte adatot kell kivinni másodpercenként, megfelelő sebességű csatornán keresztül! A raszteres megjelenítőkben ezt a problémát a képmű (ld. 12.1.1.1. alpont) oldja meg; a számítógépnek csak akkor kell a frissítőtárat feltöltenie, ha új képet akarunk megjeleníteni.

A ff (raszteres) képmonitor képernyője annyi foszforpontból áll, ahány képpontból a digitális kép. A foszforpontokat az elektronsugár a frissítőtárban tárolt világosságkódoknak megfelelő (analóg) videojel alapján gerjeszti s így "rajzolja fel" a (különböző fényességű világító fénypontokból álló) képet a képernyőre. A továbbiakban ezt látványképnek nevezzük. Az elektronsugár a bal felső sarokból indulva balról-jobbra és felülről lefelé halad. Először az első félképhez tartozó (a páros sorszámú sorokban levő) fénypontokat villantja fel, majd - az 1. sor elejére visszaugorva - ezek közé "berajzolja" a (páratlan sorszámú sorokból álló) második félképet. A kép a foszforpontok utánvilágítási képessége és a szemünk tehetetlensége következtében "áll össze".

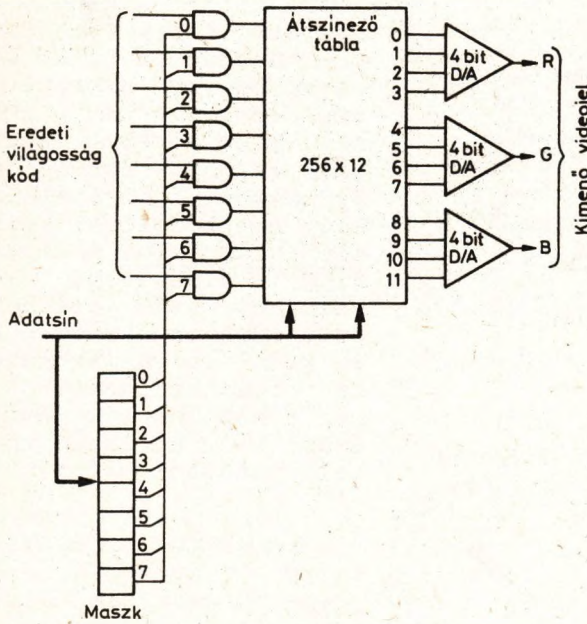
A színes (raszteres) monitoron egy képpont 3 különböző (piros, zöld, illetve kék) színben világító foszforpontból áll, amelyeket 3 együttl futó, de egymástól függetlenül vezérelt elektronsugár gerjeszt. A színárnyalatokat a foszforpontok egyidejű, de különböző mértékű gerjesztésével állítják elő (additív színkeverés).

A mai rendszerekben általánosan elterjedt a kimeneti átszínezőtábla (a továbbiakban: KAT; angol elnevezése szerint LUT = look-up-table) használata. Ennek segítségével természetesen meg lehet változtatni a látványképben a képpontok világosságát és/vagy színét anélkül, hogy a frissítőtár tartalma megváltoznék. A KAT elemeinek száma megegyezik a digitális képben megengedett világosság-, illetve színekódok számával; értékészletük pedig a megjeleníthető szürkességi szintek, illetve színek kódjainak halmaza. Jelöljük az előbbieket q -val, az utóbbiakat q' -vel, a KAT q -adik elemét pedig $T(q)$ -val, akkor a q világosságkódú képpontok a látványképben

$$q' = T(q)$$

"színben" fognak megjelenni.

A KAT előnye, hogy programból feltölthető s az átalakítás a felrajzolással egyidejűleg, idővesztés nélkül történik (ld. 1-17. ábra).



1-17. ábra
Hardver átszínezőtábla működési elve

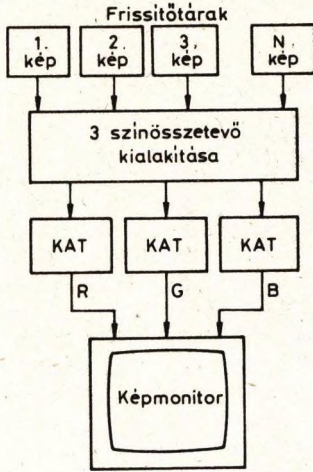
A hatékonyságot tovább lehet növelni többképsíkos képmű alkalmazásával; egy ilyen megoldást mutat az 1-18. ábra.

12.1.4.2. Képrögzítők

A kép rögzítésének legegyszerűbb módja a képernyő lefényképezése. A polaroid technika elterjedése különösen meggyorsította ezt az eljárást. A fellépő torzító hatások (horzótorzítás, fényvisszaverődés stb.) kiküszöbölésére speciális rendszereket fejlesztettek ki.

Az 1-19. ábrán látható rendszerben a különleges, sík felületű, monokromatikus képernyőn külön jelenítik meg az egyes (R, G, B) színösszetevőket, és megfelelő szűrők közbeiktatásával fényképezik le a képet.

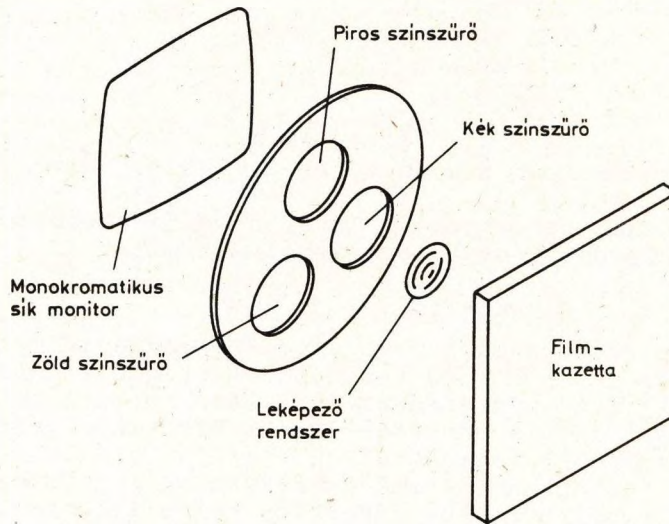
Az előző pontban ismertetett dobos fotodigitalizálókat általában úgy készítik, hogy megfelelő kiegészítésekkel alkalmasak legyenek a képek nagy pontosságú felrajzolására is. Ezekben a filmíró berendezésekben egy jól fókuszált 10-100 μm átmérőjű fénysugár fényerősségét modulálják, a (frissítő)



1-18. ábra
Többképsíkös megjelenítő
elvi felépítése

tárból kiolvasott világossággkódok szerint. A fénysugár a képsoroknak megfelelően pásztázza végig a forgódobra felerősített filmet, s így "rárajzolja" a képet, amely előhívás után válik láthatóvá.

A videoteknika elterjedésével egyre nagyobb szerepet kapnak a képfeldolgozásban a képmagnetofonok. Ezek a képet (analóg) összetett videojel formájában rögzítik. Színesképek felvétele úgy történik, hogy egy kódolóval kevert (kompozit) videojelet állítanak elő az R, G, B videojélből. (Az OIRT szabványú videojelekhez SECAM, a CCIR szabványúakhoz PAL rendszerű kódolóra van szükség.) Létezőnek címezhető változatok is, ezekben a képeket cím (azonosító sorszám) szerint lehet rögzíteni, ill. visszajátszani.



1-19. ábra
Speciális ernyőfényképező rendszer működési elve

A képmagnó előnye a nagy felvételi sűrűség és a kazetta olcsó ára. (Egy kazettán legalább 3600 képet lehet tárolni.) További előnyt jelent, hogy a felvett képet bármikor vissza lehet játszani, így bemeneti eszközként is használható. (A képmagnó kimenő videojele megegyezik a kameráéval. Színes felvételek visszajátszásakor a megfelelő dekódolóval kell visszaállítani az R, G, B videojelet.)

Hátrányt jelent, hogy egy professzionális képmagnó viszonylag drága. További problémát okoz, hogy a képmű kimenőjele a digitális kép visszaállításával (ld. 21.1.3. alpont) áll elő. Ha ezt képmagnón rögzítjük, visszajátszáskor a képmű újra-digitálja, ami jelentős torzulásokat eredményezhet.

A képrögzítők további fajtái a különböző színes plottek-rek, amelyek a képet képpontként rajzolják ki. Az egyik impakt nyomtatási módszer a görgős rajzolófejes eljárás. A rajzpapírt egy forgó hengerre erősítik fel. A képpontokat az alapszínekkel (fehér papír esetén magenta, sárga, türkiz; a jobb fedés miatt rendszerint még külön fekete is) festékezett görgők állítják elő. A dob egy körülfordulása alatt egy képsor felrajzolása történik meg úgy, hogy a vezérlő rendszer a megfelelő időpontban (elektromágnesek segítségével) a papírhoz nyomja a görgőket. Ezalatt a görgőket hordozó szán éppen egy sor szélességével mozdul el (a forgástengellyel párhuzamosan), így a következő körülfordulás során a következő képsor rajzolódik ki.

A keverékszíneket az alapszínek egymásra helyezésével állítják elő (szubtraktív színkeverés).

A tintasugaras (tintalövő = ink jet) eljárásnál az ábrát kis fúvókákból a megfelelő időpontban kilövellő festékszempcsék hozzák létre. A festékcseppeket elektrosztatikus módszerekkel egyrészt porlasztják, másrészt eltérítik a kívánt tónusok elérése céljából.

Terjedelmi korlátok miatt csak röviden említjük meg, hogy újabban elterjedtek a különböző lézer- és hőnyomtatók. A lézernyomtatók működése a hagyományos fénymásolók elvén alapul, a hőnyomtatókban pedig a felmelegített festékszalag felületéről szublimáló festék kerül át a papírra. Léteznek olyan változatok is, amelyek gyors előhívási technológiát alkalmazva fénykép minőségű képet állítanak elő.

12.1.5. Interaktív eszközök

A képfeldolgozásban általában a vizuális ellenőrzés vagy kiértékelés befolyásolja a feldolgozás további menetét, ezért a kényelmes visszacsatolási lehetőséget nyújtó, interaktív eszközök fontos szerepet töltenek be.

A billentyűzet (klaviatúra) a leggyakrabban alkalmazott interaktív eszköz; ezen keresztül lehet parancsokat és paramétereket megadni a programok vezérlésére. Gyakran találunk kiegészítő billentyűzetet, ill. - ezek hiányában - egyes billentyűkhöz programból rendelnek megkülönböztetett értelmet.

A legtöbb klaviatúrán található külön funkcionális billentyűzet, ezt használhatjuk a (grafikus) kurzor (cursor) mozgására. (A grafikus kurzor a képmonitoron megjelenő villogó jel (szálkereszt, nyíl stb.), amely pozicionálás és azonosítás céljából tetszőleges irányban mozgatható. Ettől megkülönböztetjük a program futtatásra szolgáló terminál képernyőjén látható alfanumerikus kurzort.)

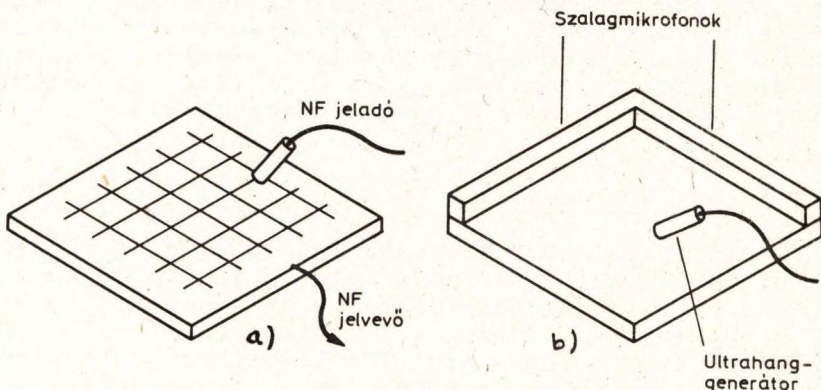
A klaviatúra hátránya, hogy a billentyűk kezelése elvonja az operátor figyelmét a képmonitorról; ennek kiküszöbölésére kapcsolókat és potenciométereket használnak. Pl. a gyakran alkalmazott botkormány (joystick), ill. a pozícionáló golyó (track ball) segítségével a felhasználó anélkül léptetheti a kurzort, hogy a szemét le kellene vennie a képernyőről.

Mindkettőben két, egymásra merőlegesen elhelyezett érzékelő jelzi az eszköz x-, ill. y-irányú elmozdulását. Ezt használjuk azután a kurzor vezérlésére.

A digitalizáló táblák kölcsönösen egyértelmű megfeleltetést létesítenek a képmonitor és a tábla felülete között, s így közvetlenül vihetünk be grafikus adatokat. Működésük elvét tekintve két alaptípust különböztethetünk meg:

- A vezetékrácsos rendszerűek elvi működését az 1-20/a. ábrán szemléltetjük. A kezelő a ceruzaformájú szondát a megfelelő pozícióra helyezi. A szonda által kibocsátott elektromágneses hullám feszültséget indukál a szigetelő rétegek között elhelyezett vezetékekben. Az ezekhez kapcsolt érzékelőkben keletkező jelekből meghatározható a szonda x-y koordinátája.

- Az 1-20/b. ábrán látható akusztikus digitalizáló tábla az oldallapos érzékelő rendszerek működését mutatja. A tábla két szomszédos oldalán elhelyezett szalagmikrofonok a szonda által keltett ultrahangot érzékelik. Az időzítési viszonyokból meghatározható a szonda helyzete.

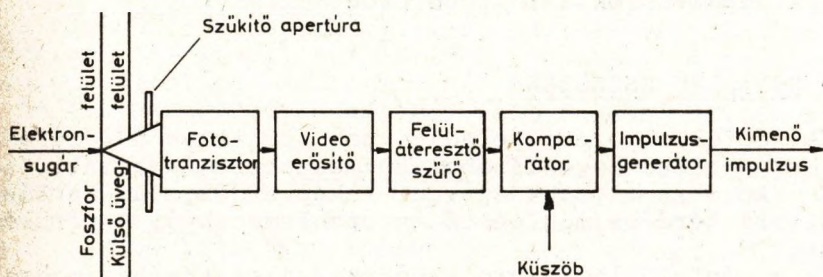


1-20. ábra
Digitalizáló táblák típusai

A digitalizáló táblákat nemcsak pozíció-meghatározásra lehet használni: a tábla szélein fenntartott funkció-mezőkhöz különböző funkciókat rendelnek, amelyeket rámutatással lehet aktiválni.

Megemlítjük még, hogy a digitalizáló táblák rajz vagy kézírás bevitelére is alkalmasak. A rajzelemek vagy a szöveg felismerésére természetesen megfelelő programokat kell kidolgozni.

A fénytollal (light pen; gyakran fényceruzának nevezik) a kezelő közvetlenül a képmonitoron keresztül teremthet kapcsolatot a rendszerrel. A kinézetre toll formájú eszköz -elnevezésével ellentétben - nem ír a képernyőre, hanem leolvassa egy képpont koordinátáit. A fénytoll felépítése az 1-21. ábrán látható.



1-21. ábra
A fénytoll működési elve

A belépő fénysugarat egy szűkítő apertúra egy fototranzisztorra fókuszálja. Az erősítőt a fénytoll kapcsolójának megnyomása aktiválja. (A kapcsoló általában úgy működik, hogy a fénytollat a képernyőhöz kell nyomni, de szoktak külön kapcsolót is beépíteni.) A képet felrajzoló elektronsugár elhaladásakor a fototranzisztor egy impulzust generál, amelyet - erősítés és szűrés után - a komparátor összehasonlít egy küszöbfeszültséggel. (A küszöbértéket úgy kell megválasztani, hogy a szórt világításból és a zajból eredő impulzusok hatástalanok maradjanak.) Ha a feszültségváltozás meghaladja a küszöbértéket, az impulzusgenerátor előállítja a kimenőjelet. A képszinkronjel ismeretében az időkülönbségből egyértelműen meghatározható a fénytoll pozíciója.

A fénytoll fő hátránya, hogy a sötét területen levő képpontokat nem tudja azonosítani; a problémát ez ideig nem sikerült kielégítően megoldani. (Egyes rendszerekben átmenetileg felvillantják a sötét képpontokat, a frissítőtár megkerülésével. Ennek - a bonyolult megoldáson túlmenően - az a hátránya, hogy zavarja az operátort.)

A fénytollat mára teljesen kiszorította az egér. Mozgását egy golyó két, egymásra merőleges tengely körül forgó mágnesre viszi át, amelyek minden körülfordulásra 1 impulzust generálnak az őket körülvevő tekercsben. Az időegységre eső impulzusok számát a vezérlőelektronika átadja a kezelőprogramnak, amely - az előre megadott lépték(ek) felhasználásával - kiszámítja az interaktív kurzor relatív elmozdulását. Ennek ismeretében az egeret különféle rajzolás- és vezérlési feladatokra lehet beprogramozni. (Ha pl. meghatározott időközönként megjelenítjük a kurzort a képernyőn az aktuális pozícióban, a felhasználó úgy érzékeli, hogy az egerrel mozgatta.)

Vezérlési célokra 2 vagy 3 nyomógombot is ráépítenek az egérre, amelyek lenyomását és/vagy felengedését a vezérlő-elektronika ugyancsak érzékeli. A gombokhoz tetszőleges interaktív funkciókat lehet rendelni; felhasználatók pl. egy menüsor vagy ikon aktiválására, képrészlet kijelölésére stb.

Programozásának megkönnyítésére a gyártók rendszerint egy sor funkcionális szubrutint is szállítanak az egérrel, amelyek meghívhatók a legtöbb programnyelvből.

12.2. SOFTWARE ESZKÖZÖK

A képfeldolgozási feladatoknak speciális hardware eszközökkel történő megoldásához természetesen speciális software eszközökre is szükség van. A képfeldolgozás területén a software kérdése meglehetősen problematikus. A probléma kettős:

- a képfeldolgozó programok kifejlesztéséhez magasan kvalifikált szakemberekre van szükség, mivel ez a munka jelentős matematikai ismereteket feltételez, és jártasságot igényel a különböző határtudományágakban;
- tovább nehezíti a helyzetet, hogy a futási idők csökkentése érdekében a programokat általában alacsony szintű (assembly) nyelven kell írni. A sokszor ellentmondó követelményeket csak különféle programozási fogásokkal és kompromisszumok árán lehet teljesíteni. Ezért és a különböző speciális periférikus eszközök használata miatt a képfeldolgozó software áttelepíthetősége (portabilitása) rossz; más gépeken nehezen adaptálható. Ez az oka annak, hogy - bár számos képfeldolgozó programrendszer létezik - szinte minden új eszközhöz többé-kevésbé új programokat fejlesztenek ki.

A képfeldolgozó software kialakításánál a felhasználási terület, a felhasználói igények jelentős szerepet kapnak; nincsenek általános, mindenre használható rendszerek. Ennek az a fő oka, hogy a képfeldolgozás általános elmélete még nincs kidolgozva. A legtöbb feladat megoldására nincs egyszakt algoritmus, vagy ha van, elviselhetetlenül hosszú futási időket eredményez. Az elmélet hiánya az oka annak is, hogy azonos feladatokra számos, más-más környezeti feltételek vagy feldolgozási célok esetén előnyös megoldás létezik.

Ílymódon az esetek többségében csak kísérleti úton lehet kiválasztani a megoldási módszert és eljárást. Egész más megközelítést igényel pl. egy ipari környezetben működő, célorientált, kamera bemenettel rendelkező, automatikus minőségellenőrzést végző rendszer kialakítása, mint egy úrfelvételek kiértékelése céljából létrehozandó képfeldolgozó központé. Ennek ellenére számos közös, vagy hasonló módszerekkel megoldható feladat jelentkezik.

A képfeldolgozó programrendszerek kidolgozása során az alábbi fő összetevők megoldási módjai határozzák meg a képfeldolgozó software végső formáját:

- a feldolgozási mód,
- a vezérlési mód,
- a képfájl struktúra,
- be/kiviteli rutinok,
- alkalmazói programok,
- numerikus programkészlet.

1. A feldolgozási mód köteget (batch), interaktív, ill. vegyes típusú lehet ugyanúgy, mint bármely más alkalmazói rendszernél. Az ipari, automatikus, valós idejű rendszereket leszámítva általában az interaktív megoldást részesítik előnyben, de rutinszerű munkáknál meg kell engedni a köteget feldolgozást is.

2. Három típusú vezérlési mód létezik: a párbeszéd (dialógus), a menütáblás és a parancslánc forma.

A párbeszéd változat a kezdő felhasználók számára előnyös, de meglehetősen idegesítő lehet a tapasztaltabbaknak.

A menütáblás vezérlés esetén a képernyőn megjelenített funkciók közül kell kiválasztani a megfelelőt. Ez a módszer hatékonyabb mint a párbeszéd vezérlés, és némi segítséget jelent a felhasználó számára.

A leggyorsabb vezérlési mód a parancslánc megadása, de ehhez a felhasználónak meg kell tanulnia a lehetséges parancsokat és paramétereik megadási módját.

3. A nagy adattömegek miatt a képfájl-ok struktúrájának és a kiegészítő információk tárolásának gondos megtervezése alapvető fontosságú a rendszer hatékonysága szempontjából. Egyszerűbb esetekben csak arról kell dönteni, hogy sor vagy oszlop szerint rendezzük-e rekordokba a világossághódokat; esetleg a teljes képmátrixot kisebb blokkokra osztva lap-szervezésben tároljuk-e inkább. Alapvető szempont a lemez-műveletek számának csökkentése, ezért lehetőleg a teljes képet, de ha a főtár kapacitása ezt nem teszi lehetővé, legalább a gyakran szükséges képrészleteket a főtárban kell tárolni a feldolgozás folyamán.

Bonyolultabb rendszerekben a képtartalmon túlmenően egyéb adatokat is tárolni kell, mint például:

- képméretek (sorok, oszlopok száma stb.);
- képjellemzők (nagyítási arány, képkezdőpont stb.);
- hozzáférési specifikáció (blokkméret, rekordméret, írható, csak olvasható stb.);
- adattípus (egész, valós, komplex, szöveges stb.);
- feldolgozási specifikáció (a kép eredete, a feldolgozás módja stb.);
- statisztikai jellemzők (átlag, szórás, minimum, maximum, hisztogram stb.);
- egyéb adatok.

Többsávú képek feldolgozása esetén dönteni kell az összetartozó képsávok egymáshoz viszonyított elhelyezéséről. Három alapvető tárolási mód létezik:

- BSQ: a sávok egymást követően szekvenciálisan helyezkednek el;
- BIL: a sávok azonos sorszámú sorai követik egymást;
- BIP: a sávok azonos koordinátájú képpontjai követik egymást.

4. A be/kiviteli rutinok szintén nagymértékben befolyásolják a hatékonyságot. Az alkalmazói programokat fejlesztők többnyire nem kívánnak foglalkozni a file-okat, a virtuális tárat, a speciális periférikus eszközöket stb. kezelő rutinok megírásával. A probléma akkor jelentkezik, ha általános célú számítógépre épülő képfeldolgozó rendszert akarunk létrehozni. A gyártó cégek ugyanis általában nem vállalják a speciális eszközök illesztését és programozását, másrészt viszont nem nézik jó szemmel sorozat-termékük módosítását. Ezért az alkalmazói rendszer fejlesztőjének sokszor nem megfelelő, hiányos dokumentációk birtokában kell megoldania ezeket a feladatokat is.

5. Az alkalmazói programok kifejlesztése során célszerű közös csoportokba összevonni az azonos elven működőket. Ez csökkenti a programrendszer méretét (közös be/kiviteli rutinok, közös számítási részek), és könnyebbé teszi a karbantartást. Egy ilyen csoportosítási lehetőséget mutatunk be az alábbiakban:

Képmegjelenítés:

interaktív eszközök kezelése; képvétel, képforgalmazás a különböző táruk között; részletkijelölés, kicsinyítés-nagyítás; megjelenítés, átszínezések; stb.

Statisztikai számítások:

átlag, szórás, kovariancia meghatározás; hisztogram készítés; stb.

Pontenkénti műveletek:

sávok közötti műveletek; világosságkód-transzformációk; stb.

Lokális műveletek:

kontrasztkiemelés, szűrések; élkirajzolás; stb.

Globális transzformációk:

Fourier-, cosinus, Hadamard-, Karhunen-Loève- stb. transzformáció; geometriai transzformációk; analitikus eljárások; referenciapontok illesztése; stb.

Osztályozási eljárások:

téglatest-módszer; Bayes-módszer; klaszterezés stb.

6. A numerikus programkészletekkel kapcsolatban annyit említünk meg, hogy sokszor jelentős munkát lehet megtakarítani, ha a géppel szállított matematikai-tudományos programkönyvtárakat használjuk, ezért a géptípus kiválasztásánál ennek kidolgozottságát is figyelembe kell venni. A kész programkönyvtárak használata előtt azonban célszerű meggyőződni arról, hogy a céljainknak megfelelő algoritmus szerint működnek-e, és a többnyire magas szintű nyelven (PASCAL, C) írt rutinok nagyobb futási ideje elviselhető-e számunkra.

13. Matematikai háttér

Ebben a fejezetben röviden összefoglaljuk a képfeldolgozási eljárásokban leginkább használt matematikai alapismerteket. A bizonyításokat általában mellőzzük és a teljességre sem törekedhettünk. A későbbiekben még mélyebben foglalkozunk néhány kérdéssel, a konkrét eljárások ismertetésekor.

13.1. MÁTRIXALGEBRA

A képfeldolgozásban a mátrixok hasznos segédeszközt jelentenek, mert segítségükkel tömören és elegánsan lehet leírni a digitális képekkel végzendő műveleteket. Az alábbiakban röviden összefoglaljuk a fontosabb ismereteket.

13.1.1. A mátrix fogalma és főbb típusai

Legyen T egy adott test (pl. a valós számtest); az

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \quad (a_{i,k} \in T; i=1, \dots, m; k=1, \dots, n)$$

táblázatot a T test feletti mátrixnak nevezzük; az a_{ik} adatok a mátrix elemei.

A sorokat felülről lefelé, az oszlopokat balról jobbra számozzuk: az i -edik sor k -adik elemét a_{ik} jelöli. Felhívjuk a figyelmet arra, hogy a képfeldolgozásban fordított indexelést használunk!

Egy mátrixot akkor tekintünk adottnak, ha egyrészt ismert sorainak és oszlopainak száma, másrészt minden eleme adott. Jelölésére az

$$\underline{A} = [a_{ik}]_{m \times n}$$

szimbólumot fogjuk használni, és - ha nem vezet félreértésre - az $m \times n$ indexet elhagyjuk.

Az $\underline{A} = [a_{ik}]$ és a $B = [b_{ik}]$ mátrix akkor és csak akkor egyenlő, ha egyrészt azonos típusúak (vagyis sor- és oszlopszámuk megegyezik), másrészt minden i, k elemükre

$$a_{ik} = b_{ik}$$

13.1.2. Mátrixok összeadása, kivonása és szorzása

Az összeadás és a kivonás csak azonos típusú mátrixokra van értelmezve. A

$$\underline{C} = [c_{ik}]_{m \times n}$$

mátrixot az $\underline{A} = [a_{ik}]_{m \times n}$ és a $\underline{B} = [b_{ik}]_{m \times n}$ ($a_{ik}, b_{ik} \in T$) mátrixok összegének (különbségének) nevezzük, ha

$$c_{ik} = a_{ik} \pm b_{ik} \quad (i = 1, 2, \dots, m; \quad k = 1, 2, \dots, n).$$

Az így értelmezett mátrixösszeadás asszociatív és kommutatív.

A $\lambda \in T$ skalár és az \underline{A} mátrix szorzatán a

$$\lambda \underline{A} = [\lambda a_{ik}]_{m \times n} \quad (a_{ik} \in T: \quad i = 1, 2, \dots, m; \\ k = 1, 2, \dots, n)$$

mátrixot értjük. A skalárral való szorzás kommutatív és disztributív.

A $\underline{C} = [c_{ik}]_{m \times p}$ mátrixot az $\underline{A} = [a_{i\ell}]_{m \times n}$ és a $\underline{B} = [b_{\ell k}]_{n \times p}$ ($a_{i\ell}, b_{\ell k} \in T$) mátrixok szorzatának nevezzük, ha

$$(1-1) \quad c_{ik} = \sum_{\ell=1}^n a_{i\ell} b_{\ell k},$$

és az \underline{AB} szimbólummal jelöljük.

Hangsúlyozni kell a következőket:

- az \underline{AB} szorzat csak akkor van értelmezve, ha \underline{A} -nak ugyanannyi oszlopa van mint ahány sora van \underline{B} -nek; vagyis \underline{A} és \underline{B} az adott sorrendben konform,
- mátrix mátrixszal való szorzása általában nem kommutatív.

A mátrixok felépítése és szimmetriatulajdonságai alapján a következő fontosabb típusokat különböztetjük meg (az elmondandókat a könnyebb megértés kedvéért az 1-22. ábrán adott példákkal illusztráljuk):

- Az oszlopmátrix vagy oszlopvektor egyetlen oszlopból ($\underline{P} = \vec{p}$), a sormátrix vagy sorvektor egyetlen sorból ($\underline{S} = \vec{s}$) áll.

- A négyzetes vagy kvadratus mátrix sorainak és oszlopainak száma egyenlő. Az n sorból és oszlopból álló négyzetes mátrixot n -ed rendűnek, a főátlóban levő a_{ii} elemeit pedig diagonális elemeknek nevezzük. (Pl. az \underline{F} harmadrendű négyzetes mátrix diagonális elemei: (1, 5, 6).)

- Diagonálmátrix az olyan négyzetes mátrix, amelynek csak a főátlóban vannak 0-tól különböző elemei, vagyis minden a_{ik} ($i \neq k$) eleme 0. Ha a főátlóban minden elem 1, -vagy-

$$\underline{\underline{F}} = \begin{bmatrix} 1 & 3 & 2 \\ 7 & 5 & 4 \\ 8 & 9 & 6 \end{bmatrix}; \underline{\underline{F}}^T = \begin{bmatrix} 1 & 7 & 8 \\ 3 & 5 & 9 \\ 2 & 4 & 6 \end{bmatrix}; \underline{\underline{F}}^{-1} = \begin{bmatrix} -0,6 & 0 & 0,2 \\ -1 & -1 & 1 \\ 2,3 & 1,5 & -1,6 \end{bmatrix};$$

$$\underline{\underline{G}} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}; \quad \underline{\underline{G}}^T = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix};$$

$$\underline{\underline{H}} = \underline{\underline{H}}^T = \begin{bmatrix} 0,6 & 0,8 \\ 0,8 & -0,6 \end{bmatrix}; \quad \underline{\underline{Z}} = \underline{\underline{Z}}^{*T} = \begin{bmatrix} \frac{\sqrt{5}}{3} & \frac{2}{3}j \\ -\frac{2}{3}j & -\frac{\sqrt{5}}{3} \end{bmatrix};$$

$$\underline{\underline{Z}}^T = \underline{\underline{Z}}^* = \begin{bmatrix} \frac{\sqrt{5}}{3} & -\frac{2}{3}j \\ \frac{2}{3}j & -\frac{\sqrt{5}}{3} \end{bmatrix};$$

$$\underline{\underline{E}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \underline{\underline{O}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix};$$

$$\underline{\underline{P}} = \underline{\underline{P}}^T = \begin{bmatrix} 2 \\ 7 \end{bmatrix}; \quad \underline{\underline{S}} = \underline{\underline{S}}^T = \begin{bmatrix} 7 & 4 \end{bmatrix}.$$

1-22. ábra
Példa mátrixok

is $a_{ii} = 1$, ($i = 1, 2, \dots, n$) - a diagonálmátrixot egységmátrixnak nevezzük (pl. \underline{E} másodrendű egységmátrix).

- Egy mátrix transzponáltját úgy nyerjük, hogy felcseréljük a sorait és az oszlopait; jelölése \underline{A}^T (pl. $\underline{F}^T, \underline{H}^T$). Az oszlopmátrix transzponáltja sormátrix és fordítva.

- Ha egy mátrixban minden elem helyére a konjugáltját írjuk, a konjugált mátrixhoz jutunk (pl. \underline{Z}^*).

- Szimmetrikusnak, ill. hermitikusnak nevezzük az olyan négyzetes mátrixot, amely egyenlő a transzponáltjával, ill. konjugált transzponáltjával (pl. $h_{ik} = h_{ki}$, ill. $z_{ik} = z_{ki}^*$, vagyis $\underline{H} = \underline{H}^T$, $\underline{Z} = \underline{Z}^{*T}$).

Ortogonalisnak, ill. unitérnek nevezzük az olyan négyzetes mátrixot, amelynek a transzponáltjával, ill. a konjugált transzponáltjával képzett szorzata egységmátrix; pl.

$$\underline{G} \cdot \underline{G}^T = \underline{E}; \quad \underline{H} \cdot \underline{H}^T = \underline{E}; \quad \underline{Z} \cdot \underline{Z}^{*T} = \underline{E}.$$

- Sormátrix és oszlopmátrix szorzata ebben a sorrendben szám, fordított sorrendben mátrix. Az előbbit skalárszorzatnak, az utóbbit diadikus szorzatnak nevezzük. A két mátrix elemszámának mindkét esetben meg kell egyeznie.

13.1.3. Osztás négyzetes mátrixszal

A számok körében az osztást az osztó reciprokával - az inverzével - való szorzásként is értelmezhetjük. Ahhoz, hogy ezt a négyzetes mátrixok körében is megtehesük, értelmeznünk kell az inverzmátrix fogalmát. Az \underline{A}^{-1} mátrixot az \underline{A} négyzetes mátrix inverzének nevezzük, ha

$$\underline{A} \cdot \underline{A}^{-1} = \underline{A}^{-1} \cdot \underline{A} = \underline{E}.$$

Ismeretes, hogy az inverzmátrix akkor és csak akkor létezik, ha az \underline{A} elemeiből alkotott

$$(1-2) \quad \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

determináns nullától különböző. A determinánst az $|\underline{A}|$ szimbólummal jelöljük, mivel a mátrixok körében ugyanazt a szerepet játssza, mint az abszolútérték a valós és a komplex számok körében. (Részletes definíciója megtalálható a megfelelő szakirodalomban; röviden: a determináns az összes olyan szorzat előjeles összege, amely a mátrix minden sorából és oszlopából pontosan egy elemet tartalmaz. Belátható, hogy ezek száma $n!$.)

Ezáltal helyettesítettük a képet egy térbeli felülettel, és lehetőségünk nyílik arra, hogy alkalmazzuk a matematikai analízis módszereit a képfeldolgozási eljárásokban. Tekintve, hogy a digitális képek csak az (x, y) sík meghatározott pontjaiban vannak értelmezve, és a világosságkódok értékkészlete is diszkrét (egész), a differenciálhányadost differenciahányadossal, az integrálokat (végtelen) sorokkal fogjuk közelíteni, ez azonban nem érinti a módszerek lényegét.

13.2.1. A Dirac-delta

A képek kényelmes leírása céljából vezessük be a pontforrás fogalmát; ez a képek egy olyan sorozatának határértéke, amelyben az információt hordozó, 0-tól különböző világosságú képpontok területe nullához tart. Ezzel a sorozattal akkor tudjuk meghatározni a kép egyes pontjainak világosságát, ha a közelítő sorozat függvényértéke egyidejűleg minden határon túl nő. Egy lehetséges egyszerű példaként induljunk ki a "körlap" függvényből, amelynek definíciója:

$$\text{körlap}_n(x, y) = \begin{cases} 1, & \text{ha } \sqrt{x^2 + y^2} \leq \frac{1}{n} \quad (n = 1, 2, \dots) \\ 0, & \text{különben} \end{cases}$$

ezzel a pontforrás közelítő függvénye:

$$\delta_n(x, y) = \frac{n^2}{\pi} \text{körlap}_n(x, y).$$

A definícióból látjuk, hogy δ_n értéke 0 az $\frac{1}{n}$ sugarú körön kívül, és $\frac{n^2}{\pi}$ a kör belsejében (ld. 1-23. ábra). Következésképpen

$$\delta(x, y) = \lim_{n \rightarrow \infty} \delta_n(x, y) = \begin{cases} \infty, & \text{ha } x=0 \text{ és } y=0; \\ 0, & \text{különben;} \end{cases}$$

továbbá formálisan

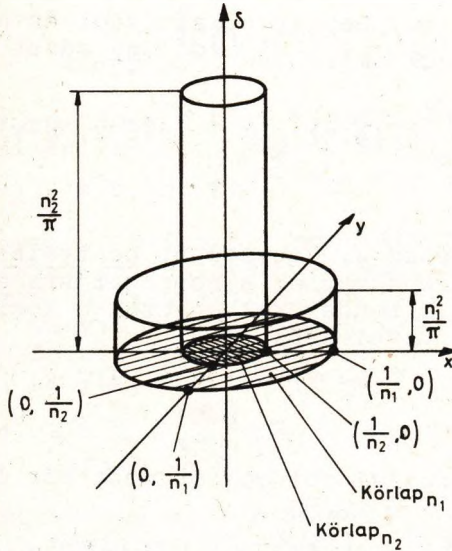
$$\iint_{-\infty}^{\infty} \delta(x, y) \, dx \, dy = \lim_{n \rightarrow \infty} \iint_{-\infty}^{\infty} \delta_n(x, y) \, dx \, dy = 1.$$

A $\delta(x, y)$ függvényt Dirac-deltának nevezik, ami a nulla alapterületű, végtelen magasságú, egységnyi térfogatú impulzus absztrakciója. (Matematikailag ez a képlet hamis, mert a baloldali integrál mindenütt 0. A problémát a funkcionálmélet oldja fel, amelyben bebizonyítható, hogy az így definiált Dirac-delta és egy tetszőleges folytonos $f(x, y)$

függvény szorzatának integrálja $f(0,0)$, vagyis éppen a függvényérték az $[x=0, y=0]$ pontban.) Könnyen be lehet látni, hogy érvényes a

$$\delta(x - \xi, y - \eta) = \begin{cases} \infty, & \text{ha } x = \xi, \quad y = \eta \\ 0, & \text{különben} \end{cases}$$

"toló" tulajdonság.



1-23. ábra
A Dirac-delta közelítése

A Dirac-delta segítségével egyszerűen leírhatjuk a mintavételezési folyamatot. Ha a képfüggvényt $f(x, y)$ jelöli:

$$(1-3) \quad \iint_{-\infty}^{\infty} f(x, y) \delta(x - \xi, y - \eta) dx dy = f(\xi, \eta)$$

A képlet azt fejezi ki, hogy az $f(x, y)$ képet "generáló" pontforrás a képtartomány egy (ξ, η) pontjában éppen az $f(\xi, \eta)$ függvényérték.

A Dirac-deltát fogjuk használni a folytonos és a digitális kép közötti összefüggések matematikai megfogalmazására.

13.2.2. Konvolúció

Vegyük észre, hogy az (1-3) képlet szerint az $f(x, y)$ képfüggvény felírható pontforrások lineáris összegeként, mivel minden (x, y) pontban

$$f(x, y) = \lim_{\substack{\xi \rightarrow x \\ \eta \rightarrow y}} \iint_{-\infty}^{\infty} f(\xi, \eta) \delta(\xi - x, \eta - y) d\xi d\eta$$

Tegyük fel, hogy a leképezőrendszer lineáris, és az eredeti képjelet valamilyen

$$g(x, y) = T \{f(x, y)\}$$

függvénybe viszi át. Behelyettesítéssel és a linearitás kihasználásával ($f(\xi, \eta) = \text{állandó egy adott pontban}$):

$$g(x, y) = \iint_{-\infty}^{+\infty} f(\xi, \eta) \lim_{\substack{\xi \rightarrow x \\ \eta \rightarrow y}} T \{ \delta(\xi - x, \eta - y) \} d\xi d\eta$$

Az integrandusz 2. tényezőjét pontválasznak nevezzük, mivel a leképezőrendszernek a pontforrásra adott válaszát írja le. Ez lineáris rendszerekben csak a koordináta-különbségtől függ, ezért vezessük be a

$$(1-4) \quad h(x - \xi, y - \eta) = \lim_{\substack{\xi \rightarrow x \\ \eta \rightarrow y}} T \{ \delta(\xi - x, \eta - y) \}$$

jelölést. A leképezési folyamatot most már leírhatjuk a

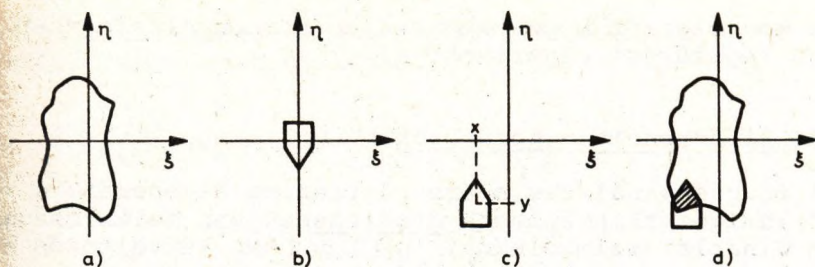
$$(1-5) \quad g(x, y) = \iint_{-\infty}^{+\infty} f(\xi, \eta) h(x - \xi, y - \eta) d\xi d\eta$$

konvolúcióintegrállal, amelyet a továbbiakban röviden a

$$g(x, y) = f(x, y) \otimes h(x, y)$$

szimbólummal fogunk jelölni.

A konvolúcióintegrált az 1-24. ábrán szemléltetjük. Az a) ábrán az eredeti képtartományt látjuk a (ξ, η) síkban, efelett helyezkedik el az $f(\xi, \eta)$ képfüggvény. A b. ábrán mutatjuk meg, hogy a leképezőrendszer milyen tartományba képezi le a $(0, 0)$ koordinátájú pontforrást. A leképezőrendszer úgy működik, hogy minden (x, y) pontot "bejár", s mindegyikben összegzi az eltolt és saját középpontján át tükrözött leképezési tartomány (ld. c. ábra), valamint az eredeti képtartomány közös része feletti (d. ábra) pontforrásokra adott pontválaszait.



1-24. ábra
Konvolúció szemléltetése

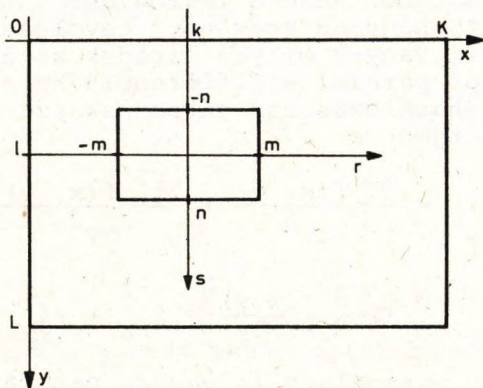
Digitális képek esetén a $g(k,l)$ képfüggvény csak a véges $K \times L$ méretű képsíkon van értelmezve, azaz

$$g(k,l) = 0, \text{ ha } k \geq K, \text{ illetve } l \geq L.$$

Az (1-4) kifejezéssel definiált pontválaszról az általánosítás megszorítása nélkül szintén feltehetjük, hogy csak egy $(2m+1) \cdot (2n+1)$ méretű, téglalap alakú tartomány felett különbözik nullától, vagyis megadható egy $T(r,s)$ mátrixszal. Ekkor az (1-5) konvolúcióintegrál az 1-25. ábra jelöléseivel az

$$(1-6) \quad r(k,l) = \sum_{r=-m}^m \sum_{s=-n}^n g(k+r,l+s) T(r,s)$$

összegzésre egyszerűsödik. A T mátrixot konvolúciós szűrőnek nevezzük; $r(k,l)$ pedig az átalakított ("szűrt") kép, amelyben minden képpont világosságszintjét csak a szűrő által meghatározott környezetébe eső képpontok befolyásolják.



1-25. ábra
Konvolúciós szűrő
illesztése egy képpontra

A konvolúciós összegzést széles körben alkalmazzák a különböző képszűrési eljárásokban.

13.2.3. Differenciál-operátorok

A függvényanalízis számos differenciál-operátora közül a képfeldolgozásban főleg a gradiensnek van jelentősége, amit a ∇ (nabla) szimbólummal jelölünk és kétváltozós esetben a

$$(1-7) \quad \nabla f(x, y) = \frac{\partial f(x, y)}{\partial x} \vec{e}_x + \frac{\partial f(x, y)}{\partial y} \vec{e}_y \\ = f'_x \vec{e}_x + f'_y \vec{e}_y$$

összefüggéssel definiálunk; feltéve, hogy a parciális differenciálhányadosok léteznek. A képletben \vec{e}_x , \vec{e}_y , az x-, ill. y-irányú egységvektort, f'_x , f'_y a megfelelő irányú parciális differenciálhányadost jelenti. (Az (x, y) argumentumot a rövidség kedvéért elhagyjuk.) A gradiens nagyságát, illetve a +x tengellyel bezárt szögét a

$$(1-8) \quad \|\nabla f\| = \sqrt{f'^2_x + f'^2_y}, \text{ ill. } \operatorname{tg} \alpha = f'_y / f'_x$$

képlet adja.

Ismeretes, hogy a gradiens jelöli ki a függvény egy adott pontjában a legnagyobb változás irányát, s ennek értékét a gradiens nagysága adja meg. (Esetünkben, szemléletesen szólva: a gradiens megmondja, hogy a képfüggvénnyel meghatározott "terepen" merre van a legmeredekebb "kapaszzkodó", vagyis merre tart a "csúcshoz vezető legrövidebb út; nagysága pedig arról tájékoztat, hogy az egységnyi távolságon mekkora a "szintkülönbség", vagyis milyen meredek az út.)

A másodrendű parciális differenciálhányadosok a szélsőérték-helyek meghatározásában játszanak szerepet. A képfeldolgozásban elterjedt a

$$(1-9) \quad \nabla^2 \{f(x, y)\} = \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2} \\ \left[+2 \frac{\partial^2 f(x, y)}{\partial x \partial y} \right] = f''_{xx} + f''_{yy} \left[+2f''_{xy} \right],$$

Laplace-operátor használata (a vegyes parciális differenciálhányadosok nélkül, vagy azokkal) az élek kiemelésére vagy kitérésére.

Digitális képek esetében a differenciálhányadosokat differenciahányadosokkal közelítjük; pl. x-irányban:

$$(1-10) \quad \frac{\partial f(x,y)}{\partial x} \approx \Delta_x q(k,l) = \frac{q(k,l) - q(k-1,l)}{k - (k-1)} = q(k,l) - q(k-1,l);$$

$$(1-11) \quad \frac{\partial^2 f(x,y)}{\partial x^2} \approx \Delta_{xx} q(k,l) = \Delta_x (\Delta_x q(k+1,l)) = \\ = \Delta_x (q(k+1,l) - q(k,l)) = q(k+1,l) - q(k,l) - q(k,l) + \\ + q(k-1,l) = q(k+1,l) + q(k-1,l) - 2q(k,l).$$

Hasonlóan számíthatók ki az y-irányú differenciahányadosok is. A digitális gradiens nagyságának közelítő kiszámításához az (1-8) képlet helyett inkább az (1-35)-tel definiált abszolútérték-normát használjuk (ld. 13.4.3. pont):

$$(1-12) \quad \|\nabla q(k,l)\| \approx |\Delta_x q(k,l)| + |\Delta_y q(k,l)| = \\ = |q(k,l) - q(k-1,l)| + |q(k,l) - q(k,l-1)|.$$

Kis számolással ellenőrizhető, hogy a fentiekkel a digitális (nem teljes) Laplace-operátor

$$(1-13) \quad \nabla^2 q(k,l) = \Delta_{xx} q(k,l) + \Delta_{yy} q(k,l) = \\ = q(k+1,l) + q(k,l+1) + q(k-1,l) + q(k,l-1) - 4q(k,l)$$

alakú. Mindezeket az 1-26. ábrán adott konvolúciós szűrőkkel lehet egyszerűen megvalósítani.

$$\begin{vmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{vmatrix}$$

 Δ_x

$$\begin{vmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{vmatrix}$$

 ∇^2

$$\begin{vmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{vmatrix}$$

 ∇_t^2

1-26. ábra

Differencia-operátorokat megvalósító konvolúciós szűrők

13.2.4. Függvénytranszformációk

A gyakorlatban előforduló képfüggvényeket analitikus formában általában nem lehet felírni, matematikai kezelésük nehézkes. A probléma megoldására kedvelt módszer a sorbafejtés valamilyen jól kezelhető és/vagy könnyen programozható bázisfüggvények szerint.

Egy $f(x, y)$ függvényt felbonthatónak nevezünk a sík egy zárt, derékszögű A tartománya ($a \leq x \leq b$; $c \leq y \leq d$) felett, ha teljesülnek rá a következő feltételek:

- A -ban majdnem mindenütt folytonos (vagyis legfeljebb megszámlálhatóan végtelen sok szakadása van);
- A -ban korlátos és A -n kívül mindenütt 0;
- abszolút integrálható, azaz

$$\iint_{-\infty}^{+\infty} |f(x, y)| dx dy < \infty$$

A képfüggvények mindig eleget tesznek ezeknek a feltételeknek.

A függvényt sor együtthatóinak meghatározását szokás transzformációnak, az eredményül kapott függvényt pedig transzformált képfüggvénynek nevezni. Nyilvánvaló, hogy az eljárás akkor vezet célhoz, ha a felbontás egyértelmű és invertálható; ez a követelmény bizonyos korlátozásokat ró a bázisfüggvényekre. Anélkül, hogy részletekbe mennénk, összefoglaljuk a főbb ismérveket.

13.2.4.1. Ortogonalis függvényrendszerek

A $\varphi_0, \varphi_1, \dots, \varphi_{n-1} = \{\varphi_n\}$ függvény sorozatról akkor mondjuk, hogy ortogonalis függvényrendszert alkot, ha a sík egy zárt, derékszögű A tartományában

$$(1-14) \quad \iint_A \varphi_k(x, y) \varphi_l(x, y) dx dy = \begin{cases} N_k \neq 0, & \text{ha } l=k; \\ 0 & \text{különben.} \end{cases}$$

(Emlékeztetünk rá, hogy két vektor skalárszorzata akkor nulla, ha merőlegesek egymásra; (1-14) ennek általánosítása a függvényekre.) A $k=l$ esetben a $\sqrt{N_k}$ számot a φ_k függvény normájának nevezzük.

A továbbiakban tételezzük fel, hogy $N_k = 1$, vagyis a függvényrendszer ortonormált.

A felbontáshoz az egyes φ_k komponensek együtthatóját az

$$a_k = \iint_A f(x, y) \varphi_k(x, y) dx dy$$

integrállal kell meghatározni. Ezekkel

$$(1-15) \quad f_k(x, y) = \sum_{i=0}^k a_i \varphi_i,$$

$$(1-16) \quad f(x, y) \sim f_k(x, y).$$

(1-16)-ot úgy olvassuk, hogy f_k átlagosan konvergál f -hez, s ezen azt értjük, hogy a négyzetes eltérése nullához tart:

$$(1-17) \quad \lim_{k \rightarrow \infty} \iint_A [f(x, y) - f_k(x, y)]^2 dx dy = 0.$$

Egy ortonormált függvényrendszert zártnak nevezünk, ha a fenti feltételek mellett bármely felbontható függvény sorfejtése (1-17) szerint átlagosan konvergál az adott függvényhez. Ez alapján kimondhatjuk a következő tételt:

Zárt, ortonormált függvényrendszer szerint bármely, a sík egy A tartományában felbontható $f(x, y)$ függvényt egyértelműen és invertálhatóan (1-15) alakú sorba lehet fejteni. A sorfejtés (1-17) értelmében átlagosan konvergál az $f(x, y)$ -hoz és a négyzetes eltérés minimális minden más típusú sorbafejtéshez képest. Egyidejűleg teljesül a

$$\sum_{i=0}^k a_i^2 \leq \iint_A [f(x, y)]^2 dx dy$$

feltétel is.

13.2.4.2. Fourier-transzformáció

Korábban leginkább a trigonometrikus függvények szerinti sorbafejtés terjedt el, ezt Fourier-transzformációnak nevezzük. Nagy számításigénye miatt azonban nehézkes az alkalmazása, ezért egy sereg más "olcsóbban" futtatható transzformációt is kidolgoztak, ezek közül egyet ismertetünk a következő alponban.

Itt és a későbbiekben is, a képfüggvényeket kisbetűvel, a transzformált függvényeket a megfelelő nagybetűvel jelöljük; az előbbiekről mindig feltesszük, hogy felbonthatók a 13.2.4.1. alpont értelmében.

Egy $f(x, y)$ függvény Fourier-transzformáltját, vagy - általánosan elterjedt szóhasználattal - frekvenciaspektrumát az

$$(1-18) \quad F(\omega_x, \omega_y) = \iint_{-\infty}^{+\infty} f(x, y) \cdot \exp[-j2\pi(\omega_x \cdot x + \omega_y \cdot y)] dx dy$$

képlet definiálja. Itt és a továbbiakban \exp az e-alapú hatványfüggvényt, $j=\sqrt{-1}$ a képzetes egységet jelöli. (Az, hogy a határok a végtelenbe vannak kitolva, nem zavaró, ha az A képtartományon kívül $f(x, y)$ -t mindenütt nullának tekintjük.) (1-18)-ből az eredeti függvényt az

$$(1-19) \quad f(x, y) = \frac{1}{4\pi^2} \iint_{-\infty}^{+\infty} F(\omega_x, \omega_y) \cdot \exp [j2\pi(\omega_x \cdot x + \omega_y \cdot y)] d\omega_x d\omega_y$$

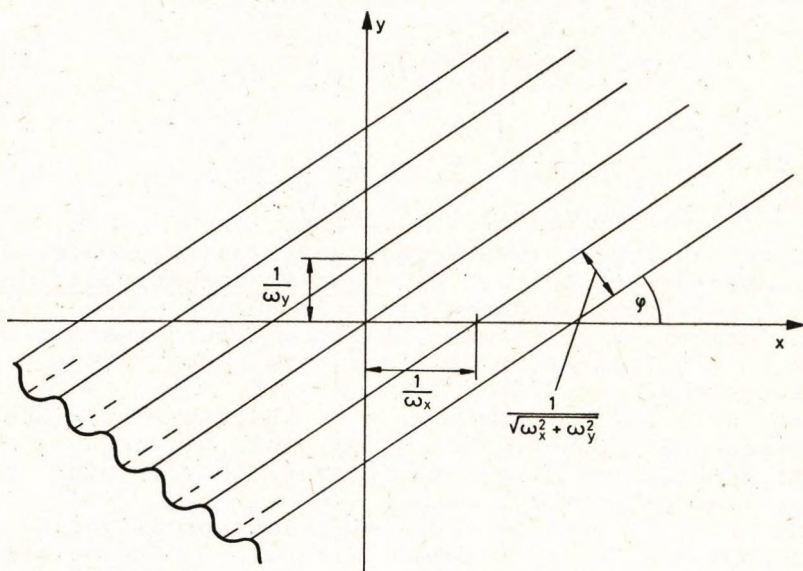
szerinti inverz Fourier-transzformációval kaphatjuk meg.

(1-18)-at az alábbiakban értelmezzük:

Ismeretes, hogy

$$e^{j2\pi(\omega_x \cdot x + \omega_y \cdot y)} = \underline{\cos}(\omega_x \cdot x + \omega_y \cdot y) + j \underline{\sin}(\omega_x \cdot x + \omega_y \cdot y).$$

Tekintsük pl. a cos $(\omega_x \cdot x + \omega_y \cdot y)$ tagot, ez egy síkhullámot ír le az $(x-y)$ síkban, amelynek x -, ill. y -irányú frekvenciája ω_x , ill. ω_y , vagyis a megfelelő hullámhosszak $\frac{1}{\omega_x}$, ill. $\frac{1}{\omega_y}$ (ld. 1-27. ábra).



1-27. ábra
Cos-típusú síkhullám

A hullámfront az x-tengellyel

$$\varphi = \underline{\text{arc tg}} \left(\frac{\omega_x}{\omega_y} \right)$$

szöveget zár be, s az erre merőleges irányban a síkfrekvencia $\sqrt{\omega_x^2 + \omega_y^2}$. Ugyanez érvényes értelemszerűen a szinuszhullámú összetevőre is.

Ezek szerint az (1-18)-cal definiált $F(\omega_x, \omega_y)$ nem más, mint az $f(x, y)$ -nak trigonometrikus függvények szerinti felbontásában a $\sqrt{\omega_x^2 + \omega_y^2}$ síkfrekvenciájú tag együtthatója: azt mondja meg, hogy az adott síkhullám mekkora amplitúdóval szerepel a felbontásban. (Pontosabban - mivel a trigonometrikus függvények nem normáltak - az amplitúdó ennek $\frac{1}{4\pi^2}$ -szerese.)

A digitális képek felbontásánál a helyzet egy kicsit más, mivel a képfüggvény csak diszkrét (k, ℓ) ($k, \ell = 0, +1, +2, \dots$) pontokban van értelmezve, és ennek megfelelően a frekvenciaspektrum sem folytonos, hanem csak a diszkrét u, v ($u, v = 0, +1, +2, \dots$) frekvenciák fordulnak benne elő.

Jelölje a felbontandó kép méretét x -, ill. y -irányban K , ill. L , akkor (1-18) a következőképpen módosul:

$$F(u, v) = \frac{1}{KL} \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} f(k, \ell) \exp \left[-j2\pi \left(\frac{ku}{K} + \frac{\ell v}{L} \right) \right].$$

Ezekkel az együtthatókkal a felbontás

$$f(k, \ell) = \sum_{u=0}^{K-1} \sum_{v=0}^{L-1} F(u, v) \exp \left[j2\pi \left(\frac{ku}{K} + \frac{\ell v}{L} \right) \right].$$

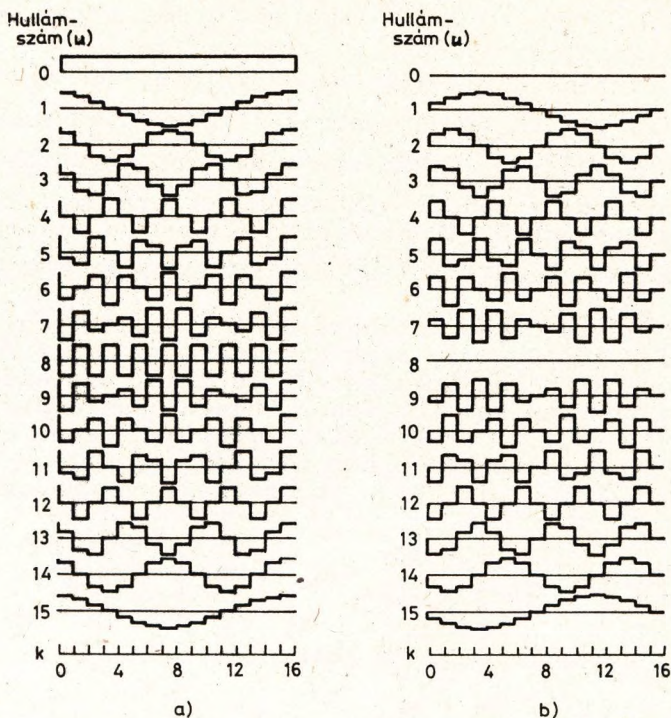
Fontos megjegyezni, hogy ebben az esetben az $F(u, v)$ együtthatójú

$$\cos \left[2\pi \left(\frac{ku}{K} + \frac{\ell v}{L} \right) \right], \text{ ill. } \sin \left[2\pi \left(\frac{ku}{K} + \frac{\ell v}{L} \right) \right]$$

bázisfüggvények csak durva közelítései a folytonos cos, ill. sin függvénynek: a magasabb frekvenciákon négyszöghullámba mennek át. Ennek érzékeltetésére megadjuk az egydimenziós

$$\cos \left(2\pi \frac{ku}{K} \right), \text{ ill. } \sin \left(2\pi \frac{ku}{K} \right)$$

bázisfüggvények alakját $K=16$ esetre. (Ld. 1-28. ábra; a valóságban az ezekhez hasonló síkhullámokkal történik a közelítés, amelyeknek ez a "keresztmetszetük" a hullámfrontra merőleges irányban.)



1.28. ábra
Fourier-bázisfüggvények

13.2.4.3. Hadamard-transzformáció

Az Hadamard-transzformáció a normalizált Hadamard-mátrixokon alapul. Ezek ± 1 elemekből álló négyzetes mátrixok, és ortonormáltak, vagyis érvényes a

$$\underline{\underline{H}} \cdot \underline{\underline{H}}^T = \underline{\underline{E}}$$

összefüggés (ld. 13.1.2. pont).

A legkisebb ortonormált Hadamard-mátrix másodrendű:

$$\underline{\underline{H}}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

A transzformációhoz csak az $N = 2^n$ ($n = 0, 1, 2, 3, \dots$) rendű mátrixokat használják; ezek rekurzívan építhetők fel a

$$\underline{H}_{2i} = \frac{1}{\sqrt{2}} \begin{pmatrix} \underline{H}_i & \underline{H}_i \\ \underline{H}_i & -\underline{H}_i \end{pmatrix}$$

képlet szerint. Így pl.:

$$\underline{H}_4 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ \hline 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}; \quad \underline{H}_8 = \frac{\sqrt{2}}{4} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ \hline 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

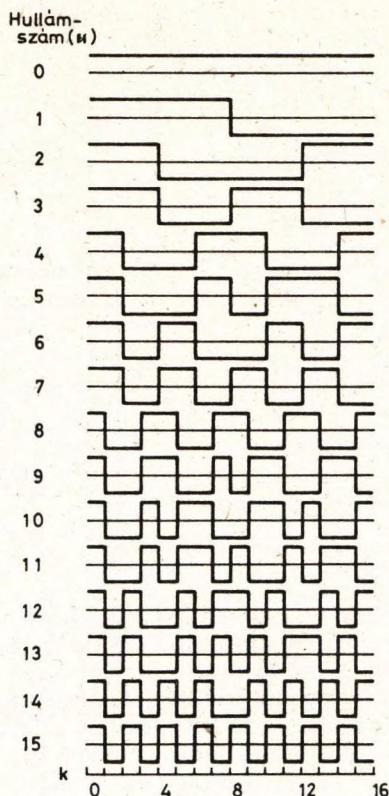
stb.

A K-ad rendű Hadamard-mátrixok sorait $u = 0, 1, \dots, K-1$ frekvenciájú négyzőghullámokkal ábrázolhatjuk, ezeket Walsh-függvényeknek nevezik. Eszerint az Hadamard-transzformáció nem más, mint egy függvény felbontása Walsh-függvények szerint. A gyakorlatban - a bemutatott helyett - a mátrixok, ill. Walsh-függvények "rendezett" sorozatát használják, amelyekben a frekvencia (ami megfelel a mátrix egyes soraiban az előjelváltások számának) soronként nő. Legyenek a képméretetek $K = 2^m$, ill. $L = 2^n$; ezekkel a sorfejtés együtthatóit az

$$F(u, v) = \frac{1}{\sqrt{KL}} \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} f(k, \ell) (-1)^{e(k, \ell, u, v)}$$

$$e(k, \ell, u, v) = \sum_{i=0}^{m-1} (u_{m-i} - u_{m-i-1}) \cdot k_i + \sum_{i=0}^{n-1} (v_{n-i} - v_{n-i-1}) \cdot \ell_i$$

képlet adja, ahol u_i, v_i, k_i, ℓ_i az adott számok 2^i helyiértékű bitjét jelöli. A bázisfüggvényeket $K=16$ esetre egydimenziósan az 1-29. ábrán mutatjuk be.



1-29. ábra
Egydimenziós Walsh bázisfüggvények

13.3. VALÓSZÍNŰÉGSZÁMÍTÁS

A digitális képeket sztochasztikus sokaságként is fel-foghatjuk, s így a képeket valószínűségyszámítási módszerekkel is kiértékelhetjük. Ebben az alpontban vázlatosan át-tekintjük az idevágó fontosabb tudnivalókat.

13.3.1. Események és eseménytér

A valószínűségyszámítás a matematika egyik legmodernebb és talán a gyakorlatban legtöbbet alkalmazott ága. Feladata a véletlen tömegjelenségek törvényszerűségeinek matematikai vizsgálata. Véletlen jelenségekről beszélünk, ha egy folyamatnak vagy kísérletnek több lehetséges eredménye van, amelyeket a tekintetbe vett körülmények nem határoznak meg egyértelműen, így előre nem tudhatjuk biztosan, hogy melyik fog bekövetkezni. Végezzünk el egy olyan kísérletet, amelynek lehetséges eredményei véletlen jelenségeknek tekinthetők (pl. kockadobás). Jelöljük Ξ -vel a lehetséges összes ered-

mény halmazát: ebben a kísérlet minden lehetséges eredményének pontosan egy halmazelem felel meg.

A Ξ alaphalmaz tetszés szerinti számú eleméből képzett E_1 részhalmazokat eseményeknek, ezek \mathcal{E} halmazát pedig eseménytérnek nevezzük.

Az eseményekkel kapcsolatban értelmezhetünk bizonyos műveleteket (eseményalgebra) és bevezethetünk bizonyos elnevezéseket, az alábbiak szerint:

- A D és az E esemény összege azt jelenti, hogy a kettő közül legalább az egyik bekövetkezik; ezt a halmazok egyesítése (úniója) fejezi ki, jelölése:

$$D \cup E.$$

- A D és az E esemény szorzata azt jelenti, hogy mindkettő egyidejűleg bekövetkezik; ezt a halmazok metszete (közös része) fejezi ki, jelölése:

$$D \cap E.$$

- Két eseményt kizáró eseménynek nevezünk, ha szorzatuk (metszetük) az üres halmaz, vagyis

$$D \cap E = \emptyset.$$

(Ilyenek pl. az elemi események.)

- A D és az E esemény különbsége azt jelenti, hogy D bekövetkezik, és egyidejűleg E nem; jelölése:

$$D \setminus E (= D \cap \bar{E}).$$

- A D esemény ellentettje azt az eseményt jelenti, hogy D nem következik be; jelölése:

$$\bar{D} (= I \setminus D).$$

(Gyakran szokás az események (ill. halmazok) egyesítését (únióját) a $D+E$, közös részét (metszetét) a $D \cdot E$, különbségét a $D-E$ szimbólummal jelölni, ez azonban félreértésekhez vezethet, ezért nem használjuk.)

Vegyük észre, hogy a fenti műveletek nem vezetnek ki az \mathcal{E} eseménytérből (még végtelen sok elem esetén sem).

Az eseménytér eseményalgebrát képez, ha teljesül az alábbi 3 feltétel:

1. Az \mathcal{E} halmaznak eleme a \emptyset üres halmaz és a $\bar{\Xi}$ alaphalmaz képlettel $\emptyset, \Xi \in \mathcal{E}$. \emptyset -t lehetetlen eseménynek nevezzük, mivel bekövetkezése azt jelentené, hogy a kísérlet eredménye nem bekövetkezik meg egyik lehetséges eredménnyel sem, ami lehetetlen. Ξ a biztos esemény, bekövetkezése ugyanis azt jelenti, hogy az eredmény a lehetséges események egyike.

2. Az \mathcal{E} elemeinek tetszés szerinti egyesítésével (uniójával; jele: \cup) képzett halmazok szintén elemei \mathcal{E} -nek (számuk végtelen sok is lehet!). Képlettel:

$$E_1, E_2, \dots \in \mathcal{E} \quad \text{esetén} \quad \bigcup_i E_i \in \mathcal{E}$$

Az eseménytér elemi eseményei azok az események, amelyek nem állíthatók elő más események uniójaként. (Kivéve: \emptyset -t nem tekintjük elemi eseménynek.)

Megjegyezzük még, hogy ha az egymást páronként kizáró események egyesítése a biztos esemény, akkor teljes eseményrendszer alkotnak. Ez biztosan teljesül bármely eseménytér elemi eseményeire, de nem csak azokra.

3. Az \mathcal{E} tetszés szerinti elemeinek különbség-halmaza (jele: \setminus) szintén eleme \mathcal{E} -nek; vagyis

$$E_1, E_2 \in \mathcal{E} \quad \text{esetén} \quad E_1 \setminus E_2 \in \mathcal{E}.$$

Az E esemény akkor következik be, ha a végzett kísérlet eredménye az E részhalmazhoz tartozik; egyidejűleg mindazok az események bekövetkeznek, amelyeknek az adott eredmény az eleme.

Egy dobókockás kísérletben jelöljük a lehetséges eredményeket $\Xi = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ -zal. (Pl. az $x_i \in \Xi$ azt a dobást jelenti, amelynek eredményeként az i szám van felül.) Álljon az \mathcal{E}_0 eseménytér a összes lehetséges részhalmazából; az \mathcal{E}_0 elemi eseményei az $\{x_i\}$ események. Az $\{x_3\}$ elemi esemény pl. azt jelenti, hogy a kockadobás eredménye 3, az $\{x_1, x_2\}$ (nem elemi) esemény pedig azt, hogy 1 vagy 2 volt. Ha azonban az

$$\mathcal{E}_1 = \{\emptyset, \{x_1, x_2\}, \{x_3, x_4, x_5, x_6\}, \{x_1, \dots, x_6\} = \Xi\}$$

eseményteret definiáljuk (vegyük észre, hogy ez megfelel a fenti 3 feltételnek), akkor pl. a 3-as dobását nem tekintjük elemi eseménynek! (Az \mathcal{E}_1 eseménytérhez - a lehetetlen és a biztos eseményen kívül - két elemi esemény tartozik: az, hogy a dobás eredménye 3-nál kisebb, ill. nem kisebb.)

Eddigi példáinkban az \mathcal{E}_0 eseménytér $\{x_1\}, \dots, \{x_6\}$ elemi eseményei teljes eseményrendszer képeznek; ugyanez áll az $\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}$ elemeire is, bár ezek nem elemi események. Ha azonban az \mathcal{E}_1 halmazt tekintjük eseménytérnek, az $\{x_1, x_2\}$ és az $\{x_3, \dots, x_6\}$ már elemi események, és nyilván szintén teljes eseményrendszer alkotnak.

13.3.2. Valószínűségek és alaptételek

Végezzünk el n -szer egy kísérletet, amelynek eredményei egymástól függetlenek (ld. alább), és tegyük fel, hogy az eseménytér egy E eseménye (példánkban mondjuk a 3 dobása) k -szor következett be. A k számot az E esemény gyakoriságának, a $0 \leq \frac{k}{n} \leq 1$ hányadost pedig relatív gyakoriságának nevezzük.

A nagy számok törvénye értelmében az utóbbi n növekedésével stabilitást mutat a

$$0 \leq P(E) = \lim_{n \rightarrow \infty} \frac{k}{n} \leq 1$$

összefüggés szerint. Legyen a $\bar{\mathcal{E}}$ eseménytérben \mathcal{E} tetszőleges eseményalgebra, ekkor a P halmazfüggvényt amely a

$$P: \mathcal{E} \rightarrow [0, 1]$$

leképezést valósítja meg valószínűségi mértéknek, vagy röviden valószínűségnek nevezzük, ha teljesül a következő három axióma:

- I. Minden $E \in \mathcal{E}$ eseményre $0 \leq P(E) \leq 1$;
- II. $P(\bar{\mathcal{E}}) = 1$;
- III. Ha E_1, E_2, \dots egymást páronként kizáró események \mathcal{E} -ben, akkor az egyesítésükkel keletkezett esemény (amely a fenti 2. feltétel szerint szintén eleme E -nek!) valószínűségét az egyes események valószínűségének összege adja. Képlettel:

$$P\left(\bigcup_i E_i\right) = \sum_i P(E_i); \text{ ha } E_i \cap E_j = \emptyset \forall i \neq j.$$

A $(\bar{\mathcal{E}}, \mathcal{E}, P)$ hármast valószínűségi mezőnek nevezzük. (A fentiekből következik, hogy $P(\emptyset) = 0$. A továbbiakban \emptyset -val illetve I -vel jelöljük és lehetetlen, illetve biztos eseménynek nevezzük mindazokat az eseményeket, amelyeknek valószínűsége 0, illetve 1. (Eddigi példánkban $\emptyset \equiv \sigma$ és $\bar{\mathcal{E}} \equiv I$, de ez nem szükségszerű.)

A mondottak alapján könnyen beláthatók a következő összefüggések:

$$\begin{array}{lll} \bar{\emptyset} = I; & \bar{I} = \emptyset; & \bar{\bar{E}} = E; \\ E \cup E = E; & E \cup \bar{E} = I; & E \cup I = I; \quad E \cup \emptyset = E; \\ E \cap E = E; & E \cap \bar{E} = \emptyset; & E \cap I = E; \quad E \cap \emptyset = \emptyset; \end{array}$$

illetve

$$\begin{aligned} P(\emptyset) &= 0; & P(I) &= 1; & P(\bar{E}) &= 1 - P(E); & P(D \cup E) &= \\ & & & & & & &= P(D) + P(E) - P(D \cap E). \end{aligned}$$

Érdeemes megjegyezni még, hogy kizáró eseményekre

$$P(D \cap E) = P(O) = 0, \text{ és így } P(D \cup E) = P(D) + P(E);$$

vagyis annak valószínűsége, hogy legalább az egyikük bekövetkezik (= az egyesített esemény) egyenlő a valószínűségük összegével.

Jelöljük $P(D|E)$ -vel a D esemény bekövetkezésének valószínűségét akkor, ha az E esemény már bekövetkezett; ezt feltételes valószínűségnek nevezzük. Definíció szerint:

$$(1-20) \quad P(D|E) = \frac{P(D \cap E)}{P(E)}.$$

(A definíciónak csak $P(E) \neq 0$ esetén van értelme.)

A gyakorlati életben két eseményt akkor nevezünk függetlennek, ha egyikük bekövetkezése sem függ a másik bekövetkezésétől. Matematikailag ezt úgy fejezhetjük ki, hogy pl.:

$$P(D|E) = P(D),$$

vagy (1-20)-ba behelyettesítve és átrendezve:

$$(1-21) \quad P(D \cap E) = P(D) \cdot P(E).$$

Szavakkal: független eseményekről beszélünk, ha együttes bekövetkezésük valószínűsége egyenlő a valószínűségük szorzatával.

A feltételes valószínűség definíciójának átalakításával megkaphatjuk a valószínűségek szorzási szabályát:

$$(1-20/a) \quad P(D \cap E) = P(D|E) \cdot P(E).$$

Alkossanak az E_1, E_2, \dots, E_n események teljes eseményrendszert és legyen $P(E_k) > 0$ ($k = 1, 2, \dots, n$; $n = \infty$ megengedett). Ekkor bármely D eseményre igaz, hogy

$$(1-22) \quad P(D) = \sum_{k=1}^n P(D|E_k) \cdot P(E_k)$$

Ez a teljes valószínűség tétele.

(Bizonyítás: Abból, hogy E_1, E_2, \dots, E_n teljes eseményrendszer, következik, hogy $D = D \cap I = D \cap (E_1 \cup E_2 \cup \dots \cup E_n) = D \cap E_1 \cup D \cap E_2 \cup \dots \cup D \cap E_n$, ahol $(D \cap E_i) \cap (D \cap E_k) = O$, $\forall i \neq k$. Ha az egyenlet jobb oldalára alkalmazzuk az egymást kizáró események összegének valószínűségére és a valószínűségek szorzására vonatkozó szabályokat (II. axióma, ill. (1-20/a) felírhatjuk, hogy

$$P(D) = \sum_{k=1}^n P(D \cap E_k) \quad \text{és} \quad P(D \cap E_k) = P(D|E_k) \cdot P(E_k);$$

amiből (1-22) következnek).

Ha E_1, E_2, \dots, E_n teljes eseményrendszer, $P(E_k) > 0$ ($k = 1, 2, \dots, n$), D pedig olyan esemény, amelyre $P(D) > 0$, akkor

$$(1-23) \quad P(E_k|D) = \frac{P(D|E_k) \cdot P(E_k)}{\sum_{i=1}^n P(D|E_i) \cdot P(E_i)},$$

ez Bayes tétele.

(Bizonyítás: A feltételes valószínűség (1-20) definíciója szerint

$$P(E_k|D) = \frac{P(E_k \cap D)}{P(D)}, \quad \text{ill.} \quad P(D|E_k) = \frac{P(D \cap E_k)}{P(E_k)}.$$

Ebből következik, hogy

$$P(E_k|D) \cdot P(D) = P(D|E_k) \cdot P(E_k)$$

vagyis

$$P(E_k|D) = \frac{P(D|E_k) \cdot P(E_k)}{P(D)},$$

ebből (1-22) behelyettesítésével (1-23)-at kapjuk.)

A tétel azt mondja ki, hogy ha ismerjük a D esemény feltételes valószínűségeit az E_1, E_2, \dots, E_n teljes eseményrendszerre vonatkozólag, továbbá ismerjük az E_k események valószínűségeit, akkor az E_k ($k = 1, 2, \dots, n$) eseményeknek a D feltételre vonatkozó feltételes valószínűségét kiszámíthatjuk. Szokás a $P(E_k|D)$ valószínűségeket "a posteriori" valószínűségeknek, a $P(E_k)$ valószínűségeket "a priori" valószínűségeknek is nevezni.

(Más szóval: következtethetünk az okok ($P(E_k|D)$) valószínűségére a megfigyelések ($P(D|E_k)$) valószínűségeiből.)

Az eddigiek képezik a klasszikus valószínűségszámítás alapjait. A képletek alkalmazása nagymértékben épül a kombinatorika képleteinek felhasználására - ezekre itt nem térünk ki. Megjegyezzük azonban, hogy nehézségek merülhetnek fel abban az esetben, ha az elemi események száma végtelen. Ilyenkor segíthet, ha a valószínűségek meghatározását geo-

metriai alakzatok mértékeinek (ívhossz, terület, térfogat) meghatározására vezetjük vissza.

13.3.2. Valószínűségi változók

Legyen adva egy (Ξ, \mathcal{E}, P) valószínűségi mező. Rendeljük hozzá a Ξ minden eleméhez egy-egy valós számot. Így egy olyan függvényhez jutunk, amelynek értelmezési tartománya a kísérlet kimeneteleinek Ξ halmaza, értékészletét pedig a hozzájuk rendelt számok alkotják.

$$\xi: \Xi \rightarrow \mathbb{R}$$

A leképezést valószínűségi változónak (VV) nevezzük (jelölésére a görög kisbetűket szokták használni), ha teljesülnek bizonyos nem túl szigorú feltételek. (Itt, és a továbbiakban \mathbb{R} a valós számok halmazát jelöli.)

Az angol nyelvű irodalomban a "random variable" (= "véletlen változó") a magyar nyelvűben a "valószínűségi változó" (az orosz nyelvűben ue., természetesen oroszul) elnevezés különösen szerencsétlen, mivel a VV sem nem valószínűségi, sem nem változó, a szokásos értelemben. Véletlen jellege abban van, hogy előre nem tudjuk: a Ξ melyik eleme fog bekövetkezni; ha azonban a kísérlet egy kimenetele bekövetkezett, akkor ez már egyértelműen meghatározza a VV értékét. Ennek megfelelően általában van értelme arról beszélni, hogy a ξ értéke milyen valószínűséggel lesz egy adott számnál kisebb: ez egyenlő azon elemi események úniojának bekövetkezési valószínűségével, amelyekhez rendelt függvényértékek nem nagyobbak a megadott számnál.

Legyen $z \in \mathbb{R}$ tetszőleges szám; ekkor az

$$(1-24/a) \quad S_{\xi}(z) = P(\{x \in \Xi \mid \xi(x) < z\}),$$

vagy röviden

$$(1-24/b) \quad S(z) = P(\xi < z)$$

valószínűséget, mint z függvényét a ξ VV valószínűség-eloszlás függvényének, vagy röviden eloszlásfüggvényének nevezzük. (A képletben az $x \in \Xi$ a vizsgált folyamat vagy kísérlet egy lehetséges eredményét jelöli. Megjegyezzük, hogy a lehetséges eredmények száma nem szükségképpen véges, sőt kontinuum-számosságú is lehet.) Az eloszlásfüggvény nyilván függ attól, hogy a VV hogyan van definiálva, ez azonban (1-24/b)-ből is egyértelműen kiderül; hasonlóan nem okoz félreértést, ha elhagyjuk a VV argumentumát.

Az eloszlásfüggvény - könnyen beláthatóan - az alábbi tulajdonságokkal rendelkezik:

- monoton növekvő;

- $\lim_{z \rightarrow -\infty} S(z) = 0$ és $\lim_{z \rightarrow \infty} S(z) = 1$;

- jobbról folytonos és balról minden pontban van határértéke.

(Az állítást meg is fordíthatjuk: bebizonyítható, hogy ha egy S függvény rendelkezik a fenti 3 tulajdonsággal, akkor létezik olyan VV , amelynek S az eloszlásfüggvénye.)

Egy VV -t diszkrétnek nevezünk, ha értékkészlete megszámlálható. Legyenek ξ lehetséges értékei z_1, z_2, \dots ,

akkor a

$$p_i = P(\xi = z_i), \quad (i=1, 2, \dots)$$

valószínűségeket a ξ eloszlásának hívjuk; az eloszlásfüggvénye pedig

$$S(z) = P(\xi < z) = \sum_{z_i < z} p_i;$$

ami szakaszosan konstans értékű lépcsősfüggvény.

Ha a VV értékkészlete kontinuum számosságú, folytonosnak, ha még az eloszlásfüggvénye a teljes számegegyenesen folytonos és majdnem mindenütt differenciálható is, akkor abszolút folytonosnak nevezzük.

Ebben az esetben az

$$(1-25) \quad s(z) = \frac{d S(z)}{dz} = S'(z)$$

függvényt sűrűségfüggvénynek nevezzük. Érvényesek a következő összefüggések:

$$\int_{-\infty}^{+\infty} s(z) dz = 1;$$

$$P(\xi < a) = \int_{-\infty}^a s(t) dt = S(a);$$

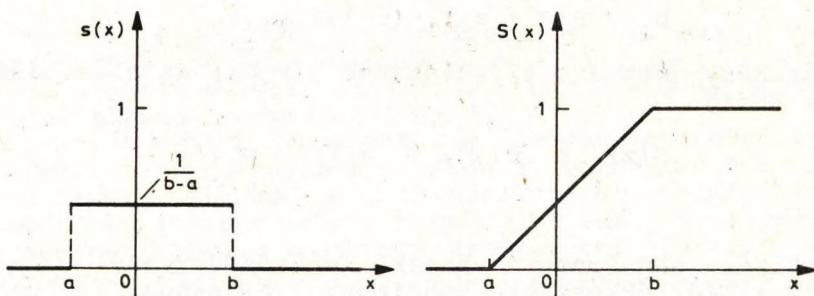
$$P(a \leq \xi < b) = \int_a^b s(t) dt = S(b) - S(a); \text{ illetve}$$

$$P(\xi=z) = \lim_{h \rightarrow 0} P(z \leq \xi < z+h) = \lim_{h \rightarrow 0} \int_z^{z+h} s(t) dt = \lim_{h \rightarrow 0} [(S(z+h) - S(z))] = 0$$

az $S(z)$ függvény jobb oldali folytonossága miatt.

Nézzük most a legfontosabb folytonos eloszlásokat.

Egyenletes eloszlás. A ξ valószínűségi változót az (a, b) intervallumon egyenletes eloszlásúnak nevezzük, ha sűrűségfüggvénye, ill. eloszlásfüggvénye (ld. 1-30. ábra):



1-30. ábra

Egyenletes eloszlás sűrűség- és eloszlásfüggvénye

$$s(z) = \begin{cases} \frac{1}{b-a}, & \text{ha } a \leq z < b; \\ 0, & \text{egyébként;} \end{cases}$$

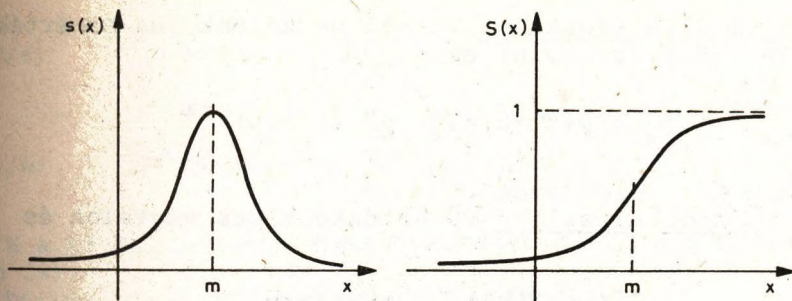
$$S(z) = \begin{cases} 0, & \text{ha } z < a; \\ \frac{z-a}{b-a}, & \text{ha } a \leq z < b; \\ 1, & \text{ha } b \leq z. \end{cases}$$

Normális eloszlás. A ξ valószínűségi változót normális eloszlásúnak nevezzük, ha sűrűségfüggvénye, ill. eloszlásfüggvénye (ld. 1-31. ábra):

$$s(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right) \quad (-\infty < t < +\infty)$$

$$S(z) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right) dt \quad (-\infty < z < +\infty)$$

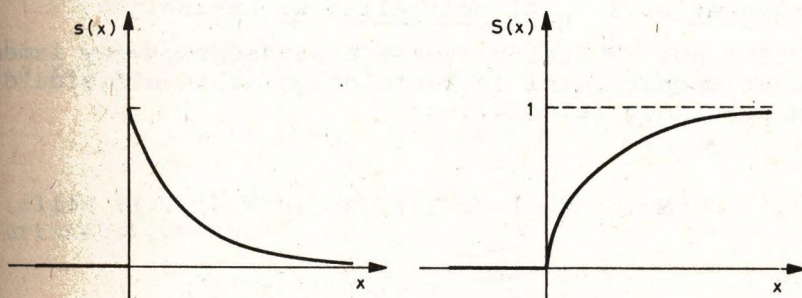
$(m \in \mathbb{R}, \sigma \in \mathbb{R}, \sigma > 0)$.



1-31. ábra

Normális eloszlás sűrűség- és eloszlásfüggvénye

Exponenciális eloszlás. A ξ valószínűségi változót exponenciális eloszlásúnak nevezünk, ha sűrűségfüggvénye, ill. eloszlásfüggvénye (ld. 1-32. ábra):



1.32. ábra

Exponenciális eloszlás sűrűség- és eloszlásfüggvénye

$$s(z) = \begin{cases} \lambda e^{-\lambda z}, & \text{ha } z \geq 0; \\ 0, & \text{egyébként;} \end{cases}$$

$$S(z) = \begin{cases} 1 - e^{-\lambda z}, & \text{ha } z \geq 0; \\ 0, & \text{egyébként.} \end{cases}$$

$(\lambda \in \mathbb{R}, \lambda > 0)$.

A diszkrét eloszlású valószínűségi változók közül azt a két esetet említjük meg, amikor az értékkészlet a nemnegatív egész számok halmaza.

Binomiális eloszlású VV-ről beszélünk, ha az értékkészlet véges $(0, 1, 2, \dots, n)$ és

$$p_k = P(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

ahol $0 \leq p \leq 1$ valós szám.

A Poisson eloszlású VV értékkészlete végtelen és

$$p_k = P(\xi = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

ahol λ tetszőleges valós szám.

13.3.3. Statisztikai jellemzők

Az eloszlás- és sűrűségfüggvények jellemzésére az $M_k(\xi)$ momentumokat és a $D_k(\xi)$ centrálisokat használjuk.

Folytonos VV esetén ezeket a sűrűségfüggvény ismeretében lehet meghatározni (a képfeldolgozásban előforduló esetekben ez mindig létezik):

$$(1-27/a) \quad M_k(\xi) = \int_{-\infty}^{+\infty} z^k s(z) dz, \quad (k=1, 2, \dots); \text{ ill.}$$

$$(1-27/b) \quad D_k(\xi) = \int_{-\infty}^{+\infty} (z - M_1(\xi))^k \cdot s(z) dz, \quad (k=2, 3, \dots).$$

Fontos speciális esetként megjegyezzük, hogy a

$$(1-28/a) \quad \mu \equiv M_1(\xi) = \int_{-\infty}^{+\infty} z \cdot s(z) dz$$

kifejezést a VV várhatóértékének, vagy átlagának a

$$(1-28/b) \quad \sigma \equiv \sqrt{D_2(\xi)} = \left[\int_{-\infty}^{+\infty} (z - \mu_1)^2 \cdot s(z) dz \right]^{1/2}$$

kifejezést pedig szórásnak is nevezik; ez utóbbi a várhatóértéktől való négyzetes eltérés várhatóértékének pozitív négyzetgyöke.

Diszkrét VV esetében az integrálok véges összegzésre egyszerűsödnek; speciálisan

$$(1-29/a) \quad \mu \equiv M_1(\xi) = \sum_{n=1}^N z_n \cdot p_n, \text{ illetve}$$

$$(1-29/b) \quad \sigma^2 \equiv D_2(\xi) = \sum_{n=1}^N (z_n - \mu)^2 \cdot p_n;$$

ahol N a kísérlet lehetséges eredményeinek száma (lehet végtelen is).

Térjünk vissza a kockadobási példánkhoz:

$\Xi = \{x_1, \dots, x_6\}$; $N=6$; ξ_0 pedig összes részhalmazainak halmaza, amelynek $\{x_n\}$ elemi eseménye azt jelenti, hogy a dobott szám n.

Mivel az elemi események egymást páronként kizárják, és egyesítjük a biztos eseményt:

$$P(\{x_1\} \cup \dots \cup \{x_6\}) = P(\{x_1\}) + \dots + P(\{x_6\}) = P(I) = 1.$$

Ha a kocka szabályos, mindegyik szám dobása egyenlő valószínűségű, s így

$$P(\{x_1\}) = \dots = P(\{x_6\}) = \frac{1}{6}.$$

Definiáljuk a VV-t úgy, hogy értéke legyen egyenlő a dobott szám értékével, vagyis

$$\xi(x_n) = n \quad (n = 1, 2, \dots, 6).$$

Nyilvánvaló, hogy ξ diszkrét VV, s így (1-29/a), ill. (1-29/b) szerint

$$\mu = M_1(\xi) = \sum_{n=1}^6 n \cdot \frac{1}{6} = 3,5; \text{ ill.}$$

$$\sigma^2 = D_2(\xi) = \sum_{n=1}^6 (n-3,5)^2 \cdot \frac{1}{6} \approx 2,9.$$

Megjegyezzük, hogy a P valószínűségi mértéket másképpen is lehet választani (ha pl. a kocka nem szabályos); a matematikai statisztika foglalkozik azzal, hogy egy eseménytérhez milyen mérték választása célszerű. Becslési módszerekkel a 31.2.3. pontban foglalkozunk; de emlékeztetünk a nagy számok már említett törvényére.

Felmerül a kérdés, hogy mit mondhatunk több VV kapcsolatról. (Szemléletes példa: egy szénrakás térfogatát és alapterületét tekintve VV-nek, nyilvánvaló, hogy egyik ismeretében lényegesen jobban következtethetünk a másikra mint

anélkül; más szóval ezek nem függetlenek egymástól.) Általában a kovarianciát és a korrelációs együtthatót szokás kiszámítani, amit két ξ, η VV esetében a

$$\text{cov}(\xi, \eta) = \mu(\xi \cdot \eta) - \mu(\xi) \cdot \mu(\eta),$$

illetve az

$$r(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{\sigma^2(\xi) \cdot \sigma^2(\eta)}$$

összefüggések definiálnak.

Ha $r(\xi, \eta) = +1$, akkor a két VV között lineáris összefüggés van. Ha $r(\xi, \eta) = 0$, a két VV csak akkor tekinthető függetlennek, ha az eloszlásfüggvényük normális. Bár egyéb eloszlások esetében ez a feltétel általában nem elegendő, általában igaz, hogy minél kisebb a korrelációs együttható, annál inkább függetleneknek tekinthetők a VV-k.

Megjegyezzük még, hogy a képfeldolgozásban többdimenziós VV-k is előfordulnak. Ezekkel bővebben nem foglalkozunk; konkrét példák a 3. részben találhatóak.

13.4. DISZKRÉT GEOMETRIA

13.4.1. Alapfogalmak és definíciók

A \mathbb{N} diszkrét képsík egy x-y tengelyű derékszögű Descartes-koordinátarendszer azon pontjainak halmaza, amelyeknek koordinátái egészszámok, továbbá y-tengelye felülről lefelé irányul.
így:

$$\mathbb{N} = \{P(x, y) \mid x \in \mathbb{N} \wedge y \in \mathbb{N}\}.$$

A diszkrét képsík elemei a képpontok. A korlátos képsíkot mindig téglalap alakúnak tekintjük és az origót a téglalap bal felső sarkába helyezük. Nyilvánvaló, hogy ebben az esetben a képpontok száma véges.

13.4.2. Topológiai jellemzők

1. Szomszédság

A diszkrét képsíkon a pontokat növekvő képkoordináták szerint (x-irányban balról jobbra, y-irányban felülről lefelé) rendezzük.

A (k, l) koordinátájú P képpontnak 4, illetve 8 szomszédja van attól függően, hogy csak a koordinátatengelyekkel párhuzamos, illetve $\pm 45^\circ$ -os haladási irányokat is megengedünk-e. A négy illetve nyolc szomszédos pont összességét (4)-, illetve (8)-szomszédságnak nevezzük (ld. 1-33. ábra):

(4)-szomszédság: $\{P_0(k+1,1), P_1(k,1-1), P_2(k-1,1), P_3(k,1+1)\};$

(8)-szomszédság: $\{P_0(k+1,1), P_1(k+1,1-1), P_2(k,1-1), P_3(k-1,1-1), P_4(k-1,1), P_5(k-1,1+1), P_6(k,1+1), P_7(k+1,1+1)\};$

A (8)-szomszédság nyilván magában foglalja a (4)-szomszédságot.

2. Útvonal

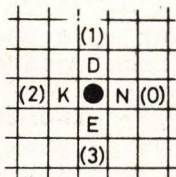
A (4/8) útvonal a (kép)-pontok véges sorozata, amelyben - az elsőt kivéve - bármelyik pont (4/8)-szomszédos az előtte állóval. A sorozat első és utolsó elemét végpontnak nevezük. Az útvonal akkor egyszeres, ha - a végpontok kivételével - minden elemének pontosan két (4/8)-szomszédja van benne (ld. 1-34. ábra).

Rendeljük hozzá egy pont szomszédságához az 1-33. ábrán zárójelben írt $\{0-3\}$, ill. $\{0-7\}$ számokat; ezeket Free-

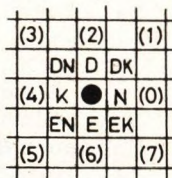
mann-féle iránykódvektornak, röviden iránykódnak nevezünk, és

$\vec{\alpha}_k$ -val jelöljük. Ha $\vec{\alpha}_k = \vec{\alpha}_m \pmod{4/8}$, akkor $\vec{\alpha}_k = \vec{\alpha}_m$.

Ábrázoljunk egy egyszeres utat úgy, hogy a - kezdőponttól kiindulva - felsoroljuk a következő pontra mutató iránykódokat; ekkor az út lánckódját kapjuk; ez a haladás irányát is előírja. (Az 1-34. ábrán mindkét út mindkét irányú lánckódját is feltüntettük.)

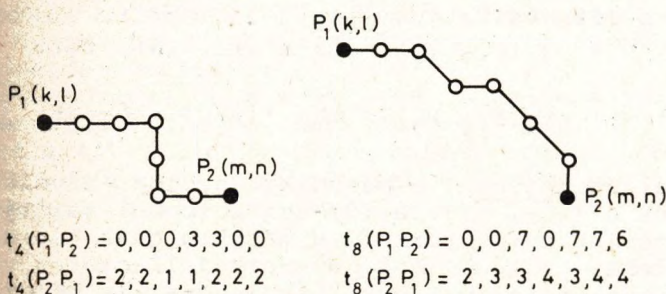


a)

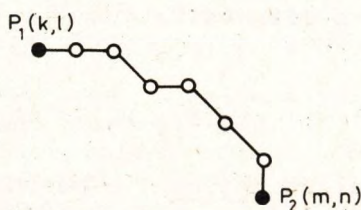


b)

1-33. ábra
Négyes (a) és nyolcas (b) szomszédság



a)



b)

1-34. ábra
Négyes (a) és nyolcas (b) útvonal

Jelölje egy egyszeres (4/8)-útvonal egyik irányú $(P_1 \rightarrow P_2)$ lánckódját az $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$, iránykódsorozatot. A fordított irányú lánckódot ekkor az

$$\vec{x}'_i = \vec{x}_{n-i} + 2 \pmod{4}; \text{ ill. } \vec{x}'_i = \vec{x}_{n-i} + 4 \pmod{8}$$

összefüggés adja (4)-, ill. (8)-útvonal esetén.

3. Összefüggőség

A diszkrét képsík pontjait értékes pontoknak nevezzük, ha valamilyen értékelhető információt hordoznak; a többi pont háttérpont. Az értékes pontokat különböző tulajdonságaik alapján osztályozhatjuk.

Két képpont egy osztályozásra nézve (4/8)-összefüggő, ha létezik olyan (4/8)-útvonal, amelynek

- a két pont a végpontja, és
- minden eleme ugyanahhoz az osztályhoz tartozik.

A tartományok az egy osztályba tartozó összefüggő pontok halmazai; egy tartomány bármely két pontja összefüggő. Minden véges tartománynak szükségképpen vannak olyan pontjai, amelyeknek legalább egy szomszédjuk nem tartozik a tartományhoz, ezek a kontúrpontok. Az összefüggő kontúrpon

thalmaza a kontúrvonal. A (4)-összefüggő tartományok kontúrvonala (8)-összefüggő és fordítva. Mivel a digitális kép véges, nyilvánvalóan vannak olyan képpontok, amelyeknek legalább egy szomszédjuk nem tartozik a képhez; ezeket (kép)határpontoknak nevezzük. A határpontokal összefüggő háttérpontok egymással is összefüggők és együttesen képezik a háttér

et. Azok a háttérpontok, amelyek nem függenek össze egyik határponttal sem, lyuk(ak)at alkotnak.

13.4.3. Geometriai jellemzők

1. Diszkrét metriák

Két $P_1(k, \ell), P_2(m, n) \in \mathcal{P}$ képpont távolságát a \mathcal{P} -ben kétféleképpen definiálhatjuk (ld. 1-35. ábra):

$$(1-30) \quad \mathcal{S}_4(P_1, P_2) = |k-m| + |\ell-n|, \text{ ill.}$$

$$(1-31) \quad \mathcal{S}_8(P_1, P_2) = \max(|k-m|, |\ell-n|).$$

Az elsőt négyes-, a másodikat nyolcas-távolságnak nevezzük, és a (4)-, ill. a (8)-szimbólummal jelöljük. (Az előbbi szerint egyik pontból a másikba csak koordinátairányokban lehet haladni, ezt New-York jellegzetes városnegyedéről manhattan-(city-block) távolságnak is nevezik. A második esetben 45°-os "utcák" is léteznek, ez megfelel egy sakktáblán a vezér mozgási lehetőségeinek, ezért a sakktávolság (chess board) elnevezés is előfordul.)

				4					
			4	3	4				
		4	3	2	3	4			
	4	3	2	1	2	3	4		
4	3	2	1	●	1	2	3	4	
	4	3	2	1	2	3	4		
		4	3	2	3	4			
			4	3	4				
				4					

a)

4	4	4	4	4	4	4	4	4	4
4	3	3	3	3	3	3	3	3	4
4	3	2	2	2	2	2	2	3	4
4	3	2	1	1	1	2	3	4	
4	3	2	1	●	1	2	3	4	
4	3	2	1	1	1	2	3	4	
4	3	2	2	2	2	2	3	4	
4	3	3	3	3	3	3	3	4	
4	4	4	4	4	4	4	4	4	4

b)

1-35. ábra
Négyes (a) és nyolcas (b) távolság

Könnyű belátni, hogy az (1-30) és az (1-31) definíció valóban távolságfüggvény (=metrika), mert minden $P_1, P_2,$

P_3 pontra teljesíti a következő feltételeket:

- nemnegatív definit: $\varphi(P_1P_2) \geq 0$ és $\varphi(P_1P_2) = 0 \leftrightarrow P_1 = P_2$;

- szimmetrikus: $\varphi(P_1P_2) = \varphi(P_2P_1)$;

- érvényes a háromszögegyenlőtlenség:

$$\varphi(P_1P_3) \leq \varphi(P_1P_2) + \varphi(P_2P_3);$$

- reguláris: $\forall P_1, P_3 \in \mathcal{P} \quad \varphi(P_1P_3) \geq 2 \quad \exists P_2 \quad \{ \varphi(P_1P_2) = 1 \text{ és} \\ \varphi(P_2P_3) = \varphi(P_1P_3) - 1 \}$

(A \forall szimbólum jelentése: "az összes"; az \exists szimbólumé: "van olyan" vagy "létezik". A regularitás azt fejezi ki, hogy a diszkrét képsík pontjai egyenletes sűrűségűek; folytonos metrikák esetében ez magától értetődik.

Könnyen belátható, hogy két pont távolsága megegyezik az őket összekötő legrövidebb útvonal leírásához szükséges iránykódok darabszámával.

Megjegyezzük, hogy a távolság fogalma egyértelmű, de a hozzá rendelhető "legrövidebb útvonal" még lokálisan sem az.

Mind a (4)-, mind a (8)-távolság erősen irányfüggő. Helyettük - bizonyos közelítéssel - értelmezhetjük két képpont euklidészi távolságát is a

$$(1-32) \quad \rho_E(P_1P_2) = \sqrt{(k-m)^2 + (\ell-n)^2} \quad (P_1, P_2 \in \mathcal{P})$$

pontos képlettel; ennek kiértékelése azonban túlságosan

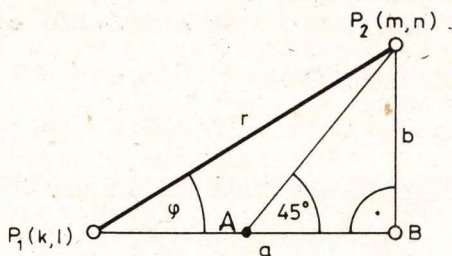
számításigényes. Helyette - két szomszédos pont (4)-távol-
ságát 1-nek, (8)-távolágát a ferde irányokban 1.5-nek
véve - két tetszőleges képpontot összekötő legrövidebb
(8)-útvonal hosszára a következő közelítő távolságfüggvényt
definiáljuk (ld. 1-36. ábra):

$$(1-33) \quad \rho_k = (a-b) + 1.5 b = |k-m| + |l-n| + 1.5 |l-n| = \\ = |k-m| + 0,5 |l-n|.$$

Bebizonyítható, hogy az így definiált és az euklidészi met-
rika átlagos aránya:

$$\rho_k / \rho_E = 1.087,$$

ami a gyakorlatban kielégítő közelítés.



1-36. ábra
Képpontok közelítő távolsá-
gának meghatározása

2. Diszkrét normák

A norma a távolságfogalom általánosítása tetszőleges
 n -dimenziós vektorterekre. Vektorterről akkor beszélünk, ha
létezik egy ortonormált $\{\vec{u}_i\}$ bázis, amelynek minden elemére
igaz, hogy

$$\vec{u}_i \vec{u}_j = \begin{cases} 0, & \text{ha } i \neq j; \\ 1, & \text{ha } i = j; \end{cases} \quad (i, j = 1, 2, \dots, n).$$

(Az $\vec{u}_i \vec{u}_j$ skalárszorzatot az (1-1) képlet definiálja.) Ekkor
a vektortérben bármely vektor a

$$\vec{z} = \sum_{j=1}^n z^{(j)} \vec{u}_j$$

alakban írható, ahol $z^{(j)}$ tetszőleges valós szám.

A norma a vektortér elemeinek általánosított hossza, amit a $\|\cdot\|$ szimbólummal jelölünk és mindig ki kell elégítenie az alábbi 3 feltételt (\vec{v}, \vec{z} a vektortér tetszőleges eleme, λ valós szám):

a) $\|\vec{z}\| \geq 0$; ($\|\vec{z}\| = 0$ akkor és csak akkor, ha $\vec{z} = 0$);

b) $\|\lambda \vec{z}\| = |\lambda| \cdot \|\vec{z}\|$,

c) $\|\vec{v} + \vec{z}\| \leq \|\vec{v}\| + \|\vec{z}\|$; (háromszögegyenlőtlenség).

A jegyzetben használt normák:

(1-34) euklidészi norma: $\|\vec{z}\|_E = \sqrt{\sum_{j=1}^n z^{(j)2}}$;

(1-35) abszolútérték norma: $\|\vec{z}\|_a = \sum_{j=1}^n |z^{(j)}|$;

(1-36) maximum norma: $\|\vec{z}\|_m = \text{maximum } (|z^{(j)}|)$.
 $1 \leq j \leq n$

13.4.4. Diszkrét egyenesek

Kimutatható, hogy - racionális meredekségű egyenesek esetében - a diszkrét egyenest reprezentáló iránykód sorozatokban

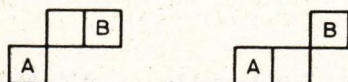
- legfeljebb kétféle iránykód fordulhat elő, és ezeknek mod(8) szomszédosaknak kell lenniük;

- kétféle iránykód esetében egyikük csak egyszeresen fordulhat elő (azaz mindkét szomszédja a másik iránykód);

- az egyszeres iránykódok sűrűségeloszlásának a lehető legegyszerűsebbnek kell lennie.

Be lehet bizonyítani, hogy ezek a feltételek általános esetben is elégéségek.

Megjegyzendő, hogy a diszkrét egyenes nem feltétlenül egyértelmű (ld. 1-37. ábra).



1-37. ábra

Diszkrét egyenes többértelmősége

Diszkrét görbéket úgy közelítünk, hogy a folytonos görbe "elég közeli" pontjait diszkrét egyenesekkel kötjük össze. A kritérium pontosabb megfogalmazásával nem foglalkozunk, mert a továbbiakban nem lesz rá szükségünk.

2. ELŐFELDOLGOZÁS

Ebben a részben olyan eljárásokkal foglalkozunk, amelyek képből képet állítanak elő. Az eljárásokat az elérendő cél szerint csoportosítjuk:

- A képkorrekciós eljárások célja a képminőség javítása. Bár ennek objektív, egzakt mértéke ez idő szerint nem ismeretes, sőt általánosan elfogadott normái sincsenek, mégis jellemezni lehet a képhűséggel és az értelmezhetőséggel. Az előbbi növelésekor egy ideális leképezőrendszer által létrehozott, hibátlan kép megközelítésére törekszünk; az utóbbi javításakor a képinformáció minél könnyebb, gyorsabb és pontosabb emberi megértése a cél. Ebben a fejezetben vizsgáljuk a digitális kép létrehozását és az analóg kép visszaállítását is (21. fejezet).

- Külön fejezetben foglalkozunk a geometriai korrekciókkal, mert ezeket speciális, az előzőektől lényegesen eltérő módszerekkel valósítják meg (22. fejezet).

- Végül a 23. fejezetben tárgyaljuk a szegmentálási eljárásokat, amelyek célja a kép jellemző sajátságainak, alakzatainak megkeresése és kiemelése, esetleg segítségükkel a kép szintaktikus (formális) leírása.

Fontos megjegyezni, hogy a képfeldolgozás nem szükségképpen jelenti az egyes képpontok világosságának, ill. színének megváltoztatását (bár az eredmény legtöbbször ez); más képparaméterek (pl. hisztogram, statisztikai jellemzők, geometria stb.) megváltoztatására is irányulhat. Az eredmény kiértékelésében a fizikai jelentést mindig figyelembe kell venni.

21. Képkorrekciós eljárások

Mielőtt egy kép feldolgozásra kerül, rendszerint több átalakításon megy keresztül: felvétel, digitalizálás, esetleg átvitel, másolás stb. Mindegyik eljárás hibával jár, ezek rontják a kép minőségét: torzításokat okoznak. Ezekon kívül hamis információk: zajok keveredhetnek a képhez.

A torzítások következtében megváltozik a kép információtartalma. Ha ez csak egyes, elszigetelt képpontokban, vagy a kép egészéhez véve kis területen jelentkezik, lokális

vagy pontszerű hibának, ha az egész képre kiterjed, globális hibának nevezzük.

A korrekciós eljárások általában a hibák hatásának megszüntetésére, az ideális kép visszaállítására irányulnak. A lokális eljárások az egyes képpontok meghatározott környezetének elemzése alapján javítanak, a globálisak az egész képet értékelik. Mivel - mint említettük - a képminőség egzakt mértéke nem ismeretes, egy-egy eljárás hatékonysága és javítóképessége szubjektív megítélés kérdése. Ezért, valamint a képjavítás átfogó elméletének hiányában, az alkalmazandó módszert és eljárást egyéb szempontok, rendszerint a könnyű programozhatóság, a rövid futási idő stb. alapján választják ki.

21.1. DIGITALIZÁLÁS ÉS VISSZAÁLLÍTÁS

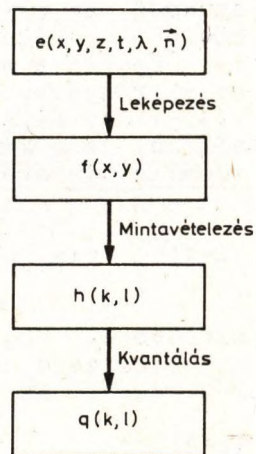
A továbbiakban képen a vizuális információt hordozó jelet (elektromágneses hullámot) értjük. (A falon függő műalkotás, amely a hétköznapi szóhasználatban képnek neveznek, ebben az értelemben nem kép, hanem képforrás.)

A természetes kép analóg, térben (x, y, z) és időben (t) folytonos, különböző színárnyalatokat tartalmazhat (λ) és függ a megfigyelés irányától (\vec{n}) is, matematikailag valamilyen $e(x, y, z, t, \lambda, \vec{n})$ folytonos függvénnyel lehet leírni.

Feldolgozás céljára a képet fel kell venni valamilyen leképező eszközzel (pl. kamerával). A felvett kép kétdimenziós, és mivel a felvétel egy meghatározott időpillanatban, - frekvenciasáv(ok)ban és -irányból történt, valamilyen $f(x, y)$ folytonos függvénnyel lehet leírni (ld. 2-1. ábra). (A kép síkjául az x - y síkot választjuk. A síkbeli koordináta-rendszert - a TV-technikában használt megjelenítési mód miatt - úgy szokás megválasztani, hogy az origó a kép bal felső sarokpontja; az x -tengely balról-jobbra, az y -tengely felülről lefelé irányított.)

Fontos megjegyezni, hogy egy tárgyról vagy jelenetről egyidejűleg gyakran több frekvenciasávban is készül felvétel (pl. az úrfelvételek esetében). Ezeket együttesen kell feldolgozni, ami - bár információnyereséget jelent - számos problémát vet fel; ezekre a későbbiekben visszatérünk.

Számítógépes feldolgozás céljából a képletet digitalizálni kell. Ez azt jelenti, hogy a képet pontokra bontjuk és az egyes képpontok világosságértékét (színét) kvantáljuk. A digitális kép - bár a vizuális információt megőrzi - "képisé-



2-1. ábra
Digitális kép létrehozása

gét" elveszti. Ahhoz, hogy ismét látható legyen, analóg kép-jellé kell visszaalakítani (= visszaállítás). A két folyamatnak felváltva ismételhetőnek kell lennie, ezért csak olyan eljárásokat szabad használni, amelyek ezt lehetővé teszik.

A képfüggvényt általában determinisztikusnak tekintjük, bár bizonyos képek (főleg a zajhatások) kényelmesebben tárgyalhatók sztochasztikus módszerekkel. Az utóbbiakkal nem foglalkozunk részletesebben.

21.1.1. Mintavételezés

A digitalizálás első lépése az (analóg) kép felbontása képpontokra. Ez úgy történik, hogy a képsík meghatározott pontjaiban mintákat veszünk a képjelből. Ezek a minták elvileg a folytonos kép valamilyen mérhető tulajdonságát reprezentálják, legtöbbször a kép világosságát (fényerősségét) vagy színét az adott pontban. (Emlékeztetünk rá, hogy a színt 3 adat jellemzi - pl. a színárnyalat, a világosság és a telítettség.) Más szóval a mintavételezés tulajdonképpen mérés; s mint minden mérés, valamilyen bizonytalansággal, mérési hibával jár. Nagyon fontos, hogy a mérési hibák mérhetőek legyenek, s így értékelni lehessen a mintavételezés "jósságát".

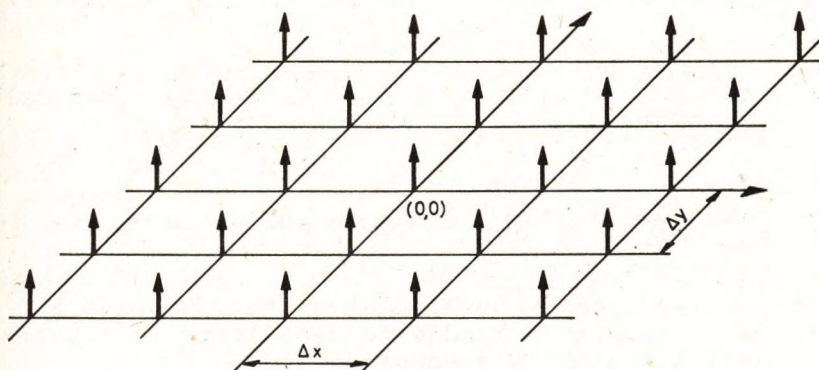
Legyen a folytonos, végtelen kiterjedésű, ideális (hibátlan) képnek a mérendő sajátságát (a továbbiakban felteszük, hogy ff képek esetében a világosságot, színes képek esetében a 3 színparamétert) leíró függvény $f(x, y)$. (Színes képek esetében 3 egymástól független függvényre van szükség. Az eljárást csak egy függvényre írjuk le, de nehézség nélkül kiterjeszthető háromra is.)

Legyen a mintavételezési pontok x -irányú távolsága Δx , az y -irányú Δy . Egészítsük ki továbbá a véges, $(K-1)\Delta x \cdot (L-1)\Delta y$ méretű képet "üres" képpontokkal a teljes x - y síkban, és a kép középpontja legyen az origóban. Ekkor az ideális mintavételt a

$$(2-1) \quad t(x, y) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \delta(x - m \cdot \Delta x, y - n \cdot \Delta y)$$

mintavételi függvényvel lehet leírni, amely a Dirac-delta kiterjesztése az összes kijelölt mintavételezési pontra (ld. 2-2. ábra). Az eredményül kapott digitalizált képet a

$$(2-2) \quad \begin{aligned} h(k, l) &= f(x, y) \cdot t(x, y) = \\ &= \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} f(m \cdot \Delta x, n \cdot \Delta y) \cdot \delta(x - m \cdot \Delta x, y - n \cdot \Delta y) \end{aligned}$$



2-2. ábra
A Dirac-delta kiterjesztése

összefüggés írja le; ami nem más mint az eredeti képfüggvényértékek halmaza a $k = m \cdot \Delta x$, $l = n \cdot \Delta y$ pontokban. (Szemléletesen: képzeljük el a képet felületként és "szűrjük át" a Dirac-deltáknaka 2-2. ábrán látható halmazával. Ekkor $h(k, l)$ a (k, l) ponthoz tartozó Dirac-delta "döféspontjának" az x - y síktól mért távolsága.) Ezzel a képet felbontottuk K.L képpontra.

Feldolgozási célokra előnyös, ha ismerjük a kép síkfrekvencia-spektrumát. Ezt - mint már említettük - a képfüggvény Fourier-transzformációjával állíthatjuk elő. Ehhez tudnunk kell, hogy egyrészt a (2-1) mintavételező függvény Fourier-transzformáltja

$$(2-3) \quad T(u, v) = \frac{4\pi^2}{\Delta x \Delta y} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \delta(u - m \cdot \Delta u, v - n \cdot \Delta v)$$

ahol $\Delta u = 2\pi/\Delta x$, ill. $\Delta v = 2\pi/\Delta y$ a mintavételezési síkfrekvencia az u , ill. a v síkfrekvenciatengely irányában. Másrészt a mintavételezési folyamatot a transzformált függvények konvolúciójával lehet leírni, vagyis (2-2) alapján

$$(2-4) \quad H(u, v) = \frac{1}{4\pi^2} F(u, v) \otimes T(u, v)$$

Helyettesítsük be (2-3)-at a (2-4)-be, és végezzük el a konvolúcióképzést:

$$\begin{aligned}
 (2-5) \quad H(u, v) &= \\
 &= \frac{1}{\Delta x \Delta y} \iint_{-\infty}^{\infty} F(u-\mu, v-\nu) \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \delta(\mu-m \cdot \Delta u, \nu-n \cdot \Delta v) d\mu d\nu \\
 &= \frac{1}{\Delta x \Delta y} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} F(u-m \cdot \Delta u, v-n \cdot \Delta v).
 \end{aligned}$$

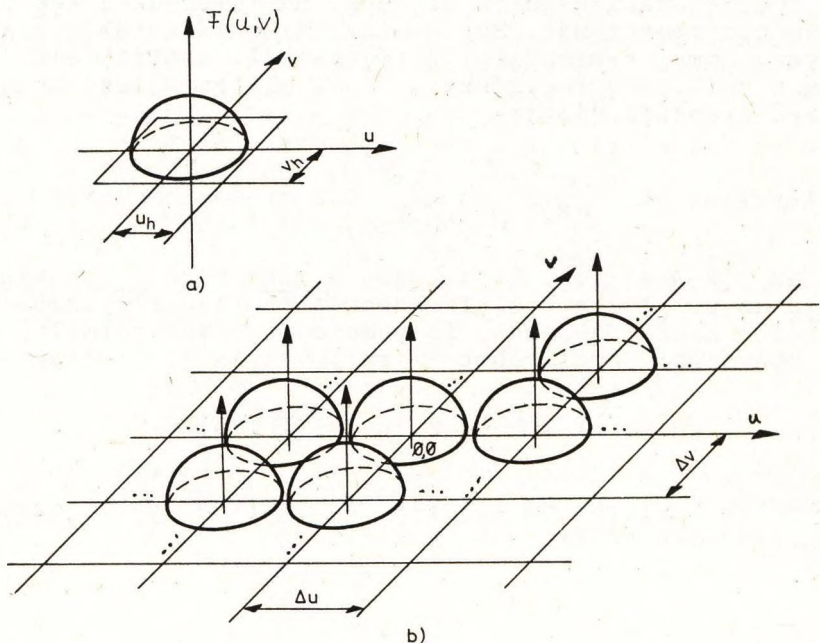
Az utolsó eredményhez felhasználtuk, hogy az integrálás és az összegzés sorrendje felcserélhető és alkalmaztuk a Dirac-delta "toló" tulajdonságát (ld. (1-3)).

A (2-5)-ből leolvashatjuk, hogy a digitalizált kép sík frekvencia-spektruma nem más, mint az eredeti $f(x, y)$ képfüggvény frekvenciaspektrumának végtelen ismétlése az u - v frekvenciasík $m \cdot \Delta u, n \cdot \Delta v$ rácspontjaiban.

Tegyük fel, hogy $F(u, v)$ sávkorlátos, azaz

$$(2-6/a) \quad F(u, v) = 0 \quad \text{ha } |u| > u_h, \quad \text{ill. } |v| > v_h$$

(ld. 2-3/a. ábra); ekkor $H(u, v)$ spektrumképét a 2-3/b. ábra



2-3. ábra
Digitális kép frekvenciaspektruma

láthatjuk. Nyilvánvaló, hogy ez utóbbi akkor és csak akkor átfedésmentes, ha teljesül az

$$(2-6/b) \quad u_h \leq \frac{\Delta u}{2} = \frac{\pi}{\Delta x}, \quad \text{ill.} \quad v_h \leq \frac{\Delta v}{2} = \frac{\pi}{\Delta y}$$

Nyquist-feltétel. Ezt a fontos felismerést fel fogjuk használni a visszaállításnál.

Anélkül, hogy részletekbe mennénk, megemlítjük, hogy a zajok síkfrekvenciaspektruma általában nem sávkorlátos, vagy legalábbis szélesebb a képénél. Mivel így nem teljesül a Nyquist-feltétel, a zajos kép alulszelektált lesz, és nem lehet egyértelműen visszaállítani (ld. 21.1.3. pont).

21.1.2. Kvantálás

A digitalizálási folyamat második lépésében az A/D átalakító összehasonlítja az egyes képpontokhoz tartozó függvényértékeket a lehetséges kimeneti szintjeivel. A folyamatot kvantálásnak nevezzük (ld. 2-1. ábra). Eredményként annak a kimenő szintnek a kódját kapjuk, amelyikhez a vizsgált függvényérték a legközelebb esik. Ílymódon minden képponthoz egy, a világosságtól függő kódszám fog tartozni, amelyet világosságkódnak nevezünk és q -val jelölünk. Ezzel létrehoztuk a digitális képet. (Színesképek esetében a digitális színkód 3 függvényérték egyidejű kvantálásával áll elő.) A világosságkód nemnegatív egészszám, és értéke a gyakorlatban $0 \leq q \leq 255$ intervallumba esik. A lehetséges világosságkódok halmazára a $\{Q\}$, az értékkészletére az I_{255} jelölést használjuk.

Tegyük fel, hogy a mintavételezéskor kapott függvényértékek a

$$h_a \leq h(k, \ell) \leq h_f$$

alsó és felső határok közé esnek. Legyen a megengedett kvantumszintek száma I , és jelöljük ezeket q_1, \dots, q_I -vel. (Következésképpen $q(k, \ell) \in \{q_i \mid i = 1, \dots, I\}$.) Végül jelöljük d_i -vel a döntési szinteket ($d_0 = h_a$ és $d_I = h_f$); úgy, hogy

$$q(k, \ell) = q_i$$

akkor és csak akkor, ha

$$d_{i-1} \leq h(k, \ell) < d_i \quad (i = 1, 2, \dots, I);$$

(ld. 2-4. ábra).

Az alábbiakban az egyszerűség kedvéért elhagyjuk a (k, l) argumentumot. Jelöljük a mintavételi érték sűrűségfüggvényét $s(h)$ -val. A kvantálás "jószágát" a négyzetes hibával fejezzük ki:

$$(2-7) \quad E = \int_{h_a}^{h_f} (h-q)^2 \cdot s(h) dh = \sum_{i=1}^I \int_{d_{i-1}}^{d_i} (h-q_i)^2 \cdot s(h) dh.$$

A kvantálást akkor tekintjük optimálisnak, ha a négyzetes hiba minimális, adott I számú szint esetén. Az optimum megkeresése céljából tekintsük (2-7) differenciálhányadosát a d_i -k, ill. q_i -k szerint:

$$(2-8) \quad \frac{\partial E}{\partial d_i} = (d_i - q_i)^2 \cdot s(d_i) - (d_i - q_{i+1})^2 \cdot s(d_i) = 0; \quad (i = 1, \dots, I-1)$$

$$(2-9) \quad \frac{\partial E}{\partial q_i} = -2 \int_{d_{i-1}}^{d_i} (h - q_i) s(h) dh = 0; \quad (i = 1, \dots, I)$$

Ha $s(d_i) \neq 0$, (2-8)-ből és (2-9)-ből következik, hogy

$$(2-10) \quad d_i = \frac{q_i + q_{i+1}}{2}, \quad (i=1, 2, \dots, I-1)$$

$$(2-11) \quad q_i = \frac{\int_{d_{i-1}}^{d_i} h s(h) dh}{\int_{d_{i-1}}^{d_i} s(h) dh} \quad (i = 1, 2, \dots, I)$$

vagyis az optimális kvantáló döntési szintjei a kvantumszintek felezőpontjaiban helyezkednek el; míg a kvantumszintek azon sűrűségfüggvény-darab alatti terület súlypontjai, amelyet az őket közrefogó döntési szintek határolnak (ld. 2-4. ábra). A (2-10) és a (2-11) egyenletrendszerből a q_i és d_i értékeket - $s(h)$ ismeretében - iteratív közelítéssel lehet meghatározni. Ehhez vegyük fel q_1 értékét önkényesen, majd rendre számítsuk ki a $d_1, q_2, d_2, \dots, q_I, d_I$ értékeket.

Ha most $d_I \neq h_f$, akkor q_1 megfelelő korrekciója után ismétljük meg az eljárást.

Könnyű belátni, hogy ha a mintavételi sűrűségfüggvény egyenletes eloszlású, akkor (2-11) a

$$(2-12) \quad q_i = \frac{d_{i-1} + d_i}{2} \quad (i = 1, 2, \dots, I)$$

középértékre egyszerűsödik, vagyis ez esetben az egyenlőközű kvantálás az optimális. (A 2-4. ábrán ezt az esetet szaggatott vonallal tüntettük fel.)

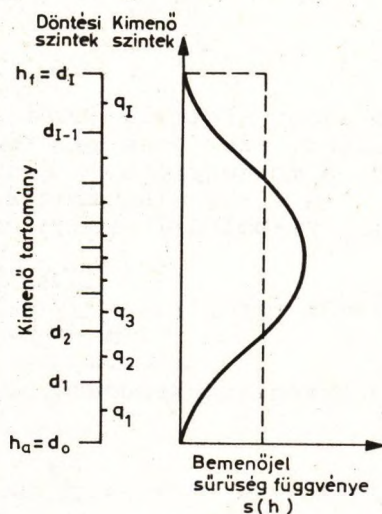
Összefoglalva megállapíthatjuk, hogy a kvantálás információvesztéssel jár. A (2-10)-ből és a (2-11)-ből kiolvasható, hogy ha a sűrűségfüggvény nem egyenletes, a kvantálást finomítani kell azokban a világosságkód-tartományokban, ahová a képpontok többségének kódja esik. Ennek az eloszlásfüggő (tapered) kvantálásnak megvalósítása azonban bonyolult rendszerekhez vezet, ezért a gyakorlatban ritkán alkalmazzák.

Ha $I = 2^b$, a K.L képpontból álló kép tárolásához K.L.b bitre van szükség. Kísérletileg bebizonyították, hogy ezt a kapacitást akkor lehetne legjobban kihasználni, ha a mintavételezési intervallum nem volna állandó. Olyan képeknél, ahol a világosság lassan változik, növelni kellene a kvantálás finomságát és csökkenteni a mintavételi pontok számát; míg a részletgazdag képeknél a felbontást kellene finomítani, s emellett megengedhető a kvantumszintek számának csökkentése. A gyakorlatban ilyen digitalizálók nem léteznek.

21.1.3. Visszaállítás (rekonstrukció)

A visszaállítás vagy rekonstrukció (reconstruction) során a digitális képből interpolációval állítjuk elő az analóg képjelet. Mivel a kvantálaskor a mintavételezéssel nyert képfüggvényértékeket a legközelebbi kvantumszinthez igazítottuk, az eredeti képet többé nem lehet rekonstruálni.

Jelöljük $\hat{f}(x, y)$ -nal azt a folytonos képfüggvényt, amely az $f(x, y)$ -től annyiban tér el, hogy a mintavételezési pontokban a kvantált függvényértékek szerepelnek. (Ezek általában nem egyeznek meg az eredetivel, hanem alatta vagy felette vannak.) A "képfelület"-folytonos lévén - emiatt természetesen a közbülső pontokban is általában el fog tér-



2-4. ábra
Kvantálási szintek megválasztása

ni az eredetitől. Erről a (2-7) alapján annyit állíthatunk, hogy

$$\int_{-\infty}^{+\infty} [f(x, y) - \hat{f}(x, y)]^2 dx dy = \text{minimum}$$

az adott digitalizálási technika esetében. (Ami persze nem jelenti azt, hogy más technikák nem adhatnak jobb közelítést; ld. a megjegyzést az előző pont végén.)

A (2-2) alapján $q(k, \ell)$ -t úgy foghatjuk fel, mintha az $\hat{f}(x, y)$ -ből állt volna elő mintavételezéssel:

$$(2-13) \quad q(k, \ell) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \hat{f}(m\Delta x, n\Delta y) \delta(k-m\Delta x, \ell-n\Delta y).$$

Frekvenciaspektrumát pedig a (2-5) alapján a

$$(2-14) \quad Q(u, v) = \frac{1}{\Delta x \Delta y} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \hat{F}(u - m \cdot \Delta u, v - n \cdot \Delta v)$$

összefüggés adja, ahol \hat{F} a \hat{f} Fourier-transzformáltja.

A visszaállított f^{-1} képfüggvényt $q(k, \ell)$ -nek, az $r(x, y)$ interpolációs függvénnyel képzett konvolúciója eredményezi:

$$(2-15) \quad f^{-1}(x, y) = r(x, y) \otimes q(k, \ell).$$

Helyettesítsük be (2-13)-at és végezzük el a konvolúciós integrálást, akkor

$$(2-16) \quad f^{-1}(x, y) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \hat{f}(m\Delta x, n\Delta y) r(x - m\Delta x, y - n\Delta y)$$

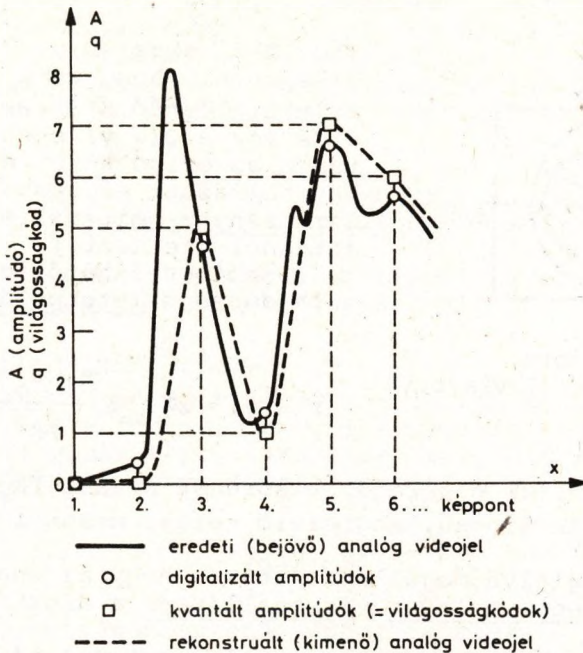
lesz. Az eredményt egydimenziósesetre a 2-5. ábrán szemléltetjük.

A visszaállított képfüggvény frekvenciaspektruma (2-15) alapján

$$F^{-1}(u, v) = R(u, v) \cdot Q(u, v),$$

vagy (2-14) felhasználásával:

$$(2-17) \quad F^{-1}(u, v) = \frac{R(u, v)}{\Delta x \Delta y} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \hat{F}(u - m\Delta u, v - n\Delta v).$$



2-5. ábra
Egydimenziós visszaállított jel

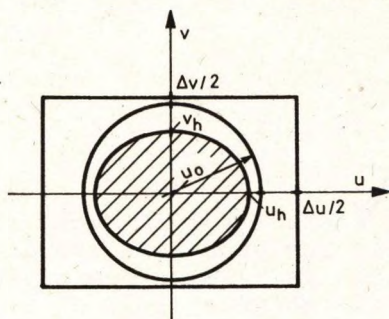
A második tényezőről tudjuk, hogy az $\hat{f}(x, y)$ síkfrekvencia-spektrumának végtelen ismétlése. Az egyszerűbb jelölések kedvéért tegyük fel, hogy a 2-3. ábrán $\hat{F}(u, v)$ -t, ill. $Q(u, v)$ -t látjuk, és a (2-6) feltételek $\hat{F}(u, v)$ -re is igazak.

Nyilvánvaló, hogy $F^{-1}(u, v)$ akkor fog megegyezni $\hat{F}(u, v)$ -vel, ha sikerül olyan $R(u, v)$ függvényt találni, amelyik a $Q(u, v)$ spektrumából kiszűri mindazokat az összetevőket, amelyekre $\{m, n\} \neq 0$, a többit pedig változatlanul áttereszti. Ez pl. akkor teljesül, ha

$$(2-18) \quad R(u, v) = \begin{cases} K, & \text{ha } \sqrt{u^2 + v^2} \leq u_0 \\ 0, & \text{különben} \end{cases}$$

és

$$\sqrt{u_h^2 + v_h^2} \leq u_c \leq \min \left(\frac{\Delta u}{2}, \frac{\Delta v}{2} \right).$$



2-6. ábra
Határfrekvenciák viszonya

(Ld. 2-6. ábra; $R(u, v)$ -t úgy kell elképzelni, hogy az u_0 sugarú kör felett állandó K értékű, azon kívül 0; s így a $Q(u, v)$ spektrumából kivágja az origó körül elhelyezkedő, bevonalkázott ellipszisz feletti tartományt; amit $K=1$ esetén változatlanul átérészt.)

Bebizonyítható, hogy az ilyen tulajdonságú interpolátor függvény

$$r(x, y) = 2\pi u_0 K \frac{B_1\left(u_0 \sqrt{x^2 + y^2}\right)}{\sqrt{x^2 + y^2}}$$

alakú, ahol $B_1(z)$ elsőfajú, elsőrendű Bessel-függvény. (Természetesen más típusú, megfelelő tulajdonságú interpolátorok is léteznek.)

Összefoglalva megállapíthatjuk, hogy az analóg képet akkor lehet egyértelműen visszaállítani a digitális képből, ha

a) a képfüggvény síkfrekvenciaspektruma sávkorlátos (2-6/a); és

b) a Nyquist-feltétel teljesül, vagyis a mintavételezési periódus $(\Delta x, \Delta y)$ nem nagyobb a képen előforduló legfinomabb képrészlet periodicitásának felénél (2-6/b); (vö. Shannon-tétele!)

Ebben az esetben - megfelelően választott interpolációs függvénnyel - elérhetjük, hogy a visszaállított és az eredeti képfüggvény síkfrekvencia-spektruma megegyezzek, a kvantálási hibáktól eltekintve. Ilyenkor a két képet azonosnak tekintjük.

21.2. KÉPJAVÍTÁS

Képjavításnak az olyan eljárásokat nevezzük, amelyekben a hibajavítás során nem vizsgálják a képnek az eredetihez való viszonyát. Az eljárás hatékonyságát különféle objektív és szubjektív becslések alapján értékelik, ezekkel azonban nem foglalkozunk.

A hibák torzításokra és zajhatásokra vezethetők vissza. A torzítás általában csökkenti a kontrasztot (vagyis kisebb lesz a különbség a legsötétebb és a legvilágosabb képpontok világossága között), és elmosódást okoz (vagyis kiszélesednek az alakzatok közötti átmeneti tartományok, "életlen" lesz a kép).

A képjavítás célja, hogy a képet az emberi szemlélő vagy a további gépi feldolgozás számára kedvezőbbé alakítsa.

Ennek megfelelően a kontraszt fokozására, a kép "élesítésére" és a zajok kiszűrésére irányul, az egyes képpontok világosságkódjának megváltoztatásával; ezáltal megváltozik a világosságkódok eloszlása. (A geometriai jellemzők megváltoztatására irányuló korrekciókkal a következő fejezetben foglalkozunk.)

21.2.1. Világosságkód-transzformációk

A transzformációk célja a kontraszt növelése a világosságkódok eloszlásának megváltoztatásával.

21.2.1.1. A hisztogram

A képjavítás során különösen - de egyéb képfeldolgozási eljárásokban is - fontos szerepet játszik a hisztogram, ami megadja a világosságkódok eloszlását az adott képben. A hisztogramtáblázat annyi elemű, amennyi a világosságkódok értékkészlete, és

$$\chi'(q_i) = N_{q_i},$$

ahol N_{q_i} a q_i világosságkódú képpontok száma a digitális képben. A hisztogramot oszlopdiaagrammal ábrázoljuk (ld. pl. 2-8. ábra); a vízszintes tengelyen a lehetséges világosságkódokat mérjük fel; a $[q_i, q_{i+1})$ intervallumhoz tartozó ordináta pedig a q_i világosságkódú képpontok relatív gyakoriságával arányos:

$$(2-19) \quad \chi(q_i) = c \frac{N_{q_i}}{N},$$

ahol N a képet alkotó képpontok száma.

A továbbiakban mindig feltesszük, hogy a hisztogram normalizált ($c = 1$), amiből következik, hogy ekkor

$$\sum_{i=0}^I \chi(q_i) = 1.$$

Ha a világosságkódot valószínűségi változóként kezeljük, akkor annak P valószínűsége, hogy értéke a q és $q+\Delta q$ közé esik, valamint a sűrűségfüggvény között fennáll az

$$s(\xi) = \lim_{\Delta q \rightarrow 0} \frac{P(q \leq \xi < q + \Delta q)}{\Delta q}$$

összefüggés, ami éppen a (2-19) hányadossal közelíthető, vagyis

$$s(\xi) \sim \chi(q_1)$$

21.2.1.2. Skálázások

A továbbiakban vezessük be a következő jelöléseket:

Legyen $\{Q_b\}$ a világosságkódok eredeti (bemenő), ill.

$\{Q_k\}$ az eredményül kapott (kimenő) értékkészlete,

($\{Q_k\}$ nem szükségképpen egyenlő $\{Q_b\}$ -vel!) és jelölje q_{bm} , ill. q_{km} a megfelelő halmaz legkisebb, q_{bM} , ill. q_{kM} a legnagyobb elemét; továbbá $q_b \in \{Q_b\}$ egy képpont eredeti, $q_k \in \{Q_k\}$ pedig a transzformált világosságkódját.

Skálázásnak nevezzük a

$$T : \{q_{ba}, q_{bf}\} \rightarrow \{Q_k\}$$

leképezést megvalósító globális transzformációkat, ahol $q_{ba} \geq q_{bm}$, ill. $q_{bf} \leq q_{bM}$ a bemenő világosságkód-intervallum alsó, ill. felső határa.

Legtöbbször szakaszonként lineáris transzformációt végzünk a következő összefüggés alapján:

$$q_k = \begin{cases} q_{km} ; & \text{ha } q_b < q_{ba} \\ \text{int} \left(\frac{q_{kM} - q_{km}}{q_{bf} - q_{ba}} (q_b - q_{ba}) \right) + q_{km} ; & \text{ha } q_{ba} \leq q_b \leq q_{bf} \\ q_{kM} ; & \text{ha } q_b > q_{bf} \end{cases}$$

ahol int a "legközelebbi egészszám" függvény.

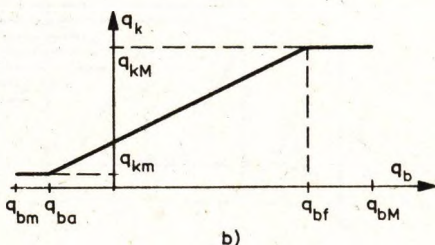
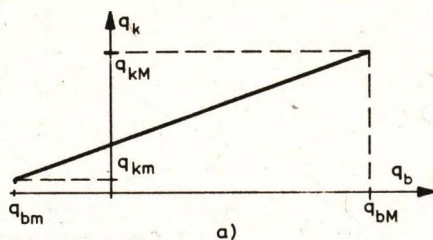
Ha a teljes bemenő intervallumot akarjuk transzformálni, akkor $q_{ba} = q_{bm}$, ill. $q_{bf} = q_{bM}$ (2-7/a. ábra), különben sávzsugorításról beszélünk (2-7/b. ábra). Az utóbbit akkor érdemes alkalmazni, ha az értékes képpontok világosságkódja a $[q_{ba}, q_{bf}]$ tartományba esik. (Gyakran előfordul, hogy egy előző feldolgozási lépésben számított világosságkódok $\{Q_b\}$)

halmaza nagyobb mint a megjeleníthető világosságokódoké; a 2-7. ábrán pl. negatív bemenő értékek is előfordulnak.) A sávzsugorítás akkor is célszerű, ha a szélsőséges világosságokódú képpontok száma kevés, és megtartásuk feleslegesen szegényítené a többség kontrasztját.

Az eljárást kontrasztkiemelésnek nevezzük, ha

$\{Q_b\} \subset \{Q_k\}$; ugyanis ilyenkor ténylegesen megnő a különbség a legsötétebb és a legvilágosabb képpontok világosságkódja között, az eredménykép hisztogramjában pedig "lyukak" keletkeznek (2-8. ábra).

Tetszés szerinti t (kvadratikus, logaritmikus stb.) transzformációt valószínűsíthetünk meg a



2-7. ábra
Szakaszonként lineáris transzformációk

$$q_k = \text{int} \left(\frac{q_{kM} - q_{km}}{q_{tM} - q_{tm}} (t(q_b) - t(q_{bm})) \right) + q_{km}$$

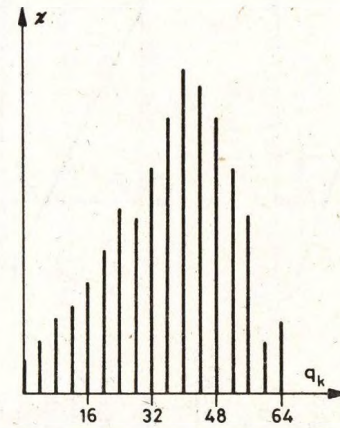
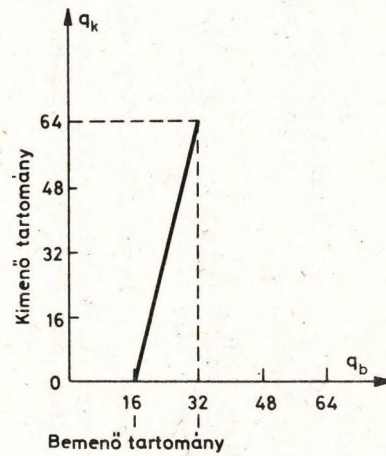
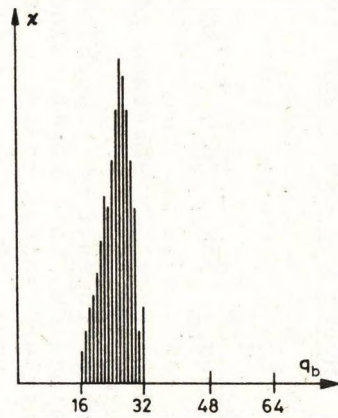
képlettel. Itt q_{tM} , ill. q_{tm} a bemenő kódokra alkalmazott t transzformáció során nyert legkisebb, ill. legnagyobb érték; monoton növekvő t függvény esetén természetesen $q_{tM} = t(q_{bM})$ és $q_{tm} = t(q_{bm})$.

Az eddig tárgyalt, monoton növekvő transzformációs függvények helyett monoton csökkenő vagy nem monoton függvények is használatosak. Inverz megjelenítést érhetünk el a 2-9/a. ábrán látható

$$q_k = \text{int} \left(\frac{q_{kM} - q_{km}}{q_{bM} - q_{bm}} (q_{bM} - q_b) \right) + q_{km}$$

invertáló transzformációval, ha pl. a sötétebb képrészek hordozzák a hasznos információt.

A 2-9/b. ábrán látható "fűrészfog" transzformációs függvényt pl. akkor alkalmazzuk, ha a képet színesben akarjuk látni, és a megjeleníthető színek száma kevesebb, mint a világosságokódoké. Ezzel ugyan $\{Q_b\}$ több eleméhez rendeljük a $\{Q_k\}$ ugyanazon elemét, de ez nem zavaró, ha az azonos színű részek geometriailag elkülönülnek. (Pl. a Balaton és kör-



2-8. ábra
Kontrasztkiemelés

nyéke megjelenítésekor lehetnek azonos színű foltok mind a vizet, mind a partot ábrázoló képrészekben; a tó jellegzetes alakja lehetővé teszi, hogy ezeket különbözőképpen értelmezzük.)

A skálázásokat átszínezésnek is nevezzük, mivel színes megjelenítés esetén megváltoztatják a képpontok színét. Legegyszerűbben Q-táblákkal valósíthatók meg. A Q-táblának annyi eleme van mint a $\{Q_b\}$ halmaznak, és az elemek értékkészlete a $\{Q_k\}$ halmaz. A skálázást egyszerűen a

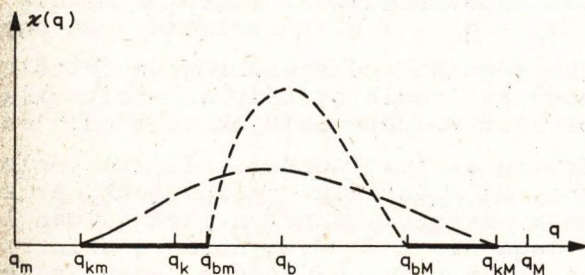
$$(2-20) \quad q_k = T(q_b)$$

kódcseréje jelenti, ahol $T(q_b)$ a Q-

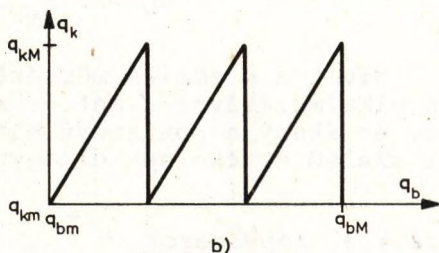
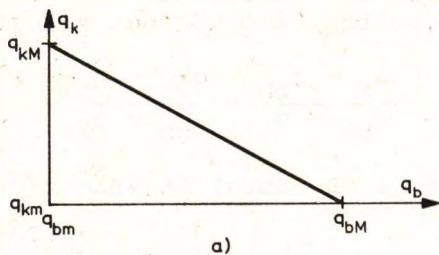
tábla q_b -edik eleme. Különösen kellemes a helyzet, ha a kódcseréje hardware úton (KAT-val, ld. 1-17. ábra) valósul meg, ilyenkor az átszínezés csak a megjelenített látványképre vonatkozik, a tárolt eredeti kép változatlanul marad.

Ha a képek háttér-megvilágítása egyenetlen, a globális eljárások általában nem vezetnek kielégítő eredményre. A 2-10. ábrán egy lokális kontrasztkiemelési eljárás elvét szemléltetjük. Feltéve, hogy a bemeneti világosságok $q_{bM} - q_{bm}$ tartománya, sőt még ennek p-szerese is kisebb a lehetséges $q_M - q_m$ tartománynál, p-szeres nyújtást végzünk, vagyis

$$q_{kM} - q_{km} = p \cdot (q_{bM} - q_{bm}).$$



2-10. ábra
Lokális kontrasztkiemelés



2-9. ábra
Speciális skálázó
függvények

A kimeneti intervallumot azonban úgy csúsztatjuk el, hogy a végpontok távolságának aránya ne változzék, vagyis

$$\frac{q_M - q_{bM}}{q_{bm} - q_m} = \frac{q_{kM} - q_{bM}}{q_{bm} - q_{km}} \quad \text{és} \quad \frac{q_b - q_{bm}}{q_{bM} - q_{bm}} = \frac{q_k - q_{km}}{q_{kM} - q_{km}}$$

(ld. 2-10. ábra). A transzformáció képlete:

$$q_k = p \cdot q_b - \frac{(p - 1)(q_M \cdot q_{bm} - q_{bM} \cdot q_m)}{(q_M - q_m) - (q_{bM} - q_{bm})}$$

Mivel a kiemelés mértéke a környezettől függ, az eljárás alkalmazásával a háttér egyenletesebbé válik. Hátránya, hogy érzékeny a pontszerű hibákra; ezt a bemeneti intervallum szélső értékeinek elhagyásával lehet kompenzálni.

21.2.1.3. Képvágások

Vágásnak a képpontok küszöbök szerinti osztályozását nevezzük. A küszöb az osztályozás szempontjából figyelembe vett - mindig mérhető és ezért számszerűsíthető - tulajdonság határértéke.

Legyen

$$\{x_0 < x_1 < \dots < x_n\}$$

a megadott $n+1$ küszöb halmaza; ezekkel a képpontokat n osztályba sorolhatjuk. Jelöljük a vizsgált képjellemzőt α -val. Egy képpont akkor tartozik a i -edik osztályba, ha teljesül a

$$(2-21) \quad x_{i-1} \leq \alpha < x_i$$

feltétel.

A legegyszerűbb vágási eljárásban a képpontokat világosságkódjuk szerint osztályozzuk. Ilyenkor mindig érvényes a $x_0 = q_m$, és a $x_n = q_M + 1$ alapértelmezés. Ha egyben átszínezést is akarunk végezni, célszerű rögtön feltölteni egy Q -táblát is úgy, hogy az i -edik osztályba sorolt világosságkódnak megfelelő táblaelem helyére beírjuk az i -edik osztály q_i színkódját. Ezután az átszínezést a (2-20) képlet alapján lehet elvégezni. Ha n osztályt definiálunk, az eljárást n -szintre vágásnak nevezzük. Ehhez nyilvánvalóan legalább $n-1$ küszöböt kell megadni. A 2-11/a. ábrán 4-szintrevágást mutatunk be. Természetesen nem szükséges, hogy az egy osztályba sorolt világosságkódok egy összefüggő intervallumot alkossanak.

(Az ábrán pl. a 2. és a 4. osztály azonos.)

Igen gyakori eset a két-szintrevágás, amikor a képpontokat - egy küszöb megadásával - értékes, ill. háttér- (vagy nem értékes) képpontoknak minősítjük, s a továbbiakban csak az értékes képpontokat dolgozzuk fel.

További speciális eset a sávkivágás, amihez két küszöbre (x_1, x_2) és két osztályszínre (q_1, q_0) (értékes pontok; q_0 : háttérpontok) van szükség. Az eljárás a

$$x_1 \leq q(k, \ell) < x_2$$

világosságkódú képpontokat q_1 színűre színezi át. A többi pont vagy "háttérszínű" lesz (ld. 2-11/b. ábra), vagy - ha nem adtunk meg háttérszín-kódot - változatlan színű marad (ld. 2-11/c. ábra.)

Zajos képekben rendszerint mind a háttérben, mind az értékes képpontok között vannak azonos világosságkódúak. Ilyenkor az egyszerű vágás nem vezet eredményre. A 2-12/a. ábrán olyan kétszínű tesztábrát mutatunk, amelyben a fekete pontok mindentől előfordulnak, de a sűrűségük nagyobb a keresett objektumban. A 2-18. ábrán megadott T_{a_3}

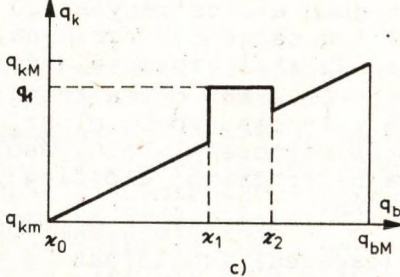
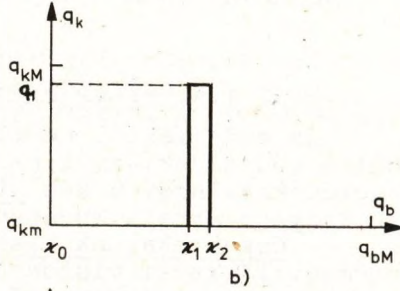
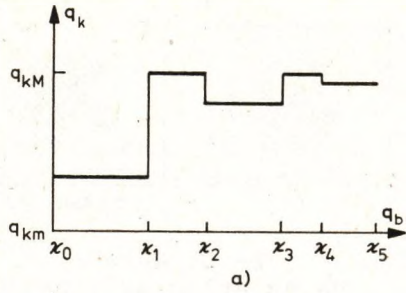
után kétszintrevágás a 2-12/b. ábrán látható eredményre vezet.

Az elvégzett transzformáció:

$$q'(k, \ell) = q(k, \ell) \otimes T_{a_3}(r, s); \text{ majd}$$

$$q''(k, \ell) = \begin{cases} 0 \text{ (fekete)}, & \text{ha } q'(k, \ell) \leq 31; \\ 63 \text{ (fehér)}, & \text{ha } q'(k, \ell) > 31. \end{cases}$$

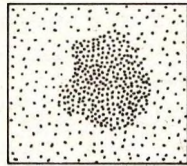
A 2-13/a. ábra tesztábráján a világosságkódok átlaga is azonos. Itt a különböző gradiensű képpontok szétválogatásával lehet a 2-13/b. ábrán látható eredményt elérni.



2-11. ábra

Szintrevágások

nagyobb a keresett objektum simítószűrő alkalmazása

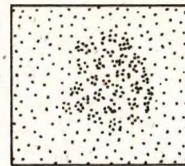


a)



b)

2-12. ábra
Vágás átlag szerint



a)



b)

2-13. ábra
Vágás gradienseloszlás
szerint

21.2.1.4. Hisztogram transzformációk

Az eddigiektől némileg eltérnek azok az ugyancsak globális módszerek, amelyekben úgy transzformáljuk a világosságkódokat, hogy a kép hisztogramja előre meghatározott alakú legyen. Erre a következő esetekben lehet szükség:

- Össze akarunk hasonlítani két képet, amelyek különböző megvilágítási viszonyok mellett készültek, s így a valószínűségben azonos fényességű pontokhoz a két képből különböző világosságkódok tartoznak. Ilyenkor az egyik hisztogramot a másikkal azonos alakúra kell hozni.

- A kép olyan tulajdonságait akarjuk elemezni, amelyek a világosságkódok eloszlásától függenek. Ekkor a megvilágítás változásai befolyásolják az eredményt; ezért célszerű a hisztogramot előzőleg valamilyen "normalizált" alakra hozni.

- Utólag meg akarjuk valósítani az eloszlásfüggő (tapered) kvantálást, a kvantálási hiba csökkentése érdekében rendszerint adott számú kvantumszint megtartásával. Mint láttuk (21.1.2. pont) e célból sűríteni kell a kvantumszinteket azokban a világosságkód-tartományokban, ahová a képpontok nagyobb része tartozik, míg a többi tartományban a szinttávolságok nőnek. Az eljárást közvetlen hisztogram-specifikációnak is nevezzük, mivel hatására az eredménykép hisztogramja előre meghatározott - a további feldolgozás szempontjából kedvező lefutású - függvény közelítése lesz.

Jelölje a bemeneti hisztogram ordinátáit $\chi(q)$, a kimenőt $\tau(r)$. A (2-19)-ből következik, hogy ezek arányosak a q -adik, illetve az r -edik kvantumszinthez tartozó képpontok számával, és

$$\sum_{q=0}^J \chi(q) = \sum_{r=0}^I \tau(r) = 1,$$

ahol J -vel, illetve I -vel a bemenő, illetve a kimenő világosságkódok maximális értékét jelöltük. A továbbiakban csak monoton transzformációkkal foglalkozunk, ezért a halmozott részhisztogramoknak minden lépésben meg kell egyezniük:

$$(2-22) \quad \sum_{q=0}^{m_n} \chi(q) = \sum_{r=0}^n \tau(r) ; (m_n \leq J, n \leq I).$$

Tekintsük a legegyszerűbb esetet, amikor azt akarjuk elérni, hogy a kimenőkép hisztogramja a vízszintes tengelyvel párhuzamos egyenes (közelítése), azaz:

$$\tau(r) \approx \frac{1}{I}$$

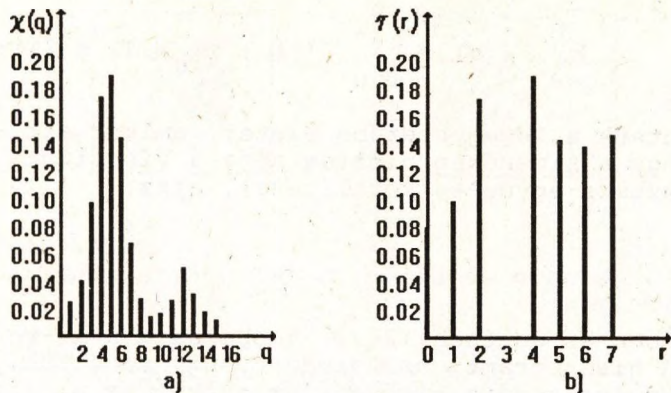
legyen. Ebben az esetben (2-22) jobboldala n/I -re egyszerűsödik. A hisztogramtranszformációt úgy végezzük, hogy $n=1$ -től indulva rendre kiszámítjuk azokat az m_n értékeket, amelyekre:

$$\sum_{q=0}^{m_n} \chi(q) \leq I < \sum_{q=0}^{m_{n+1}} \chi(q).$$

Az egyenlőtlenség alapján adódó n értékek kijelölik a kimenő hisztogramban a kvantumszinteket, az m_n index alapján pedig meghatározhatók az adott kvantumszinthez tartozó bemenő világosságkódok. Mivel a kiinduló feltevés teljesülése esetén a kimenő hisztogramban az egyes világosságkód-intervallumokba közel egyenlő számú képpont esik, ezt az eljárást hisztogramkiegyenlítésnek nevezzük. A 2-14/a. ábrán a bemenő-, a b)-n a kimenőkép hisztogramját látjuk, ahol a lehetséges világosságkódok számát a felére csökkentettük. Teljes kiegyenlítés csak akkor volna lehetséges, ha a bemenőképen azonos világosságkódú képpontok halmazát felbontanánk, és a kimenőképben különböző világosságkódú halmazokba sorolnánk. Önkényesen végrehajtott felbontás természetesen értelmetlen és hamis eredményhez vezet, csak akkor célszerű, ha valamilyen módon a lokális tulajdonságokat fel tudjuk használni.

A probléma általános megoldásához vegyük figyelembe hogy - a világosságkódot folytonos valószínűségi változónak tekintve - a hisztogram a sűrűségfüggvény, míg a halmozott részhisztogram az eloszlásfüggvény közelítésekként értelmezhető. Ez alapján a (2-22) feltételt az alábbiak szerint írhatjuk át:

$$\int_{q_{\min}}^q s(\zeta) d\zeta = \int_{r_{\min}}^r t(\rho) d\rho$$



2-14. ábra
Hisztogramkiegyenlítés hatása a hisztogramra

ahol $s(\xi)$ a bemenő, $t(\xi)$ pedig a kimenő sűrűségfüggvény.
Hisztogramkiegyenlítés esetén

$$t(\xi) = \frac{1}{r_{\max} - r_{\min}} = \text{állandó},$$

amit az eloszlásfüggvényekre felírt összefüggésbe beírva az alábbi eredményre jutunk:

$$\int_{q_{\min}}^q s(\xi) d\xi = \frac{1}{r_{\max} - r_{\min}} (r - r_{\min}).$$

Ha a bemenő eloszlásfüggvény helyére behlyettesítjük a diszkrét közelítésnek megfelelő halmozott bemenő részhisztogramot, minden q_b bemenő világosságkódhoz explicit módon meghatározhatjuk az r kimenő világosságkód közelítő értékét:

$$(2-23) \quad r \approx (r_{\max} - r_{\min}) \sum_{q=0}^{q_b} \chi(q) + r_{\min}.$$

Az előbbihez hasonló módon más kimenő hisztogramot eredményező átviteli függvényt is meghatározhatunk. Gyakran használatos eljárás pl. a hisztogram hiperbolizáció; amikor is

$$t(\xi) = \frac{c}{\xi}.$$

Ha az állandót

$$c = \frac{1}{\ln r_{\max} - \ln r_{\min}}$$

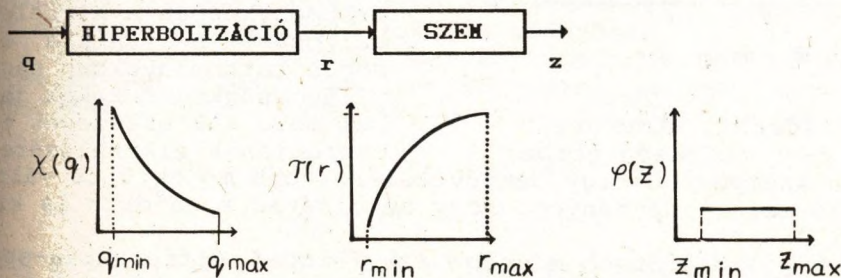
értékűre választjuk, akkor

$$\int_{q_{\min}}^q s(\xi) d\xi = \frac{\ln r - \ln r_{\min}}{\ln r_{\max} - \ln r_{\min}} = \frac{\ln \frac{r}{r_{\min}}}{\ln \frac{r_{\max}}{r_{\min}}};$$

amiből az r közelítő értéke:

$$(2-24) \quad r \approx r_{\min} \left(\frac{r_{\max}}{r_{\min}} \right)^{\frac{\sum_{q=0}^{qb} \chi(q)}{qb}}$$

A hisztogram hiperbolizációjával azt szeretnénk elérni, hogy a kimenőkép hisztogramja hiperbolikus lefutású legyen. Ennek azért van különös jelentősége mert - mint említettük (ld. 11.2.2. pont) - az emberi szem a világosságkódokat logaritmikusan képezi le, így a hiperbolizált hisztogramú képről az agyban eredőül kiegyenlített képet kapunk (ld. 2-15. ábra).

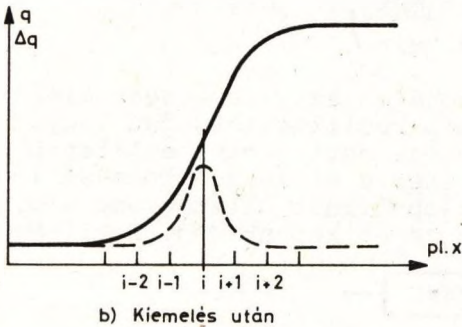
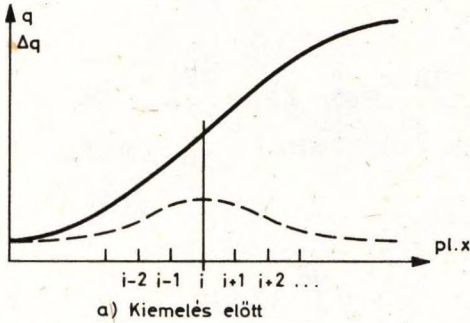


2-15. ábra
Hisztogram hiperbolizáció hatása

Az előzőekhez hasonlóan más kimenő hisztogram esetére is meghatározható a kimenő világosságkódokat megadó képlet. Egy adott képet természetesen nem pontról pontra transzformálunk, hanem az összes bemenő világosságkódhoz egyszer kiszámítjuk a megfelelő kimenő értéket és az így elkészített hozzárendelési táblázattal dolgozunk.

21.2.2. Élkiemelés (élesítés)

A torzítások második csoportja a határátmenetek kiszélesedésében, az élek elmosódásában jelentkezik. Pszichofizikai kísérletekkel bebizonyították, hogy szubjektíve előnyösebb érzetet kelt a túlhangsúlyozott élekkel rendelkező kép, mint a valósághű ábrázolás.



2-16. ábra

Ideális átmeneti-
és meredekség-görbe

egyéb szempontok figyelembevételére több módosított változat létezik.) Hátrányos, hogy az eljárás a zajokat is kiemeli.

Legelterjedtebb módszer a különböző differencia-operátorok használata; ezeket rendszerint konvolúciós szűrővel realizálják.

Az eljárás (1-6) alapján azt jelenti, hogy minden képpont világosságkódját helyettesítjük a szűrő által kijelölt környezetében levő képpontok világosságkódjának valamilyen súlyozott átlagával.

Talán a leggyakrabban alkalmazott élkiemelő függvény az

$$(2-25) \quad f(x, y) - \nabla^2 f(x, y)$$

kifejezés, ahol a ∇^2 az (1-13)-ban definiált digitális Laplace-operátor.

Ezen túlmenően az élkiemelést a szegmentálásnál tárgyalt élkitűzés (ld. 23.2 alfejezet) előkészítő fázisaként is gyakran alkalmazzák.

Az eljárások célja az egyes képrészletek közötti átmeneti tartomány szűkítése; a határátmenetek meredekebbé tétele (ld. 2-16. ábra); az elmosódások korrigálása; következésképpen általában lokális jellegűek.

Élkiemelést lehet megvalósítani a statisztikus differencia-operátorral. Az eljárás során egy képpont korrigált világosságkódját úgy számítják ki, hogy elosztják a környezetében levő képpontok világosságkódjának eloszlása alapján meghatározott szórással. Az eredményképben a hirtelen változások helyén kis, a homogén tartományokban nagy világosságkód-értékek jelennek meg. (Az esetleges túlcsoportulások elkerülésére és

Így a (2-25) függvényt a 2-17. ábrán megadott T_{f_1} konvolúciós szűrő realizálja (ld. még 1-26. ábra).

$$T_{f_1} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} T_{f_2} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix} T_{f_3} = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 5 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

2-17. ábra
Felületáteresztő (élkiemelő) szűrők

Számos más eljárás is létezik; további 2 példát láthatunk a 2-17. ábrán (T_{f_2} és T_{f_3}). Figyeljük meg, hogy a szűrőelemek összege (= a szűrő súlya) mindig 1!

21.2.3. Zajelnyomás (simítás)

A képhibák harmadik csoportját a zajok alkotják. A felvétel, a digitalizálás, az átvitel stb. során a képet különféle zavaró hatások érik, amelyek megváltoztatják a képpontok eredeti világosságkódját, hamis információkat visznek be. A zajokat az érdemi feldolgozás előtt el kell távolítani a képről, de legalábbis csökkenteni kell a hatásukat. Az erre irányuló eljárásokat nevezzük zajelnyomásnak vagy simításnak.

A zaj általában véletlenszerűen előforduló, izolált pontokban jelentkezik, amelyeknek világosságkódja lényegesen eltér a környezetüktől. Ezen a megfigyelésen alapulnak a zajok elnyomására szolgáló lokális módszerek.

A legegyszerűbb esetben minden képpont világosságkódját összehasonlítjuk valamilyen környezetében levő képpontok világosságkódjának átlagértékével (\bar{q}). A korrekció képlete:

$$q_k(k, \ell) = \begin{cases} \bar{q}, & \text{ha } |q_b(k, \ell) - \bar{q}| > \varepsilon; \\ q_b(k, \ell), & \text{különben;} \end{cases}$$

ahol ε a kép jellegétől függő, kísérleti úton meghatározandó küszöbérték. A \bar{q} 3x3 méretű környezet esetében pl. a 2-18. ábrán látható T_{a_1} konvolúciós szűrővel számítható ki.

$$T_{a_1} = \frac{1}{8} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} T_{a_2} = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} T_{a_3} = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

2-18. ábra
Aluláteresztő (simító) szűrők

A 2-18. ábrán további két (T_{a_2} és T_{a_3} aluláteresztő szűrőt mutatunk. Ezek elemei mindig pozitívak, és összegük ugyancsak mindig 1! Többnyire négyzetesek, és ritkán nagyobbak 5×5 méretűnél, mert akkor már túlságosan "elmosnák" a képet.

Kedvező tulajdonságai miatt szívesen alkalmazzák zajelnyomásra a mediánszűrést, amely ugyan lokális jellegű, de nem konvolúciós típusú. (Egy digitális minta mediánján a rendezett minta középső elemét értjük.) A mediánszűrés során a figyelembe vett mintát a szűrőbe eső képpontok világosságcódja alkotja. A szűrést úgy valósítják meg, hogy minden képpont világosságcódját kicserélik a ráillesztett szűrő által meghatározott minta mediánjával.

A 2-19. ábra az egyszerű átlagoló (2-18. ábra T_{a_2}) és a mediánszűrő hatását mutatja be lépcső, lejtő, háromszög és impulzus jellegű határátmenetre 5 képpontból álló ablak esetén (egyszerűség kedvéért egydimenziós mintára). Az ábráról leolvashatók a mediánszűrő kedvező tulajdonságai: a lépcsős- és a lejtő-átmenet jellegét nem változtatja meg; az impulzusszerűen jelentkező zajt elnyomja, ha az impulzus periódusa kisebb mint az ablak szélességének fele. A háromszög-átmenet csúcsát viszont kissé letompítja.

A mediánszűrő általában négyzet alakú, de használnak pl. a képaránynak megfelelő méretű téglalapot is. Ha a szűrőméret páros, rendszerint a két középső elem számtani közepét tekintik mediánnak.

A simítószűrők gyakorlati használatánál ügyelni kell arra, hogy nem megfelelő alkalmazásuk több kárt okozhat, mint amennyi hasznot várunk tőle. Általában célszerű figyelembe venni a kép jellegét, és az eredményt vizuálisan is ellenőrizni.

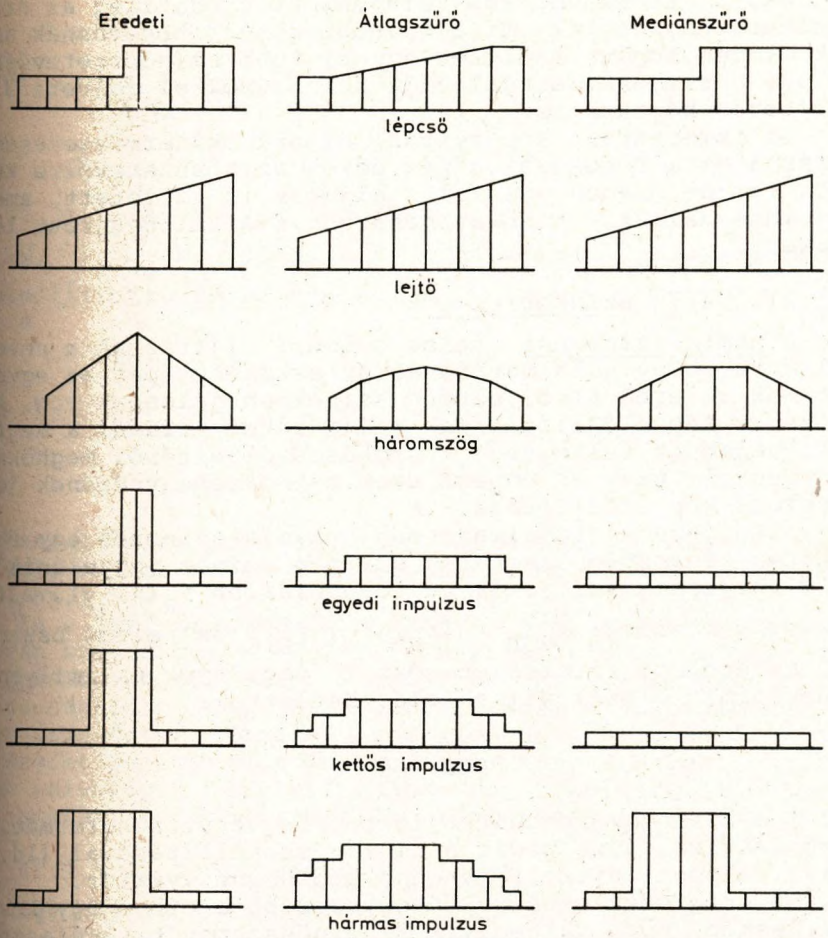
Bizonyos esetekben nem egyetlen képen belüli, hanem időbeli átlagolással javíthatjuk a kép minőségét. A többszörös képátlagolási eljárás akkor alkalmazható, ha az egy-egy más után felvett képekre additív zaj rakódik, azaz

$$q_i(k, l) = q(k, l) + z_i(k, l);$$

ahol i a képvétel sorszáma, és feltesszük, hogy a $z_i(k, l)$ zaj minden pontban független és nulla átlagú. Ebben az esetben a zajos képek $q_i(k, l)$ sorozatát véve és képpontonként átlagolva az

$$r(k, l) = \frac{1}{n} \sum_{i=1}^n q_i(k, l)$$

képet kapjuk (mivel feltevésünk szerint a zaj átlaga minden képpontban nulla). Az $r(k, l)$ kép annál jobban megközelíti a zajmentes $q(k, l)$ képet, minél több képre végezzük el az összegzést.



2-19. ábra
 Átlag- és mediánszűrő hatása

21.2.4. Többsávós képek feldolgozása

A többsávós (multispektrális) kép ugyanarról a jelenetről egyidejűleg több frekvenciasávban készített felvételtől áll. Ezt a

$$(2-26) \quad \vec{q}(n)^T = \{q_1(k, \ell), q_2(k, \ell), \dots, q_n(k, \ell)\}$$

szimbólummal jelöljük; $\vec{q}(n)$ -et $n \times 1$ elemű oszlopvektorként kezeljük. A $q_i(k, \ell)$ függvénnyel leírt digitális képet (i -edik) színösszetevőnek nevezzük.

Bár a "többsávós kép" elnevezést eredetileg az űrfelvételekre használták, mi a továbbiakban többsávósnak nevezünk minden képet, amelynek egynél több színösszetevője van, így pl. a színeskamerával vagy színszűrőkkel felvett 3 színösszetevős képeket is.

Az ismertetett képjavítási eljárások természetesen alkalmazhatók a többsávós képek egyes színösszetevőire külön-külön is, de számos speciális eljárás is született, amelyekben kihasználják a több színösszetevő által nyújtott lehetőségeket.

21.2.4.1. Színkorrekciók

A hamis-színezési (false colour) eljárásnak a neve is utal arra, hogy alkalmazásával az eredményképen az egyes objektumok az eredetitől eltérő színekben jelennek meg. A hamisszínes kép előállításának célja lehet például a megfigyelő figyelmének felkeltése a szokásostól eltérő, meghökkentő színezéssel; vagy az emberi szem színérzékenységének jobban megfelelő kép előállítása.

Általános megfogalmazásban a hamis-színezés egy függvénytranszformáció, amelynek során a színes megjelenítőt vezérlő folytonos színjelek (a továbbiakban $f_R(x, y)$, $f_G(x, y)$, $f_B(x, y)$; röviden: $\vec{f}_{RGB}^T = \{f_R, f_G, f_B\}$, 3×1 elemű oszlopvektor) az eredeti színösszetevőkből, vagy ezek valamilyen transzformációjából állnak elő. Képlettel:

$$\vec{f}_{RGB} = \underline{T} \vec{f}(n)$$

ahol \underline{T} a színhozzárendelést leíró függvénytranszformáció 3×1 elemű mátrixa, $\vec{f}(n)$ pedig a $\vec{q}(n)$ visszaállításával (ld. 21.1.3. alpont) nyert folytonos képfüggvényvektor.

Az űrfelvételekből a legegyszerűbb esetben úgy készül a színeskép, hogy valamelyik 3 színösszetevőt kiválasztjuk és megfeleltetjük a 3 színjelnek. A problémát az okozza, hogy a LANDSAT MSS rendszerében csak két színösszetevő készül a látható sávban (zöld és piros), további kettő a közeli és távoli infravörös tartományban. Így valódiszínes képet nem lehet létrehozni, de 24-féle hamisszínes képet állíthatunk elő; ezeket kompozit képnek szokták nevezni. (Többnyire valamelyik infravörös színösszetevőt feleltetjük meg a piros színnek, az így készült kompozit képeken a növényzet élénkpiros színekben jelenik meg.)

A természetes színesképek lineáris leképezése esetén \underline{T} 3×3 elemű mátrix:

$$\vec{f}_{RGB} = \underline{T} \cdot \vec{f}$$

vagy részletesen kiírva:

$$(2-27) \quad \begin{bmatrix} f_R \\ f_G \\ f_B \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \cdot \begin{bmatrix} f_r \\ f_g \\ f_b \end{bmatrix}$$

ahol $f_r = f_r(x, y)$, $f_g = f_g(x, y)$, $f_b = f_b(x, y)$ az eredeti $q_r(k, l)$, $q_g(k, l)$, $q_b(k, l)$ digitális színösszetevőkből visszaállított, folytonos színjel, és $\vec{f}^T = \{f_r, f_g, f_b\}$.
(Pl. a

$$\underline{T} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

mátrixszal leírt transzformáció hatására egy családi házat ábrázoló színes képen a fű piros, az ég zöld és a háztető kék lesz.)

Ha a "színösszetevők" száma 1, azaz monokromatikus képről van szó, akkor a (2-27) transzformáció eredményét álszínes (pseudo colour) képnek nevezzük. Ez a leképezés minden világosságkódhoz egy helyvektort rendel a háromdimenziós (RGB) színtérben. Hangsúlyozni kívánjuk, hogy az álszínes kép eredetije egysávos (nem színes) kép, esetleg nem is kép, hanem valamilyen mátrix formában rendelkezésre álló adathalmaz, amelyet a háromdimenziós színtérbe transzformálunk.

21.2.4.2. Pontonkénti műveletek

Így nevezzük azokat az eljárásokat, amelyekben két vagy több színösszetevő azonos geometriai koordinátájú képpontjaihoz tartozó világosságkódokból állítjuk elő az eredményképet. Gyakran végeznek kivonást a

$$q_{mn}(k, l) = q_m(k, l) - q_n(k, l)$$

összefüggés szerint az additív zajok vagy a leképezőrendszer helytől függő additív jellegű torzításainak kiszűrésére.

A másik egyszerű művelet az osztás a

$$q_{mn}(k, l) = \frac{q_m(k, l)}{q_n(k, l)}, (q_n(k, l) \neq 0)$$

képlet szerint, a megvilágítási hibákból eredő világosságkód-torzítások kiegyenlítésére. (A hányadosképzés azon a modellen alapszik, hogy egy színösszetevő világosságkódjai

a megvilágított felület objektív $R(k, \ell)$ visszaverőképességének (amely képpontonként változik) és a képpont $M(k, \ell)$ -lel jellemezhető megvilágítás-erősségének szorzataként állnak elő; ahol az $M(k, \ell)$ ugyan képpontonként változhat, de egy képpontra közel azonos minden frekvenciasávban.) Ily módon az eredményképből ki lehet szűrni az egyenetlen háttér megvilágítás torzító hatását.

A színösszetevőknek különböző súlyozó konstansokkal és normalizáló tényezőkkel képzett lineáris kombinációit és hányadosait lehet definiálni. A lehetséges változatok nagy száma miatt véletlenszerű alkalmazásuk értelmetlen: kielégítő eredményeket csak tudatos tervezéssel, a vizsgált objektumok és az érzékelőrendszer paramétereinek ismeretében kialakított formulákkal lehet elérni.

Megjegyezzük még, hogy a többsávós képfeldolgozás alapvető problémája a feldolgozandó adatok nagy tömege (egyes rendszerekben már 7-13 sávban készítenek felvételeket). Ezek egyidejű feldolgozása még a ma rendelkezésre álló nagykapacitású számítógépek segítségével is nehézkes. E probléma feloldása úgy lehetséges, hogy egy-egy konkrét feladatra csak 1-2 színösszetevőt használunk. Ezek megfelelő kiválasztásán túlmenően célszerű kiszűrni (és egy lezármaztatott képben tömöríteni) a számításigényes feldolgozások előtt a különböző színösszetevőkben levő hasznos információkat. Ennek szokásos megoldása a főkomponens transzformáció (amit a 31.1.2.1. alpontban tárgyaltunk), azonban gyakran eredményesen alkalmazhatók a színösszetevők közötti egyszerű alapműveletek is.

21.2.5. Képjavítások frekvenciatartományban

Az előző pontokban ismertetett képjavítási eljárások a képpontok eredeti világosságkódjával végzett műveleteken alapultak. A képtartománybeli eljárások előnye, hogy könnyen programozhatók, hatékonyak és viszonylag kicsi a gépidőigényük. A számítógépek teljesítményének növekedése ma már lehetővé teszi, hogy frekvenciatartományban működő algoritmusokat is megvalósítsunk elfogadható válaszidők mellett; különösen amióta megjelentek a Fourier-transzformációt (oda-viszsa) végrehajtó speciális processzorok.

A frakvenciatartományban végzett (globális) javítások a konvolúcióelméleten alapulnak (ld. 13.2.2. pont és 13.2.4.2. alpont), amikor is a képet kétváltozós függvényként kezeljük. Képlettel:

$$g(x, y) = f(x, y) \otimes h(x, y),$$

ahol $f(x, y)$ az eredeti, $g(x, y)$ a javított képet, $h(x, y)$ a javítást végző, eltolásinvariáns "szűrőfüggvényt", \otimes pedig a konvolúcióintegrált jelöli. Ismeretes, hogy ez utóbbival egyenértékű eredményhez jutunk, ha az integranduszok Fourier-transzformáltjának szorzatát képezzük, majd erre inverz Fourier-transzformációt hajtunk végre.

A képjavítás ezek szerint most úgy történik, hogy először előállítjuk a kép $F(u,v)$ Fourier-transzformáltját, majd megszorozzuk egy alkalmasan választott $H(u,v)$ szűrőfüggvény-nyel:

$$G(u,v) = F(u,v) * H(u,v).$$

Ha most $G(u,v)$ -re inverz Fourier-transzformációt hajtunk végre, megkapjuk a javított $g(x,y)$ képet.

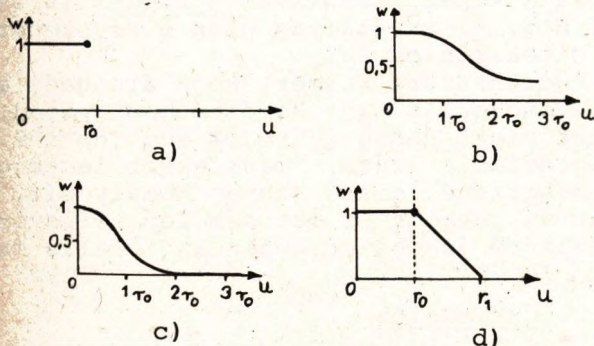
21.2.5.1. Zajsűrés

A zajok éles átmenetek, "ugrások" a világosságkódokban, így a kép Fourier-transzformáltjának magasfrekvenciás összetevőiben jelentkeznek. Ebből következik, hogy a frekvencia-tartományban a nagyfrekvenciás összetevők kiszűrése a zajok elnyomását eredményezi. Zajsűrésre tehát aluláteresztő szűrőket (lowpass filter) kell használnunk.

A legegyszerűbb, "ideális" aluláteresztő szűrő átviteli függvénye a következő:

$$H(u,v) = \begin{cases} 1, & \text{ha } \rho_k(u,v) \leq r_0; \\ 0, & \text{különben;} \end{cases}$$

itt $\rho_k(u,v)$ az (1-33)-mal definiált, az euklidészi közelítő távolságfüggvény (ld. 2-20/a. ábra). Mind az ideális, mind a folytonos átmenetű szűrők körszimmetrikusak, azaz a w -tengely-re illeszkedő bármely síkban azonos metszetet kapunk. (Ha a $H(u,0)$ függvényt ábrázoljuk az u - v síkban majd ezt a w -tengely körül körbeforgatjuk, akkor megkapjuk a szűrő-függvény felületét.)



2-20. ábra
Aluláteresztő szűrők

Látható, hogy az ideális szűrő változatlanul átengedi az r_0 sugarú kör belsejébe eső kisfrekvenciás összetevőket, a körön kívüli nagyobb frekvenciájúakat pedig teljesen ki-

szűri. Azokat a frekvenciapárokat, ahol a $H(u,v)$ függvénynek szakadása van vágási frekvenciáknak nevezzük.

Nézzünk néhány további, gyakran alkalmazott szűrőfüggvényt.

Butterworth-szűrő (2-20/b. ábra):

$$H(u,v) = \frac{1}{1 + \left(\frac{\rho_k(u,v)}{r_0} \right)^{2n}} ;$$

Exponenciális szűrő (2-20/c. ábra):

$$H(u,v) = \exp \left[- \left(\frac{\rho_k(u,v)}{r_0} \right)^n \right];$$

Trapézszűrő (2-20/d. ábra):

$$H(u,v) = \begin{cases} 1, & \text{ha } \rho_k(u,v) < r_0; \\ \frac{\rho_k(u,v) - r_1}{r_0 - r_1}, & \text{ha } r_0 \leq \rho_k(u,v) \leq r_1; \\ 0, & \text{ha } \rho_k(u,v) > r_1. \end{cases}$$

Az ideális szűrő a zajokkal együtt a magasabb frekvencia-tartományba eső értékes képi összetevőket is kiszűri, ezért elhomályosítja a képet. Ezenkívül gyűrődés is fellép, ami azt jelenti, hogy visszaállítás után a képen periodikusan jelentkező foltok tűnnek fel.

A Butterworth-szűrő kismértékben átengedi a nagyfrekvenciás összetevőket is, így kevésbé homályosodik el a kép; a síma átmenet miatt pedig gyűrődés sem lép fel.

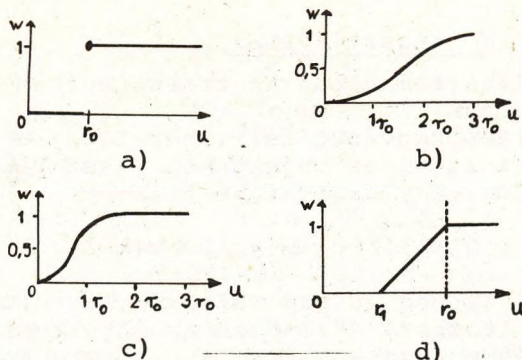
Az exponenciális szűrő - meredekebb lecsengésű lévén - a Butterworth-szűrőnél kissé jobban homályosít, de korántsem annyira, mint az ideális és itt sem lép fel gyűrődés.

A trapézszűrő átmenetet képez az ideális és a másik két szűrő között.

21.2.5.2. Élkiemelés

Az élek - a zajokhoz hasonlóan - nagyobb változást jelentenek a világosságkódokban, vagyis szintén a kép Fourier-transzformáltjának nagyobb frekvenciájú komponenseit adják. Ezért felüláteresztő szűrők (highpass filter) alkalmazásával tudjuk az élesítést elvégezni, amelyek kiszűrik a kisfrekvenciás komponenseket és változatlanul átengedik a többi.

A zajelnyomásnál megismert aluláteresztő szűrők felüláteresztő megfelelőit az alábbiakban adjuk meg.



2-21. ábra
Felüláteresztő szűrők

Ideális szűrő (2-21/a. ábra):

$$H(u, v) = \begin{cases} 0, & \text{ha } \rho_k(u, v) < r_0; \\ 1, & \text{különben.} \end{cases}$$

Butterworth-szűrő (2-21/b. ábra):

$$H(u, v) = \frac{1}{1 + \left(\frac{r_0}{\rho_k(u, v)} \right)^{2n}};$$

Exponenciális szűrő (2-21/c. ábra):

$$H(u, v) = \exp \left[- \left(\frac{r_0}{\rho_k(u, v)} \right)^n \right];$$

Trapézszűrő (2-21/d. ábra):

$$H(u, v) = \begin{cases} 0, & \text{ha } \rho_k(u, v) < r_1; \\ \frac{\rho_k(u, v) - r_0}{r_1 - r_0}, & \text{ha } r_1 \leq \rho_k(u, v) \leq r_0; \\ 1, & \text{ha } \rho_k(u, v) > r_0; \end{cases}$$

Az ideális felüláteresztő szűrő értelemszerűleg levág egy adott frekvencia alatt minden összetevőt, azaz teljesen kiszűri a kép kisfrekvenciás összetevőit; különben mindegyikük "mellékhatásai" megegyeznek a megfelelő aluláteresztő szűrőével.

21.2.5.3. Kontrasztjavítás

Frekvenciatartományban kontrasztjavítás is végezhető, az alábbi elv alapján.

A $q(k,l)$ képfüggvényt fel lehet fogni a megvilágítástól függő $m(k,l)$ és az egyes objektumok visszaverőképességétől függő $v(k,l)$ függvény szorzataként is:

$$q(k,l) = m(k,l) * v(k,l).$$

Az előbbi térben lassan változik, az utóbbi azonban hirtelen is változhat, különösen az objektumok határán. Durva közelítésben azt mondhatjuk, hogy a Fourier-transzformált kisfrekvenciás összetevői a megvilágítással, a nagyfrekvenciásak a visszaverőképességgel hozhatók kapcsolatba.

A kép tulajdonságait kedvezően befolyásolhatjuk, ha olyan sávszűrőt alkalmazunk, amelyik csökkenti a kisfrekvenciás és növeli a nagyfrekvenciás összetevők hatását.

Mivel két függvény szorzatának Fourier-transzformáltja nem szeparálható, egy ilyen eljárás előtt célszerű áttérni a képfüggvény logaritmusára, hogy függvényösszegekkel dolgozhassunk. A szűrés és az inverz transzformáció elvégzése után természetesen exponenciális leképezést kell végrehajtani. A megfelelő szűrő kialakítására a fentiek ismeretében tapasztalati úton kerülhet sor.

21.3. HELYREÁLLÍTÁS (RESTAURÁCIÓ)

A helyreállítási eljárások során arra törekszünk, hogy a rendelkezésre álló képből előállítsuk azt az ideális képet, amely megfigyelhető lett volna, ha a különböző zavaró, torzító hatások nem léptek volna fel a leképezés során. Hatékony helyreállítási eljárás tervezéséhez alapvetően fontos a torzító hatásokat leíró matematikai modell megalkotása. Ezt követően meg kell határozni azt az inverz transzformációt, amely a modell által leírt hatásokat kiküszöböli, és ezt végre kell hajtani az adott képre.

Nyilvánvaló, hogy ezektől az eljárásoktól csak jó modell esetén várható kielégítő eredmény.

21.3.1. Elvi és gyakorlati nehézségek

Bár a képhelyreállítás elvei nagyon kézenfekvőnek és egyszerűnek tűnnek, a valóságban már a zavaró hatásokat leíró modell megalkotása is számos problémát vet fel. Figyelembe véve a potenciális zavaró tényezőket - például az optikai leképező rendszer hibáit, az érzékelők nemlineáris

voltát, az atmoszférikus hatásokat, az elmozdulásokból adódó életlenséget, a geometriai torzulásokat stb. - érthető, hogy reménytelen vállalkozás lenne univerzális modell megalkotása. A gyakorlatban többé-kevésbé kielégítő modellek vannak az egyes hibatípusokra.

A modell létrehozása érdekében méréseket kell végezni a vizsgált részrendszeren (pl. a felvevő-, a digitalizáló vagy megjelenítő rendszeren). Néha lehetséges a részrendszer hatásának determinisztikus leírása, máskor csak sztochasztikusan határozható meg.

Egyszerű példaként tekintsük a kameraképeknél jelentkező árnyékhatást (shading); nevezetesen, hogy a kép szélei felé sugárirányban csökken a kép fényessége. Ez a hatás elméletileg determinisztikusan leírható az optikai törvények alapján. A gyakorlatban azonban sokszor nem ismerjük a rendelkezésre álló eszközök pontos specifikációját, vagy az egyéb zavaró tényezők oly mértékben befolyásolják a rendszert, hogy a determinisztikus modell használhatatlan. Ekkor tesztábrákon végzett mérések alapján lehet a modell paramétereit meghatározni.

A képhelyreállítási eljárások alkalmazásának másik problémája, hogy az elvégzendő inverz transzformáció legtöbbször olyan bonyolult képletekkel írható csak le, hogy még a mai számítógépekkel sem lehet elfogadható idő alatt végrehajtani.

A helyreállítási eljárások részletesebb tárgyalása messze túlvezet a jelen jegyzet keretein, ezért példaként csak az inverz szűrő alkalmazását ismertetjük. Ezenkívül a következő fejezetben részletesen foglalkozunk a geometriai torzulások helyreállításával.

21.3.2. Inverz szűrés

Egyik gyakori hibajelenség a kép elmosódása, "életlensége". Ezt a hibatípust lineáris szűrőkkel lehet javítani. Az elmosódottságot úgy lehet modellezni, hogy a mintavételezést nem Dirac-deltával, hanem véges szélességű és magasságú impulzussal végezzük, amelyet a $p(x,y)$ impulzusfüggvény ír le. Ebben az esetben a digitalizált kép (2-2)-től eltérően

$$(2-28) \quad h(k, \ell) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} f(m \cdot \Delta x, n \cdot \Delta y) \cdot p(x-m \cdot \Delta x, y-n \cdot \Delta y) + z(k, \ell)$$

alakú lesz, ahol a $z(k, \ell)$ függvénnyel az additív zajt is figyelembe vettük.

Kísérreljük meg visszaállítani az eredeti képet egy $r(x,y)$ lineáris, eltolás-invariáns interpolátorfüggvény konvolúciójával:

$$(2-29) \quad f^{-1}(x, y) = h(k, \ell) \otimes r(x, y)$$

Az interpolátorfüggvény meghatározása céljából Fourier-transzformációval vizsgáljuk (2-29) síkfrekvencia-spektrumát:

$$(2-30) \quad F^{-1}(u, v) = [F(u, v) \cdot P(u, v) + Z(u, v)] \cdot R(u, v)$$

Itt felhasználtuk (2-28)-at és azt az ismert szabályt, hogy a konvolúciós integrálás helyett a Fourier-transzformáltak szorzatával lehet számolni. Válasszuk úgy az interpolátorfüggvényt, hogy a Fourier-transzformáltjára teljesüljön az

$$(2-31) \quad R(u, v) = \frac{1}{P(u, v)}$$

összefüggés (innen van az "inverz szűrő" elnevezés). Ha (2-31)-et behelyettesítjük (2-30)-ba, a visszaállított kép Fourier-transzformáltja

$$(2-32) \quad F^{-1}(u, v) = F(u, v) + \frac{Z(u, v)}{P(u, v)}$$

alakú lesz. Végezzük el (2-32)-re az inverz Fourier-transzformációt, ekkor a visszaállított képre az

$$(2-33) \quad f^{-1}(x, y) = f(x, y) + e(x, y)$$

végeredményt kapjuk, ahol $e(x, y)$ a $\frac{Z(u, v)}{P(u, v)}$ függvény inverz Fourier-transzformáltja.

(2-33)-ból látható, hogy ha a képen nem lett volna zaj, az inverz szűrővel visszaállított kép megegyeznék az eredetivel. Sajnos a magasabb síkfrekvenciákon mind az $F(u, v)$, mind a $P(u, v)$ kicsi. Így egyrészt a zaj elnyomja a kép magasfrekvenciás összetevőit (azaz a kép finomabb részleteinél jelentős torzulás következhet be), másrészt már kis zaj hatása igen erősen jelentkezik a visszaállított képben.

Lényegesen jobb eredményt lehet elérni a Wiener-szűrő-nek nevezett speciális inverz szűrővel, amelyik a zaj hatását is figyelembe veszi. Elve az, hogy a visszaállított és az eredeti kép négyzetes eltérésének hibáját minimalizálja. Hatása csak akkor kielégítő, ha a zaj spektruma sokkal szélesebb a kép spektrumánál és ebben a tartományban a sűrűségfüggvénye állandónak tekinthető ("fehérzaj"), továbbá nincs korreláció a kép és a zajfüggvény között (ami teljesen sohasem igaz).

Az inverz szűrők alkalmazása rendkívül számításgényes: a műveletszám $K \times K$ méretű kép esetén K^4 -nel arányos. A számítások csökkentésére rekurzív szűrőket dolgoztak ki.

22. Geometriai korrekciók

A digitális képfeldolgozásban sűrűn előforduló, nagy számításigényű feladat a képek geometriai korrekciója. Ez általános esetben egy (vagy több, geometriailag összetartozó) bemenőkép átranzformálását jelenti kimenőkép(ek)re úgy, hogy a képi információ a megadott mértéknél kevésbé sérüljön, míg a kívánt geometriai összefüggés a megadott mértéknél pontosabban teljesüljön.

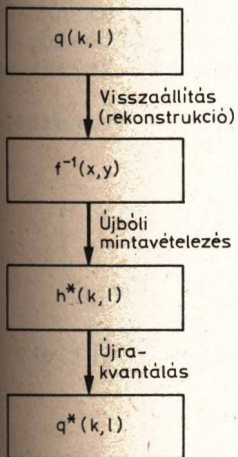
A digitális kép általában a folytonos, háromdimenziós világnak valamely képfeltevőrendszer segítségével történő méréseként jön létre. Geometriai korrekcióra pl. a következő okok miatt lehet szükség:

- a képfeltevő rendszer torzít, ill. pontatlan;
- a háromdimenziós világ objektumainak (kétdimenziós) síkra vetítése torzít (perspektivikus torzítás);
- a mérés folyamata közben az objektum és a mérőeszköz közti geometriai összefüggés változó (elmozdulás);
- az objektum különböző képei közt kell megfeleltetést létrehozni.

A geometriai korrekció modelljét a 2-22. ábrán szemléltetjük.

Visszaállításkor a $q(k, \ell)$ digitális képhez az $f^{-1}(x, y)$ folytonos képfüggvényt rendeljük hozzá, amelytől elvárjuk, hogy az eredeti $f(x, y)$ képet a lehető legjobban megközelítse. Ezzel a digitális kép geometriai korrekciója visszavezethető a folytonos képfüggvényből való újbóli mintavételezésre ($h^*(k, \ell)$) és új kvantálásra ($q^*(k, \ell)$). Ez utóbbi, geometriailag korrigált képnek (lehetőleg) ugyanazt a képi információt kell tartalmaznia, mint a korrigálatlan $q(k, \ell)$ képnek, a kívánt geometriai összefüggés teljesítése mellett.

A korrekció geometriai jellegét az eredeti képkalkotáskor használt, ill. az újbóli mintavételezés különbözősége adja. Az újramintavételezés rendjét a korrekciós összefüggés szabja meg.



2.22. ábra

A geometriai korrekció modellje

22.1. LINEÁRIS GEOMETRIAI KORREKCIÓK

Tegyük fel, hogy a digitális képkalkotásnál alkalmazott mintavételezés a valós, háromdimenziós világ valamely téreleméhez a $P(x, y)$ képpontot rendelte hozzá. Ha a kapott kép geometriai korrekciójával azt akarjuk elérni, hogy ugyanen-

nek a térelemnek a korrigált képen a $P'(x', y')$ képpont feljeljen meg, akkor szükség van az

$$(x, y) \longrightarrow (x', y')$$

koordinátatranszformációra. A korrekció végrehajtásakor így határozhatjuk meg, hogy a P' képpont előállításához a korrigálandó kép melyik P pontjának környezetében kell visszaállítani az eredeti képet, amelyet azután majd újradigitalizálunk.

Tekintsük először a lineáris koordinátatranszformációkat. Legyen V és W két vektortér a T test felett. Ismeretes, hogy az

$$\mathcal{L} : V \longrightarrow W$$

leképezést akkor nevezzük lineáris transzformációnak, ha az $\vec{x}_1, \vec{x}_2 \in V$ és $\lambda \in T$ esetén teljesül az

$$\mathcal{L}(\vec{x}_1 + \vec{x}_2) = \mathcal{L}(\vec{x}_1) + \mathcal{L}(\vec{x}_2), \text{ és az } \mathcal{L}(\lambda \vec{x}_1) = \lambda \mathcal{L}(\vec{x}_1).$$

Adott bázis (lineárisan független vektorrendszer) esetén az n -dimenziós vektortér lineáris transzformációi és az $n \times n$ -es mátrixok közt egy-egy értelmű megfeleltetés létesíthető. Így kétdimenziós esetben a lineáris transzformációkat 2×2 -es mátrixok alkalmazásával lehet megvalósítani.

A $P(x, y)$ és $P'(x', y')$ pontok közti lineáris transzformáció

$$(2-34) \quad [x \quad y] \cdot \underline{\underline{A}} = [x' \quad y'], \quad \underline{\underline{A}} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

alakban írható, ami a mátrixszorzás szabályai szerint ekvivalens az

$$x' = a_{11}x + a_{21}y,$$

$$y' = a_{12}x + a_{22}y$$

összefüggésekkel.

22.1.1. Elemi koordinátatranszformációk

(2-34) alapján röviden összefoglaljuk a lineáris transzformációk speciális típusait:

1. Ha $a_{21} = a_{12} = 0$, akkor léptékváltásról beszélünk, mégpedig

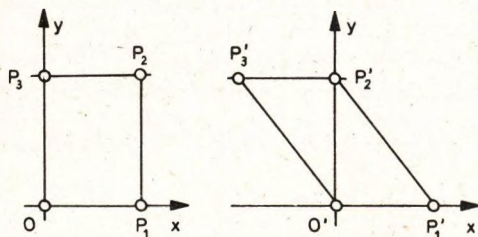
- $|a_{11}| > 1$, ill. $|a_{22}| > 1$ esetén nagyításról,
 - $0 < |a_{11}| < 1$, ill. $0 < |a_{22}| < 1$ esetén kicsinyítésről
 van szó. Ha $a_{11} < 0$, ill. $a_{22} < 0$, akkor a léptékváltás az ordináta-, ill. az abszcisszatengelyre való tükrözéssel is párosul. Ha $|a_{11}| = 1$, ill. $|a_{22}| = 1$, akkor a léptékváltás mérettartó; ha $a_{11} = 0$, ill. $a_{22} = 0$, akkor eltüntető. Az utóbbi eset a projekció, amely a kép elfajulását eredményezi.

A két koordináta transzformációja egymástól független.

2. Ha $a_{11} = a_{22} = 1$, de $a_{12} \neq 0$, ill. $a_{21} \neq 0$, akkor a transzformációt nyírásnak nevezzük. Pl. az

$$\underline{N}_x = \begin{bmatrix} 1 & 0 \\ -0.75 & 1 \end{bmatrix}$$

mátrix a 2-23. ábrán látható transzformációt valósítja meg.



2-23. ábra
X-irányú nyírás

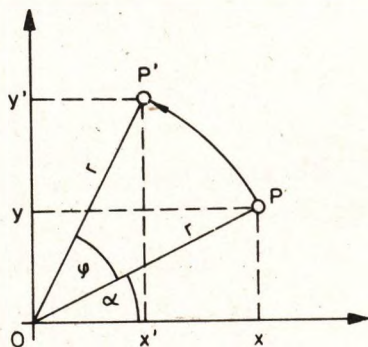
3. Az

$$\underline{A} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}$$

mátrix φ szögű elforgatást eredményez az origó körül. Ez könnyen belátható a 2-24. ábra alapján. Az ábra jelöléseivel:

(2-35/a) $x = r \cdot \cos\alpha$, $y = r \cdot \sin\alpha$;

(2-35/b) $x' = r \cdot \cos(\alpha + \varphi)$,
 $y' = r \cdot \sin(\alpha + \varphi)$;



2-24. ábra
Forgatás az origó körül

$$(2-35/c) \quad r = \sqrt{x^2 + y^2} \quad \left(= \sqrt{x'^2 + y'^2} \right).$$

Fejtsük ki 2-35/b-t és helyettesítsük be 2-35/a-t, ekkor az

$$x' = x \cdot \cos \varphi - y \cdot \sin \varphi, \quad y' = x \cdot \sin \varphi + y \cdot \cos \varphi,$$

összefüggéshez jutunk, amiből az \underline{A} mátrix elemei kiolvashatók.

Megjegyezzük még a következőket:

- Az általános 2×2 -es mátrixszal leírt síkbeli lineáris transzformáció egyedül az origót hagyja változatlanul;

- a síkbeli eltolás 2×2 -es mátrix segítségével nem valósítható meg;

- tetszőleges T területű síkbeli alakzat transzformációja után az új területet a mátrix determinánsa (ld. 1-2) ismeretében a

$$T' = T|\underline{A}| = T \cdot (a_{11}a_{22} - a_{12}a_{21})$$

kifejezésből kapjuk meg.

- A lineáris transzformációk egymásutánját az egyes transzformációk mátrixainak szorzataként fejezhetjük ki. Mivel a mátrixszorzás nem kommutatív művelet, a tényezők sorrendje lényeges. Ennek szemléltetésére lássuk a következő példát:

90° -os forgatás, majd az y -tengelyre való tükrözés eredménye:

$$[x \ y] \cdot \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = [-y \ x] \cdot \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = [y \ x].$$

(tükrözés a 45° -os egyenesre). Az y -tengelyre való tükrözés, majd 90° -os forgatás eredménye:

$$[x \ y] \cdot \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = [-x \ y] \cdot \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} = [-y \ -x]$$

láthatóan más.

22.1.2. Homogén koordináták

Ha $h \neq 0$, akkor az (x, y, h) számhármast a $P\left(\frac{x}{h}, \frac{y}{h}\right)$ síkbeli pont homogén koordinátáinak nevezzük, és ezt a továbbiakban az

$$(2-36) \quad (x^*, y^*) \sim (x, y, h)$$

szimbólummal jelöljük, ahol az (x^*, y^*) az euklidészi sík tetszőleges pontja. Beláthatjuk, hogy ugyanannak a síkbeli pontnak végtelen sokféle homogén-koordinátás megfelelője van, (2-36) alapján ugyanis

$$(x^*, y^*) \sim (\lambda x, \lambda y, \lambda h) \quad (\lambda \in \mathbb{R})$$

is áll, hiszen

$$(\lambda x, \lambda y, \lambda h) \sim \left(\frac{\lambda x}{\lambda h}, \frac{\lambda y}{\lambda h} \right) = \left(\frac{x}{h}, \frac{y}{h} \right) = (x^*, y^*).$$

Az (x^*, y^*) síkbeli normalizált homogén koordinátáinak nevezzük az (x, y, h) számhármast, ha a homogenitási reláció mellett a $h=1$ feltétel is teljesül. A (2-36) szerint minden síkbeli pontnak létezik normalizált, homogén koordinátás megfelelője:

$$(x^*, y^*) \sim (x^*, y^*, 1).$$

Megjegyezzük, hogy az origótól a síkon az (x^*, y^*) irány mindkét értelmében a végtelen távoli pont homogén koordinátái $(x^*, y^*, 0)$.

A homogén koordináták transzformációi is leírhatók mátrixokkal. Mielőtt áttekintenénk ennek előnyeit, lássuk be, hogy az összes lehetőség továbbra is rendelkezésre áll, ami a 2×2 -es mátrixműveletekkel adódott. Ugyanis:

$$(x^*, y^*) \stackrel{\underline{A}}{\sim} (x^*, y^*, 1) \cdot \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

hiszen a "homogén 1" a 3×3 -as mátrixművelet után is megmarad.

22.1.3. Eltolás és perspektív transzformáció

A sík pontjainak megfeleltetett homogén koordináták 3×3 -as transzformációs mátrixa általános esetben négy fő részre osztható:

$$\left[\begin{array}{cc|c} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \end{array} \right] \left[\begin{array}{c|c} \underline{A} & \begin{matrix} a_{13} \\ a_{23} \end{matrix} \\ \hline a_{31} & a_{32} & a_{33} \end{array} \right]$$

Mint azt korábban láthattuk, az a_{11} , a_{12} , a_{21} , a_{22} elemek segítségével léptékváltást, nyírást, forgatást és tükrözést végezhetünk.

Vizsgáljuk meg a többi rész szerepét külön-külön:

1. Az a_{31} , a_{32} elemek segítségével eltolás valósítható meg, ugyanis:

$$(x^*, y^*, 1) \cdot \left[\begin{array}{cc|c} \underline{\underline{E}}_2 & & 0 \\ & & 0 \\ \hline & & 1 \\ a_{31} & a_{32} & \end{array} \right] = (x^* + a_{31}, y^* + a_{32}, 1)$$

($\underline{\underline{E}}_2$ a másodrendű egységmátrixot jelöli.)

2. Ha $a_{33} \neq 1$, akkor

$$(x^*, y^*, 1) \cdot \left[\begin{array}{cc|c} \underline{\underline{E}}_2 & & 0 \\ & & 0 \\ \hline & & a_{33} \\ 0 & 0 & \end{array} \right] = (x^*, y^*, a_{33}).$$

A homogén koordináták definíciójából következik, hogy ez a választás mindkét koordináta irányban $(1/a_{33})$ arányú egyidejű léptékváltást ("zoom") eredményez.

3. Ha $a_{13} \neq 0$, és $a_{23} \neq 0$, akkor a transzformáció

$$(x^*, y^*, 1) \cdot \left[\begin{array}{cc|c} \underline{\underline{E}}_2 & & a_{13} \\ & & a_{23} \\ \hline & & 1 \\ 0 & 0 & \end{array} \right] = (x^*, y^*, (a_{13}x^* + a_{23}y^*) + 1),$$

alakú, és a transzformált homogén koordináták az

$$\left(\frac{x^*}{a_{13}x^* + a_{23}y^* + 1}, \frac{y^*}{a_{13}x^* + a_{23}y^* + 1} \right)$$

síkbeli pontnak felelnek meg. Belátható, hogy a_{13} , ill. a_{23} növelésével a transzformált pont egyre közelebb kerül az origóhoz, azaz a transzformáció centrális projekciót eredményez; vagyis az a_{13} és a_{23} elem változtatásával síkbeli perspektív transzformáció valósítható meg.

Láthatjuk, hogy a homogén koordináták segítségével az eltolás és a perspektív transzformáció is megfogalmazható mátrixszorzással.

Összetettebb transzformációra példaként tekintsük a sík tetszőleges pontja körüli forgatást. Legyen a forgatás centruma (m, n) , a forgatás szöge pedig φ . A kívánt transzformációt három lépésben valósítjuk meg:

- a forgatás centrumát eltoljuk az origóba;
- forgatást végzünk az origó körül a megadott szöggel;
- az origót eltoljuk a forgatás centrumába.

Ez homogén koordináták felhasználásával így írható fel:

$$(x, y, 1) \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -m & -n & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos \varphi & \sin \varphi & 0 \\ -\sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m & n & 1 \end{bmatrix}$$

Ha elvégezzük a mátrixszorzásokat, a transzformáció mátrixát a

$$\begin{bmatrix} \cos \varphi & \sin \varphi & 0 \\ -\sin \varphi & \cos \varphi & 0 \\ -m(\cos \varphi - 1) + n \cdot \sin \varphi & -m \cdot \sin \varphi - n(\cos \varphi - 1) & 1 \end{bmatrix}$$

alakba írhatjuk.

22.1.4. Gyakorlati megvalósítás

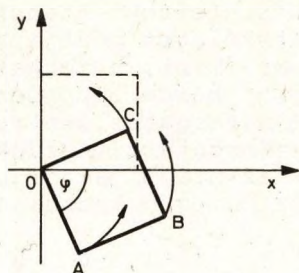
A korrekció gyakorlati megvalósításánál nagy jelentősége van a hatékonyságnak. Ezért pl. nagy adattömeg feldolgozására nem célszerű az elegáns homogén koordinátákat használni, alkalmazási területük inkább a számítógépes grafika.

A gyakorlati megoldásokban a rekonstrukció - újramintavételezés - kvantálás folyamatot a lehető legkisebb műveletigénnyel igyekszünk megvalósítani, ezért ezek a lépések nem is különülnek el élesen egymástól.

Az alkalmazott algoritmusok többsége a diszkrét egyenes létrehozásának törvényszerűségein alapul. A korrekciós összefüggés inverz értelmű, azaz a kívánt pozíciókhoz határozza meg a korrigálatlan képi rendszerbeli pozíciókat. A korrekt újramintavételezési eljárások műveletigénye elviselhetetlenül nagy, ezért a gyakorlatban kompromisszumra van szükség a hatékonyság és a pontosság között. Mivel a képi adatok többnyire viszonylag lassú hozzáférésű háttértárak használatát teszik szükségessé, a korrekciókat úgy kell szervezni, hogy minél kevesebbszer kelljen ezekhez fordulni.

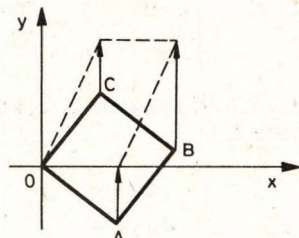
Példaként tekintsünk egy olyan gyakorlati megoldást, amely a mátrixműveletek dekompozícióján alapul. A feladat a kép φ szögű elforgatása az origó körül. A kívánt transzformációt a 2-25/a. és b. ábrán követhetjük végig. Mivel a leghatékonyabban a koordinátatengelyekkel párhuzamos nyírás és a léptékváltás valósítható meg, az eljárás a forgatást 2 nyí-

$$\begin{bmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{bmatrix}$$



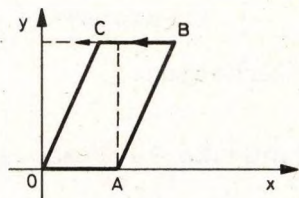
$$\begin{bmatrix} 1 & \operatorname{tg} \varphi \\ 0 & 1 \end{bmatrix}$$

I. : nyírás



$$\begin{bmatrix} 1 & 0 \\ -\sin \varphi \cos \varphi & 1 \end{bmatrix}$$

II. : nyírás

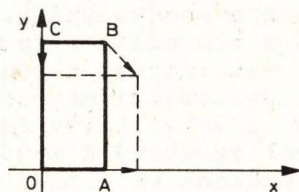


$$\begin{bmatrix} \frac{1}{\cos \varphi} & 0 \\ 0 & \cos \varphi \end{bmatrix}$$

III. : léptékváltás

a)

b)



2-25. ábra
Forgatás visszavezetése nyírásokra

rásra és 1 léptékváltásra bontja. A leírt lépéseket fixpontos műveletekkel meg lehet valósítani, emellett nagymértékben párhuzamosíthatók.

22.2. NEMLINEÁRIS TRANSZFORMÁCIÓK

A geometriai transzformációk általános esetben globálisan (a teljes képre vonatkoztatva) nem lineárisak, és emiatt, valamint a nagy adatmennyiség miatt is jelentős a számításigényük. Ennek ellenére a digitális képfeldolgozás terjedésével jelentőségük egyre nő.

A geometriai korrekciók végrehajtásának hatékonysága (elsősorban gyorsasága) alapvető fontosságú kérdés. Szerencsére bizonyos belső összefüggések (pl. invertálható korrekciós összefüggés esetén a szomszédos pontok szomszédosak maradnak; lineáris korrekciók esetén egyenesek képe is egyenes stb.) felhasználásával hatékonyabb algoritmusokat lehet kidolgozni.

22.2.1. Felületelemes közelítés

Az általánosság megsértése nélkül fel lehet tételezni, hogy kis tartományban minden szóba jöhető geometriai korrekció lineárisan viselkedik. Ezért a globálisan nemlineáris korrekciók megvalósításánál is indokolt a lineáris korrekciós algoritmusok használata, megfelelő módosításokkal. Ez a módosítás történhetne pontról-pontra is, de "sima" korrekciós összefüggés esetén ennél lényegesen ritkábban is. Elvileg jó megoldás lenne, ha a lassan változó helyeken ritkábban, a gyorsan változó helyeken sűrűbben kerülne sor módosításra.

A szükséges módosítások szempontjából nem optimális, de számítástechnikai megfontolásból célszerűbb a következő megoldás alkalmazása:

- Fedjük le az egyik (A) rendszerbeli képet szabályos rácsozatban elhelyezkedő négyzetekkel (ezeket a felületelemeket a továbbiakban rácselemeknek nevezzük), amelyekben belül a korrekciót elegendő pontosan megvalósíthatjuk lineáris transzformációkkal. (Minél "simább" a korrekciós összefüggés, annál több képpont tartozhat egy rácselemhez.)

- Számítsuk ki a korrekciós összefüggés segítségével a rácselemek sarokpontjainak a másik (B) rendszerbeli pozícióját. Ezeket a továbbiakban vezérlőadatoknak nevezzük.

- A tényleges korrekciót hajtsuk végre külön-külön minden egyes rácselemre a sarokpontok (B) rendszerbeli pozíciójának ismeretében, mint általános négypontos transzformációt (ld. a következő alpontot). Invertálható korrekciós összefüggés esetén ezek a (B) rendszerben is diszjunktak és teljesen lefedik a síkot.

A megoldás mellett a következők szólnak:

- a változó sűrűségű módosítás vezérléséhez külön számításokra lenne szükség; ez szabályos esetben elmarad;

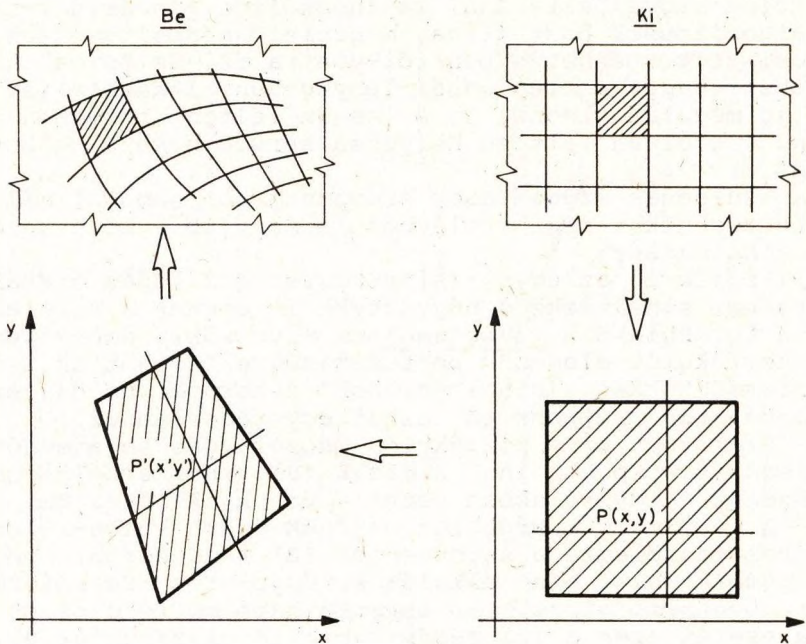
- a vezérlőadatok létrehozására ugyan a szabályos esetben többször van szükség, mint amennyiszer valóban indokolt lenne (hiszen a leggyorsabban változó részre kell felkészülni), de a korrekció tényleges számításigényét végső soron a képpontokkal kapcsolatos, sokkal nagyobb mennyiségű művelet szabja meg.

Számítástechnikai megfontolásból legyen az előbb leírtakban szereplő (A) rendszer a korrigált kép (= kimenő rendszer), a (B) rendszer pedig a korrigálandó kép rendszere (= bemenő rendszer). Ekkor a korrekciós összefüggést az in-

verz korrekciónak megfelelően kell megállapítani. A sorfolytonos adatfeldolgozás érdekében a vezérlőadatok segítségével kijelölt általános négyponos korrekciók közül azokat, amelyek egyazon sort érintenek, egyidejűleg kell végrehajtani. Így a tényleges korrekció a vezérlőadatokból alkotott vezérlősáv segítségével történik.

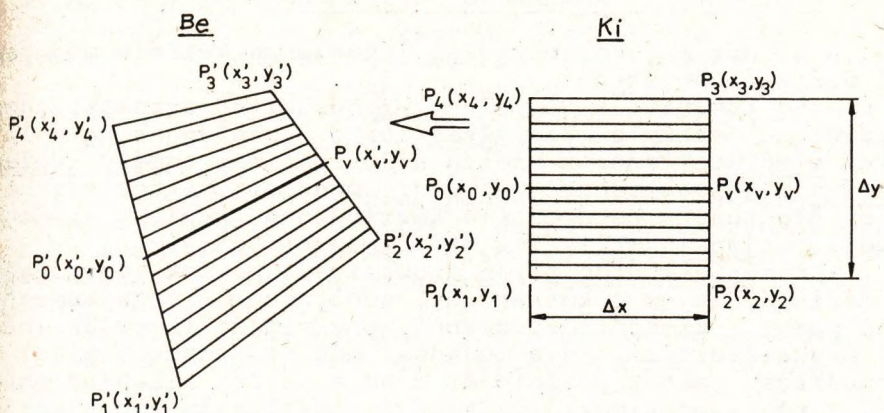
22.2.2. Felületelem transzformációja

Az előzőekben leírtak szerint a globálisan nemlineáris geometriai korrekció felületelemes approximációját a 2-26. ábrán szemléltetjük.



2-26. ábra
Nemlineáris transzformáció közelítése

Projektív geometriai megközelítésben az általános négyponos korrekció kollineáció, és a projektív leképezések alaptörvénye értelmében kettősviszonytartó. Ennek többek között az is a következménye, hogy a kimenő rendszerben a koordinátatengelyekkel párhuzamos egyenesek képe a bemenő rendszerben is egyenes lesz. Ha tehát a kimenő rendszerben sorfolytonosan kezeljük a képpontokat, akkor ezekhez a bemenő rendszerben is egyenes szakaszok mentén elhelyezkedő pontok tartoznak. Ettől a megfigyeléstől azt várjuk, hogy az általános négyponos korrekciót vissza tudjuk vezetni diszkrét egyenest létrehozó algoritmusokra. Tekintsük a 2-27. ábrát.



2-27. ábra
Felületelem transzformációja

A négyszögeken belüli szakaszok a kimenő rendszerben a képsorokat, a bemenő rendszerben pedig ezek ősképet jelölik. A $\overrightarrow{P'_O P'_V}$ kimenő rendszerbeli vektornak a bemenő rendszerben a $\overrightarrow{P'_O P'_V}$ vektort kell megfeleltetni. Mivel $y_O = y_V$ és $x_V - x_O = \Delta x$ (= a rácselem x-irányú mérete), ez a megfeleltetés azt jelenti, hogy a transzformáció elvégzéséhez a $\overrightarrow{P'_O P'_V}$ vektort kell egyenletesen Δx részre felosztani.

A teljes rácselem transzformálásához a P_O pontot a $\overrightarrow{P_4 P_1}$ vektoron, a P_V pontot a $\overrightarrow{P_3 P_2}$ vektoron kell végigléptetni, Δy lépésben; ennek megfelelően a P'_O , ill. a P'_V pont a bemenő rendszerben a $\overrightarrow{P_4 P_1}$, ill. a $\overrightarrow{P_3 P_2}$ vektoron halad végig.

Belátható, hogy a $(\Delta x) \cdot (\Delta y)$ méretű, kimenő rendszerbeli rácselem összes pontjának pozícióját a fenti módszerrel

$$\ell_{\Delta x \Delta y} = 4 (\Delta y) + 2 \Delta x (\Delta y + 1)$$

lépésben kapjuk meg, ahol az első tag a sorok kezdő- és végpontjainak, a második a szakaszpontoknak a számát jelenti. A teljes műveletszám egyenesen arányos az egyes interpolációs lépések átlagos műveletszámával. Könnyen belátható, hogy a diszkrét lineáris interpoláció és a diszkrét egyenes létrehozása ekvivalens feladatok, így a nemlineáris geometriai korrekciót ennek alkalmazására vezettük vissza.

22.3. A GEOMETRIAI KORREKCIÓ DEFINIÁLÁSA

A geometriai korrekciókat lényegében kétféle módszerrel definiálhatjuk:

- Ha ismerjük a képfelvevő rendszer torzításait, az ábrázolt objektum elmozdulását leíró összefüggéseket, ill. egyéb elméleti ismeret esetén a korrekciós szabály képlet-szerűen megfogalmazható (elméleti módszer). Ekkor egzakt módon meg tudjuk határozni a korrigálandó kép(ek) minden egyes pontjának korrigálás utáni pozícióját.

- Kísérleti úton olyan azonosítási pontok segítségével is definiálhatjuk a korrekciót, amelyeknek mind a bemenő, mind pedig a kimenő rendszerbeli pozícióját ismerjük (mérés-sel meghatároztuk). Ez a megadási mód nem egzakt, mivel az azonosítási pontok pozícióját csak a mérési hibahatáron belül, a többi pont pozícióját pedig egyáltalán nem ismerjük.

Az elméleti és a kísérleti megadási mód keveredhet egymással. Pl.: az elméleti úton meghatározott típusú korrekciós összefüggést azonosítási pontok felhasználásával is paraméterezhetjük.

22.3.1. Elméleti módszerek

A geometriai korrekciót akkor tudjuk kizárólagosan elméleti úton definiálni, ha egyrészt ismerjük a függvénykapcsolatot a bemenő és a kimenő koordinátarendszer között, másrészt ismerjük e függvénykapcsolat összes paraméterét. Ellenkező esetben a konkrét korrekciós összefüggést kísérleti úton kell megállapítani (ld. következő alpontot).

A különböző leképezési hibákkal, ill. torzítást okozó jelenségekkel itt nem kívánunk foglalkozni. Megjegyezzük, hogy általában legfeljebb harmadfokú korrekciós összefüggéseket szokás használni.

Gyakorlati követelmény, hogy a lineáris korrekciót implicit formában (léptékváltási arány, forgatási szög, eltolás) is meg lehessen adni. Bonyolultabb esetekben az inverz korrekciós összefüggés explicit megadása célszerűbb.

22.3.2. Kísérleti módszerek

A geometriai korrekciót olyan azonosítási pontok pozíciójának megadásával is definiálhatjuk, amelyeknek mind a bemenő-, mind a kimenőképen ugyanaz a jelentésük. Mivel ezeket a pozíciókat legalább az egyik kép koordinátarendszerében mérésekből kapjuk, korlátozott pontosságúak. (Még a feltűnő képi alakzatok kiemelt pontjai sem azonosíthatók eléggé egyértelműen, másrészt a pontosságnak elvi korlátot szab a pozíció kvantálása.) Ezek figyelembevételével logikus, hogy a korrekciós összefüggést a következő feltételek alapján keressük:

- az azonosítási pontok összerendelésénél elkövetett hiba legyen minimális;
- a meghatározott korrekciós összefüggés a teljes képtartományban a lehető "legsímábban" viselkedjék;
- az eleve szükségesnél nagyobb számú azonosítási pont felhasználásával a korrekció legyen pontosítható.

A korrekciós összefüggést előállító algoritmussal kapcsolatos igények:

- az azonosítási pontok pozícióin kívül ne legyen szükség más összefüggés ismeretére, de ha ilyenek rendelkezésre állnak, azokat fel lehessen használni (pl. elméleti úton meghatározott korrekció paraméterezése azonosítási pontok segítségével);

- az algoritmus egyaránt tartalmazzon egyszerűbb és bonyolultabb közelítéseket, de a szükséges bonyolultságot elő lehessen írni;

- az azonosítási pontok összerendelésénél elkövetett hiba nagyszámú azonosítási pont esetén is elég gyorsan konvergáljon zérushoz;

- legyen lehetőség az elért pontosság, ill. az elkövetett hiba kvantitatív ellenőrzésére.

Legyen N pontból álló azonosítási ponthalmazunk, amely az (A) és (B) koordinátarendszer közt kölcsönösen egyértelmű leképezést definiál:

$$P_s : (x_{sA}, y_{sA}) \rightarrow (x_{sB}, y_{sB}) \quad (s = 1, 2, \dots, N),$$

ahol P_s az s -edik azonosítási pont, $(x_{sA}, y_{sA}, x_{sB}, y_{sB})$ pedig az (A), ill. a (B) rendszerbeli pozíciók koordinátái. Mivel ezeket a mérésekből ismerjük, tegyük fel, hogy az elkövetett hibák normális eloszlásúak, zérus várható értékkel.

Keressük külön az

$$\{ f_x : (x_{sA}, y_{sA}) \mid s = 1, 2, \dots, N \} \rightarrow x_{sB}$$

és az

$$\{ f_y : (x_{sA}, y_{sA}) \mid s = 1, 2, \dots, N \} \rightarrow y_{sB}$$

folytonos függvényeket, amelyek a megfeleltetést a lehető legkisebb hibával valósítják meg a $P_s (s = 1, 2, \dots, N)$ pontokban, és elegendően "sima" függvényként viselkednek. (Így az elkövetett hiba nem izotróp, viszont a szeparálás jelentősen csökkenti a műveletszámot.)

Keressük a korrekciós összefüggést szabadon megválasztható függvények lineáris kombinációjaként. (A képfeldolgozás területén fellépő geometriai transzformációkat, illesztési, ill. torzítási jelenségeket többnyire le lehet írni alacsony fokszámú polinomokkal.) Tételizzük fel, hogy a keresett f_x

és f_y közelítő függvény előállítható lineárisan független folytonos függvények összegeként:

$$(2-37) \quad f(x_A, y_A) = \sum_{i=1}^m a_i g_i(x_A, y_A),$$

ahol f vagy az f_x vagy az f_y függvényt közelíti. Ekkor az azonosítási pontokra felírhatjuk, hogy

$$\vec{f} = \sum_{i=1}^m a_i \vec{g}_i,$$

ahol az \vec{f} , ill. a \vec{g}_i (N elemű) oszlopvektorok elemei az f , ill. a g_i függvény helyettesítési értékei az (x_{SA}, y_{SA}) pontban (az előbbieket nyilván az x_{SB} vagy az y_{SB} függő változók közelítő értékei); a_i pedig az alkalmasan megválasztott \vec{a} együtthatóvektornak a g_i függvényhez tartozó eleme.

A korrekciós összefüggés keresésének első (pontossági) feltétele kielégíthető, ha az eltéréseket alkalmas módon leíró

$$(2-38) \quad \Phi(\vec{a}) = \sum_{s=1}^N \varphi_s D(\vec{a}), \text{ ill.}$$

$$(2-39) \quad D(\vec{a}) = \left(f - \sum_{i=1}^m a_i g_i \right)^2,$$

függvényt minimalizáljuk a_1, a_2, \dots, a_m szerint.

(A képletben φ_s súlytényezőket; f tényleges függvényértéket jelöl a fentiek szerint.)

A φ_s súlyokat célszerű a P_s pontok eloszlási sűrűségétől függően választani; ahol P_s pontok sűrűbben helyezkednek el, ott φ_s legyen kisebb.

A korrekciós összefüggés keresésének második (simasági) feltétele értelmében célszerű az $m \ll N$ választás. Ekkor ui. az f függvénynek viszonylag kevés paramétere van, és megfelelő g_i függvények esetén várható, hogy a közelítés a mérési pontok között is javul.

A simaságot alkalmasan választott regularizációval is elő lehet segíteni. Ekkor az a_i együtthatókat $\Phi(\vec{a})$ helyett $\Phi'(\lambda, \vec{a})$ minimumából lehet meghatározni a

$$(2-40) \quad \min_{\vec{a}} \phi'(\lambda, \vec{a}) = \min_{\vec{a}} (\phi(\vec{a}) + \lambda \psi(\vec{a}))$$

összefüggés szerint. Itt a $\psi(\vec{a})$ funkcionál a függvény "simaságát" fejezi ki. A λ paraméter megválasztásától függ, hogy a közelítés pontosabb lesz a megadott pontokban, vagy "simább".

Keressük a megoldást $\varphi_s = 1$ és $\lambda = 0$ esetre. Ekkor (2-39) a következőképpen alakítható át:

$$(2-41/a) \quad D(\vec{a}) = \vec{f}^T \vec{f} - 2 \sum_{i=1}^m a_i (\vec{f}^T \vec{g}_i) + \sum_{i=1}^m \sum_{k=1}^m a_i a_k (\vec{g}_i^T \vec{g}_k),$$

ahol a T felsőindex a transzponált vektort jelöli. Vezessük be a

$$\vec{z}_{m \times 1} = \left[\vec{f}^T \vec{g}_i \right], \text{ ill. a } \underline{G}_{m \times m} = \left[\vec{g}_i^T \vec{g}_k \right] \\ (i, k = 1, 2, \dots, m)$$

oszlopvektort, ill. mátrixot; ezekkel (2-41/a) a

$$(2-41/b) \quad D(\vec{a}) = \vec{f}^T \vec{f} - 2 \vec{a}^T \vec{z} + \vec{a}^T \underline{G} \vec{a}$$

alakba írható. A minimalizáláshoz fejtsük sorba $D(\vec{a})$ -t tetszőleges (m -dimenziós térbeli) \vec{a}^* pont körül:

$$D(\vec{a}) = D(\vec{a}^*) + \left[\frac{\partial D(\vec{a})}{\partial \vec{a}} \right]_{\vec{a}=\vec{a}^*} (\vec{a} - \vec{a}^*) + \\ + \frac{1}{2} (\vec{a} - \vec{a}^*)^T \left\{ \frac{\partial}{\partial \vec{a}} \left[\frac{\partial D(\vec{a})}{\partial \vec{a}} \right] \right\}_{\vec{a}=\vec{a}^*} (\vec{a} - \vec{a}^*) + \dots$$

ahol a parciális deriváltak (2-41/b) alapján:

$$\left[\frac{\partial D(\vec{a})}{\partial \vec{a}} \right] = -2 \vec{z}^T + 2 \cdot \vec{a}^T \underline{G}; \\ \left\{ \frac{\partial}{\partial \vec{a}} \left[\frac{\partial D(\vec{a})}{\partial \vec{a}} \right] \right\} = 2 \underline{G};$$

és - mivel \underline{G} független (\vec{a}) -tól - az összes magasabb rendű parciális derivált zérus. Így (2-41/b)-t végül is a

$$D(\vec{a}) = D(\vec{a}^*) + (-2 \vec{z}^T + 2 \vec{a}^T \underline{G}) (\vec{a} - \vec{a}^*) + \\ + \frac{1}{2} \left[(\vec{a} - \vec{a}^*)^T \cdot 2 \underline{G} \cdot (\vec{a} - \vec{a}^*) \right]$$

alakra hoztuk. A $\phi(\vec{a})$ függvényre előírt szélsőérték teljesülésének feltétele (2-38) értelmében az, hogy

$$\left[\frac{\partial D(\vec{a})}{\partial \vec{a}} \right] = -2 \vec{z}^T + 2 \vec{a}^T \underline{G} = 0$$

legyen, amiből az

$$(2-42) \quad \vec{a}^T \underline{G} = \vec{z}^T \implies \sum_{k=1}^m a_k \cdot \vec{g}_k^T \cdot \vec{g}_k = \vec{f}^T \cdot \vec{g}_i \quad (i=1, 2, \dots, m)$$

végeredmény adódik. A közelítés hibája is könnyen kifejezhető:

$$(2-43) \quad D(\vec{a}) = \vec{f}^T \cdot \vec{f} - \sum_{i=1}^m \vec{f}^T \cdot \vec{g}_i$$

A leírt megoldás műveletigénye erősen függ m -től, mivel (2-42) szerint m db approximációs függvény alkalmazása esetén $m \times m$ -es mátrixot kell invertálnunk. Így nagy m esetén a g_i bázisfüggvények cserélgetése, ill. a közelítésben részt vevő függvények számának változtatása nehézkes, lassú.

Részletesebb ismertetés nélkül megjegyezzük, hogy rekurzív algoritmus alkalmazásával a probléma áthidalható. Ekkor az adott m -tagú közelítésben részt vevő függvényeket egy $M \gg m$ függvényt tartalmazó bázisfüggvény-könyvtárból választja ki a program az adott választáshoz tartozó $D(\vec{a})$ hiba minimális értékének megfelelően. Így egyrészt nem kell előre meghatározni a közelítő függvényvektor tagszámát, másrészt minden lépésnél a legjobb eredményt adó bázisfüggvények kerülnek felhasználásra. Az algoritmus alkalmazásakor a $k+1$ tagú közelítés számításainak nagy része támaszkodik a k tagú közelítés számításainak részeredményeire, vagyis az eljárás folytatásakor a műveletigény nagyságrendje nem nő. Az egyes közelítések közül a legmegfelelőbbet a pontossági és a simasági tulajdonságok együttes figyelembevételével lehet kiválasztani.

23. Szegmentálás

A szegmentálás során először osztályozzuk a képpontokat valamilyen sajátságvektor szerint, majd megkeressük a kapott osztályozásra nézve összefüggő képponthalmazokat, azaz tartományokat.

A továbbiakban az értékes tartományokat nevezzük objektumoknak. Ezekről feltesszük, hogy megfelelnek a leképezett

valóságos tárgyakat határoló felületeknek, illetve metszész-vonalaiknak; bár a sajátságvektorok kialakításánál erre nem voltunk tekintettel. A problémát a harmadik dimenzió, a térbeliségre vonatkozó információ hiánya jelenti; emiatt az eredmények néha erősen torzulhatnak. (Pl. a felhőzet és egy víztükör képe egybeolvad; az árnyékok miatt megváltozik az objektumok területe; stb.) Rekonstruálásra - és ezzel az eredmények javítására - számos módszert dolgoztak ki, amelyekkel helyszűke miatt nem foglalkozhatunk.

A szegmentálás sikere döntően az előre kialakított definíciókon múlik. A következőket kell meghatározni:

- a sajátságvektorokat; dimenziószámuk 1-4; legtöbbször a világosság- vagy színekódokat, illetve az ezekből leszármaztatott sajátságokat (hisztogram, gradiens, súlyozott átlagok stb.) elemzik;
- a távolságfüggvényt; ezzel mérjük a sajátságvektorok távolságát - az s-távolságot - a sajátságtérben (amelynek dimenzióit a választott sajátságok képezik);
- a küszöbököt, esetleg valamilyen összetettebb döntésfüggvényeket; s-távolságuk ismeretében ezek alapján sorolunk két képpontot azonos vagy különböző osztályba;
- az osztályok jellemzőit: sajátságaikat, számukat, azonosítójukat, esetleg sajátszínüket stb.; egyes esetekben a pontos osztálydefiníciókat tanítók (mintaelemek) megadása helyettesíti.

A szegmentálás tulajdonképpen úgy történik, hogy meghatározzuk minden képpont sajátságvektorát, majd mindegyiket abba az osztályba soroljuk, amelyiktől (illetve amelynek (egyik) tanítójától) elég kicsi s-távolságra van a választott küszöbök (v. döntésfüggvény) szerint. Végül megkeressük az egy osztályba sorolt, összefüggő ponthalmazokat. A harmadik dimenzió már említett hiányán és a definíciók pontatlanságán túl (pl. az osztályok alul- vagy túlhatározottsága, a feladathoz nem adekvát döntésfüggvény stb.) téves besorolást eredményezhetnek a képhibák vagy az ellentmondó követelmények is. Ezért szükség lehet utólagos korrekcióra újabb, addig figyelembe nem vett szempontok (pl. környezet, folytonosság stb.) szerint.

Mivel a szegmentálás minden további feldolgozási lépés alapja, nagyon fontos, hogy minél pontosabban végezzük el. Evégett számos módszert és mindegyiken belül rengeteg eljárást dolgoztak ki, teljes körű áttekintésükre már csak nagy számuk miatt sem kerülhet sor. Itt érezhető talán legfájóbban az elméleti alapok tisztázatlansága: egy konkrét feladat megoldásához rendszerint hiányzik Ariadné fonala. Jobb híján anyagunkat a bemutatott módszerek szerint csoportosítottuk.

23.1. FOLTELEMZÉS

Az egy osztályba tartozó (4)-összefüggő tartományokat (ld. 13.4.2. pont) foltnak nevezzük, ha a sajátságvektorok hasonlósági jellemzőket mérnek. A továbbiakban a foltnak kijelölésére, illetve a megtalált foltnak elemzésére irányuló eljárásokkal foglalkozunk.

23.1.1. Foltnak kijelölése

A szegmentálás "jóságát" az egyes foltnak alkotó képpontok homogenitása jellemzi vagyis az, hogy mennyire "hasonlók". Ideális esetben a belső pontok sajátságvektorának közel egyenlőknek kell lenniük, vagyis nem lehetnek "lyukak" a foltnak. A jó megkülönböztethetőség érdekében viszont fontos, hogy a különböző foltnakhoz tartozó képpontok sajátságvektora minél eltérőbb legyen és rendszerint az is igény, hogy a kontúrvonal ne legyen zezugos.

Mindezeket a követelményeket egyidejűleg szinte lehetetlen kielégíteni. Ha pl. szigorítjuk a homogenitási feltételt (vagyis csökkentjük a foltnakhoz tartozó képpontok sajátságvektorának s-távolságát), rendszerint lyukak keletkeznek és a határvonal töredezett lesz; ha viszont a különböző foltnak s-távolságát növeljük, elveszíthetjük a határvonalak egy részét. Elfogadható eredményeket csak kompromisszumok árán lehet elérni és csak akkor, ha a szegmentálás előtt megfelelő képpontkorrekciót (ld. előző fejezet) végzünk.

23.1.1.1. Globális módszerek

A gyakorlatban figyelembe vehető globális (= az összes képpont egyidejű elemzésével meghatározott) képjellemező a hisztogram (ld. 21.2.1.1. alpont). Ezen esetben tehát a sajátságvektor egydimenziós és értékét az adott világosságkódú képpontok relatív gyakorisága adja. A módszerek közös alapelve, hogy megkeresik a hisztogram csúcspontjait - a módusokat, illetve a közöttük levő völgypontokat, majd ez utóbbiakat küszöbként használva képvágást (ld. 21.2.1.3. alpont) végeznek. Az egy osztályba sorolt pontokat ezután át lehet színezní a módosukhoz tartozó világosságkód által meghatározott osztályszínre, ez azonban a további gépi feldolgozáshoz nem szükséges.

Ha a hisztogram kétmódusú (vagyis két móduspontja között egy völgypont van), a foltnak két szintre vágással alakíthatjuk ki. Ehhez a K vágási küszöböt az

$$\rightarrow (2-44) \quad S(1-S) (\mu_1 - \mu_2)^2 = \text{maximum}$$

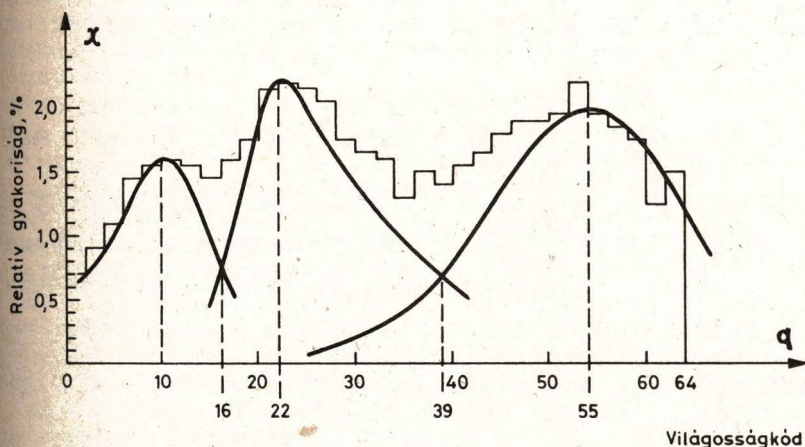
képletből számíthatjuk ki, ahol a szokásos jelölésekkel (q a világosságkód, $J = 2^b - 1$ a maximális értéke, b a kódhossz

(8 bites kódok esetében $J=255$); $\chi(q)$ relatív gyakoriság, vagyis a hisztogram megfelelő ordinátája):

$$(2-44) \quad s = \sum_{q=0}^J \chi(q); \quad \mu_1 = \frac{\sum_{q=0}^{\kappa} q \cdot \chi(q)}{s}; \quad \mu_2 = \frac{\sum_{q=\kappa+1}^J q \cdot \chi(q)}{1-s}.$$

(A maximumot könnyű kiszámítani: $\kappa = 0$ -tól indulva értékét egyesével addig növeljük, amíg a (2-44) kifejezés értéke növekszik.)

Kettőnél több módusú hisztogram esetében a vágási köszöbök a hisztogramot közelítő - megfelelően transzformált - normális sűrűségfüggvények (ld. 13.3.2. pont) metszéspontja határozza meg. (A közelítő sűrűségfüggvények alatti terület összege közel egyenlő a hisztogramgörbe alatti területtel.) A módszert a 2-28. ábrán szemléltetjük; itt a két küszöb és a 3 osztályszín: $\kappa_1=16$, $\kappa_2=39$, illetve $q_1=10$, $q_2=22$, $q_3=55$.



2-28. ábra
Hárommódusú hisztogram

Ha a hisztogramgörbe "lapos", a völgypontok helyzete bizonytalan. Ilyenkor jó eredményt érhetünk el tanítók interaktív megadásával. Ezt egyszerűen úgy tesszük, hogy körükerítjük a képen azokat a tartományokat, amelyekről biztosan megítélhető, hogy egy foltot alkotnak és megadunk egy osztályszínt. Ezután a program erre a színre színezi át mindazokat a képpontokat, amelyek világosságkódja előfordul a megadott mintaterületben. (Egy osztályhoz természetesen több mintaterületet is ki lehet jelölni. Ha egy világosságkód több, különböző osztályt reprezentáló mintaterületben is előfordul, a program ahhoz az osztályhoz rendeli hozzá, amelyikhez tartozó mintában legnagyobb a relatív gyakorisága.)

23.1.1.2. Lokális módszerek

A globális eljárások közös gyengéje, hogy nem tudják figyelembe venni a geometriai követelményeket; ezért a vágás után a foltok gyakran "kilyukadnak", a határvonalak "zajosak" vagy kiszélesednek.

A lokálitás viszonylagos fogalom. Általában lokálisnak nevezzük azokat a módszereket, amelyekben az aktuális képpontnak 3×3 vagy 5×5 , de legfeljebb 15×15 képpont méretű környezetét elemezzük. Ennél nagyobb környezetek vizsgálata esetén regionális módszerekről beszélünk. (Egy régió akár 128×128 képpontból is állhat.) A figyelembe vett környezet technikai okokból legtöbbször négyzetalakú és páratlan oldalhosszú, de egyes speciális eljárásokban ettől eltérnek.

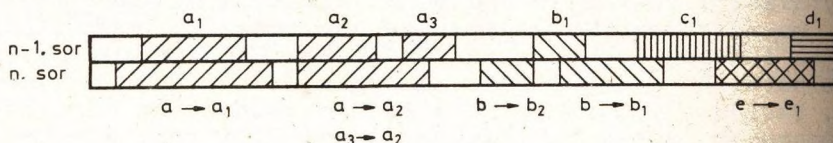
A legkedveltebb lokális foltkereső módszer az összefűzés. Legegyszerűbb esetben egy tetszőlegesen választott képpontból kiindulva addig fűzzük hozzá a (4)-szomszédait (majd ezek (4)-szomszédait stb.), amíg világosságkódjuk különbsége "elég kicsi" (kisebb egy előre adott küszöbnél). Sajnos ez az eljárás igen érzékeny a képhibákra; ezért inkább többdimenziós sajátsgvektorral szokás dolgozni. Háromdimenziós esetben pl. a távolságfüggvény

$$d = s_1 |q_i - q_j| + s_2 |\bar{q}_j - q_i| + s_3 \bar{q}_i - q_j|$$

alakú lehet; ahol q_i , ill. q_j a vizsgált két képpont világosságkódja, \bar{q}_i , ill. \bar{q}_j pedig a figyelembe vett környezetük átlaga; $s_1 + s_2 + s_3 = 1$, alkalmasan választott súlytényezők.

Jó futási időket lehet elérni, ha a soronkénti növesztési (growing) technikát alkalmazzuk. (Feltéve, hogy - mint az a TV-rendszerű megjelenítés miatt szokásos - a digitális kép a főtárban sorfolytonosan helyezkedik el.)

Az eljárást a 2-29. ábrán illusztráljuk. Minden sor feldolgozása úgy kezdődik, hogy meghatározzuk, melyik képpont melyik osztályba tartozik, ill. melyek a háttérpontok. Az egy osztályba tartozó, összefüggő képpontok metszeteket alkotnak. (Az ábrán ezeket azonos vonalkázással jelöltük.)



2-29. ábra
Soronkénti növesztés

Tegyük fel, hogy az n -edik képsort dolgozzuk fel, és az $(n-1)$ -edik 6, az n -edik 5 foltot metszett át, amelyek rendre

az a-, a-, a-, b-, c-, d-, ill. az a-, a-, b-, b-, e-osztályba tartoznak az ábra szerint.

A feldolgozás során csak (4)-szomszédságot vizsgálunk. Az aktuális (a feldolgozás alatt álló sorban éppen vizsgált) metszet az előző sorban levő metszethez képest a következő helyzetekben lehet:

- nem szomszédos egyikkel sem, vagy csak más osztály-(ok)ba tartozókkal szomszédos: ekkor új folt kezdődik ($b \rightarrow b_2, e \rightarrow e_1$);

- egyetlen vele azonos osztályúval szomszédos: a folt kiegészül az új metszettel ($a \rightarrow a_1$);

- legalább kettővel szomszédos, és mindegyik vele azonos osztályú: az eddig különböző foltok egy folttá egyesülnek, s ehhez tartozik az új metszet is ($a \rightarrow a_2, a_3 \rightarrow a_2$);

- legalább kettővel szomszédos, és nem mindegyikkel azonos osztályú: a vele azonos osztályú folt kiegészül az új metszettel (ha több ilyen volt, ezek most egyesülnek) ($b \rightarrow b_1$).

A feltérképezett foltokról lista készül, amely az azonosításhoz szükséges paramétereken kívül a területüket is tartalmazza.

Az eljárást úgy is lehet módosítani, hogy csak előre megadott pontokat tartalmazó foltok kerüljenek osztályozásra.

A területnövesztő és az összefűzési módszerek kombinálásával nagy pontosságú és a hibákra kevésbé érzékeny eljárásokat alakíthatunk ki.

Megemlítjük még, hogy igazán jó eredményeket csak többlé-
péses hibrid módszerekkel lehet elérni, amelyekben felváltva alkalmaznak globális és lokális eljárásokat. Erre a leg-
egyszerűbb példa, ha egy globális vágás után lokális többségi szűrést végzünk: minden képpont világosságkódját helyettesítjük egy adott környezetében legtöbbször előfordulóval.

Számos más, bonyolultabb eljárás létezik; ezek számítás-
igénye azonban rendszerint nagy és ezért a gyakorlati alkalmazásuk ritka.

23.1.2. Foltvizsgálat

A továbbiakban a foltokra szegmentált - esetleg át-
színezett - képből indulunk ki. Igen gyakori feladat a megtalált foltok terület szerinti osztályozása, illetve eloszlásuk meghatározása. A folt területe egyenlő a hozzá tartozó képpontok számával; ezt könnyen megállapíthatjuk; pl. az említett foltnövesztési eljárás során "melléktermékként" megkapjuk. Az osztályozás rendszerint azt jelenti, hogy a foltokat zónákba soroljuk, amelyeket előre adott területhatárok jelölnek ki.

A nagyszámú elemzési módszer átfogó ismertetésére sajnos nincs terünk; mindössze két gyakori probléma megoldására mutatunk egy-egy eljárást.

23.1.2.1. Hámozás (erózió)

A képhibák miatt a foltkeresési eljárással nyert foltok - noha kielégítik az osztályozás kritériumait - sokszor nem feleltethetők meg egyik objektumtípusnak sem azok közül, amelyeket a képen keresünk.

Legtöbbször arról van szó, hogy a térben közeli, vagy a felvétel irányából egymást takaró objektumok képe egybeolvad, - esetleg csak a rossz megvilágítás miatt egymásra vetett árnyékok következtében.

Ha konvex objektumokat keresünk, a túl nagy területű konkáv foltokat a hámozási (vagy eróziós) technikával bonthatjuk szét.

Az eljárás neve onnan ered, hogy minden lépésben "le-hámozzuk" (vagy "lemorzsoljuk") a folt legkülső rétegét - a határpontokat - mindaddig, amíg a további zsugorítás már nem lehetséges (ld. 2-30. ábra).

Az eljárás akkor vezet kielégítő eredményre, ha az egyes lépésekben felváltva (4)-, ill. (8)-szomszédságot vesszük figyelembe, mert a zsugorodás így lesz a leegyenletesebb (ld. 2-31. ábra: itt a páratlan számú pontok az elődeik a (4)-, a párosak (8)-szomszédai: ezt kombinált vagy nyolcszögű szomszédságnak is szokták nevezni.)

Ha elértük a folt csomópontjait, az ehhez szükséges lépésszám, valamint a távolságuk ismeretében lehet meghatározni a vágóvonala(ka)t. (A csomópontok minden pontja határpont; ezek a 2-30. ábrán a vastagon bekeretezett pontok. Ezeket kívül bevonalkáztuk azokat a pontokat, amelyek a vágóvonalhoz tartozhatnak.)

23.1.2.2. Mintaillesztés

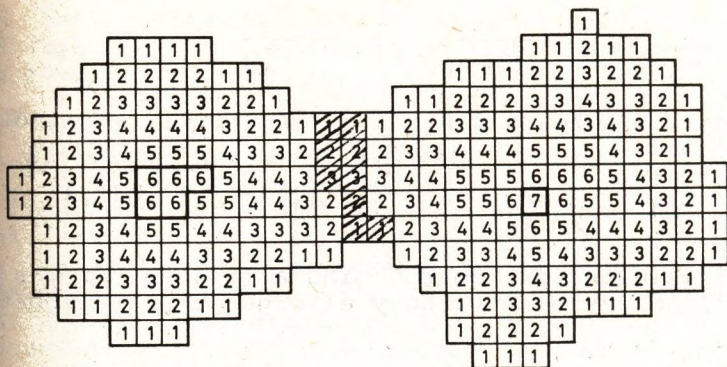
A valóságos objektumokra sok esetben a megfelelő foltok geometriai alakjának vizsgálatával következtethetünk. Gyakori eset, hogy ismerjük a lehetséges objektumok "igazi" alakját, előre megadott minták formájában, amelyek meghatározzák a lehetséges osztályokat. Az egyes foltokat abba az osztályba soroljuk, amelyiknek mintájához a legjobban "hasonlítanak".

A kellő hasonlóságot numerikus vagy grafikus eljárásokkal állapítjuk meg.

Foglaljuk be a mintákat egy-egy $(2m+1) \cdot (2n+1)$ képpont méretű ablakba, ekkor az illeszkedés pontosságát az

$$(2-45) \quad e_i(k,l) = \sum_{r=-m}^m \sum_{s=-n}^m \rho(q(k+r,l+s) - m_i(r,s))$$

menyiség jellemzi egy adott képpontban, ahol ρ tetszőleges metrika, $m_i(s,t)$ pedig az i -edik mintát leíró függvény.



2-30. ábra
Foltok szétválasztása hámozással

			4	4	4	4	4		
		4	4	3	3	3	4	4	
	4	4	3	2	2	2	3	4	4
	4	3	2	2	1	2	2	3	4
	4	3	2	1	●	-1	2	3	4
	4	3	2	2	1	2	2	3	4
	4	4	3	2	2	2	3	4	4
	4	4	3	3	3	4	4		
		4	4	4	4	4			

2-31. ábra
Kombinált szomszédság

A (k, l) pont környezetében akkor "találtuk meg" az i -edik mintát, ha ott az e_i hibafüggvény a legkisebb és kisebb egy adott küszöbnél. (Így működnek pl. a postai irányítószámokat leolvasó automaták. Ezesetben 10 osztály van, amelyeket a megadott minta-számjegyek reprezentálnak.)

Az eljárás nem vezet eredményre, ha a képen előforduló objektumok mérete és állása különböző lehet. Egy egyszerű grafikus eljárásban a problémát úgy oldják meg, hogy kiszámítják az objektumok digitális nyomatékait az

$$N_{mn} = \sum_{k, l \in A} k^m l^n q(k, l)$$

képletből, ahol A a minimális befoglaló négyzetet jelenti és $q(k, l) \in \{0, 1\}$. (Vagyis bináris képpel dolgozunk, amelyben az értékes pontokat 1, a háttérét 0 jelzi. Ha a képen több, különböző osztályhoz tartozó folt van, az eljárást minden osztályra külön kell végrehajtani.) A folt területét az M_{00} nyomaték, súlypontjának koordinátáit pedig a

$$k_0 = \frac{M_{10}}{M_{00}}, \quad l_0 = \frac{M_{01}}{M_{00}}$$

képlet adja. Ezekkel a centrális (súlypontra vonatkoztatott) nyomatékok az

$$(2-46) \quad I_{mn} = \sum_{k, l \in A} (k - k_0)^m (l - l_0)^n q(k, l)$$

képletből számíthatók. Hozzuk mind a foltokat mind a mintákat "standard" helyzetbe úgy, hogy elforgatjuk a súlypontjuk körül

$$\alpha = \arctg \frac{2 I_{11}}{|I_{20} - I_{02}|}$$

szöggel ($-90^\circ \leq \alpha \leq 90^\circ$), majd kicsinyítést vagy nagyítást végzünk úgy, hogy minden ablak azonos méretű legyen. Ezután az illeszkedés jóságát úgy határozzuk meg, hogy minden mintát "ráhelyezünk" minden egyes foltra és megszámláljuk a fedőpontokat.

23.2. ÉLKITŰZÉS

Ebben a pontban azokat a nem kevésbé fontos szegmentálási eljárásokat tekintjük át, amelyek a képjellemzők hirtelen változása (i) alapján osztályozzák a képpontokat. Az ilyen hirtelen (ugrásszerű) változásokat éleknek nevezzük; akkor is, ha nem a világozásgódok (vagy színek) változásairól van szó. Az osztályozáshoz a (8)-szomszédságot vesszük figyelembe.

Élkitűzésen az élekhez tartozó képpontok megkeresését értjük. Ilyenkor általában csak két osztályt különböztetünk meg: az egyiket az élpontok alkotják (most ezek az "értékes" pontok), az összes többi képpont háttérpont. Megjelenítéskor az éleket az élosztály saját színére színezzük át; a háttérpontok vagy változatlanul maradnak ("rárajzolás"), vagy mind háttérszínűek lesznek. A további feldolgozáshoz nem szükséges a képet átszínezni. Ilyenkor inkább egy bittérkép készül, amelyben minden képpontnak egy bit felel meg; az élpontokat az 1-bitiek jelzik.

Az élkitűzésre is vannak globális és lokális eljárások, de a gyakorlatban szinte kizárólag csak az utóbbiakat használják, ezért csak ezeket tekintjük át.

A zajos képek feldolgozásakor a legnagyobb nehézséget az okozza, hogy egyrészt az élpontnak minősített ("élgyanús") képpontok nem alkotnak folytonos vonalat, nem (8)-összefüggők; másrészt olyan pontok is élgyanúnak minősülhetnek, amelyek nem tartoznak egyetlen folt kontúrvonalához sem. Emiatt az élkitűző eljárások rendszerint kétmeneteselek:

az első menetben történik meg az élgyanús pontok kijelölése; a másodikban ezek szelektálásával, ill. kiegészítésével kapjuk meg a valódi élpontokat.

23.2.1. Él(gyanús) pontok kijelölése

A leggyakoribb feladat az élek meghatározása a világosságkódok (színek) hirtelen változásaiból. Az átmenet ritkán ideális lépcső (ez csak szintrevágott képekben fordul elő). A természetes képekben az ideális átmeneti tartomány síkmetszete a 2-16. ábra szerinti. A zajos képekben az arányok eltolódnak, erős aszimmetriák lépnek fel. Az élgyanús pontokat ott szokás kijelölni, ahol az átmeneti görbe "inflexiós pontja", vagyis (az ábrán szaggatott vonallal rajzolt) differencia-görbe maximuma van. Megkeresésükre természetesnek tűnik a differencia-hányadosok elemzése, ezek azonban nagyon erősen irányfüggők. Így a gyakorlatban csak legalább két (x-y) irányt "figyelő", gradiens-jellegű operátorok terjedtek el.

Egy képpont akkor élgyanús, ha ott a gradiens nagyobb egy küszöbértéknél. Értékét a pontos összefüggés helyett inkább az (1-12)-ben adott közelítő képlet alapján számoljuk ki, de közkedvelt a

$$(2-47) \quad |\nabla q(k, \ell)| \approx |q(k, \ell) - q(k+1, \ell+1)| + |q(k, \ell+1) - q(k+1, \ell)|$$

Roberts-operátor is. Egyéb eljárások is léteznek; mindegyik azon alapul, hogy a gradienst bármilyen két, egymásra merőleges irányban képzett differenciahányadossal lehet közelíteni.

A Roberts-operátor legfőbb hibája, hogy a vizsgált pontra nézve aszimmetrikus. Ennek kiküszöbölésére számos konvolúciós szűrőt dolgoztak ki az élpontok kijelölésére, s a gradienst a

$$(2-48) \quad |\nabla q(k, \ell)| \approx |q(k, \ell) \otimes E|$$

alakú kifejezéssel közelítik, ahol \otimes a (1-6)-ban definiált konvolúciós összegzést, E pedig valamilyen élkijelölő szűrőt jelent. Lényeges különbség azonban, hogy az élkijelölő szűrők súlya mindig 0 (különben homogén képen is "találnának" élpontokat!), ezért a "szűrés" eredménye nem tekinthető világosságkódnak.

Egyszerűsége miatt nagyon elterjedt az (1-13)-ban kiszámított Laplace-operátor, amelyet az 1-26. ábrán szemléltetünk. Be lehet bizonyítani, hogy az optimális gradiens-közelítő szűrő a detektálási irányra merőlegesen tükrözhető, azzal párhuzamosan szimmetrikus, és - a zajhatások csökkentése érdekében - a középső eleme 0. Ezeket a feltételeket jól teljesíti pl. a Sobel-operátor, amely külön x-, illetve y-irányú éleket detektáló szűrőből áll (ld. 2-32. ábra); S_x és

S_y elemei pedig úgy vannak megválasztva, hogy a gradienst széles szögtartományban jól közelítse.

$$S_x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}; \quad S_y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix};$$

2-32. ábra
Söbel-operátor

A gradiens közelítési hibája 1 szűrő használata esetén 50% is lehet; a kétkomponensű szűrők ebből a szempontból is kedvezőbbek (a hiba közel 0-ra is csökkenhet).

A fentiektől némileg eltér a legjobb illesztési módszer. Elve a következő:

- Tekintsük a képet térbeli felületnek. Fektessünk át egy síkot az aktuális képponton és a Roberts-operátor szerinti 3 szomszédján.
- Mivel ez egzaktul általában nem lehetséges, határozzuk meg a sík paramétereit úgy, hogy a négyzetes hiba minimális legyen.
- Közelítsük a gradienst ezen sík normálisával.

Nem túl bonyolult számolással a következő végeredményre jutunk:

$$(2-49) \quad |\nabla q(k, \ell)| \approx \frac{1}{2} (|q(k+1, \ell+1) + q(k+1, \ell) - q(k, \ell+1) - q(k, \ell)| + |q(k+1, \ell+1) + q(k, \ell+1) - q(k+1, \ell) - q(k, \ell)|).$$

Az eddigieknél finomabb módszer, ha a differenciahányadosokat 4 irányban ($x, y, +45^\circ, -45^\circ$) képezzük, és ezek alapján keressük meg a 2-16. ábra szerinti meredekség-görbe maximumhelyeit. A szimmetriatulajdonságok javítása érdekében célszerű a másodrendű differenciákkal számolni; pl.

$$(2-50) \quad |^2\Delta_x q(k, \ell)| = |q(k-1, \ell) - q(k+1, \ell)|; \text{ stb.}$$

Élgyanús pontot ott jelölünk ki, ahol legalább két irányban találtunk maximumot, s ez nagyobb egy megadott küszöbértéknél.

23.2.2. Élkorrekció

Mint már említettük, egyik élkijelölő eljárás sem ad a szem számára kifogástalan eredményt; pusztán a világosságkódok elemzésével illet nem is lehet elérni. Kellően sima

tok eloszlásának vizsgálatával és szükség szerinti korrigálásával kaphatunk. A cél a valódi élekben (kontúrvonalakban) keletkezett szakadások összekötése és a hamis élpontok törlése.

A görbeillesztési eljárásokban valamilyen

$$p(x) = a_0 + a_1 x + \dots + a_n x^n$$

alakú polinomot illesztenek $(n+1)$ élpontra, s ezt a görbét tekintik tényleges élnek; a "kilógó" élgyanús pontokat törlik. Az illeszkedés "jóságát" rendszerint a szokásos

$$h = \sum_{i=0}^n (\ell_i - p(k_i))^2$$

négyzetes hibafüggvénnyel mérik, s a polinom együtthatóit a

$$\frac{\partial h}{\partial \ell_i} = 0$$

egyenletrendszerből határozzák meg. (Az illesztésre kiválasztott pontsorozatban legalább az egyik koordinátának monoton növekvőnek vagy csökkenőnek kell lennie, különben a polinom nem határozható meg egyértelműen.) A módszer a metsző élekre érzékeny, ezek környezetében "el tud tévedni".

A gyakorlatban csak első fokú polinomokat használnak, ezt sokszöges közelítésnek nevezik. Ha az élek görbülete és a hézagok hossza nem túl nagy (az élhosszakhoz képest), az eljárás kielégítő eredményeket ad.

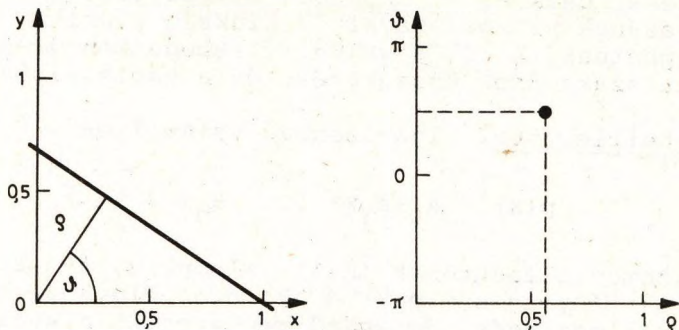
Egy másik módszer a Hough-transzformáción alapszik, és ugyancsak egyenesekkel közelíti az éleket. A Hough-transzformáció egy Descartes-koordinátarendszerbeli egyenest egy polárkoordinátarendszerbeli pontba visz át a

(2-51)

$$\rho = x \cos \vartheta + y \sin \vartheta$$

összefüggés szerint (ld. 2-33. ábra). Egy ponton átmenő egyeneseknek (sugársor) megfelelő pontok egy összefüggő görbét eredményeznek, ez a hordozó pont képe.

Az élkorrekcióhoz egy kis módosítást vezettek be: a szög egy éldarab gradiensét jelenti (a ρ értelmezése változatlan). Ezután a ρ - ϑ síkot a képpontok sűrűségének megfelelően cellákra bontják, és meghatározzák az egyes éldarabok képét. Annak a cellának a számlálóját, amelyikbe a transzformált pont esik, megnövelik a gradiens értékével. Ha minden éldarabot áttranszformáltak, elemzik a cellaszámlálókat;
 - nagy érték azt jelenti, hogy - a cella $(\rho; \vartheta)$ koordi-



2-33. ábra
A Hough transzformáció elve

nátáinak megfelelő egyenes jól illeszkedik az éldarabokra, ezt visszaállítják;

- kis érték elszigetelt éldarabokra utal, ezt törlik az élképből.

Kevésbé számításigényesek és egészen jó eredményeket adnak az olyan heurisztikus algoritmusok, amelyek az élgyanús pontok meghatározott környezetét elemzik. Alapelvük általában az, hogy a megkezdett irányok folytatására törek-szenek; az elszigetelt vagy a megtalált irányokból kieső élgyanús pontokat pedig törlik.

Egészen más elven alapszik a fogyasztásos élkorrekció. A fogyasztás (thinning) olyan művelet, amely az összefüggő



2-34. ábra
Fogyasztási eljárás
eredménye

képponttartományokat azonos topológiai tulajdonságú, (8)-összefüggő, digitális ívdarabokká vagy görbék-ké zsugorítja. Algoritmusok részletezése helyett az eljárás hatását a 2-34. ábrán szemléltetjük: a bepon-tozott eredeti tartományból a fo-gyasztás után a vastagon pontozott (1 képpont szélességű) digitális görbevonallal marad. Az élkorrekcióra úgy kerül sor, hogy lecsökkentik az élkijelölő eljárás küszöbértékeit, miáltal a valódi élek közelében sok hamis élpontot generálnak. Az így kapott "szétkent" éleket azután egy fogyasztó eljárással redukálják vonalakra. Ílymódon lényegesen meg-javul az élek folytonossága és zárt-sága.

Ez idő szerint nem ismeretes olyan megoldás, amelyik minden szempontból kielégítő élképet eredményezne. Újabban kísérletek folynak környezetfüggő (contextual) eljárások ki-dolgozására, amelyek a gradiens elszigetelt vizsgálata he-

lyett az aktuális képpont egy meghatározott környezetét
elemzik (vö. az említett heurisztikus módszerekkel!). Ezál-
tal csökken a hamis élpontok kijelölésének vagy az értékes
élpontok elvesztésének valószínűsége.

3. KÉPOSZTÁLYOZÁS

A digitális képfeldolgozás második nagy területe az alakfelismerés (pattern recognition). Célja, hogy a képben tárolt információt adekvátan és tömören leírja (kép → leírás leképzés: ld. 1-1. ábra); rendszerint azért, hogy az automatikus "látási" folyamat döntő lépését, a képfelismerést elősegítse, ill. lehetővé tegye.

Ezt a tevékenységet agyunk is szüntelenül végzi, amikor a látott információkból kiszűri a lényeges jellemzőket - ezeket a továbbiakban sajátságoknak nevezzük -, majd osztályozza a képrészleteket, vagyis felismeri és azonosítja valamilyen, korábban már látott kép(elem)mel. Fontos megjegyezni, hogy sem a sajátságok, sem az osztályok nem eleve adottak; tanulással és absztrakcióval alakulnak ki. Az értékelésben nemcsak a hasonlóságot, hanem a különbözőséget jellemző sajátságok is fontos szerepet játszanak.

Lényegében ezt a folyamatot modellezzik az alakfelismerés során is. Az elnevezés nem túl szerencsés, mert olyan módszereket és eljárásokat takar, amelyeket számos egyéb területen is sikerrel lehet alkalmazni; elterjedtsége miatt mégis megtartottuk. Egységes elmélete ezideig nem alakult ki, részben talán éppen az alkalmazások sokrétősége miatt. A továbbiakban természetesen csak a képfeldolgozási alkalmazásokról lesz szó. Módszereit tekintve két, lényegesen eltérő ága van, ezeket statisztikus (vagy döntéselméleti), ill. szintaktikus (vagy struktúrális) alakfelismerésnek nevezik. Az előzővel részletesen, a másodikkal csak érintőlegesen foglalkozunk. A továbbiakban feltételezzük, hogy az előkészítő képfeldolgozás (pl. hibajavítás, szegmentálás stb.) megtörtént, és így a felismerendő objektum sajátságait meg lehet mérni.

31. Statisztikus alakfelismerés

A statisztikus (vagy döntéselméleti) módszereken alapuló alakfelismerés általában két fázisban történik (ld. 3-1. ábra). Először meghatározzuk, "kinyerjük" az objektumok sajátságait (feature extraction). (Noha a sajátság elvileg kvalitatív tulajdonság is lehet (pl. szín), a továbbiakban számunkra csak a valamilyen módon mérhető sajátságoknak van

jelentőségük. Ekkor a sajátság "kinyerése" azt jelenti, hogy meghatározzuk a vizsgált objektumra vonatkozó konkrét mérőszámát).

Ezután az osztályozás következik, amikor is az objektumot annak a sajátság-osztálynak az elemeként ismerjük fel, amelytől a mért sajátsági alapján a legkevesébé tér el; leírására pedig az osztály általános leírását (nevét stb.) használjuk.

Az egy osztályba sorolt objektumok bizonyos értelemben "hasonlítanak" egymásra, míg a különbözőkbe soroltak eltérők. Ez nem feltétlenül szembeötlő az egyszerű megfigyelőnek, mivel pl. az adathalmaz áttekinthetetlenül nagy és/vagy bonyolult. Így azt is állíthatjuk, hogy az alakfelismerés segítségével új, eddig ismeretlen összefüggéseket tárunk fel.

A következőkben a döntésfüggvényeken alapuló eljárások alapelveit és néhány egyszerű, gyakran előforduló eljárást ismertetünk. Ezen belül példát adunk arra az esetre, amikor az osztályok teljesen meg vannak határozva (Bayes-módszerek) és arra is, amikor a leírásnak hiányos (távolságmérésen alapuló módszerek) vagy egyáltalán nincs (klaszterezés). Természetesen számos más módszer is lehetséges (és létezik is), de ezek közül itt csak néhányat áll módunkban ismertetni; ilyen például a lineáris diszkrimináns függvények módszere.

31.1. A DÖNTÉSI FOLYAMAT ELŐKÉSZÍTÉSE

Ebben a fejezetben a döntést (vagyis a tulajdonképpeni osztályozást) megelőző feldolgozási lépéseket tekintjük át.

Ahhoz, hogy az előkészített képen található objektumok sajátságvektorait létrehozassuk, de még sokszor a lehetséges osztályok definiálása céljából is, megfelelő méréseket kell végeznünk a képen. Ezt követően pedig lehetőleg ki kell szűrni az adatokból a redundáns információt, hogy meggyorsítsuk a döntési folyamatot.

A következőkben erre szolgáló eljárásokat ismertetünk; végül egzaktul megfogalmazzuk az osztályozási feladatot.



3-1. ábra
Statisztikus alakfelismerési folyamat

31.1.1. A sajátságvektorok előállítás

Az alábbiakban röviden megmutatjuk, hogy az alakfelismerési eljárásokat hogyan lehet felhasználni a digitális képek kiértékelésére.

31.1.1.1. Mérhető képjellemzők keresése

Mint már említettük, a sajátságvektorok összetevői a feldolgozandó kép valamilyen mérhető jellemzőinek konkrét mérőszámai. Az első lépés természetesen a tulajdonképpeni feldolgozandó kép kijelölése, ami lehet a teljes digitális kép, vagy ennek tetszés szerint kiválasztott egy vagy több (esetleg átfedő) részlete.

A sajátságvektorok meghatározására alapvetően kétféle módszer létezik:

1. A strukturális (makro) jellemzők mérésekor a képnek bizonyos jellegzetes alakzatait: a foltokat, ill. kontúrvonalait vizsgáljuk (ld. 23. fejezet). Néhány példa a mérhető jellemzőkre:

- ívhossz vagy kerület (L),
- terület (A),
- alaktényező (A/L^2),
- maximális/minimális átmérő (D, d);
- elnyújtottság (D/d),
- Fourier-jellemzők (pl. a kerület Fourier-transzformáltja),
- inerciatengelyek és szimmetriatengelyek; területosztás az inerciatengelyek alapján,
- a belső pontok átlagos távolsága a kontúrvonalaktól (Danielsson-tényező),
- konvexitás (a foltot befoglaló legkisebb területű konvex síkidom területének aránya A -hoz),
- vázkitűzés (skeleton); a váz és az eredeti alak viszonya,
- sugárfüggvény, polárkoordinátás leírás,
- véletlenszerű szelők metszeteinek eloszlása stb.

A lehetőségek száma, így a sajátságvektor mérete (dimenziószáma) szinte korlátlan, bár ezek a jellemzők egymástól nem függetlenek.

2. A texturális jellemzők a kép mikroszerkezetét írják le. Mint már említettük, a képeken gyakran találunk statisztikusan értékelhető, többé-kevésbé periódikusan előforduló mintákat, amelyeknek mérete kicsi a vizsgált képrészlethez viszonyítva. Ezt nevezzük textúrának (ld. 11.2.4. pont).

Néhány eljárás a sajátságvektorok összetevőinek meghatározására a texturális jellemzők alapján:

- Ortogonalis transzformációk (ld. 13.2.4. pont). Pl. a Fourier-transzformáltból kiválaszthatjuk a

$$\sum |F(u, v)|$$

$$k_1 \leq u^2 + v^2 \leq k_2$$

$$v_1 \leq \operatorname{tg}^{-1}\left(\frac{u}{v}\right) \leq v_2$$

részletet, ahol k_1, k_2, v_1, v_2 a feladatnak megfelelő konstansok.

- Futási hosszak meghatározása. Valamilyen elemi alakzatok (pl. konkrét világosságkódok) segítségével generáljuk a képet. A létező periodicitások miatt ezek nem véletlenszerűen követik egymást.

- Keresztkorreláció. Toljuk el a képet ciklikusan valamilyen távolságra, és képezzük a normalizált keresztkorrelációt az eredetivel. A keletkező korrelációs értékek - ugyancsak a periodicitás következtében - jellemző módon változnak.

- A hisztogram alapján a világosságkódok relatív gyakoriságának statisztikai jellemzőit is (átlag, szórás, maximum, minimális értékek, terjedelem stb.) választhatjuk saját-sávvektornak.

- Az együtt-előfordulási mátrix (co-occurrence matrix) b_{ik} elemét úgy is definiálhatjuk, mint azon q_k világosságkódú képpontok számát, amelyektől megadott távolságban található q_i kódú képpont. A lehetséges jellemzők pl.

- kontrasztosság:
$$\sum_{i=1}^I \sum_{k=1}^I (q_i - q_k)^2 b_{ik};$$

- uniformitás:
$$\sum_i \sum_k b_{ik}^2;$$

- entrópia:
$$\sum_i \sum_k -b_{ik} \log b_{ik};$$

- korreláció:
$$\sum_i \sum_k \frac{q_i \cdot q_k \cdot b_{ik} - \bar{\mu}_i \cdot \bar{\mu}_k}{\bar{\sigma}_i \bar{\sigma}_k};$$

(ahol $\vec{\mu}_i$ és $\vec{\sigma}_i$, ill. $\vec{\mu}_k$ és $\vec{\sigma}_k$ a mátrix i -edik sorának, ill. k -edik oszlopának átlag-, ill. szórásvektora).

A felsorolt és a további jellemzők részletes vizsgálatára nincs lehetőségünk, de megjegyezzük, hogy nagyon jól használhatók a textúrák leírására.

31.1.1.2. Többszörös felvételek jellemzése

Gyakran előfordul, hogy egy adott területről több, geometriailag egybevágó felvétel áll rendelkezésünkre. Ezek készülhetnek azonos időben, de különböző spektrumban (multi-spektrális képek), ill. adott időközökben (multitemporális képek). Ekkor arra is lehetőség van, hogy az egyes képpontokat és az azokhoz tartozó méréseket tekintsük osztályozandó objektumoknak. A sajátságvektort képezhetjük egyszerűen a rendelkezésünkre álló mérésekből. Többsávós képek legjobb példáját az erőforráskutató, meteorológiai és egyéb célú műholdak által készített (3,4, ill. 7 sávós) felvételek szolgáltatják. Gyakorlatban ezeket az információkat általában 2 sávba össze lehet tömöríteni, miáltal az osztályozás jelentősen gyorsítható.

Azonos területről különböző időben készített felvételek általában eltérők. Ezt a felvétel körülményeinek és magának az objektumnak a változásai magyarázzák. (Mezőgazdasági példánál maradva: a vegetáció fejlődése során egy adott vetemény másként néz ki júniusban mint júliusban stb.). Ha azonos képpontokhoz tartozó, de különböző időpontokban nyert méréseket mint egy hisztogramot tekintjük, akkor a korábbiakban már ismertetett eljárásokkal új jellemzőket képezhetünk. A hangsúly ismét az adatvektor hosszának csökkentésén van. Más jellegű, de ugyancsak többsávós képet kapunk, ha valószínűsítés (RGB) képműkártyával vagy szkennelrel digitalizálunk, ilyenkor általában sávonként 256 szintes kép jön létre. Osztályozási feladatra vezet, ha a háromsávós képből egyszávós - újabb terminológia szerint palettás - színes képet akarunk létrehozni. A feladat nehézségét mutatja a számos, nagyon gyenge megoldást tartalmazó, mégis elterjedt program.

31.1.1.3. Számítástechnikai megfontolások

Miután kialakítottuk a sajátságvektorokat, válasszuk ki azt a módszert, amelyik a legjobban megfelel a sebességi és a pontossági követelményeknek. Ideális esetben előre ismerjük az osztályainkat, pontos paramétereik rendelkezésre állnak. Egyébként paraméterbecslést kell végeznünk (ld. 31.2.3. pont). A megbízható paraméterbecsléshez szükséges adathalmaz nagysága a sajátságvektor dimenziószámával gyorsan nő. Így esetleg szükség lehet lényegtömörítés (ld. 31.1.2. pont) ismételt alkalmazására is. Ha nincs elég tanítónk, klaszterezést végezhetünk, igen nagy körültekintéssel (ld. 31.4. alfejezet).

Az osztályozás előtt igen lényeges, hogy megvizsgáljuk az osztályok statisztikai jellemzőit. Az osztályozási módszerek túlnyomó többsége egymódusú (unimodális) (esetleg

többdimenziós), normális eloszlással jól közelíthető osztályok esetén működik kielégítően. Ezek a feltevések azonban általában nem teljesülnek. Ilyenkor nincs más lehetőség, mint új osztályokat képezni (akár klaszterezéssel, akár más módon), amelyek már kielégítik a fenti követelményeket.

Ezek után kétféle osztályfogalmat célszerű megkülönböztetni:

- az informatikai osztályt, ill. a
- spektrális osztályt.

Az előbbi olyan osztályozást jelent, ami a feladat szempontjából fontos, a mérések közti megkülönböztetésre alkalmas. Ilyen fogalom lehet például a "mezőgazdasági hasznosítású terület", "sűrűn beépített terület" stb. Nyilvánvaló, hogy az informatikai osztály általában nem teljesíti az említett elvárásainkat. Ezért például az alábbi bontást lehet elképzelni: a "mezőgazdasági hasznosítású terület" osztályt felbonthatjuk pl. "búza", "kukorica", "legelő", "vízzel borított terület", "egyéb" spektrális osztályokra, ahol hallgatólagosan feltesszük, hogy az új osztályok már unimodális eloszlásúak.

Minden esetben fontos az osztályozási eredményeink értékelése. Mivel az egzakt eredményt általában nem ismerjük, megint csak közelítő becslésekre kell hagyatkoznunk. Még az osztályozás elvégzése előtt, az osztályok közti távolságok vizsgálatával (ld. 31.1.2. pont) kaphatunk hozzávetőleges képet a várható eredményről. Osztályozás után - az eredményünket a hasonló adatokkal összevetve - utólagos becslést végezhetünk.

A fentiekből világosan kitűnik, hogy az osztályozás feladata igen összetett. A felhasználónak tiszta elképzeléssel kell munkához látnia, és a feldolgozás megfelelő fázisaiban az aktuális eredménytől függően beavatkoznia. Az eljárások ugyan nagymértékben automatizálhatók, de az emberi közreműködést egyelőre nem lehet nélkülözni.

31.1.2. Lényegtömörítés

Az alakfelismerés során végzendő osztályozásoknál nagy jelentősége van az információ tömör tárolásának. Arról van szó, hogy a módszerek hatékonysága erősen függ a sajátságvektorok dimenziószámától; minél kisebb dimenziószámú sajátságvektorok alapján tudunk eredményes döntést hozni, annál gyorsabbak lesznek az algoritmusaink, és egyidejűleg a tárigényük is csökken.

A dimenziócsökkentést két tényező korlátozza: egyrészt ezáltal csökken a döntés pontosága, ami bizonyos határon túl nem megengedhető; másrészt az adattömörítésnek is van valamekkora műveletigénye. Ennek ellenére nyilvánvaló a dimenziószám csökkentésének gyakorlati szükségessége. Feltehető ugyanis, hogy általában - egy adott osztályozási feladat szempontjából - nem minden mérési adat egyformán fontos, egyesek elhanyagolhatók. Léteznek továbbá belső összefüggések is az adatok között, emiatt a mérési információ redundáns.

A csökkentésre lényegében két lehetőségünk van:

- lényegkiemelésnek (feature selection) nevezzük azt az eljárást, amikor az új sajáttságvektort a réginek bizonyos összetevőiből állítjuk össze, a többit egyszerűen elhagyjuk;

- lényegtömörítésről (feature extraction) pedig akkor beszélünk, ha az új vektort a réginek tetszőleges transzformációjával állítjuk elő, amelyben minden összetevőt figyelembe veszünk. (Jól látható, hogy ez az eset az előbbit is magában foglalja.)

Jelképesen a lényegtömörítésen olyan

$$(3-1) \quad T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

operátort értünk, amely az n dimenziós teret $m < n$ dimenziós térbe képezi le. A feladat ennek az operátornak a meghatározása.

Látszólag a lényegkiemelés az egyszerűbb feladat, hiszen csak válogatni kell, és az eredményt kell kiértékelni. A gyakorlatban azonban a lehetséges részhalmazok száma a dimenziószámmal nő, így itt is csak közelítő eljárások jöhetnek szóba. Ezért a továbbiakban csak lényegtömörítő eljárásokat ismertetünk.

31.1.2.1. Karhunen-Loève transzformáció

Legyen A egy N elemű méréshalmaz \mathbb{R}^n -ben és az $m < n$ adott szám. Keressük a (3-1) szerinti T transzformációt a lineáris leképezések körében. Vezessük be az alábbi jelöléseket:

- $\vec{z}^T = (z_1, z_2, \dots, z_n)$ a sajáttságvektor (a mérési értékekből képzett sorvektor);

- $s(\vec{z})$ az A halmazban előforduló mérések sűrűségfüggvénye;

- $\vec{\mu} = M(\vec{z}) = 0$ (a $\vec{\mu} = 0$ feltétel nem jelent lényeges megkötést: egyszerű kivonással mindig elérhető);

- $\underline{\Omega} = M(\vec{z} \vec{z}^T)$ a kovariancia-mátrix.

Ekkor az

$$\underline{\Omega} \vec{u} = \lambda \vec{u}, \quad (\vec{u} \in \mathbb{R}^n, \lambda \in \mathbb{R})$$

egyenletrendszer $\vec{u}_i \in \mathbb{R}^n$ megoldásait az $\underline{\Omega}$ sajátvektorainak, a hozzájuk tartozó λ_i értékeket pedig sajátértékeinek nevezük. Bizonyítható, hogy minden $\underline{\Omega}$ kovariancia mátrix esetén található n db. (\vec{u}_i, λ_i) pár úgy, hogy az

$$\vec{u}_i^T \cdot \vec{u}_k = \begin{cases} 0, & \text{ha } i \neq k \\ 1, & \text{ha } i = k \end{cases}$$

egyenlőségrendszer fennáll, és a λ_i sajátértékek nemnegatívok. Sorszámozzuk az \vec{u}_i vektorokat a hozzájuk tartozó sajátértékek monoton csökkenő sorrendjében. Ekkor a Karhunen-Loève transzformáció ("főkomponens-transzformáció") mátrixát a

$$(3-2) \quad \underline{K} = [\vec{u}_1^T, \dots, \vec{u}_m^T]$$

kifejezés definiálja. Látjuk, hogy \underline{K} az $\underline{\Omega}$ sajátvektoraiból mint sorokból képzett $m \times n$ méretű mátrix. Be lehet bizonyítani, hogy - ha a (3-1)-ben $T = \underline{K}$ választással élünk, akkor a

$$(3-3) \quad \vec{v} = \underline{K} \vec{z}$$

transzformáció olyan \vec{v} -t eredményez, amelynek - mint m -dimenziós vektornak - hossza átlagosan a lehető legnagyobb, minden más lineáris transzformációval nyerhető vektorral szemben.

Eszerint a (3-3) transzformáció optimális abban az értelemben, hogy átlagos négyzetes hibája minimális a lineáris leképezések között. Megjegyezzük, hogy a \vec{v} koordinátái közötti korreláció 0, azaz az új "mérések" korrelálatlanok lesznek.

A transzformáció fontos tulajdonsága még, hogy érvényes az alábbi összefüggés

$$(3-4) \quad M(|\vec{v}|^2) = \sum_{i=1}^m \lambda_i.$$

Ezek szerint, ha m értéke nincs előre megadva, akkor érdemes úgy megválasztani, hogy a

$$\frac{\sum_{i=1}^m \lambda_i}{\sum_{m=1}^n \lambda_i} \approx 1$$

egyenlőtlenség teljesüljön.

Előnyös tulajdonságai mellett fel kell hívnunk a figyelmet a Karhunen-Loève transzformáció két hibájára is:

- a K megadásához ismerni kell az $\underline{\Omega}$ kovariancia-mátrixot, amely feladatról feladatra változhat;
- az optimalitási kritérium a transzformált v közelítő tulajdonságaira vonatkozik, és nem az osztályok közti eredeti különbségekre, márpedig minket nyilván az utóbbi érdekel jobban.

31.1.2.2. Tömörítés célfüggvény alapján

A következőkben olyan célfüggvényeket tárgyalunk, amelyek szóba jöhetnek alternatív megoldásként:

1. Entrópia

Legyen

$$E(\vec{z}) = - \int_{\mathbb{R}^n} \log(s(\vec{z})) s(\vec{z}) d\vec{z}$$

egy n -dimenziós eloszlás entrópiája. Olyan $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineáris leképezést keresünk, amely az $E(\vec{z})$ entrópiát maximalizálja.

Bizonyítható, hogy a megoldást a Karhunen-Loève transzformáció szolgáltatja, ha tudjuk, hogy az eredeti vektorok 0 várható értékű normális eloszlást követnek.

2. Divergencia

Az egyszerűség kedvéért $p=2$ osztályra szorítkozunk. Legyen $s_1(\vec{z})$ és $s_2(\vec{z})$ a két osztályhoz tartozó feltételes sűrűségfüggvény, ekkor az

$$U(s_1, s_2) = \int_{\mathbb{R}^n} \left(s_1(\vec{z}) \log\left(\frac{s_1(\vec{z})}{s_2(\vec{z})}\right) + s_2(\vec{z}) \log\left(\frac{s_2(\vec{z})}{s_1(\vec{z})}\right) \right) d\vec{z}$$

összefüggéssel megadott értéket a két sűrűségfüggvény divergenciájának nevezzük. Minél nagyobb a divergencia, annál nagyobb az eltérés az osztályok között, annál kisebb a várható hiba.

Megmutatható, hogy $m < n$ esetén a $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ lineáris leképezést használva $U > U_T$, ahol U_T az m -dimenziós térbeli eloszlás divergenciája. A megfelelő T transzformációt a $\max_T (U_T)$ optimalizálás révén lehet megtalálni.

3. Hellinger-integrál

Szintén a $p=2$ esetre definiáljuk a Hellinger-integrált:

$$H(s_1, s_2) = \int_{\mathbb{R}^n} \sqrt{s_1(\vec{z}) s_2(\vec{z})} d\vec{z}$$

ahol s_1 és s_2 a két osztály feltételes sűrűségfüggvénye. Mivel $0 \leq H(s_1, s_2) \leq 1$ érvényes, ezért szokták a

$$B(s_1, s_2) = -\log H(s_1, s_2)$$

értéket is tekinteni, amely a divergenciához hasonlóan annál nagyobb, minél jobban különbözik a két eloszlás.

$B(s_1, s_2)$ -t Battacharayya-távolságnak is nevezzük. A lényegtömörítés most is úgy zajlik, hogy adott $s_1(\vec{z})$ és $s_2(\vec{z})$ sűrűségek esetén a T transzformációt úgy választjuk meg, hogy az m -dimenziós térben az átalakított sűrűségek $B(\hat{s}_1, \hat{s}_2)$ távolsága maximális legyen.

Az érdekesség kedvéért megjegyezzük, hogy a Hellinger-integrál segítségével megbecsülhetjük az optimális osztályozás hibáját is.

Az eddig ismertetett eljárásokban a T transzformáció nem az osztályok különbségi jellemzőit maximalizálja; ez legalábbis megkérdőjelezi a hasznosságukat. Ennek ellenére az esetek túlnyomó többségében kielégítő eredményt adnak.

Léteznek egyéb, az osztályok szeparálhatóságát közvetlenül figyelembe vevő eljárások is (pl. a Foley-Sammon módszer); nemlineáris, bizonyos feltételek mellett alkalmazható algoritmusok (pl. Koontz-Fukunaga módszer); ezek elemzése azonban már meghaladja lehetőségeinket.

31.1.3. A döntési feladat megfogalmazása

Az alábbiakban precízen megfogalmazzuk a statisztikus alakfelismerés alapfeladatát.

Tegyük fel, hogy bizonyos objektumokról mérések állnak rendelkezésünkre. A mérési eredményeket n -dimenziós vektoroknak tekintjük és a

$$(3-5) \quad \vec{\xi}^T = \{ \xi^{(1)}, \dots, \xi^{(n)} \}$$

szimbólummal jelöljük. A $\xi^{(r)}$ ($r=1, \dots, n$) függvényeket valószínűségi változóként kezeljük, amelyek valamely $(\Xi, \xi P)$ valószínűségi mezőről képeznek le az \mathbb{R} valós számegyenesre. (A feladat szempontjából nem szükséges, hogy pontosan ismerjük a Ξ halmazt, az ξ eseményrendszert és a P valószínűségi mértéket; esetenként szükségünk lesz viszont az eloszlás-, ill. a sűrűségfüggvényre.) A ξ vektort sajátságvektor-nak (vagy tünetvektorok, feature-vektorok), az \mathbb{R}^n teret mintatérnek nevezzük.

Feltesszük, hogy a $D = \{d_1, \dots, d_p\}$ döntéshalmaz létezik, és definiálja a mért objektumok lehetséges osztályait. Az osztályozási vagy felismerési feladat egy

$$(3-6) \quad \delta: \mathbb{R}^n \rightarrow D \quad \text{vagy} \quad \delta: \xi \rightarrow \delta(\xi)$$

leképezés megtalálása.

A döntésfüggvény tehát a mintatér egy-egy vektorát rendel hozzá a döntéstér elemeihez. (Általában mégis az objektumok osztályozásáról beszélünk, mivel az objektumokról csak a mintatérbeli vektorok formájában van információnk. A δ segítségével éppen ebből következtetünk vissza az objektum tulajdonságaira, azaz, hogy milyen osztályhoz is tartozott.)

A mondottak értelmezésére a következő megjegyzéseket tesszük:

1. A mintatér gyakorlatilag lehetne sokkal általánosabb, tehát az objektum leírása lényegesen összetettebb is. (Ha például egy test hőmérsékletét folytonosan mérjük egy időintervallumon keresztül, leírásának (legalábbis egy része) a hőmérsékletfüggvény lehet.) A gyakorlatban azonban az \mathbb{R}^n választás elegendő.

2. A D halmaz lehet végtelen is (akár kontinuum számosságú). (Ha pl. egy objektum színképéből próbálunk a sebességére következtetni, a sebesség gyakorlatilag egy intervallum tetszőleges értéke lehet.)

3. Ha D elemeit nem, vagy nem mindegyiket ismerjük előre, klaszterezésről beszélünk. Ekkor - valamilyen szempontok alapján - az osztályokat is ki kell jelölni.

4. A δ leképezésről általában feltesszük, hogy teljes, azaz minden \mathbb{R}^n -beli vektorhoz hozzárendel valamilyen döntést (D legalább egy elemét). Ez nem jelent tényleges megszorítást, ugyanis ellenkező esetben legyen

$$\bar{D} = D \cup \{d_0\},$$

ahol d_0 a "nincs döntés" osztályt jelenti, és

$$\delta: \begin{cases} \bar{\delta}(\xi) = d_0, & \text{ha addig a } \delta \text{ döntés nem} \\ & \text{volt értelmezve, ill.} \\ \bar{\delta}(\xi) = \delta(\xi), & \text{egyébként.} \end{cases}$$

5. Feltesszük azt is, hogy δ egyértelmű. Valójában észszerű lenne megengedni, hogy egy objektumot akár több osztályba is besoroljunk. Lehet ugyanis, hogy az osztályok nem kizáró sajátosságokat jelentenek; vagy a rendelkezésre álló információ alapján csak azt lehet eldönteni, hogy az objektum egyáltalán beletartozhat-e valamelyik osztályba. (Szemléletes példaként gondoljunk meg, hogy pusztán a lábainak számából nem tudjuk eldönteni, hogy milyen állatról van szó.) Ténylegesen azonban a fenti feltételezés sem jelent megszorítást. Tekintsük ugyanis a $\tilde{D} = 2^D$ (azaz a D részalmazzaiból álló) halmazt: ekkor már lehet egyértelmű a döntésfüggvényünk.

6. A δ függvénnyel szemben két ellentmondó követelmény, hogy

- kellően pontos (ez azt jelenti, hogy az objektum besorolása alapján nyert osztály-sajátságvektorból, ill. a mért sajátosságvektorból visszakövetkeztetett eredmény lehetőleg megegyezzek); illetve

- jól programozható és "olcsón" futtatható (kis helyfoglalás, rövid futási idők) legyen.

A döntések pontatlanságának két legfőbb oka:

- Egyrészt az objektumokra vonatkozó mérési eredmények nem egy-egyértelműek. Előfordulhat, hogy több (különböző osztályba tartozó) objektum esetén is ugyanaz a sajátosságvektor adódik (lásd a fenti példát 4 láb esetén).

- Másrészt ismét ki kell hangsúlyozni, hogy a mérési eredmények valószínűségi változók. Eszerint a mérés, különböző okok miatt, azonos osztályba tartozó objektumok esetén is eredményezhet különböző sajátosságvektor(oka)t. Mindezek alapján világos, hogy a pontosság kérdését valószínűségi alapon kell megközelíteni.

7. Az osztályozást halmazelméleti eszközökkel is meg lehet fogalmazni. Tulajdonképpen arról van szó, hogy az \mathbb{R}^n mintateret diszjunkt alterekre (= partíciókra) bontjuk fel.

$$\mathbb{R}^n = \bigcup_{k=1}^p D_k, \quad (D_i \cap D_k = \emptyset, \quad \forall i \neq k);$$

(3-7)

$$D_i = \{ \vec{s} \in \mathbb{R}^n \mid \delta(\vec{s}) = d_i \} \quad (i = 1, \dots, p).$$

Látjuk, hogy egy alteret azok a vektorok jelölnék ki, amelyeket ugyanahhoz a döntéshez (osztályhoz) rendeltünk hozzá.

8. Tegyük fel, hogy a (3-6) döntésfüggvényt úgy lehetett megadni, hogy léteznek a

$$g_\ell : \mathbb{R}^n \rightarrow \mathbb{R} \quad (\ell = 1, \dots, p)$$

függvények, amelyekre igaz, hogy

$$(3-8) \quad \delta(\vec{\xi}) = d_i \Leftrightarrow g_i(\vec{\xi}) \geq g_\ell(\vec{\xi}) \quad \forall \ell\text{-re.}$$

Ezeket a függvényeket diszkrimináns függvényeknek fogjuk nevezni. Azokat a $\vec{\xi} \in \mathbb{R}^n$ vektorokat, amelyekre igaz, hogy $\exists i, \ell \quad g_i(\vec{\xi}) = g_\ell(\vec{\xi})$, a diszkrimináns-felület elemeinek hívjuk. (Az \exists szimbólum szokásos jelentése: "van olyan", vagy "létezik".)

A $\{g_\ell\}$ ($\ell=1, 2, \dots, p$) függvényrendszer egyértelműen meghatározza a hozzá tartozó δ döntésfüggvényt, fordítva azonban ez nem áll. Tetszőleges monoton (egyenlőtlenség-tartó) transzformáció alkalmazása esetén újabb diszkrimináns-függvényekhez jutunk, amelyek ugyanazt a döntésfüggvényt határozzák meg; miközben a diszkrimináns-felületek nem változnak.

31.2. BAYES-DÖNTÉSELMÉLET

A továbbiakban feltételezzük, hogy a szegmentálás végrehajtása után meghatároztuk az objektumok sajátosságait. Így elegendő ismeret áll rendelkezésünkre ahhoz, hogy el tudjuk végezni az optimális osztályozást.

Tegyük fel, hogy ismerjük az alábbiakat:

- a $D = \{d_1, \dots, d_p\}$ döntéshalmazt, amely definiálja a lehetséges osztályokat;

- az egyes osztályok előfordulásának $P(d_i)$ ($i=1, \dots, p$) "a priori" (előre ismert) valószínűségét;

- az $s(\vec{\xi} | d_i)$ feltételes sűrűségfüggvényeket; amelyek a $\vec{\xi}$ valószínűségi változó előfordulási valószínűség-eloszlását adják meg, ha tudjuk, hogy d_i osztályhoz tartozó objektumról van szó; végül

- az $L_{ik} \geq 0$ ($i, k = 1, \dots, p$) veszteségi súlyokat; ezek a számok azt a veszteséget fejezik ki, amivel a téves besorolást büntetjük. (Megjegyezzük, hogy ezek az adatok rendszerint nem ismeretesek, becslésekkel lehet közelítő értékeket kapni. Ezekre a későbbiekben visszatérünk.)

Jelölje továbbá $P(d_i | \vec{\xi})$ annak az "a posteriori" (megfigyelés utáni) valószínűségét, hogy egy konkrét $\vec{\xi}$ mérési eredmény esetén éppen a d_i döntést fogjuk hozni (vagyis, hogy ekkor az objektumot a d_i -vel jellemzett osztályba soroljuk.)

A következőkben megadjuk a várható veszteséget tetszőleges δ döntésfüggvény esetén. A fenti jelölések felhasználásával legyen

$$(3-9) \quad R(d_i | \vec{\xi}) = \sum_{k=1}^p L_{ik} \cdot P(d_k | \vec{\xi}) \quad (i = 1, 2, \dots, p)$$

az a feltételes döntési kockázat, amelyet akkor vállalunk, ha ξ mérés esetén a d_1 döntést fogadjuk el. Ekkor nyilvánvalóan

$$(3-10) \quad R_{\delta} = \int_{\mathbb{R}^n} R(\delta(\vec{\xi}) | \vec{\xi}) \cdot s(\vec{\xi}) d\vec{\xi}$$

lesz a δ döntésfüggvény esetében várható átlagos veszteség.

31.2.1. A Bayes-féle döntési eljárás

Bayes-féle döntésről beszélünk, ha a (3-10) kifejezés minimális:

$$(3-11) \quad R_{\delta^*} = \min_{\delta} (R_{\delta}).$$

Az így kapott R_{δ^*} értéket Bayes-veszteségnek nevezzük, és úgy is tekinthetjük, mint az adatainkban levő "belső" bizonytalanság mértékét: semmiféle más döntéssel nem tudunk ennél kisebb hibát elérni.

(3-10) vizsgálata alapján kézenfekvő, hogy a kockázat akkor minimális, ha a (3-9) szerinti kockázat-függvény tetszőleges $\vec{\xi}$ -ra minimális. Ezek alapján $\delta^*(\vec{\xi}) = d_1$, vagyis a $\vec{\xi}$ méréssel jellemzett objektumot akkor soroljuk minimális veszteséggel a d_1 osztályba, ha

$$(3-12) \quad \sum_{k=1}^p L_{ik} \cdot P(d_k | \vec{\xi}) \leq \sum_{k=1}^p L_{lk} \cdot P(d_k | \vec{\xi}) \quad (l=1, \dots, p).$$

Ha az egyenlőség több i értékre is teljesül, akkor az objektumot ezek közül bármelyik osztályba sorolhatjuk.

A fenti általános eset speciális változatában legyen

$$(3-13) \quad L_{ik} = \begin{cases} 1 & \forall k \neq i, \\ 0 & \forall k = i, \end{cases} \quad (i, k = 1, \dots, p).$$

Mivel eszerint a hibás döntés egységnyi, a helyes döntés 0 veszteséggel jár, a várható átlagos veszteség egyenlő lesz a várható hibaszámmal. A hibát minimalizáló Bayes-döntés ekkor

$$(3-14) \quad P(d_i | \vec{\xi}) \geq P(d_k | \vec{\xi}) \quad (k = 1, \dots, p).$$

Mind a (3-12), mind a (3-14) összefüggésből látjuk, hogy a Bayes-döntésfüggvényt meg lehet adni diszkriminációs-függvények segítségével is. (Pl. (3-14)-ben egyszerűen a $g_i = -P(d_i | \vec{\xi})$ választással élhetünk.)

Alkalmazzuk most az (1-23)-ból ismert Bayes-tételt

$$(3-15) \quad P(d_1 | \vec{\xi}) = \frac{s(\vec{\xi} | d_1) \cdot P(d_1)}{s(\vec{\xi})}, \quad \text{ahol}$$

$$(3-16) \quad s(\vec{\xi}) = \sum_{k=1}^p P(d_k) \cdot s(\vec{\xi} | d_k).$$

A diszkrimináns-függvények átalakíthatóságáról mondatokat felhasználva a $\delta^*(\vec{\xi}) = d_1$ döntést akkor hozzuk, ha

$$(3-17) \quad s(\vec{\xi} | d_1) \cdot P(d_1) \geq s(\vec{\xi} | d_k) \cdot P(d_k) \quad (k=1, \dots, p)$$

(Az $s(\vec{\xi})$ tényező a nevezőben d_k -ktől független, és ezért elhagyható!) A várható döntési hibára pedig rövid számolás után az

$$(3-18) \quad R_{\delta^*} = 1 - \int_{\mathbb{R}^n} \max_{1 \leq i \leq p} (P(d_i | \vec{\xi}) \cdot s(\vec{\xi})) \, d\vec{\xi}$$

eredményt kapjuk.

Különösen fontos a $p=2$ eset, mikor csak két lehetséges kimenete van a döntésnek. Érdeemes ezért ezt a speciális esetet részletesebben is szemügyre venni. A (3-12) összefüggést a (3-15) és a (3-16) figyelembevételével

$$(3-19) \quad \delta^*(\vec{\xi}) = \begin{cases} d_1, & \text{ha } \frac{L_{12} - L_{22}}{L_{21} - L_{11}} \cdot \frac{P(d_2)}{P(d_1)} < \frac{s(\vec{\xi} | d_1)}{s(\vec{\xi} | d_2)} \\ d_2, & \text{egyébként} \end{cases}$$

alakúra hozhatjuk. Ebből a (3-13)-nak megfelelő, minimális hibájú döntés:

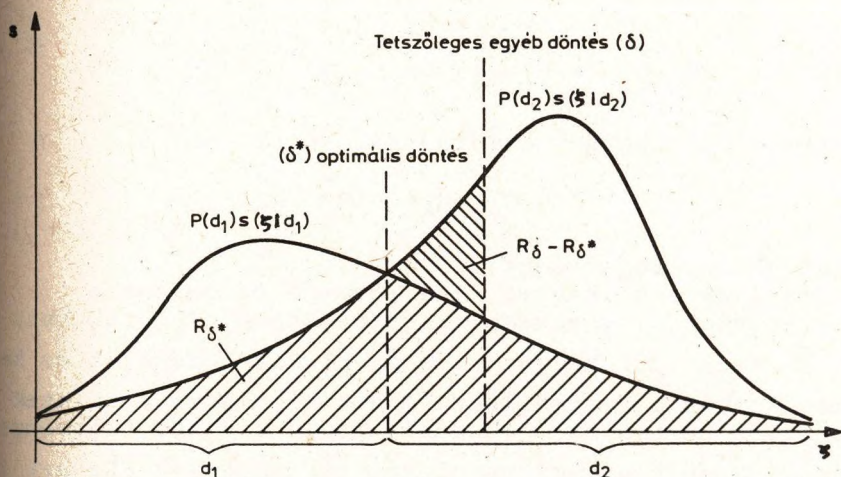
$$(3-20) \quad \delta^*(\vec{\xi}) = \begin{cases} d_1, & \text{ha } \frac{P(d_2)}{P(d_1)} < \frac{s(\vec{\xi} | d_1)}{s(\vec{\xi} | d_2)} \\ d_2, & \text{egyébként} \end{cases}$$

alakot ölt, vagyis a valószínűségek hányadosát egy jól meghatározott, $\vec{\xi}$ -től független küszöbértékkel kell összehasonlítani. Eredményeinket 1-dimenziós esetre a 3-2. ábrával illusztráljuk.

A veszteség értékére hasonlóan elegáns, egyszerű eredményt kapunk. Hibaminimalizálás esetén a (3-18) összefüggésből kiindulva rövid számolás után az

$$(3-21) \quad R_{\delta^*} = \frac{1}{2} - \frac{1}{2} \int_{\mathbb{R}^n} \sqrt{1 - 4 \cdot s(d_1 | \vec{\xi}) \cdot s(d_2 | \vec{\xi})} \, d \vec{\xi}$$

illetve az



3-2. ábra

Kétesélyes Bayes-féle döntés egydimenziós modellje

$$(3-22) \quad R_{\delta^*} = \frac{1}{2} - \frac{1}{2} \int_{\mathbb{R}^n} |D^*(\vec{\xi})| \cdot s(\vec{\xi}) \, d \vec{\xi}$$

kifejezéseket kapjuk, ahol

$$D^*(\vec{\xi}) = P(d_2 | \vec{\xi}) - P(d_1 | \vec{\xi})$$

31.2.2. Speciális esetek

A gyakorlat szempontjából legfontosabb esetek egyikéhez akkor jutunk, ha feltesszük, hogy a mérési eredmények (n-dimenziós) normális (Gauss-féle) eloszlást követnek. Ekkor a sűrűségfüggvény:

$$(3-23) \quad s(\vec{\xi} | d_i) = \frac{1}{(2\pi)^n \cdot |\underline{\Omega}_i|^{1/2}} \exp\left(-\frac{1}{2} (\vec{\xi} - \vec{\mu}_i)^T \underline{\Omega}_i^{-1} (\vec{\xi} - \vec{\mu}_i)\right)$$

(i=1, ..., p)

alakú lesz, ahol $\vec{\mu}_i$ a d_i osztályhoz tartozó valószínűségi változók várhatóértékéből képzett vektor, $\underline{\Omega}_i$ pedig a megfelelő kovariancia-mátrix.

A fejezet további részében csak a hibaminimalizálás esetét vizsgáljuk, a (3-17) szabályból kiindulva.

A diszkrimináns-függvényeken azonos átalakításokat végezve vehetjük az egyenlőtlenség mindkét oldalának logaritmusát, a konstans tagokat pedig elhagyhatjuk. Végül is a $\sigma^*(\vec{\xi}) = d_i$ döntést akkor hozzuk, ha a

$$(3-24) \quad \tilde{g}_k(\vec{\xi}) = c_k - (\vec{\xi} - \vec{\mu}_k)^T \underline{\Omega}_k^{-1} (\vec{\xi} - \vec{\mu}_k)$$

diszkrimináns-függvény $k=i$ -re maximális, ahol a

$$c_k = 2 \log P(d_k) - \log |\underline{\Omega}_k|$$

értékek $\vec{\xi}$ -től függetlenek.

Számítástechnikailag ez a következőt jelenti:

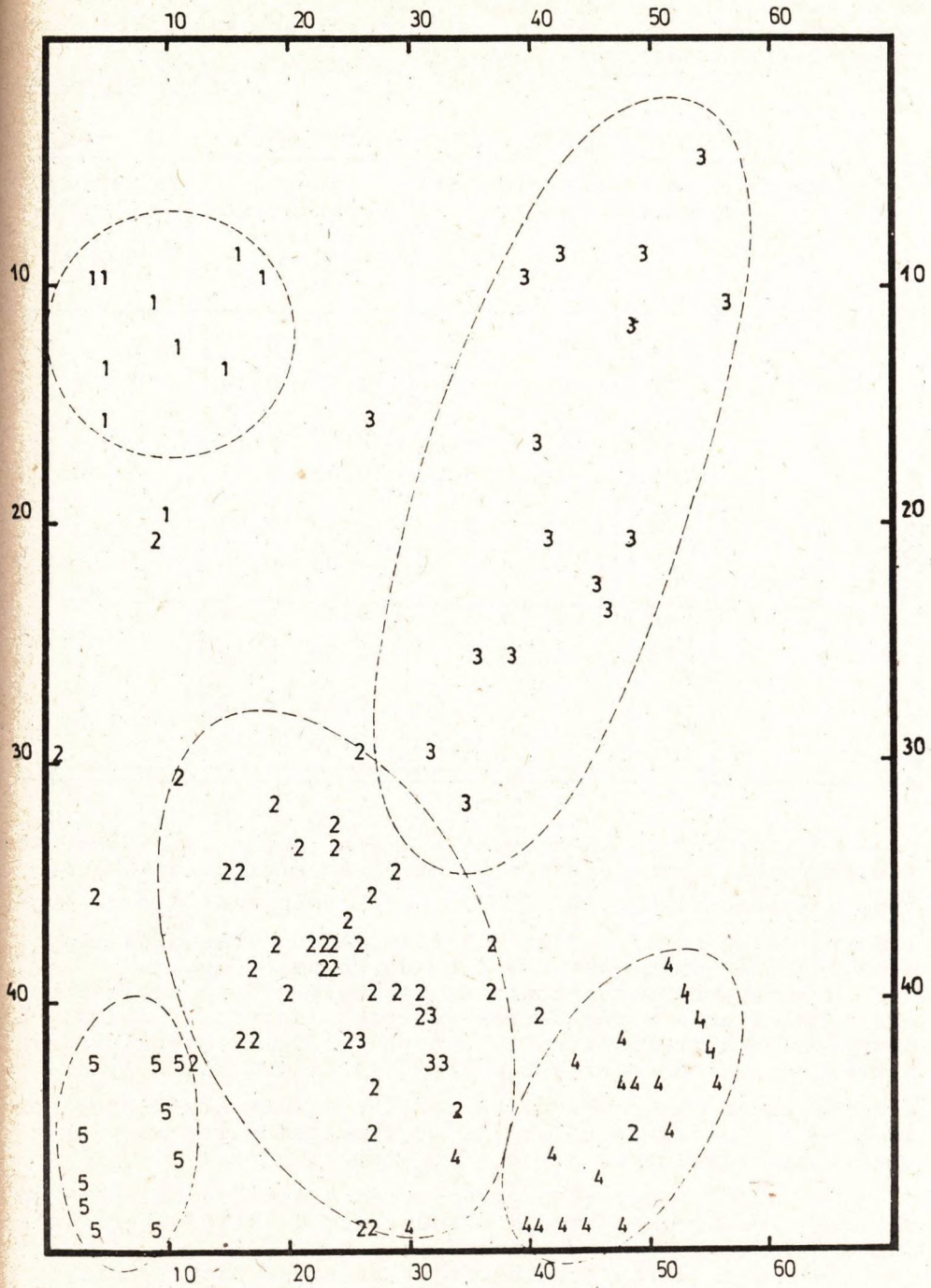
Számítsuk ki egyszer (és tároljuk) a c_k értékeket, majd adott $\vec{\xi}$ esetén értékeljük ki a $\tilde{g}_k(\vec{\xi})$ kifejezéseket és keressük meg köztük a maximálisat. A diszkrimináns felületek az n -dimenziós térben másodrendűek, mivel a $\tilde{g}_k(\vec{\xi})$ függvények tulajdonképpen kvadratikus alakot tartalmaznak. Ez azt jelenti, hogy a számítási igény a mintatér dimenziójával négyzetes arányban növekszik.

Az elmondottakat egy számpéldával illusztráljuk.

Számítógéppel generáltunk 100 db kétdimenziós (x, y) sajtáságvektort, s ezeket a 3-3. ábrán ábráztuk úgy, hogy az egyik sajtáságot az x -, a másikat az y -koordináta reprezentálja. (Az y -tengely felülről lefelé van irányítva, és nyomtatási okokból az y -lépték nagyobb.) Az ábrán egy osztályba tartozó vektorokat a végpontjukban feltüntetett, azonos szimbólumok jelzik.

Az egyes osztályok generálási paramétereit a 3-1. táblázat tartalmazza.

Ezek után elvégeztük a sajtáságvektorok osztályozását, a Bayes-féle döntéelmélet alapján. Ehhez kiindulásul a 100 db $\xi(x, y)$ sajtáságvektor végpontjának koordinátáit és az osztályoknak a 3-1. táblázatban megadott jellemzőit használtuk fel; (vagyis "elfelejtettük", hogy eredetileg melyik vektor melyik osztályba tartozott). A (3-23) szerinti normális sűrűségfüggvény feltételezésével a (3-17) képletből kiindulva és a (3-13) szerinti "büntetőfüggvényt" felhasználva (miszerint a hibás döntés 1, a helyes döntés 0 hibapontot eredményez), azonos átalakításokkal a (3-24) szerinti döntési függvényhez jutunk, amelynek kiértékeléséhez minden adat rendelkezésünkre áll. Ennek alapján mindegyik (x, y) sajtáságvektorhoz kiszámítottuk a diszkrimináns-függvény ér-



3-3. ábra
Számítógéppel generált fiktív sajáttságvektorok

tékét mindegyik osztályra vonatkozóan ($k = 1, 2, \dots, 5$) és az egyes vektorokat abba az osztályba soroltuk, amelyekre nézve ez az érték maximális volt.

3-1. táblázat

Mintapéllda generálási paramétereit

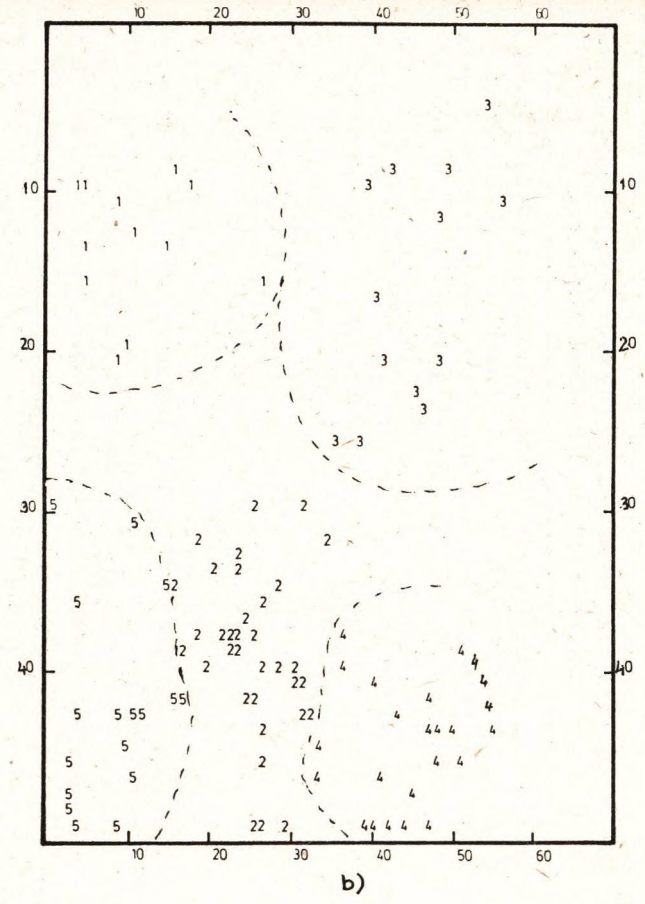
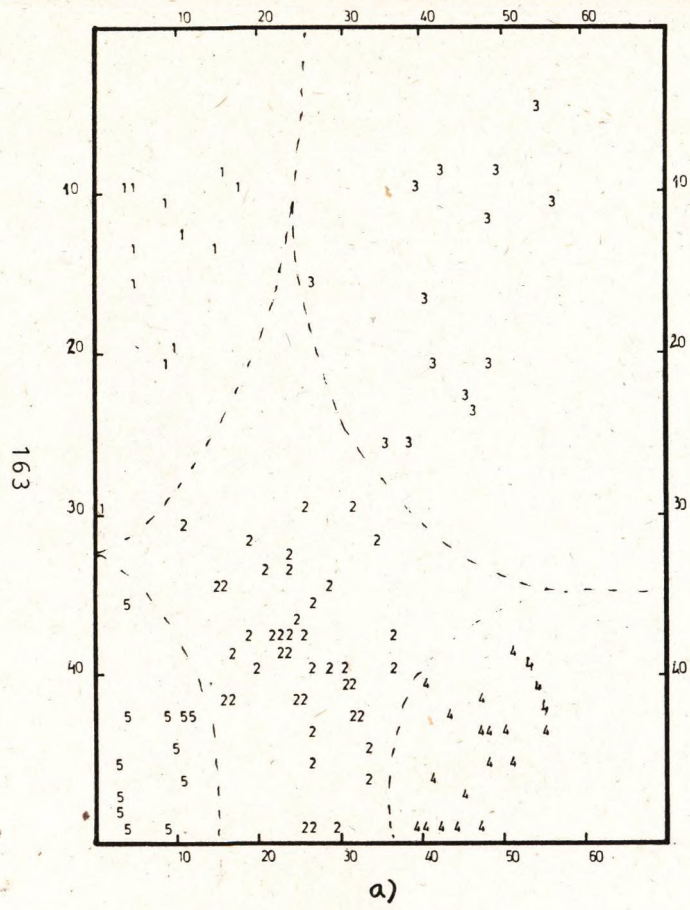
jellemzők oszt.szám	várható érték μ	kovariancia mátrix $\underline{\Omega}$	inverz kovariancia mátrix $\underline{\Omega}^{-1}$	"a priori" valószínűség $P(d_i)$
1	$\begin{bmatrix} 10 \\ 12 \end{bmatrix}$	$\begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix}$	$\begin{bmatrix} 0,14 & 0 \\ 0 & 0,14 \end{bmatrix}$	0,1
2	$\begin{bmatrix} 26 \\ 35 \end{bmatrix}$	$\begin{bmatrix} 12 & -2,7 \\ -2,7 & 8,3 \end{bmatrix}$	$\begin{bmatrix} 0,09 & 0,03 \\ 0,03 & 0,13 \end{bmatrix}$	0,4
3	$\begin{bmatrix} 40 \\ 20 \end{bmatrix}$	$\begin{bmatrix} 8,3 & -2,7 \\ -2,7 & 12 \end{bmatrix}$	$\begin{bmatrix} 0,13 & 0,03 \\ 0,03 & 0,09 \end{bmatrix}$	0,2
4	$\begin{bmatrix} 48 \\ 45 \end{bmatrix}$	$\begin{bmatrix} 6 & 4,5 \\ 4,5 & 12 \end{bmatrix}$	$\begin{bmatrix} 0,23 & 0,09 \\ -0,09 & 0,12 \end{bmatrix}$	0,2
5	$\begin{bmatrix} 6 \\ 47 \end{bmatrix}$	$\begin{bmatrix} 5 & -1,6 \\ -1,6 & 4,5 \end{bmatrix}$	$\begin{bmatrix} 0,23 & 0,08 \\ 0,08 & 0,25 \end{bmatrix}$	0,1

Az 1. iteráció eredményét a 3-4/a. ábrán láthatjuk. Az így létrejött osztályozás szerint újraszámoltuk az osztályok paramétereit ($\mu, \underline{\Omega}, \underline{\Omega}^{-1}$), s ezek felhasználásával megismételtük az osztályozási eljárást. A 10. iterációs lépés után nyert eredményeket a 3-4/b. ábrán adjuk meg.

A könnyebb összehasonlítás kedvéért a 3-2. táblázatban összefoglaltuk a várhatóérték-vektorok (vagyis az osztályközpontok) alakulását, a 3-5. ábrán pedig az osztályozási mátrixokat. (A B_i mátrix b_{kl} eleme azt fejezi ki, hogy az eredetileg az i -edik osztályba tartozó elemek közül hány került be a k -adik osztályba; i az iteráció sorszáma. Így a k -adik sor elemszámát az $e_k = \sum_i b_{kl}$ összeg adja.)

Láthatjuk, hogy az eljárás során azonos osztályokba kerülnek a ténylegesen közeli elemek, míg a "kilógók"-ra vonatkozó döntések bizonytalanok.

Ha a 3-4. ábrán körülkerítjük az egyes osztályokat, a javulás - összehasonlítva az eredetivel - szembetűnő.



3-4. ábra
Osztályozás Bayes-féle döntéssel

Várhatóérték-vektorok (osztályközeppontok) alakulása
Bayes-döntési eljárás esetén

Osztály- szám	Kiindulási értékek	1. iteráció után	10. iteráció után
1	$\begin{bmatrix} 10 \\ 12 \end{bmatrix}$	$\begin{bmatrix} 8,47 \\ 14,47 \end{bmatrix}$	$\begin{bmatrix} 10,58 \\ 13,30 \end{bmatrix}$
2	$\begin{bmatrix} 26 \\ 35 \end{bmatrix}$	$\begin{bmatrix} 25,16 \\ 38,67 \end{bmatrix}$	$\begin{bmatrix} 25,30 \\ 38,47 \end{bmatrix}$
3	$\begin{bmatrix} 40 \\ 20 \end{bmatrix}$	$\begin{bmatrix} 43,82 \\ 15,91 \end{bmatrix}$	$\begin{bmatrix} 45,17 \\ 15,95 \end{bmatrix}$
4	$\begin{bmatrix} 48 \\ 45 \end{bmatrix}$	$\begin{bmatrix} 45,53 \\ 45,64 \end{bmatrix}$	$\begin{bmatrix} 43,79 \\ 45,07 \end{bmatrix}$
5	$\begin{bmatrix} 6 \\ 47 \end{bmatrix}$	$\begin{bmatrix} 6,93 \\ 44,68 \end{bmatrix}$	$\begin{bmatrix} 7,88 \\ 41,96 \end{bmatrix}$

$$\underline{\underline{B}}_0 = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 \\ 0 & 40 & 0 & 0 & 0 \\ 0 & 0 & 20 & 0 & 0 \\ 0 & 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 0 & 10 \end{bmatrix}; \quad \underline{\underline{B}}_1 = \begin{bmatrix} 10 & 2 & 0 & 0 & 0 \\ 0 & 34 & 6 & 2 & 0 \\ 0 & 0 & 14 & 0 & 0 \\ 0 & 2 & 0 & 18 & 0 \\ 0 & 2 & 0 & 0 & 10 \end{bmatrix};$$

$$\underline{\underline{B}}_{10} = \begin{bmatrix} 10 & 1 & 1 & 0 & 0 \\ 0 & 27 & 6 & 1 & 0 \\ 0 & 0 & 13 & 0 & 0 \\ 0 & 5 & 0 & 19 & 0 \\ 0 & 7 & 0 & 0 & 10 \end{bmatrix}$$

3-5. ábra
Bayes-döntési eljárás
osztályozási mátrixai

Megvizsgálunk még további 3 speciális esetet:

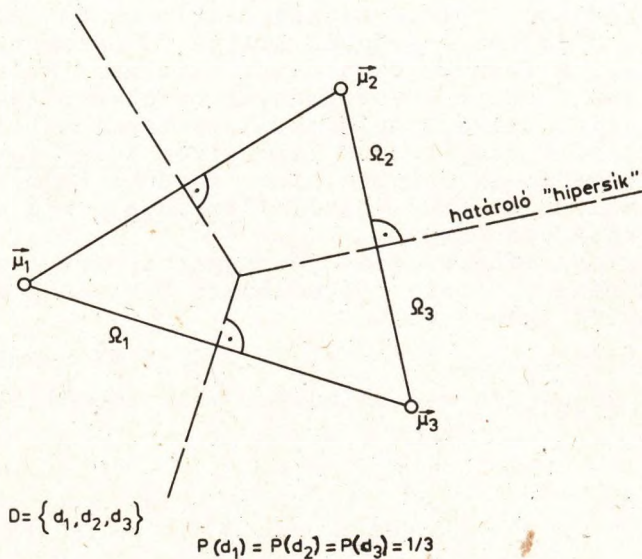
1. Legyen minden osztály kovarianciamátrixa egyenlő, jelöljük ezt $\underline{\Omega}$ -val. A (3-24) diszkriminánsfüggvény ekkor átírható az alábbi formába:

$$\begin{aligned}
 (3-25) \quad \tilde{g}_k(\vec{\xi}) &= c_k - (\vec{\xi} - \vec{\mu}_k)^T \underline{\Omega}^{-1} (\vec{\xi} - \vec{\mu}_k) = \\
 &= c_k - \vec{\xi}^T \underline{\Omega}^{-1} \vec{\xi} + 2\vec{\mu}_k^T \underline{\Omega}^{-1} \vec{\xi} - \vec{\mu}_k^T \underline{\Omega}^{-1} \vec{\mu}_k = \\
 &= c'_k + W_k \vec{\xi}
 \end{aligned}$$

ahol

$$W_k = 2\vec{\mu}_k^T \underline{\Omega}^{-1} \quad \text{és} \quad c'_k = c_k - \vec{\mu}_k^T \underline{\Omega}^{-1} \vec{\mu}_k.$$

A $\vec{\xi}^T \underline{\Omega}^{-1} \vec{\xi}$ kifejezés elhagyható, mivel osztálytól független konstans. Jól látható, hogy ebben az esetben már csak lineáris kifejezést kell számítanunk. (A 3-6. ábra egy példát mutat a diszkrimináns felületre, amely 2-dimenzióban egyenes.)



3-6. ábra
Normális eloszláshoz tartozó
lineáris döntésfüggvény

2. További egyszerűsítés lehetséges, ha a mérések korrelálatlanok és szórásuk minden koordinátában megegyezik, minden osztályban. Más szóval

$$(3-26) \quad \underline{\underline{\Omega}}_i = \sigma^2 \underline{\underline{E}}_n \quad (\underline{\underline{E}}_n \text{ az } n\text{-dimenziós egységmátrix}).$$

Ekkor a (3-25)-tel meghatározott diszkrimináns felületek (hipersíkok) merőlegesek lesznek az osztályok várhatóértékvektorait összekötő szakaszokra (ld. 3-6. ábra).

3. Befejezésül megemlítünk még egy érdekes speciális esetet. Tegyük fel, hogy semmi információnk nincs a döntések eloszlásáról, azaz $P(d_i) = 1/p$ egyenletes sűrűségfüggvényt célszerű választani. Ekkor a (3-17) képletben $P(d_i)$ -vel egyszerűsíthetünk, és a döntési szabály így módosul: egy objektumot akkor sorolunk az i -edik osztályba, ha ebben az esetben a legnagyobb a ξ mérési eredmény előfordulási valószínűsége. Ezt a döntési szabályt a legnagyobb valószínűség (maximum likelihood) döntésnek is nevezik, és igen gyakran összetévesztik az általános Bayes-féle döntéssel.

31.2.3. Becslési módszerek

Az osztályozás végrehajtásához elengedhetetlenül szükséges, hogy a döntésfüggvények alakja és paraméterei ismertek legyenek. A függvények alakját általában külső (pontosági, sebességi stb.) követelmények befolyásolják. A paramétereket (mint például átlagok, kovarianciamátrixok stb. a Bayes-döntésnél) vagy tudjuk valamilyen külső forrásból (mérésből), ebben az esetben nincs további tennivalónk; vagy pedig valamilyen mintaadatokból ("tananyag"-ból kell) becsléssel meghatározzunk.

A becslési módszereknek jó magyar nyelvű irodalmuk van, ezért csak néhány, fontosabb módszert tekintünk át röviden, és a teljesség igénye nélkül.

Tegyük fel, hogy a $Z = \{\xi_1, \dots, \xi_t\} \in \mathbb{R}^n$ halmaz normális eloszlású, független valószínűségi változókból áll. Ekkor a

$$\hat{\underline{\underline{\mu}}} = \frac{1}{t} \sum_{i=1}^t \vec{\xi}_i$$

$$\hat{\underline{\underline{\Omega}}} = \frac{1}{t} \sum_{i=1}^t (\vec{\xi}_i - \hat{\underline{\underline{\mu}}}_i) (\vec{\xi}_i - \hat{\underline{\underline{\mu}}}_i)^T \quad \text{és} \quad \tilde{\underline{\underline{\Omega}}} = \frac{t}{t-1} \hat{\underline{\underline{\Omega}}}$$

képleteket használhatjuk a várhatóértékek és a szórások becslésére. A $\hat{\underline{\underline{\mu}}}$ és az $\hat{\underline{\underline{\Omega}}}$ legnagyobb valószínűségű (maximum likelihood) becslései $\tilde{\underline{\underline{\mu}}}$ -nak és $\tilde{\underline{\underline{\Omega}}}$ -nak: $\tilde{\underline{\underline{\Omega}}}$ pedig a korrigált empirikus kovarianciamátrix.

Bebizonyítható, hogy $M(\hat{\mu}) = \bar{\mu}$ és $M(\hat{\Omega}) = \underline{\Omega}$ a lehető legjobb becslések, pontosabbakat (kisebb szórásúakat) nem lehet megadni.

Nehezebb a helyzet, ha nem ismerjük a ξ_1 változók eloszlásfüggvényét. Próbáljuk meg felírni az $s(\vec{\xi})$ sűrűségfüggvényt ekkor Z alapján!

Egyik lehetőség az empirikus sűrűségfüggvény meghatározása. Az \mathbb{R}^n teret felosztjuk (nem szükségképpen egyenletesen) R_k ($k=1, \dots, m$) régiókra. Legyen V_k az R_k régió térfogata, és $N(R_k) = |\{\vec{\xi}_i \in Z \mid \vec{\xi}_i \in R_k\}|$. Ekkor a közelítő sűrűségfüggvényt

$$\hat{s}(\vec{\xi}) = \frac{N(R_k)}{mV_k}, \quad \forall \vec{\xi} \in R_k$$

alakban írhatjuk fel. A módszer nem problémamentes: R_k , k stb. meghatározására nincs semmilyen támpontunk.

Másik lehetőség a Parzen-Rosenblatt féle eljárás alkalmazása. Legyen $h = \mathbb{R}^n \rightarrow \mathbb{R}$ olyan szimmetrikus sűrűségfüggvény, amely kétszer korlátosan differenciálható, és

$$\int_{\mathbb{R}^n} h(\vec{\xi}) d\vec{\xi} = 1;$$

$$\int_{\mathbb{R}^n} \xi_k h(\vec{\xi}) d\vec{\xi} = 0; \quad (k = 1, \dots, n)$$

$$\int_{\mathbb{R}^n} \|\vec{\xi}\|^2 h(\vec{\xi}) d\vec{\xi} < +\infty;$$

akkor az

$$(3-27) \quad \hat{s}(\vec{\xi}) = \frac{1}{ta_t^n} \sum_{i=1}^t h\left(\frac{\vec{\xi} - \vec{\xi}_i}{a_t}\right)$$

becslés (ahol az $a_t > 0$ együtthatóra $\lim_{t \rightarrow \infty} a_t = 0$) jól közelíti $s(\vec{\xi})$ -t.

Az így kapott becslés kisebb hibát ad, mint az empirikus sűrűségfüggvény, de számítása nehezekebb.

A Parzen-Rosenblatt módszer speciális változata lehet a sorfejtés. Tegyük fel, hogy léteznek olyan

$$\varphi_k(\vec{\xi}) : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \psi_k(\vec{\xi}) : \mathbb{R}^n \rightarrow \mathbb{R}, \quad (k=1, \dots, m)$$

függvények, amelyekre teljesül a

$$h\left(\frac{\vec{\xi} - \vec{\xi}_i}{a_t}\right) = \sum_{k=1}^m b_k \varphi_k(\vec{\xi}) \psi_k(\vec{\xi}_i)$$

összefüggés. Ha (3-27)-be behelyettesítünk, a Parzen-Rosenblatt féle becslést az

$$\hat{s}(\vec{\xi}) \approx \sum_{k=1}^m c_k \varphi_k(\vec{\xi}),$$

alakban lehet írni, ami φ_k -ban lineáris; ahol a

$$c_k = \frac{b_k}{t a_t^n} \sum_{\ell=1}^t \psi_k(\vec{\xi}_\ell)$$

súlytényezőket csak egyszer kell kiszámítani.

31.3. TÁVOLSÁGMÉRÉSEN ALAPULÓ MÓDSZEREK

Az előző alfejezettel szemben most abból indulunk ki, hogy csak részlegesen ismerjük az osztályok valószínűségi jellemzőit. Jelöljük a továbbiakban $\vec{x}^T = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$ -nel az azonosítandó objektumra mért sajátosságvektort. Most nem tudjuk, hogy az osztályok milyen gyakorisággal fordulnak elő, sem azt, hogy a mérési eredmények milyen eloszlást követnek.

A döntéseket a mért sajátosságvektorok és bizonyos, előre megadott mintavektorok távolságának meghatározásával fogjuk hozni. (Mindkét vektor az n -dimenziós tér helyvektora, így a távolságukon a végpontjaik távolságát értjük, ami egyenlő a vektori különbségükkel.)

Erre a célra különféle távolságfüggvényeket szoktak használni, mi azonban az alábbiakban csak a 13.4.3. pontban megismert (8)- és euklidészi-távolságfüggvényt alkalmazzuk. Emlékeztetünk rá, hogy minden távolságfüggvénynek ki kell elégítenie az ott felsorolt 3, illetve 4 követelményt (nem-negatív definit, szimmetrikus és kielégíti a háromszögegyenlőséget, ill. még reguláris); ilyenek konstruálásával azonban nem foglalkozunk.

31.3.1. Legközelebbi szomszéd módszerek

Induljunk ki abból, hogy rendelkezésünkre áll egy

$$(3-28) \quad T = \{ \vec{z}_m, \nu_m \mid \vec{x}_m \in \mathbb{R}^n, \nu_m \in D, m=1, 2, \dots, t \}$$

"tananyag", amely döntési minták halmaza. Minden $\{ \vec{z}_m, \nu_m \}$ tanítópár egy döntési példának felel meg, ahol a $\vec{z}_m \in \mathbb{R}^n$ mintavektor (feltehetően) egy ismert objektumra vonatkozó korábbi mérésből származik, a ν_m pedig a neki megfelelő osztályt jelképezi. Feltesszük, hogy a megadott tanítópárok teljesen függetlenek és azonos eloszlásúak. (Meggjegyezzük, hogy - bár itt és a későbbiekben is felteszünk, hogy a "tananyag" előre ismert - a gyakorlatban a tanítópárok kijelölése, a "tanulás" gyakran összefonódik az osztályozással, és a jó "tananyag" csak fokozatosan alakul ki.)

A legközelebbi szomszéd (a továbbiakban: LKS) megkeresésén alapuló döntési szabályt ezek után a következő módon fogalmazzuk meg:

Egy mérési eredmény (sajátságvektor) alapján akkor sorolunk egy objektumot az i -edik osztályba, vagyis $\sigma_{LKS}(\vec{z}) = d_i$, ha

$$(3-29) \quad \varphi(\vec{z}, \vec{z}') \leq \varphi(\vec{z}, \vec{z}_m) \quad (m=1, \dots, p) \quad \text{és a} \\ (\vec{z}', \nu') \text{ pár esetén } \nu' = d_i;$$

ahol $\varphi: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ tetszőleges távolságfüggvény.

A megadott euklidészi, négyes-(manhattan) és nyolcas-távolságfogalmon kívül megemlítjük még a

$$\varphi_Q(\vec{z}, \vec{v}) = (\vec{z} - \vec{v})^T \underline{Q}^{-1} (\vec{z} - \vec{v}),$$

kifejezéssel definiált Mahalanobis-távolságot, ahol \underline{Q} n -edrendű pozitív definit négyzetes mátrix.

A (3-29) szerinti \vec{z}' -t a \vec{z} legközelebbi szomszédjának nevezzük. Ha a (3-29) feltétel több osztályra teljesül, az objektumot tetszés szerint bármelyikbe besorolhatjuk.

Elemezzük röviden a módszer hibaválószerűségét. Döntésünk akkor lesz hibás, ha a \vec{z} minta a d_i osztályba tartozik, de a (3-29) szabály értelmében a d_k ($i \neq k$) osztályba soroltuk.

A hiba valószínűsége nyilvánvalóan függ a "tananyag jóságától". Bizonyítható, hogy a tanítópárok alapján hozott döntések hibájának várható értéke konvergens, ha a tanítópárok számát növeljük; ezt a határértéket tekintjük az eljárás veszteségének:

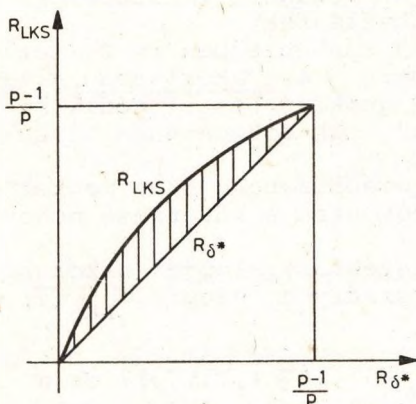
$$(3-30) \quad R_{LKS} = M \left(1 - \sum_{i=1}^p P^2(d_i | \vec{z}) \right).$$

Erre a veszteségre fennáll az

$$(3-31) \quad R_{\delta^*} \leq R_{LKS} \leq R_{\delta^*} \left(2 - \frac{p}{p-1} R_{\delta^*} \right)$$

egyenlőtlenség, vagyis az LKS módszerrel hozott döntés hibája (aszimptotikusan) legfeljebb a Bayes-döntés hibájának

kétszerese (ld. 3-7. ábra). Figyelemre méltó, hogy az optimális döntéshez szükséges ismeretnek legalább felét a legközelebbi szomszéd hordozza!



3-7. ábra

A Bayes-féle és az LKS-döntés hibafüggvénye

Az eljárás természetes általánosítása a (k, ℓ) legközelebbi szomszéd módszer, amelyet a következő módon definiálunk: a \vec{z} sajáttságvektor alapján akkor soroljuk az objektumot a d_k osztályba, ha a \vec{z} méréshez legközelebb eső k db tanítópár közül legalább ℓ tartozik a d_k osztályhoz. Ha egyetlen i -re sem teljesül ez az eset, akkor nem hozunk döntést.

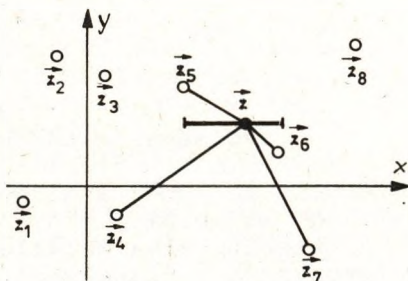
Részletesebb elemzés nélkül megjegyezzük, hogy a (k, ℓ) módszer gyengébb eredményt is adhat, mint a (3-29)-cel definiált LKS eljárás.

31.3.1.1. Gyors LKS-kereső algoritmusok

A δ_{LKS} döntésfüggvényhez igen könnyű diszkriminánsfüggvényeket megadni: $g_i(\vec{z})$ -ként egyszerűen a \vec{z} sajáttságvektor és δ_i döntéshez tartozó mintavektorok távolságának minimumát kell vennünk. Ezzel visszavezettük a döntést távolságok kiszámítására. A szükséges műveletszám (4)- és (8)-távolság esetén lineárisan függ a dimenziószámtól, de az euklidészi távolságfüggvényt is vissza lehet vezetni lineáris kifejezésekre. Ez lényeges nyereséget jelent a Bayes-módszerhez képest. (Vegyük azért észre, hogy a (3-25) összefüggésben is a várhatóérték-vektortól való euklidészi-távolságokat kell kiszámítani, és erre lineáris diszkrimináns-függvényeket tudunk megadni. Ez ismét arra utal, hogy szoros kapcsolat van a Bayes- és az LKS-módszer között.)

Első pillantásra úgy tűnik, hogy ki kell számítani a mért \vec{z} sajátsgvektorok távolságát minden mintavektortól, vagyis a számolási igény a "tananyag" méretével egyenesen arányos. Valójában azonban a helyzet ennél jobb. Szellemes eljárások léteznek, amelyek bizonyos metrikák speciális tulajdonságait kihasználva már néhány mintavektor vizsgálata után megtalálják a legközelebbit. Ennek természetesen ára van, nevezetesen arról van szó, hogy a tananyagot struktúrálni kell, és ekkor már hatékony keresést lehet végrehajtani. Az alábbiakban két módszer alap gondolatát ismertetjük:

1. A tananyagban szereplő vektorokat sorbarendezzük az \mathbb{R}^n tér valamely alkalmasan választott egyenesé (például egyik koordinátatengelye) mentén. Ezután csak a mért \vec{z} vektor és az adott egyenesen hozzá legközelebb fekvő mintavektorok távolságát számítjuk ki. Az eljárás akkor ér véget, amikor a rendező egyenes menti távolság nagyobb lesz, mint az addigi minimális távolság. (A módszert kétdimenziós esetre a 3-8. ábrán szemléltetjük. Itt a rendező egyenes az X-tengely, és az x-irányú távolságokat vastag vonallal jelöljük. Mivel



3-8. ábra
Távolságszámítás rendezett "tananyag" esetén

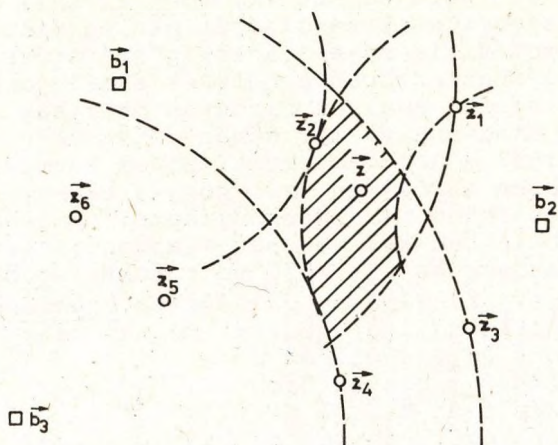
$$\rho_x(\vec{z}, \vec{z}_6) < \rho_x(\vec{z}, \vec{z}_5),$$

ezért a döntést a \vec{z}_5 és a \vec{z}_6

mintavektor y-irányú távolságának vizsgálatával már meg tudjuk hozni.) A módszer viszonylag kis előfeldolgozást, de jelentős tárkapacitást igényel, s a rendező egyenes rossz megválasztása esetén a távolságszámítások száma nagy lehet.

2. Bázisvektorok módszere. A mintavektorok közül kijelölünk 3 bázisvektort és kiszámítjuk ezek távolságát az összes többi mintavektortól. A mintavektorokat három, a kapott távolságok szerint növekvő sorba rendezzük. A döntéshez kiszámítjuk a konkrét \vec{z} vektor és a 3 bázisvektor távolságát. Ezek alapján a listáinkban keresve találhatjuk meg a legközelebbi szomszéd(ka)t. (A kereséshez csak a bázisvektorokat és a távolságlistákat kell tárolni, a mintavektorokat nem, ezért a módszer mérsékelt tárigényű.) (Ld. 3-9. ábra. Itt - ugyancsak kétdimenziós esetben - a bázisvektorokat \vec{b}_1, \vec{b}_2 , ill. \vec{b}_3 jelöli; a koncentrikus körök a listákban tárolt növekvő távolságokat jelzik. A \vec{z} -hez legközelebbi mintavektorokat az egyre táguló körgyűrűk \vec{z} -t is tartalmazó metszete jelöli ki. Az ábrán ez a bevonalkázott terület és a \vec{z}_2 .)

Mivel döntésenként csak 3 távolságszámítást kell elvégezni, és a keresés a listákban gyakorlatilag elhanyagolható



3-9. ábra
Távolságszámítás bázisvektorok segítségével

időt vesz igénybe, a döntési idő majdnem független a mintavektorok számától.

Összefoglalva megállapíthatjuk, hogy az LKS módszerek hatékonysága a "tananyag" növelésével (aszimptotikusan) nem rossz, annak ellenére, hogy minimális kezdeti ismeretből indulnak. Ezenkívül számításigényük is kedvezően alacsony (lineáris távolságszámítás, speciális gyors algoritmusok). Ebből következik, hogy még részletesebb információk birtokában is érdemes lehet az LKS módszert használni, ha a sajátságvektorok dimenziószáma magas és/vagy sok osztályunk van.

31.3.2. A téglatest-módszer

Jellegében az LKS módszerre hasonlító, mintákon alapuló, egyszerű döntési eljáráshoz juthatunk a téglatest (box)-módszer alapján. A "tananyagot" ezúttal a

$$(3-32) \quad T = \{ \vec{z}_m, \psi_m, \vec{a}_m \mid \vec{z}_m \in \mathbb{R}^n, \psi_m \in D, m = 1, 2, \dots, t \}$$

mintahármasok jelentik, ahol az

$$\vec{a}_m = (a_m^{(1)}, a_m^{(2)}, \dots, a_m^{(n)})$$

olyan n -dimenziós, $2a_n^{(r)}$ oldalhosszúságú (a koordinátatengelyekkel párhuzamos oldalú téglatestet határoz meg, amelynek középpontja illeszkedik a \vec{z}_m -re.

Az \vec{a}_m eltérésvektor lehet tetszőleges is, de gyakran megegyezik az m -edik osztályhoz tartozó szórásvektorral (vagy többszörösével).

Ezek alapján akkor hozzuk a $\sigma_{TT}(\vec{z}) = d_i$ döntést, ha van olyan $m \in \{1, 2, \dots, t\}$, amelyre teljesülnek a

$$(3-33) \quad \begin{cases} z_m^{(\ell)} - a_m^{(\ell)} \leq z^{(\ell)} \leq z_m^{(\ell)} + a_m^{(\ell)} & (\ell=1, 2, \dots, n) \\ v_m = d_i \end{cases}$$

relációk. Más szóval: mért \vec{z} sajáttságvektor a \vec{z}_m mintavektorra illeszkedő $2a_m^{(r)}$ oldalhosszúságú (n -dimenziós) téglatestbe esik.

Előfordulhat, hogy (3-33) egyetlen m -re sem teljesül; ill. - a paraméterek megválasztásától függően - esetleg több mintavektorra is fennáll, de ezek különböző döntésre vezetnek. Ekkor valamilyen egyéb módon hozunk döntést, pl. többségi elven vagy véletlenszerűen stb.

A döntés számításgénye a tanítóhármak számával egyenesen arányos, hiszen minden téglatestet meg kell vizsgálnunk. Ezen belül azonban már csak összehasonlításokat teszünk, amit nagyon gyorsan és egyszerűen lehet végrehajtani.

31.3.3. Lineáris diszkriminánsfüggvények

31.3.3.1. Döntés diszkriminánsfüggvények alapján

Diszkriminánsfüggvények szerinti döntésről akkor beszélünk, ha a σ döntésfüggvényt valamilyen

$$g_i : \mathbb{R}^n \rightarrow \mathbb{R} \quad (i=1, 2, \dots, h)$$

valós értékű függvények definiálják úgy, hogy a $\sigma(\vec{z}) = d_i$ döntést akkor és csak akkor hozzuk, ha

$$g_i(\vec{z}) > g_m(\vec{z}) \quad (m=1, 2, \dots, h).$$

Az előzőekben (ld. 31.1.3. pont) beszéltünk már a diszkriminánsfüggvényekről és tulajdonságaikról. E fejezet keretében a csak a lineáris esetre szorítkozunk, azaz, ha a fenti függvények felírhatók

$$(3-34) \quad g_i(\vec{z}) = \vec{a}_i^T \vec{z} + b_i \quad (i=1, 2, \dots, h)$$

alakban, ahol $\vec{a}_i \in \mathbb{R}^n$ adott súlyvektorok, $b_i \in \mathbb{R}$ pedig adott

valós küszöbsúlyok. Egy lineáris diszkriminánsfüggvény-rendszer esetén a létrejövő döntési régiókat hipersíkok határolják, amelyeket az

$$(\vec{a}_i - \vec{a}_m)^T \vec{z} + (b_i - b_m) = 0 \quad (i, m=1, 2, \dots, h; i \neq m)$$

egyenletek definiálnak. Így az $(\vec{a}_i - \vec{a}_m)$ különbségvektor merőleges a döntési felületre; továbbá egy tetszőleges \vec{z} vektornak a döntési felülettől mért (előjeles) távolsága

$$g(\vec{z}) = (g_i(\vec{z}) - g_m(\vec{z})) / \|\vec{a}_i - \vec{a}_m\|.$$

Kétosztályos esetben a probléma egyszerűsödik, a döntést a

$$D(\vec{z}) = g_1(\vec{z}) - g_2(\vec{z})$$

diszkriminánsfüggvénnyel az alábbiak szerint definiáljuk:

$$\sigma(\vec{z}) = \begin{cases} d_1, & \text{ha } D(\vec{z}) > 0; \\ d_2, & \text{ha } D(\vec{z}) < 0. \end{cases}$$

31.3.3.2. Lineáris szétválaszthatóság

Egy adott mintavektor-rendszerrel, amelynek ismert az osztályozása, azt mondjuk, hogy lineárisan szeparálható (szétválasztható), ha létezik olyan lineáris diszkriminánsfüggvény rendszer, amelyre teljesül a (3-34) feltétel. Kétosztályos esetben - felhasználva a korábbi egyszerűsítéseket - a $\vec{w}_1, \vec{w}_2, \dots, \vec{w}_t$ mintavektorok akkor alkotnak szétválasztható rendszert, ha létezik olyan lineáris diszkriminánsfüggvény, amelyre

$$(3-35) \quad \begin{aligned} g(\vec{w}_j) &> 0, & \text{ha } \sigma(\vec{w}_j) &= d_1; \\ g(\vec{w}_j) &< 0, & \text{ha } \sigma(\vec{w}_j) &= d_2; \end{aligned} \quad (j = 1, 2, \dots, t).$$

Ha a második osztályhoz tartozó minták helyett a \rightarrow negáltjukat tekintjük, akkor a (3-35) második része $g(-\vec{w}_j) > 0$ feltétellel helyettesíthető. Ha tehát az eredeti minták helyett a

$$(3-36) \quad \vec{v}_j = \begin{cases} \vec{w}_j, & \text{ha } \sigma(\vec{w}_j) = d_1; \\ -\vec{w}_j, & \text{ha } \sigma(\vec{w}_j) = d_2; \end{cases} \quad (j=1, 2, \dots, t)$$

vektorokkal dolgozunk, akkor a (3-35) definíció helyett a $g(\vec{v}_j) > 0$ feltételt minden j -re megkövetelhetjük. A g lineáris diszkriminánsfüggvény (3-34) szerint előáll az $\vec{a}^T \vec{z} - b$ alakban, s ezzel a (3-36) feltétel

$$\vec{a}^T \vec{v}_j \geq b \quad (j=1,2,\dots,t)$$

alakú lesz. Az \vec{a} vektort szeparáló- vagy megoldásvektornak, a b számot küszöbnek is nevezik.

A megoldásvektor (ha egyáltalán létezik) nem egyértelmű; általában valamilyen szempontból legjobb megoldásvektort szokás keresni. A "jóságot" legtöbbször valamilyen célfüggvény méri. A következő részben bemutatunk néhány olyan módszert, amelyek egy-egy célfüggvényt optimalizálnak, vagyis az adott szempontból "legjobb" megoldásvektort adják.

31.3.3.3. Lineáris diszkriminánsfüggvények előállítása

Lineáris diszkriminánsfüggvények előállítására számos módszer létezik, itt ezek közül csak néhányra tudunk kitérni. A módszerek alapját általában a gradiens-módszer képezi.

Legyen adott a $J(\vec{a})$ kritériumfüggvény, keressük azt az \vec{a} megoldásvektort, amely kielégíti az

$$\vec{a}^T \vec{v}_j > 0 \quad (j=1,2,\dots,t)$$

egyenlőtlenségeket, és minimalizálja a kritériumfüggvényt. Ha a J függvény minden változója szerinti parciális differenciálhányadosa létezik és folytonos, jelölje $\nabla J(\vec{a})$ a gradiensvektort. Egy tetszőleges $\vec{a}^{(1)}$ kiindulási vektorral keressük a megoldást az alábbi rekurzív algoritmus szerint:

$$\vec{a}^{(k+1)} = \vec{a}^{(k)} - \alpha_k \nabla J(\vec{a}^{(k)}) \quad (k=1,2,\dots)$$

ahol az α_k együtthatók adott pozitív számok, amelyek a konvergenciát befolyásolják.

A gradiens-módszeren alapul a perceptron-módszer is. A perceptron-kritérium függvény a hibásan osztályozott mintapontoknak a szeparáló hipersíktól mért távolságát összegezi:

$$J_p(\vec{a}) = \sum_{\vec{a}^T \vec{v} < 0} (-\vec{a}^T \vec{v})$$

A célfüggvény minimalizálásával azt az elválasztó hipersíkot keressük, amelytől az esetleges "rosszul" osztályozott mintapontok a lehető legkisebb távolságra vannak. A gradiens módszer alapján a következő eljárás adódik:

$$\vec{a}^{(k+1)} = \vec{a}^{(k)} + \sum_{\vec{a}^{(k)\top} \vec{v} < 0} \vec{v} \quad (k=1, 2, \dots).$$

A rekurzió során tulajdonképpen annyi történik, hogy a megoldásvektor új közelítését a korábbiából a hibásan osztályozott mintavektorok hozzávételével állítjuk elő. Ha létezik megoldásvektor, véges számú lépésben megkapjuk; ha nem, az eljárás divergál. A perceptron-módszer általánosítható többosztályos esetre is.

A Kashyap-Ho módszer esetén a minimalizálandó célfüggvény a

$$J_s(\vec{a}, \vec{b}) = \|\underline{V} \vec{a} - \vec{b}\|^2,$$

ahol $\underline{V} = [\vec{v}_1, \vec{v}_2, \dots, \vec{v}_t]^T$ a (3-36) szerinti módosított mintákból alkotott mátrix, \vec{a} és \vec{b} súlyvektorok. A J_s célfüggvény a mintavektorok szeparáló hipersíktól mért távolságának négyzetét összegzi. A J_s gradiensvektorai:

$$\nabla_a J_s = 2 \underline{V}^T (\underline{V} \vec{a} - \vec{b}), \text{ illetve}$$

$$\nabla_b J_s = -2 (\underline{V} \vec{a} - \vec{b}).$$

A \vec{b} egy értéke mellett a J_s az \vec{a} -ra nézve akkor minimális, ha

$$\vec{a} = \underline{V}^{-1} \vec{b} = (\underline{V}^T \underline{V})^{-1} \underline{V}^T \vec{b}$$

ahol \underline{V}^{-1} a $(\underline{V}^T \underline{V})^{-1} \underline{V}^T$ általánosított inverz mátrixot jelöli.

Ekkor az \vec{a} szerinti gradiens 0, \vec{b} -ra pedig a következő rekurzió adódik:

$$\beta^{(1)} > 0;$$

$$\vec{b}^{(k+1)} = \vec{b}^{(k)} + 2 \beta \vec{e}^+(k);$$

ahol

$$\vec{e}^+(k) = - (|\vec{e}^-(k)| + \vec{e}^-(k));$$

$$\vec{e}^-(k) = \underline{V} \vec{a}^{(k)} - \vec{b}^{(k)};$$

$$\vec{a}^{(k)} = \underline{V}^{-1} \vec{b}^{(k)};$$

Ha a minták szeparálhatók, $0 < \beta < 1$ esetén a módszer véges számú lépésben konvergál.

31.3.4. Egyéb módszerek

A fentiekben ismertetett módszerektől jelentősen különbözik a szekvenciális és a hierarchikus módszer.

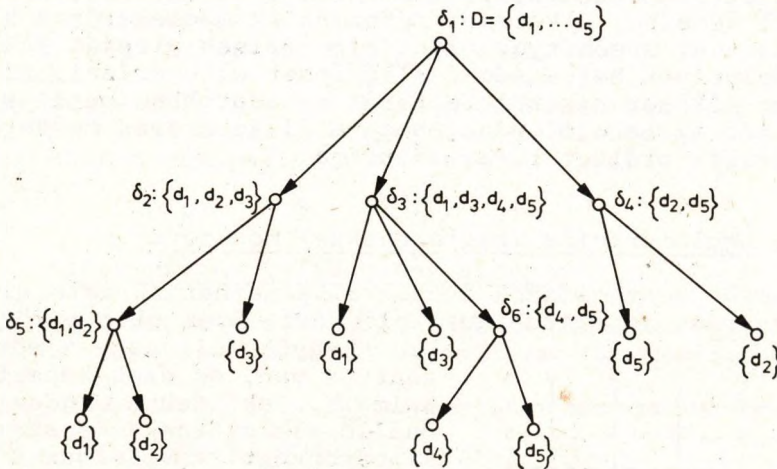
A szekvenciális döntés esetén (kissé leegyszerűsítve) a sajátságvektort koordinátáinként vizsgáljuk, és az aktuális koordináta értékétől függően, vagy

- besoroljuk a mérést az egyik osztályba, vagy
- vesszük a következő koordinátát.

A módszer különösen nagy dimenziószám esetén hasznos, hiszen sokszor már töredékvektor vizsgálata esetén is döntésre juthatunk. Fontosabb szempont lehet az, hogy maguknak a méréseknek is van költségük. Elképzelhető, hogy a méréseket az osztályozás során végezzük, és így esetleg néhány mérést megtakaríthatunk.

A hierarchikus eljárás egy döntési fa konstruálását jelenti, mint azt a 3-10. ábra is mutatja. A fa csúcspontjai D részhalmazaival vannak megcímkézve, és döntésfüggvényeket is rendelünk hozzájuk úgy, hogy

- az egy csúcspontból kiinduló leszármazott döntések uniója megegyezik a csúcsponthoz tartozó döntéssel;
 - a levelekhez (olyan csúcspontok, amelyeknek nincs kiinduló élük) egyetlen döntés tartozik.
- (Egy adott csúcsban az \mathbb{R}^n alteréből $D' \in D$ részhalmazba képező döntésfüggvényt tekintünk. A D' halmazok metszete nem szükségszerűen üres.)



3-10. ábra
Hierarchikus döntési fa

A döntési eljárás ezek alapján a következő.

A \vec{z} sajátságvektort a fa gyökerére helyezzük és az ott érvényes döntési függvény alapján valamelyik irányba továbblépünk. Itt az új döntésfüggvényt alkalmazzuk, és így tovább

addig, míg végeredményre nem jutunk. Egy mérés osztályozása során így több döntésfüggvényt is ki kell értékelnünk, de ezek általában csak a sajáttságvektor egy-egy részletét használják. Összességében ezáltal gyors eljáráshoz jutunk.

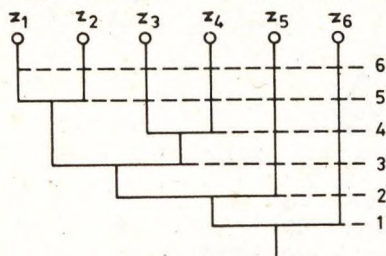
31.4. KLASZTEREZÉS

A klaszterezés (clustering) tanító nélküli osztályozás. Akkor alkalmazzuk, ha még tanítópárok sem állnak rendelkezésünkre az osztálybasoroláshoz, hanem a sajáttságvektorokból kell kitalálnunk, hogy milyen csoportosítást lehet végezni. (Ebben az értelemben a klaszterezés az osztályozás megfordításának is tekinthető: eddig ismert osztályokhoz kerestük meg a megfelelő méréseket, adatokat; most pedig az adatokhoz keressük megfelelő osztályokat.)

Mielőtt még a részletekbe mennénk, megjegyezzük, hogy egzakt klaszterfogalmat nem adunk. Jelentése eredetileg csomó, nyaláb. Szemléletesen úgy képzeljük, hogy az azonos osztályba (klaszterbe) kerülő adatok bizonyos értelemben erősen hasonlítanak egymásra, viszont a többitől különböznek. A megfogalmazás lazaságából rögtön az is következik, hogy ugyanazon adathalmaznak akár több klaszterezése is lehetséges. Nehéz felismerni, hogy ezek közül melyik megoldás felel meg leginkább a feladatunknak, sőt egyáltalán létezik-e megoldás. Nem lehet eléggé hangsúlyozni, hogy a klaszterezési eljárások eredményeit kritikusan, megfelelő óvatossággal kell kezelni. A kezdeti információk meglehetősen hiányosak, ezért az eredmények olyan hipotézisek alapján születnek, amelyeknek helyességét alig lehet ellenőrizni. Ennek ellenére sokszor sikerül feltárni az adatokban rejlő struktúrát, bár az is előfordul, hogy a klaszterezés mesterséges osztályozást erőltet rá az adatokra.

31.4.1. Agglomerációs (hierarchikus) módszerek

Kisebb adathalmazok esetén alkalmazhatjuk valamelyik agglomerációs (hierarchikus) eljárást. Ezek elve a következő:



3-11. ábra
Agglomerációs
klaszterezési eljárás

Tegyük fel, hogy N mérési adatunk van, és ezek képezik az A halmazt. Tekintsünk minden adatot önálló klaszternek, és számítsuk ki páronkénti távolságukat. Vonjuk össze a két legközelebbit és számítsuk most ki a maradék $N-1$ klaszter páronkénti távolságát. Ismét összevonva a két legközelebbit, az eljárást addig ismételjük, amíg a kívánt számú klasztert el nem érjük.

Az eljárást a 3-11. ábrán követhetjük nyomon.

Az \mathbb{R}^n -beli diszjunkt, véges Z_1, Z_2 halmaz távolságának kiszámítására az alábbi képleteket használhatjuk:

$$(3-37/a) \quad \varphi_{\min}(Z_1, Z_2) = \min_{\vec{z} \in Z_1, \vec{v} \in Z_2} (\varphi(\vec{z}, \vec{v}));$$

$$(3-37/b) \quad \varphi_{\max}(Z_1, Z_2) = \max_{\vec{z} \in Z_1, \vec{v} \in Z_2} (\varphi(\vec{z}, \vec{v}));$$

$$(3-37/c) \quad \varphi_{\text{átl}}(Z_1, Z_2) = \frac{1}{n_1 n_2} \sum_{\vec{z} \in Z_1} \sum_{\vec{v} \in Z_2} \varphi(\vec{z}, \vec{v});$$

$$(3-37/d) \quad \varphi_{\text{centr}}(Z_1, Z_2) = \varphi(\vec{\mu}_1, \vec{\mu}_2)$$

ahol:

$$n_1 = |Z_1|; \quad n_2 = |Z_2|; \quad \vec{\mu}_1 = \frac{1}{n_1} \sum_{\vec{z} \in Z_1} \vec{z}; \quad \vec{\mu}_2 = \frac{1}{n_2} \sum_{\vec{v} \in Z_2} \vec{v}$$

és φ tetszőleges távolságfüggvény.

A távolságfüggvény megválasztásától függően a módszernek más elnevezései is használatosak:

- a (3-37/a) definíció alkalmazása esetén legközelebbi szomszéd eljárásról;

- a (3-37/b) esetén minimax eljárásról beszélünk.

Megjegyezzük, hogy ezeknek a módszereknek igen nagy a tárigényük, mivel a klaszterek közti távolságokat meg kell őrizni; ezek száma pedig négyzetesen arányos a klaszterek számával.

31.4.2. Klaszterezés célfüggvény alapján

A következőkben olyan eljárásokat ismertetünk, amelyekben az osztályozás "jóságát" valamilyen célfüggvénnyel jellemezzük.

Legyen $Z = \{\vec{z}_1, \dots, \vec{z}_m\}$ a mérések halmaza és képezzük egy tetszőleges particionálását úgy, hogy

$$\bigcup_{i=1}^p Z_i = Z \quad \text{és} \quad Z_i \cap Z_l = \emptyset \quad \text{ha } i \neq l$$

($i, l = 1, \dots, p$).

Jelölje a mérések átlagát

$$\bar{\mu}_i = \frac{1}{|Z_i|} \sum_{\vec{z} \in Z_i} \vec{z}_k$$

ekkor a

$$(3-38) \quad W_{\ell knh} = \sum_{i=1}^p \sum_{\vec{z} \in Z_i} \varphi(\vec{z}_k, \bar{\mu}_i)^2$$

kifejezés a klaszterközéppontok és a klaszterelemek távolságát fogja jellemezni (ℓknh a legkisebb négyzetes hibát jelent). Írjuk (3-38)-at

$$W_{\ell knh} = \sum_{i=1}^p n_i u_i, \quad u_i = \frac{1}{n_i} \sum_{\vec{z}, \vec{v} \in Z_i} \varphi(\vec{z}, \vec{v})^2$$

alakba, ami alapján számos hasonló célfüggvény definiálható u_i alkalmas behelyettesítésével.

Kissé más jellegű eredményt kapunk, ha bevezetjük az empirikus kovariancia mátrixot az

$$\underline{\Omega}_i = \sum_{\vec{z} \in Z_i} (\vec{z} - \bar{\mu}_i)(\vec{z} - \bar{\mu}_i)^T$$

összefüggéssel és a célfüggvényt a

$$(3-39) \quad W_d = \left| \sum_{i=1}^p \underline{\Omega}_i \right|$$

kifejezéssel adjuk meg.

További összetettebb, különböző szempontból előnyös (pl. az adatok lineáris transzformációjával szemben invariáns) célfüggvényeket is lehetne definiálni, ezekkel azonban nem foglalkozunk.

A fentieket felhasználva egy klaszterezést (vagyis az adathalmaz particionálását) akkor tekintünk optimálisnak, ha minimalizálja az előírt célfüggvényt.

Az eddigi célfüggvényekben p értéke adott, azaz tudjuk előre, hány klasztert keresünk. Ha p nem lenne ismert, akkor lehetne pl.

$$W' = W + c \cdot p$$

alakú célfüggvényt választani, ahol c konstans és W a (3-38)-ban vagy a (3-39)-ben definiált célfüggvény.

31.4.3. Iteratív módszerek

31.4.3.1. A négyzetes hiba minimalizálása

Először a 31.4.2. pontban bevezetett négyzetes hiba célfüggvény optimalizálására mutatunk egy módszert. Megtartva a jelöléseket, írjuk fel a célfüggvényt a

$$W_{1knh} = \sum_{i=1}^p J_i = \sum_{i=1}^p \sum_{z_j \in Z_i} \|\vec{z}_j - \vec{\mu}_i\|^2$$

alakban. Vizsgáljuk meg mi történik, ha a egy klaszterhez hozzá, illetve belőle elveszünk egy újabb \vec{z}^* sajátsgvektort. (Jelölje c_m az m -edik klaszter elemszámát.) Egyszerűen kiszámolható, hogy

$$\vec{\mu}_{m^*} = \frac{\sum_{z_j \in Z_m} \vec{z}_j + \vec{z}^*}{c_m + 1} = \vec{\mu}_m + \frac{\vec{z}^* - \vec{\mu}_m}{c_m + 1} \quad \text{és}$$

$$J_{m^*} = \sum_{z_j \in Z_m} \|\vec{z}_j - \vec{\mu}_{m^*}\|^2 + \|\vec{z}^* - \vec{\mu}_{m^*}\|^2 =$$

$$= J_m + \frac{c_m}{c_m + 1} \|\vec{z}^* - \vec{\mu}_m\|^2, \quad \text{illetve}$$

$$\vec{\mu}_{i^*} = \frac{\sum_{z_j \in Z_i} \vec{z}_j - \vec{z}^*}{c_i - 1} = \vec{\mu}_i - \frac{\vec{z}^* - \vec{\mu}_i}{c_i - 1} \quad \text{és}$$

$$J_{i^*} = J_i - \frac{c_i}{c_i - 1} \|\vec{z}^* - \vec{\mu}_i\|^2$$

Egy vektort akkor érdemes átsorolni az i -edik klaszterből az m -edikbe, ha a célfüggvény értéke csökken (= az m -edik komponens növekedése kisebb mint az i -edik csökkenése); másszóval ha teljesül a

$$(3-40) \quad J_m^* - J_m = \frac{c_m}{c_m + 1} \|\vec{z}^* - \vec{\mu}_m\|^2 < \frac{c_i}{c_i - 1} \|\vec{z}^* - \vec{\mu}_i\|^2 =$$

$$= J_i - J_i^*$$

feltétel. A legnagyobb csökkenést akkor érjük el, ha (3-40) baloldala minimális.

A fentiek alapján megfogalmazhatunk egy iteratív eljárást a célfüggvény minimalizálására.

- 1) Válasszunk egy tetszőleges induló partíciót, számítsuk ki a W_{1kn} célfüggvény értékét és az egyes klaszterek átlagait: $\vec{\mu}_1, \vec{\mu}_2, \dots, \vec{\mu}_p$.
- 2) Válasszuk ki a következő vizsgálandó pontot és tegyük fel hogy ez az i -edik osztályhoz tartozik.
- 3) Ha az i -edik osztálynak ez az egyetlen eleme, akkor folytatás 5)-től (hiszen ezt úgysem tudjuk átsorolni). Egyébként számítsuk ki az

$$\vec{e}_m = \begin{cases} \frac{c_m}{c_m + 1} \|\vec{z} - \vec{\mu}_m\|^2, & \text{ha } m \neq i, \text{ illetve} \\ \frac{c_i}{c_i - 1} \|\vec{z} - \vec{\mu}_i\|^2, & \text{ha } m = i; \end{cases}$$

eltéréseket.

- 4) Ha van olyan index, amelyre $\vec{e}_r < \vec{e}_m$ ($m=1,2,\dots,h$) és $r \neq i$, akkor soroljuk át \vec{z} -t az r -edik osztályba, majd számítsuk ki az új osztályparamétereket.
- 5) Ha a célfüggvény értéke az utolsó körben nem változott, az eljárás befejeződött, egyébként folytatás 2)-től a következő sajáttságvektorral.

31.4.3.2. AZ ISODATA módszer

A következőben egy, az előbbinél valamivel összetettebb, rugalmasabb módszert ismertetünk.

Legyen σ tetszőleges döntésfüggvény, L pedig olyan leképezés, amely előállítja az adathalmaznak a döntésfüggvényhez szükséges leírását. A klaszterezés nagyvonalú algoritmusa ekkor a következő:

1. Válasszunk tetszőleges $i_0 \geq 1$ számú osztályt $\vec{a}_1^{(0)}, \dots, \vec{a}_{i_0}^{(0)}$ leíró adatokkal. Legyen $k=0$.
2. A σ döntésfüggvény sorolja be az adatokat a megadott i_k számú osztályba.
3. Az egymáshoz (valamilyen távolságfüggvény szerint) túl közel levő klasztereket vonjuk össze, a túl kicsiket szüntessük meg, a (valamilyen értelemben) túl "lazákat" vágjuk részekre.
4. Az L leképezés segítségével számítsuk ki az új klaszterek alapján az $\vec{a}_1^{(k+1)}, \dots, \vec{a}_{i_{k+1}}^{(k+1)}$ jellemzőket.

5. Ha az előző osztályleírókhoz képest nem túl nagy a változás, az eljárást befejezzük.

6. Egyébként legyen $k=k+1$ és folytassuk a 2. lépéstől.

A fenti váz igen sokféle algoritmust foglal magában. Legismertebb talán az ISODATA eljárás, amely az elsők között volt a maga nemében. Ebben a (3-37/a) szerinti legközelebbi szomszéd távolságfüggvényt és a klaszterek jellemzésére az egyszerű átlagot használták. Az eredeti változatban a klaszterek száma állandó (vagyis a 3. lépést kihagyták). Bizonyítható, hogy ekkor a módszer konvergál és a W_{knh} célfüggvény lokális minimumhelyhez vezet.

A klaszterek összevonásánál gyakran a centrumaik távolságát vizsgálják és aránytalanul közeli klasztereket vonnak össze. A kicsinységüket az egymáshoz viszonyított arányuk alapján szokás eldönteni. Nem egyszerű feladat megfogalmazni a "lazaságot". Gyakori, de nem kizárólagos módszer szerint a klaszter szórása vagy maximális mérete alapján döntenek.

A kezdeti értékek megválasztása gyakran véletlenszerű, s arra irányul, hogy a mintatér lehetőség szerint egyenletesen legyen kifestve. A kérdés ilyen megoldása nem feltétlenül megnyugtató, mivel a különböző kiindulási adatokhoz eltérő végeredmények tartozhatnak, amelyek közül nem tudunk választani.

Az ISODATA eljárást végrehajtottuk a 3-3. ábrán megadott sajátsgvektorokra, csak a vektorok (x,y) végpontjait tekintve ismertnek.

Az eljárás menete a következő volt:

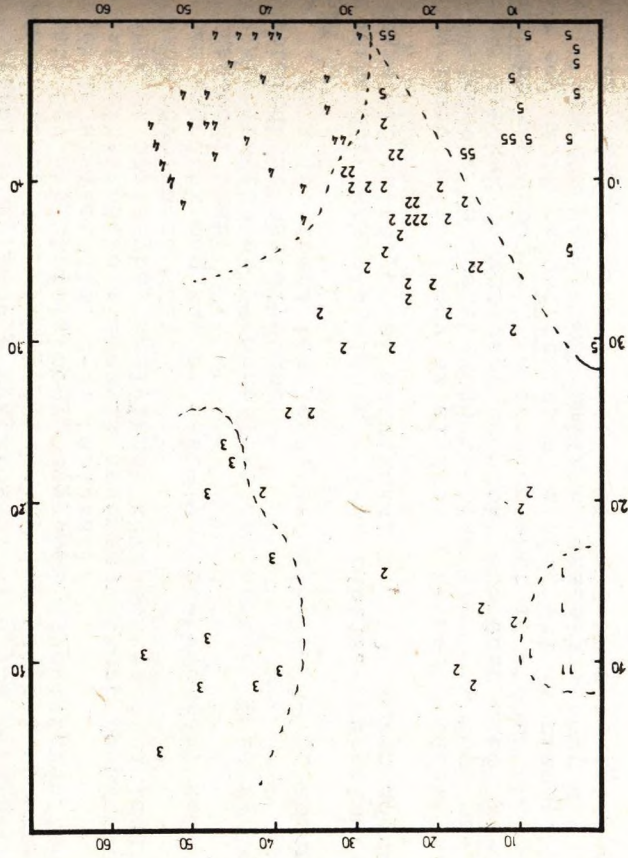
- feltételeztük, hogy a vektorok 5 osztályba sorolhatók, és a kiindulási várhatóértékeket (osztályközpontokat) úgy vettük fel, hogy nagyjából egyenletesen fedjék le a síkot (ld. 3-3. táblázat);
- ezt felhasználva mindegyik sajátsgvektort besoroltuk abba az osztályba, amelyiknek középpontjától a legkisebb távolságra volt;
- nem engedjük meg sem az osztályok szétvágását, sem összevonásukat;
- kiszámítottuk az új osztályközpontokat, majd megismételtük az osztályozást.

Az 1. és a 10. iterációs lépés eredménye a 3-12. ábrán látható.

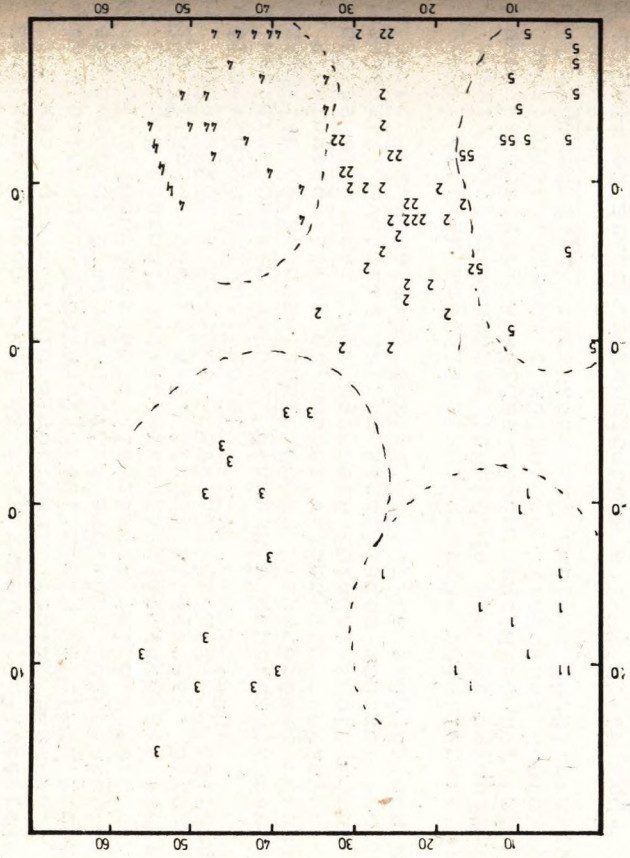
A várhatóértékek alakulását a 3-3. táblázat tartalmazza; a megfelelő osztályozási mátrixokat a 3-13. ábrán adjuk meg.

Figyelemre méltó, hogy az eljárás a Bayes-döntéssel azonos eredményre vezetett. Megjegyezzük azonban, hogy más kiindulási értékek (osztályközpontok) esetében ettől kisebb-nagyobb mértékben eltérő eredményeket is kaphatunk, mint azt egy másik próbafuttatásunk eredményei is bizonyították. Ez aláhúzza azt a már említett szabályt, hogy a klaszterezési eljárás eredményeit megfelelő körültekintéssel kell kezelnünk.

a)



b)



Várhatóérték-vektorok (osztályközpontok) alakulása
 klaszterezési eljárás során

Osztály- szám	Kiindulási értékek	1. iteráció után	10. iteráció után
1	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 4,93 \\ 11,86 \end{bmatrix}$	$\begin{bmatrix} 10,58 \\ 13,30 \end{bmatrix}$
2	$\begin{bmatrix} 20 \\ 20 \end{bmatrix}$	$\begin{bmatrix} 23,28 \\ 31,70 \end{bmatrix}$	$\begin{bmatrix} 25,30 \\ 38,47 \end{bmatrix}$
3	$\begin{bmatrix} 55 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 47,18 \\ 13,46 \end{bmatrix}$	$\begin{bmatrix} 45,17 \\ 15,95 \end{bmatrix}$
4	$\begin{bmatrix} 55 \\ 50 \end{bmatrix}$	$\begin{bmatrix} 42,39 \\ 45,04 \end{bmatrix}$	$\begin{bmatrix} 43,79 \\ 45,07 \end{bmatrix}$
5	$\begin{bmatrix} 1 \\ 50 \end{bmatrix}$	$\begin{bmatrix} 10,40 \\ 43,99 \end{bmatrix}$	$\begin{bmatrix} 7,88 \\ 41,96 \end{bmatrix}$

$$K_{\underline{1}} = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \\ 5 & 27 & 8 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 \\ 0 & 5 & 2 & 20 & 0 \\ 0 & 8 & 0 & 0 & 10 \end{bmatrix}; \quad K_{\underline{10}} = \begin{bmatrix} 10 & 1 & 1 & 0 & 0 \\ 0 & 27 & 6 & 1 & 0 \\ 0 & 0 & 13 & 0 & 0 \\ 0 & 5 & 0 & 19 & 0 \\ 0 & 7 & 0 & 0 & 10 \end{bmatrix}$$

3-13. ábra
 Klaszterezési eljárás osztályozási mátrixai

31.5. KÖRNYEZETFÜGGŐ MÓDSZEREK

A statisztikai alakfelismerésben alapvető szerepet játszik a mérések függetlenségéről szóló feltevés. Jól tudjuk azonban, hogy a képfeldolgozás esetében ez csak korlátozott mértékben állja meg a helyét. Bármekkora képrészlettel is foglalkozunk, mint felismerendő egységgel, igaz az, hogy az értelmezése nem lehet független a környezetének az értelmezésétől. A "képszerűség" éppen abban rejlik, hogy a képpon-

tok világosságát a szemünk egymáshoz viszonyítva érzékeli. Következésképpen várható, hogy sikeresebbek lesznek az alakfelismerési döntéseink, ha figyelembe vesszük ezt a környezeti információt is. Számos újabb, erre irányuló módszer létezik, ezek közül (a teljességre való törekvés igénye nélkül) megemlítünk néhányat:

1. Utófeldolgozási módszerek

Végrehajtjuk az osztályozást a hagyományos módon. Ezután elhagyjuk a sajátságvektorokat és - most már csak az osztályokkal törődve - felülbíráljuk a nyert eredményeket.

Tipikus példa erre a többségi szűrés. Tegyük fel, hogy osztályoztuk a képpontokat. Ezután végigvizsgálunk a képen pl. egy 3x3-as ablakot. Ha az ablakban levő képpontok egy megadott hányada egy osztályhoz tartozik, de az aktuális képpont (a középpont) nem, akkor felülbíráljuk a korábbi döntést. A módszer finomítható: lehetnek kikötések a környezetben előforduló mintákra, minden osztályra más arányt lehet előírni stb.

Lényegesen jobb eredményeket szolgáltatnak (nagyságrenddel nagyobb számítástechnikai kapacitás felhasználásával) a relaxációs módszerek. Ezek lényege az, hogy kompatibilitási tényezőket adunk meg, amelyek azt jelzik, hogy két (vagy több) objektum osztályai milyen mértékben "férnek össze" egymással. Ezek figyelembevételével minden objektumhoz "esélyeket" rendelünk, amelyek jellemzik az egyes osztályokhoz való tartozását. Ezután iteratív módon, minden objektumhoz meghatározzuk azt, hogy a környezetében levő objektumok lehetséges osztályozásai mennyire férnek össze a saját lehetséges osztályaival, és ezek alapján módosítjuk az esélyeket.

A módszer igen hatékony, és ennek megfelelően népszerű. Használata mégis óvatosságot igényel, mivel egy idő után az összeférhetőségi paraméterek döntően befolyásolják az eredményt, a képtől majdnem függetlenül.

2. Előfeldolgozás

Az előfeldolgozó módszerek közös alap gondolata az, hogy valamilyen módon bele kell foglalni a sajátságvektorba a környezeti információt is, amelyet így az osztályozás során automatikusan figyelembe veszünk. A befoglalás megvalósítását illetően már nagy eltérések vannak.

Egy lehetséges módszer például a sajátságvektorok egyszerű kibővítése a szomszédokéival. Az egyszerűségért azonban a vektor megnövekedésével fizetünk súlyos árat.

Javíthatjuk az eljárást, ha a környező sajátságvektorok valamilyen függvényét (átlagát stb.) vesszük csak hozzá a vektorunkhoz.

3. Előszegmentálás

Mivel az osztályozandó objektumokat a képen ábrázolt valóságos tárgyak tényleges ismerete nélkül határoztuk meg, a köztük fennálló tényleges kapcsolatok felismerése nagymértékben a szerencsés véletlenek múlik. Kézenfekvő gondolat, hogy valamilyen módon előre szegmentáljuk a képeket, majd a spektrálisan homogénnek tekinthető, és feltehetően egy valóságos tárgyat ábrázoló objektumot egyszerre soroljunk be valamelyik osztályba. A szegmentálásra számos módszer létezik (ld. 23. fejezet), de önmagában egyik sem ad teljes megoldást. Többnyire a megtalált élek alapján osztják részekre a képet.

4. Sztochasztikus módszerek

Valószínűségi modellek segítségével lehetséges a kompromisszum megteremtése az előző módszerek között. Olyan eljárások léteznek, amelyek egyszerre kezelik az objektumhoz tartozó sajátságvektor(oka)t, ill. a környezeti információt. Jellegükben a relaxációs módszerekhez hasonlítanak, azzal a döntő különbséggel, hogy a környezeti információ nem kerül túlsúlyba az algoritmus előrehaladottabb fázisában sem.

32. Szintaktikus alakfelismerés

A statisztikus módszerek tárgyalásánál többé-kevésbé világossá váltak az eljárások korlátai is; legjelentősebbek az összetett alakzatok kezelésének nehézségei. Ezekre ugyanis nehezen alkalmazható az az alapfeltevés, hogy leírhatók valamilyen - alkalmas módon megválasztott - sajátságvektorral.

Ez vezetett ahhoz a felismeréshez, hogy a megoldást az objektumok fokozatos felbontásával kell megkeresni. Az eljárás gondolatmenete a következő:

Bontsuk fel a vizsgált objektumot olyan egységekre, amelyek egyszerűbb struktúrájúak; majd ezeket még egyszerűbbekre stb., amíg olyan elemi objektumokhoz (a továbbiakban: képelemekhez) nem jutunk, amelyeket biztosan felismerhetünk a korábban tárgyalt módszerekkel. Állapítsuk meg az összetevők közti kapcsolatokat, és ezek alapján adjuk meg az objektum leírását.

A végső osztályozás, ill. felismerés ezek után abból áll, hogy egy osztályt az oda sorolt objektumok leíróinak halmazával definiálunk.

A statisztikus alakfelismeréssel összehasonlítva két lényeges eltérést találunk:

- az objektumokat statisztikai jellemzők (sajátságvektorok) helyett jelsorozattal írjuk le; és

- az osztályok jellemzésére jelsorozatok halmazát használjuk.

Tekintsünk egy egyszerű példát:

Legyenek a leírás alapegységei pl. a következő alapvektorok: $\uparrow \nearrow \rightarrow \searrow \downarrow \swarrow \leftarrow \nwarrow$. Egy (digitális) görbét most a

fenti jelek egymás után írásával adhatunk meg (ld. 3-14. ábra). Egy vízszintes alapú négyzet fogalmának pedig a

$$(\rightarrow)^a (\downarrow)^a (\leftarrow)^a (\uparrow)^a$$

jelsorozat felel meg, ahol a az oldal hosszúsága. (Hallgatólagosan feltettük, hogy a leírás mindig az objektum bal felső sarkából indul.)

Jól látható az is, hogy az osztályok ilyenképpen megadása hamar nehézségekre vezet, hiszen az oldalhossz tetszőleges pozitív szám lehet, azaz végtelen sok jelsorozatot kellene nyilvántartanunk. Célszerű ezért a formális nyelvtanok bevezetése.

A $G = \langle T, N, S, P \rangle$ négyest formális nyelvtannak nevezük, ahol

T , ill. N véges, diszjunkt halmazok (pl. T a képelemek, N a szimbólumok halmaza);

$S \in N$ a kezdő jel, és

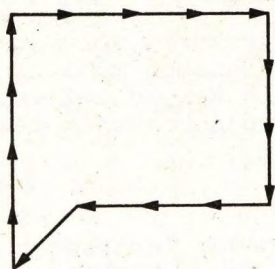
P az átalakítási szabályok véges halmaza.

Egy-egy szabály (TUN)-beli elemekből összeállított, két (véges hosszúságú) jelsorozatból áll, és úgy értelmezendő, hogy ha egy leírásban megtaláljuk a szabály első jelsorozatát, akkor ott a másodikkal helyettesíthetjük.

A szabályok alakjától függően számos nyelvtanosztály létezik. Ezek gyakorlati jelentősége abban tér el, hogy milyen egyszerűen és gyorsan elemezhető a segítségükkel egy jelsorozat.

Azt mondjuk, hogy egy T -beli jelsorozat (szó) levezethető a G nyelvtan segítségével, ha létezik olyan szabálysorozat P -ben, amelyet felhasználva - az S kezdőjelből kiindulva - egymás utáni helyettesítésekkel megkapjuk a T -beli jelsorozatot.

Egy G nyelvtan által definiált (G) nyelven azokat a T -beli jelsorozatokat értjük, amelyek a G segítségével levezethetők. Tekintsük a következő példát:



3-14. ábra
Digitális görbe
szintaktikus leírása

$G = \langle T, N, S, P \rangle$

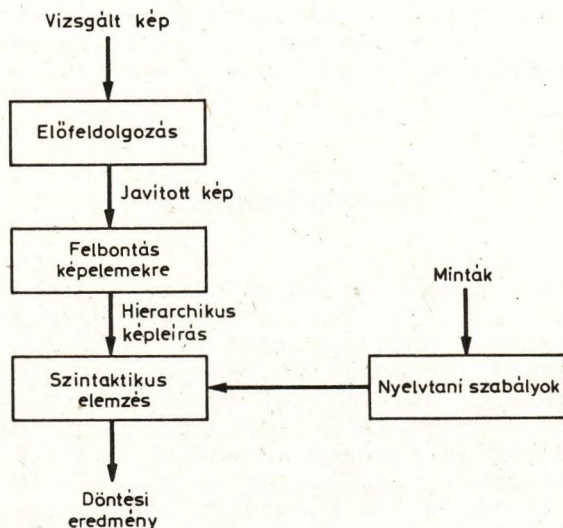
$T = \{ \uparrow \rightarrow \downarrow \}$

$N = \{ S, B \}$

$P = \{ S \Rightarrow \downarrow S \uparrow, S \Rightarrow \downarrow B \uparrow, B \Rightarrow B \rightarrow, B \Rightarrow \rightarrow \}$

(Itt a kettős nyíl a helyettesítendő jelsorozatra mutat.)
Ellenőrzéssel látható, hogy ez az $L(G)$ "nyelv" U-betűk hal-
maza.

A mondottak alapján a szintaktikus alakfelismerési el-
járás a következő lépésekből áll (ld. még a 3-15. ábrát is):



3-15. ábra
Szintaktikus alakfelismerési folyamat

1. A képelemek definíciója.
2. Az osztályokat definiáló jelsorozatok formájának meghatározása.
3. Ezek alapján minden osztályhoz egy-egy nyelvtan megadása, amely lehetőség szerint minél pontosabban generálja az osztályhoz tartozó jelsorozatokat.
4. Egy adott objektum leírásának meghatározása.
5. Elemzés, hogy melyik (ha egyáltalán valamelyik) nyelvtan generálja ezt a jelsorozatot.

Bár a módszer elvben világos és ésszerű, a gyakorlati megvalósítása előtt alapvető kérdéseket kell tisztáni. Ezek:
- Hogyan kell jó képelemkészletet kialakítani?

- Hogyan kell a nyelvtant megadni úgy, hogy ne legyen túl bonyolult, de az egyes osztályokhoz tartozó jelsorozatokat azért pontosan adja vissza?

- Hogyan kell az elemzést hatékonyan elvégezni?

A fenti és egyéb kérdésekre adandó válaszok a konkrét alkalmazástól is függnnek, általános megoldásuk ez idő szerint nem ismeretes.

A szintaktikus alakfelismerés viszonylag új módszer, mindössze néhány éves múltra tekinthet vissza. Ennek ellenére meglehetősen jól kidolgozott az elmélete, és számos sikeres alkalmazása is van. Intenzív továbbfejlesztése folyik, mert úgy látszik, hogy a statisztikus módszerekkel kombinálva jelentősen megjavítja a döntések hatékonyságát.

4. KÉPTÖMÖRÍTÉS

A (számítógépes) képfeldolgozás lényegében nagy adathalmazon (képen) végzett adatfeldolgozás. Mivel az adatok tárolása és átvitele a képek esetében is egyike az alapfeladatoknak, az adattömörítés (itt: képtömörítés) fontos, külön tárgyalást érdemlő részprobléma.

Ebben a részben a képtömörítés alapelveit, legfontosabb típusait és gyakorlati alkalmazásait ismertetjük. A csoportosítás az elméleti rendszerezésen és az eljárások alapelvein alapult.

4.1. Alapfogalmak

A hagyományos képtárolás adathordozói a fotóanyagokon (film) alkalmazott ezüstvegyületek, illetve - különösen mozgóképek rögzítésére - a képmagnetofon szalagja. A számítógépes képfeldolgozás és ezen adattárolási formák telálkozására számos példa ismeretes (pl. mikrofilm-alapú, nagykapacitású képviszszakereső rendszer; számítógépes vezérlésű tv-képparchiváló rendszer). Sajnos a tapasztalat azt mutatta, hogy ezek a rendszerek sem megbízhatóságban (időbeli stabilitásban), sem adathűségben, sem pedig adatkezelési sebességben nem váltották be a hozzájuk fűzött reményeket, s általában az áruk is igen magas. A kiutat a korszerű digitális adathordozók (pl. a nagy tárolókapacitású mágneslemez) megjelenése jelentette.

Hasonló a helyzet a képek átvitele (= átküldése hírközlő csatornán) terén is. Bebizonyosodott, hogy az adatok digitális kezelése a tárolási adatsűrűségeen kívül gyakorlatilag minden paraméter (adathűség, átviteli sebesség, tömörítési lehetőség) tekintetében előnyösebb az analóg megoldásoknál. Mivel azonban mind az adattárolók, mind pedig a hírközlési csatornák használata költséges, természetes, hogy a digitális képtárolás és képátvitel megjelenésével egyidejűleg jelentkezett az az igény is, hogy a képi információt minél hűbben, de ugyanakkor minél kisebb adatmennyiséggel lehessen ábrázolni; erre irányul a képtömörítés.

A képtömörítés iránti érdeklődés több mint 25 éves múlt-
ra tekinthet vissza. A kezdeti időszak erőfeszítései első-
sorban az analóg módszerekre korlátozódtak; tipikus példájuk
a képi jelekre vonatkozó sáv szélesség-csökkentés. (Ez meg-
fordítható jelszint- és frekvenciátranzformációkat jelen-
tett, amelyek azonban mindig információvesztéssel is jár-
tak; így a tömörítés végül is a jelhűség és a zajmenteség
rovására történt. Napjainkban minden analóg hírközlési rend-
szerben végeznek ilyen jellegű jeltranzformációkat.)

A digitális képfeldolgozás terjedésével az érdeklődés
eltolódott a digitális képtömörítés felé. Csak példaképpen
említünk meg két fontos alkalmazási területet:

- A telefax a kétszintes képek igen hatékony CCITT Gx
($x=3$ ill. 4) kódolásán alapul; így vált lehetővé a
képátvitel a szokásos telefonhálózaton keresztül.
(Az eljárás a fekete-fehér futamhosszakat kódolja.
Ez egyszerűen azt jelenti, hogy a képen sorfolytonosan
haladva a kódok felváltva adják meg az egymást követő
fekete, majd pedig fehér képpontokból álló szakaszok
hosszát.)
- A születőben levő HDTV szabvány a műsorszórásra jelen-
leg használatos tv-kép közvetítési csatornakiosztás
mellett teszi lehetővé a kb. hússzor akkora adatmennyi-
ség átvitelét. (A kódolás veszteséges transzformáción
alapul, a sáv szélességet az egyes képfázisok közti
összefüggések kihasználásával csökkentik.)

A korszerű képbeviteli eszközök igen nagy mennyiségű
adatot szolgáltatnak; pl. a szokásos felbontású színes opti-
kai lapolvasók egy A4 méretű oldalról 20 - 30 Mbyte adatot
olvasnak be. (Megemlítjük, hogy analóg jeltovábbítás esetén
1200 baud átviteli sebesség mellett ekkora képek átviteléhez
kb. 40 órára volna szükség; az átviteli idő korszerű digitá-
lis telefonhálózat - pl. ISDN - esetén is kb. 1 óra.) A tömörí-
tés tehát alapvető igényként jelentkezik.

A napjainkban elterjedőben levő multimédia eszközök
működése szintén elképzelhetetlen lenne a kép és a hang rögzí-
tése terén egyaránt alkalmazott adattömörítés nélkül.

A mindennapi élettől távolabbi területeken is igen nagy
a képtömörítés jelentősége. Ilyen területek pl.: a katonai
hírközlés, az űrtávközlés és a jelenleg még kísérleti stádium-
ban levő képteleson.

41.1. REDUNDANCIÁK

Ugyanazt az információt különböző fajtájú és mennyiségű
adat hordozhatja. Az adathalmaz redundáns, ha mennyisége nem
a lehető legkevesebb. Adattömörítéssel azt a folyamatot ért-
jük, amelynek során csökkentjük a bizonyos információt repre-
zentáló adatok mennyiségét.

A számítógépes képfeldolgozás területén háromféle redun-
danciát szokás megkülönböztetni.

41.1.1. Kódolási redundancia

Tekintsünk egy bináris képet, amelyben mindössze kétféle világosságkód fordul elő, azonban ezeket 1 byte-on tároljuk; pl. 0 a fekete és 255 a fehér képpontok kódja. Világos, hogy ez az információ egy biten is ábrázolható (pl 0: fekete, 1: fehér).

Ezesetben a kódolás redundanciája ad módot a tömörítésre.

41.1.2. Képi redundancia

A képen lehetnek olyan belső összefüggések (context) amelyek kihasználása esetén az ábrázolásához kevesebb adat is elegendő. Néhány példa:

- ha a képen csak azonos színű objektumok vannak homogén háttér előtt, hatékonyan kódolható az objektumok határolóvonalainak megadásával; (pl. differentiális iránykóddal - amikor úgy járjuk végig a határolóvonalat, hogy mindig csak az előző iránytól való eltérést kódoljuk - átlagosan 1.5 bit/határolópont érhető el);
- ha az egymás után következő képek csak kismértékben különböznek, gazdaságosan kódolhatók a képek közötti változások megadásával;
- ha a képen szabályos alakzatok vannak, hatásosan kódolható az alakzatokat generáló függvénykapcsolatok leírásával.

41.1.3. Pszichovizuális redundancia

A képek sok olyan információt tartalmaznak, amelyet az emberi látórendszer nem érzékel. Ha csak a kép megjelenítése a feladat, de egzakt feldolgozására nincs szükség, a felesleges információ kiszűrésével adattömörítés érhető el. Fontos megjegyeznünk, hogy az előző két esettel ellentétben a pszichovizuális redundancia kihasználása mindig információvesztéssel jár. Néhány példa:

- ha a képpontokat 24 biten ábrázoljuk, elvileg közel 17 millió féle színt tudnánk megjeleníteni; mivel egyrészt a jelenlegi képmegjelenítő eszközök erre nem képesek, másrészt az emberi látórendszer felbontását is messze meghaladja, nyilvánvaló, hogy elég csak a legjellemzőbb és egymástól érzékelhetően különböző színeket megtartani;
- ha a képen látható objektumok (pl. a képvétel gyenge fényviszonyai miatt) erősen zajosak, zajszűréssel csökkenthetjük a látás szempontjából felesleges információ mennyiségét, s így a kép egyéb módszerek alkalmazásával jobban tömöríthető;

- ha a kép "kevés láttnivalót" tartalmaz túl nagy "méretben", akkor a kép mérete arányosan csökkenthető, amíg a csökkentés nem megy a látvány rovására.

A képek tömörítése az előbbieken felsorolt redundanciák kihasználásán alapul.

Egy adattömörítési eljárástól általában a következőket várjuk el:

- a tömörítési arány (= a tömörített és az eredeti adatmennyiség hányadosa) lehető legkisebb legyen; évégett az algoritmus használja ki a tömörítendő adathalmaz sajátos szerkezetét, illeszkedjék hozzá; (képek esetében kihasználható például, hogy olyan kétdimenziós adathalmazról van szó, amelyben a geometriailag szomszédos adatok között általában erős korreláció van, továbbá a többsávos képek egyes sávjai, mozgó képek sorozatában pedig az egymás utániak nem függetlenek egymástól);
- az algoritmus legyen hatékonyan implementálható; (vagyis a tömörítést és a helyreállítást - akár programmal akár pedig hardver úton - egyszerűen, gyorsan, kis ráfordítással lehessen a gyakorlatban megvalósítani);
- illeszkedjék a meglévő rendszerek kötött lehetőségeihez (mind szoftver, mind pedig adatformátum tekintetében).

Képek tömörítésekor bizonyos feltételek teljesülése esetén általában megengedhető az információvesztés. Ilyen feltétel pl.: hogy a látványkép ne romoljék, hogy statikus képek átvitelekor az átviteli eljárás tartalmazzon az átviteli hibákat periodikusan korrigáló mechanizmust stb.

42. Képtömörítési eljárások

42.1. VESZTESÉGMENTES TÖMÖRÍTÉS

A veszteségmentes képtömörítés lehetővé teszi, hogy a helyreállítás során az eredeti képpel tökéletesen azonos képet kapjunk vissza. Ez azt jelenti, hogy a bemenő- és a kimenőkép megfelelő képpontjainak világosságkódja azonos.

Ilyen eljárások akkor indokoltak, ha a későbbi felhasználás során nincs semmilyen lehetőség a képtartalom módosítására. Számos olyan alkalmazási terület van, amelynél a tömörítés során veszteség nem engedhető meg; az úrfelvételek vagy az orvosi röntgenfelvételek esetében például csak a kódolási-, illetve a képi redundanciát lehet csökkenteni.

Sajnos, a szokásos összetettséű tónusos (és a felvételi eszközök zaját is tartalmazó) képek esetében csak nagyon kis tömörítés érhető el; az 1/2 arány már jónak mondható. A zajszegény vagy a számítógéppel létrehozott képek esetében a helyzet valamivel jobb.

A veszteségmentes tömörítés egyszerű esete, ha a sorfolytonosan tárolt kép leírására a világosság kódok helyett a pontonkénti változásukat használjuk. Ugyanis míg a világosság kódok széles tartományban vehetnek fel értékeket, a változások normális esetben (zajszegény képen) 0 átlagértékű, nagyon keskeny értékészletű függvénnel jellemezhetők. Ez abból következik, hogy a szomszédos képpontok nagy valószínűséggel közel azonos fényességűek, illetve színűek. A leggyakoribb változás-értékek tehát a 0, +1, +2, ...; minél nagyobb a különbség abszolútértéke, annál kisebb valószínűséggel fordul elő. Ha most a felsorolt gyakoribb értékekhez rövid, a ritkábban előfordulókhöz hosszabb kódot rendelünk, a tömörített kép tárigénye kisebb lesz, mint az eredetié, ez mégis maradéktalanul helyreállítható. Az eljárás hatékonyságát tovább növelhetjük, ha először meghatározzuk a változásértékek statisztikus eloszlását, s a változó hosszúságú kódokat az eloszlás ismeretében generáljuk. Ekkor természetesen a helyreállításához is ismernünk kell ezt a megfeleltetést.

A teljességre való törekvés nélkül lássunk néhány módszert a veszteségmentes képtömörítésre.

42.1.1. Változó hosszúságú kódolás

Változó adathosszúságú kódolás (VLC = variable length coding) során a tömörítés kizárólag a kódolási redundancia csökkentésén alapul. A szokásos esetekben 1/2, nagyon jó esetben - pl. szkennelvel beolvasott kétszintes szöveges oldalak esetén - az 1/30 tömörítési arány is elérhető. Az alábbiakban két eljárást ismertetünk.

42.1.1.1. Huffman-kódolás

Az egyik legismertebb és leggyakrabban használt eljárás A. Huffmantól származik, 1952-ből. Lényege, hogy az egyes adatokhoz rendelt kód hossza (bitszáma) fordítottan arányos előfordulásuk gyakoriságával. A kódok megállapítására pl. a következő egyszerű eljárás szolgálhat:

- írjuk fel a bemenő adatokat egymás alá az előfordulási gyakoriságuk (darabszámuk) szerint növekvő sorrendben;
- induljunk el a tábla két legkisebb gyakoriságú eleménél és vonjuk össze őket úgy, hogy egyben adjuk össze a hozzájuk tartozó gyakoriság-értékeket is;
- ezt a páronkénti összevonást mindaddig folytassuk, amíg az összeggel el nem érkezünk a teljes képpontszámig;

- ezután minden egyes adathoz olyan kódot rendeljünk, melynek bitszáma megegyezik az adatot tartalmazó összevonások számával. A kód azon bitjei legyenek "1" értékűek, amelyek a hozzájuk tartozó összevonásokban a párjuknál feljebb voltak.

A statikus Huffman-kódtábla az adott kép statisztikai jellemzőitől független, előre elkészített táblázat, amely pl. nagyszámú hasonló kép elemzése útján alakult ki. Erre példa az egydimenziós facsimile kódolás.

A globálisan adaptív Huffman-kódtábla az adott (teljes) kép hisztogramjának elkészítése után, a fentebb leírt generálási algoritmussal készül; ilyenkor a képet sajnos kétszer kell végigolvasni (először a hisztogramkészítéshez, másodszer a tényleges kódoláshoz). Az eljárás előnye, hogy a statikusnál kisebb tömörítési arányt eredményez.

A lokálisan adaptív Huffman-kódtábla a kódgenerálás során nem állandó, mivel a kódolást az adatokhoz tartozó gyakoriságértékek kísérik, s a kódok jelentése is ennek megfelelően alakul. A legkisebb tömörítési arányt ez adja, azonban hátránya, hogy a folyamatos táblafelújítás számításigénye meglehetősen nagy. Alkalmazása pl. igen drága adatátviteli csatorna jobb kihasználása érdekében indokolt.

42.1.1.2. Aritmetikai kódolás

Bár az aritmetikai kódolást a változó hosszúságú eljárások közé soroljuk, valójában egy nagy bitfolyam áll elő, amelyben a bemenő és a kimenő adatok részletei között nincs egy-egy értelmű megfeleltetés. Az eljárás a bemenő adatok sorozatának egy részletéhez rendel kódot. (Minden egyes kimenőkód 0 és 1 közé eső valós szám, amelynek a bitfolyam csak egy szeletét ábrázolhatja, a számábrázolási korlátok következtében.)

Anélkül, hogy a működést részletesebben ismertetnénk, megemlítjük az eljárás két fő előnyét:

- a lokálisan adaptív változata is könnyen programozható s így igen kis tömörítési arányok érhetőek el (ilyen például az IBM-ben kidolgozott ún. Q-kóder);
- ellentétben a Huffman kódolással, az adatokhoz nemcsak egész hosszúságú kód tartozhat, ezáltal a tömörítés tovább növekszik.

42.1.2. Bitsík-kódolás

A bitsík-kódolás szintén a kódolási redundanciára épül, de kihasználja a világosságkódok közötti korrelációt is.

Lényege, hogy a bemenőképet kódolás előtt felbontjuk annyi független képre, ahány bitek a világosságkódok; így minden részkép egy-egy bitsík adatait tartalmazza. A kapott egybites képeket ezután tömörítjük valamilyen jó hatásfokú veszteségmentes eljárással.

A szokásos képek esetében ez a fentebbi felbontás különböző összetettségű képeket eredményez; a nagyobb helyiértékű bitekhez tartozók lassan változó, azaz könnyen kódolható foltokat tartalmaznak. De a gyorsan változó, kisebb helyiértékű bitekből álló képek kódolása is előnyösebbé válik, mert a szétválasztás általában tömörebb kódtábla használatát teszi lehetővé.

Megjegyezzük, hogy a helyreállításnál a kép fokozatosan alakul ki: először a durva részletek tűnnek elő, majd a kép felbontása egyre finomabb lesz. Ez a tulajdonság számos alkalmazásnál előnyös.

42.1.3. Homogén foltok kódolása

Az alábbi - négyesfa (quad-tree) generálás néven ismert - eljárással túlnyomórészt nagy homogén foltokat tartalmazó, bináris képeket - vagy bitsíkokat - lehet hathatósan tömöríteni, mert lényegében egy speciális fára fűzi fel a képi adathalmazt.

Legyen a bemenőkép mérete $m \times n$, ahol m és n egyaránt 2 hatványa. Az algoritmus röviden a következő:

- vizsgáljuk meg az összes világosságkódot a képen;
- ha minden képpont világosságkódja azonos (azaz vagy "0", vagy "1"), akkor a képhez rendeljük a 'homogén0' vagy 'homogén1' - kódot, és ezzel a kódolás véget is ér;
- ha a világosságkódok nem azonosak, bontsuk a képet két vagy négy, azonos méretű részre (subimage) az eredeti (> 1) oldalak megfelelésével; (a k -adik lépésben a részképek mérete $\max(m/2^k, 1)$ illetve $\max(n/2^k, 1)$ lesz);
- egymás után vizsgáljuk meg minden részkép a homogenitását; amelyik homogén, ahhoz rendeljük a megfelelő homogén-kódot, amelyik nem, azt bontsuk tovább oldalfelezéssel és vizsgáljuk tovább részenként.

Az eljárás akkor ér véget, ha minden részkép homogén.

42.1.4. Határvonal-kódolás

A homogén foltok kódolásának másik gazdaságos módja, hogy a foltokat határoló vonalakat kódoljuk, és kijelöljük azokat a zárt tartományokat, amelyek nem a háttérhez, hanem az objektumokhoz tartoznak. A sokféle változat közül kettőt ismertetünk röviden:

1) Futamhossz-kódolás (run-length encoding)

A képet sorfolytonosan dolgozzuk fel, s az egyszínű képsor-szakaszok (metszetek) hosszát kódoljuk. Ez lényegében a foltok határolóvonalainak pontonkénti kódolását jelenti úgy, hogy a szomszédsági összefüggést a pontok között nem használjuk ki. Míg az 1D (egydimenziós) változat minden metszetnek önálló kódot feleltet meg, addig a 2D (kétdimenziós) változat kihasználja, hogy a szomszédos sorok közt jelentős a korreláció és a kismértékű változásokra külön kódokat használ. (Ilyen kódok pl.: a szakasz kezdetének, ill. végének helyzete nem változott; vagy balra- illetve jobbra tolódott 1,2 vagy 3 képponttal; a szakasz helyzete jelentősen változott; stb.) Tipikus példa a CCITT Gx facsimile kódolás.

2) Kontúrkövetés (contour tracing)

Hardver megvalósítás szempontjából a futamhossz kódolásnál előnytelenebb, viszont bizonyos típusú képekre jobb hatásfokkal alkalmazható tömörítési eljárás. Lényege: használjuk ki a szomszédsági összefüggést az objektumok határpontjai között, azaz menjünk végig a határvonalon pontról pontra. Egy adott határpontból a szomszédosba hétféleképpen juthatunk el, ezt (a vonal-elágazásoktól, vonalvégződésektől stb. eltekintve, azaz csak egyszeres zárt görbékelt feltételezve) három biten kódolhatjuk; ezek az iránykódok (ld. 13.4.2. pont).

A 1-34. ábrán a $P1(k,1)$ pontokból $P2(k,1)$ pontokba olyan útvonalakon lehet eljutni, melyek (4)-, illetve (8)-szomszédos képpontokon keresztül vezetnek. Az ábrák alatti kódsorozat a megfelelő iránykódokat mutatja.

További redundanciát jelent, ha kihasználjuk, hogy a szomszédos iránykódok között is van korreláció. A határvonal ugyanis általában bizonyos "tehetetlenséggel" változtat irányt, vagyis a szomszédos iránykódok különbsége általában 1 (mod 4/8). Célszerű ezért csak az irányváltoztatást kódolni, s minél kisebb, annál rövidebb kódot használni. Az így kapott kódokat nevezik differenciális iránykódoknak.

42.1.5. Kódolás előrebecsléssel

A képi redundancia kihasználásának egyik egyszerű módszere az előrebecslés (prediction). Ez azon alapul, hogy kép közeli részletei között nagymértékű a korreláció, így a kép adott részlete alapján a közeli környezete többé-kevésbé megbecsülhető.

Ha sorfolytonosan haladunk balról jobbra, és ismerjük az aktuális, valamint az előző világosságkódot, akkor a sorban következőről feltételezhető, hogy az elődeinek megfelelő

tendencia rá is érvényes. Ha pl. a megelőző világosság kódok azonosak voltak, akkor a következő világosság kód is nagy valószínűséggel ugyanakkora lesz.

Ebből kiindulva a képet alapvetően előrebecsléssel állítjuk helyre; a kódolás során pedig a kép helyett csak azokat a korrekciós adatokat tároljuk el, melyek az előrebecslés tévedéseit kijavítják. Mivel az előrebecsléseket nagy valószínűséggel csak kismértékben kell korrigálni, a gyakoribb korrekciós adatokat rövidebb kóddal lehet ábrázolni. A módszer nagy homogén foltok és lassú átmenetek esetén működik különösen jól.

42.2. VESZTESÉGES TÖMÖRÍTÉS

A veszteséges képtömörítés értelemszerűleg nem teszi lehetővé az eredeti képtartalom maradéktalan helyreállítását. A leggyakoribb követelmény a látvány változatlansága, ami (mint később erre részletesebben is kitérünk) megengedi, hogy a bemenő- és a kimenőkép megfelelő világosság kódjai eltérők legyenek.

A veszteséges eljárásokkal viszonylag nagy tömörítés érhető el, a megengedett minőségromlás mértékétől függően. Tipikus az 1/20, de ha csak a kép felismerhetősége a cél, az 1/100 arány sem lehetetlen. Különböző összetett eljárások esetében (pl.: az alakzatok kontúrjának elkülönített kódolásával, valamint a foltok textúrájának az emberi látás sajátos törvényszerűségeit is figyelembe vevő kódolásával) az 1/70 arány még jó minőségű helyreállítást tesz lehetővé.

Számos veszteséges tömörítési eljárást dolgoztak ki; ezeket két fő csoportba sorolhatjuk:

- Az előrebecslésen alapuló módszerek általában kevés információ pl. egy kis képrészlet - ismeretében nagyszámú világosságkódot becsülnek meg előre. A becsléssel előállított és a tényleges értékek közti eltéréseket a tömörítés pontossági igényétől függő mértékben korrigálni kell. A kezdőfeltételek rögzítése után az eljárás tulajdonképpen ezeket a korrekciós adatokat állítja elő. Jó esetben a korrekció mértéke is viszonylag kicsi, ekkor a korrekciós paraméterek gazdaságosan adhatók meg változó hosszúságú kódokkal.
- A matematikai transzformáción alapuló módszereknél a képet a világosság kódok helyett bázisfüggvények együttműködéssel írjuk le, illetve - értelemszerűleg - ugyan-ezen bázisfüggvények adott pontbeli súlyozott összegzésével állítjuk helyre. A tömörítésre az adja meg a lehetőséget, hogy ugyanannak a látványnak az előállításához nem feltétlenül szükséges, hogy az együttműködő változatlanok legyenek. Az emberi látórendszer sajátosságai nagymértékű pontatlanságot, elhanyagolásokat engednek meg az együttműködő kódolásában.

A következőkben néhány egyszerű módszert érintünk, majd a napjainkra szabványossá vált JPEG eljárást ismertetjük részletesebben.

42.2.1. Tömörítés előrebecsléssel

Az előző pontban ismertetett előrebecsléses módszer jó példa arra, hogyan lehet megengedni a veszteséget a kódolás során.

A lényeg rendkívül tömören: csak akkor kerüljön az előrebecslést korrigáló adat a kódok közé, ha az elkövetett hiba egy bizonyos küszöböt meghalad. A küszöb értékének változtatása befolyásolja a helyreállítás hűségét, s egyben a tömörítési arányt is.

42.2.2. Transzformációs kódolás

Az előző pontokban ismertetett módszerek a képtartományban dolgoztak (erre utal közös angol megnevezésük is: spatial domain methods). A transzformációs kódolás olyan megfogható lineáris transzformáció, mely a képet (illetve általában kisebb részleteit, a blokkokat) transzformációs együtthatókra képezi le, majd ezeket kvantálja és kódolja. A legtöbb természetes kép transzformációs együtthatói között sok a kis abszolútértékű; ezeket torzítás megengedése esetén elhanyagolhatjuk.

A sokféle transzformációs eljárás (például: Karhunen-Loève (KLT) transzformáció, diszkrét Fourier (DFT) transzformáció, Wals-Hadamard (WHT) transzformáció) közül a hatékonyság és a használhatóság alapján lehet választani.

Megállapítható, hogy az előállított együtthatók függetlensége szempontjából a Karhunen-Loève transzformáció optimális, de sajnos nem ismeretes a gyakorlati felhasználásához megfelelő hatékonyságú implementációja.

Közel ugyanilyen jó tulajdonsággal rendelkezik a diszkrét koszinusz (DCT) transzformáció is, amelyhez ráadásul igen gyors számítási eljárások ismeretesek. Érthető ezért, hogy ennek a transzformációnak a szerepe az elmúlt időszakban megnőtt.

42.2.3. JPEG szabvány

A JPEG (Joint Photographic Experts Group) ajánlások - amelyek az ISO (International Standards Organization) és a CCITT (Consultation Committee on International Telephone and Telegraph) szervezet által 1986-ban felállított munkacsoportban születtek - napjainkra az állóképek veszteséges tömörítésének szabványaiává váltak. (Így például a közismert TIFF képfelformátumban külön jelölőkód szolgál a JPEG eljárással tömörí-

tett képek jelzésére. Itt jegyezzük meg, hogy mozgóképek tömörítésére is születőben van egy szabvány: ennek kidolgozója MPEG (Motion Picture Experts Group) munkacsoport feladata.)

Az eljárás (adaptív) diszkrét koszinusz transzformáción (DCT) alapul, és eleve úgy alakították ki, hogy mind hardverben, mind programmal egyaránt könnyen lehessen megvalósítani. Segítségével - a látvány kismértékű romlása árán - 1/30 tömörítési arány is elérhető; a megengedett minőségromlás mértékét a felhasználó paraméterezheti.

A JPEG tömörítési eljárás egymástól függetlenül kezeli a színösszetevőket, vagyis a színes képek esetében mind a háromra végre kell hajtani. Ha a tömörítés veszteségmentes volna, (eltekintve az esetleges kódolási redundanciától) nem lenne túl nagy jelentősége annak, hogy a világosságkódokat milyen színtérben (color space) ábrázoljuk. Mivel azonban a JPEG szabvány megengedi a veszteséget, célszerű olyan világosságkód-ábrázolást választani, amelyik a legkevésbé érzékeny az elkövetett hibákra, illetve amelyben az adatok belső összefüggései a legnagyobb tömörítést teszik lehetővé.

Pszichovizuális kísérletek bebizonyították, hogy többféle előnnyel jár, ha a szokásos RGB helyett az YUV színrendszert használjuk, mert ezáltal a színösszetevők a látás szempontjából fontosabb és kevésbé fontos adatokra válnak szét. Az emberi látórendszer ugyanis sokkal kevésbé érzékeny az U- és a V- (krominancia-) összetevőkre, mint az Y- (luminancia-) összetevőre; ezért az előbbieket sokkal pontatlanabb helyreállítása is elfogadható. Ebből következően célszerű a színes képet tömörítés előtt transzformálni az YUV rendszerbe.

Az eljárás első lépésében az adott (R, G ill. B) színösszetevőt (a továbbiakban: képet) egymástól függetlenül, 8*8 képpontból álló blokkokra bontjuk, amelyeket ezután (csaknem teljesen) függetlenül kezelünk. (Megjegyezzük, hogy a blokkok közötti összefüggést kizárólag a világosságkód-átlaguk különbségi kódolása jelenti.)

A transzformáció végrehajtásával blokkonként (8*8=) 64 együtthetőt kapunk a kétdimenziós (2D) diszkrét bázisfüggvényekhez. Az eljárás képlettel a következőképpen fejezhető ki:

$$\begin{aligned} \text{DCT}(k, l) &= \\ &= \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} q(m, n) * \cos((2m+1) 2\pi k/4N) * \cos((2n+1) 2\pi l/4N) \end{aligned}$$

ahol

- N a blokk mérete (A JPEG szabványban rögzítetten 8);
- q(m, n) a világosságkód az (m, n) blokk-relatív koordinátájú képpontban;
- k, l blokk-relatív koordináták.

A fenti képlet közvetlen számításokra alkalmatlan, mivel valós adatokkal kellene nagyszámú műveletet elvégezni. A transzformáció igen jó közelítése pl. az RVFFT (Real Valued Fast Fourier Transform) eljárás alapul.

A DCT transzformáció a kiindulási adatoknál inkább többet eredményez, mint kevesebbet, hiszen a kapott együtthatókat nagyobb pontossággal kell ábrázolni, mint az eredeti világosságkódokat. Előnye mégis abban van, hogy a 64 transzformációs együttható gyakorlatilag nem mutat belső összefüggést, szemben a blokk többnyire erősen korreláló 64 képpontjával. (A diszkrét koszinusz transzformáció a Karhunen-Loève transzformációt megközelítő mértékben tudja az adatokat független paraméterekké alakítani.) Pszichovizuális kísérletek bebizonyították, hogy ez a 64 együttható nem azonos mértékben fontos az eredetit közelítő látvány létrehozásához. Kiderült, hogy a nagyobb frekvenciájú képtartalomváltozást reprezentáló bázisfüggvények ebből a szempontból még akkor sem játszanak túl nagy szerepet, ha a hozzájuk tartozó együttható értéke nagy. (Megjegyzendő, hogy a kísérleteket megadott képernyőméret-látótávolság összefüggések mellett végezték, s a DCT tömörítési eljárás alkalmazása esetén is feltételezik a képhelyreállítás bizonyos standard körülményeit.)

Az előbbiekből tehát az következik, hogy a nagyobb frekvenciájú bázisfüggvények együtthatói sokkal durvábban (pontatlanabban) kódolhatók, mint a kis frekvenciájúakéi. A kódolás pontosságát úgy is előírhatjuk, hogy megszabjuk, melyik együttható hányféle értéket vehet fel, azaz milyen pontosságú a kvantálása. (A legegyszerűbb kvantálás, ha az együtthatót elosztjuk egy adott számmal és a hányados egészrészét vesszük. Minél nagyobb az osztó, annál kevesebb különböző eredményt kaphatunk, így visszaszorzás után annál durvább lesz a "lépcsőzés". Ha pl. egy együttható értéke 11 és a kvantálási osztó 16, akkor az eredmény 0 lesz. Ugyancsak nulla lesz az eredmény, ha az együttható értéke a $\{-15, \dots, 15\}$ tartományban bármilyen más értéket vesz fel.)

Tipikus kvantáló- (azaz 64, együtthatóról-együtthatóra változó osztót tartalmazó) mátrixokat mutatunk a 4-1. ábrán a luminanciához (a), illetve a krominanciához (b).

A szokásos JPEG-alapú rendszerekben a képminőséget (s ezzel a tömörítési arányt is) pl. a kvantálási szintek változásával lehet befolyásolni. Minél "feljebb" toljuk a kvantálási szinteket annál durvább lesz a kép, viszont annál kisebb lesz a tömörítés aránya.

A kvantálás után kapott 64 együttható túlnyomó része 0 vagy igen kis abszolútértékű szám lesz. (Már jeleztük, hogy a blokk eredeti képtartalmát - pontosabban közelítését - úgy lehet előállítani, hogy az előbbiekből kapott számokat megszorozzuk a megfelelő kvantálási együtthatókkal, s ezekkel végezzük el az inverz transzformációt.) A kvantálási adatok megválasztásából következik, hogy a legtöbb nem-zérus értékű együttható a blokk bal felső sarkának közelében

található, s attól távolodva, egyre több lesz a nulla. Ez a tény igen jól kihasználható a kódolás során.

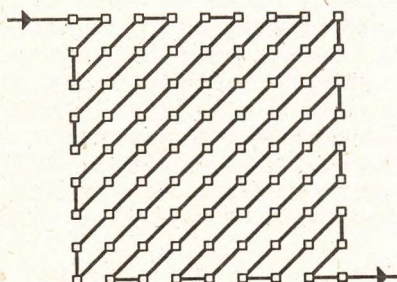
16	11	10	16	24	40	51	61	17	18	24	47	99	99	99	99
12	12	14	19	26	58	60	55	18	21	26	66	99	99	99	99
14	13	16	24	40	57	69	56	24	26	56	99	99	99	99	99
14	17	22	29	51	87	80	62	47	66	99	99	99	99	99	99
18	22	37	56	68	109	103	77	99	99	99	99	99	99	99	99
24	35	55	64	81	104	113	92	99	99	99	99	99	99	99	99
49	64	78	87	103	121	120	101	99	99	99	99	99	99	99	99
72	92	95	98	112	100	103	99	99	99	99	99	99	99	99	99

a)

b)

4-1. ábra
Kvantáló mátrixok

Évégett rendezzük át az együtthatókat abban a sorrendben, ahogyan a blokk bal felső sarkától a jobb alsó sarok felé haladó törtvonalon elhelyezkednek (ld. 4-2/a. ábra; a 4-2/b. ábrán az egyes együtthatók helyén szereplő szám megmondja, hogy az átrendezés után hányadik helyre kerülnek).



a)

1	2	6	7	15	16	28	29
3	5	8	14	17	27	30	43
4	9	13	18	26	31	42	44
10	12	19	25	32	41	45	54
11	20	24	33	40	46	53	55
21	23	34	39	47	52	56	61
22	35	38	48	51	57	60	62
36	37	49	50	58	59	63	64

b)

4-2. ábra

Az első együttható a konstans-függvényt reprezentálja, amely a blokkon belüli világosság kódok átlagértéke. Ezt az együtthatót (az elektronika "egyenáram" (direct current) elnevezésének analógiájára) DC együtthatónak nevezik. A többi együttható a nem-konstans bázisfüggvényekhez tartozik; ezek az AC - (v.ö. alternating current) együtthatók. A JPEG szabvány szerint a blokkok DC együtthatóit az AC együtthatóktól függetlenül kódolják, és helyettük a blokkonkénti változásukat tárolják. Az AC együtthatók kódolására változó hosszúságú kódokat (adaptív VLC eljárással) használnak.

A tömörítés eredménye az előbbieken leírtak szerint olyan bitfolyam, amelyik közvetlenül felhasználható az eredeti képtartalom rekonstrukciójára; ekkor sorfolytonos kép-felépítésről beszélünk. Ez a JPEG eljárás leggyakoribb felhasználási módja; pl. képi másolat készítéséhez erre van szükség.

Egyes felhasználási területeken kiemelt fontosságú, hogy a kép helyreállítása során minél hamarabb kerüljön felismerhető állapotba, még ha ennek az is az ára, hogy a teljes rekonstruk-

ció az előbbinél kissé tovább tart (progresszív képfelépítés). Előbb a DC együtthetők alapján áll össze egy igen durva kép, melynél minden 8×8 pontos blokkot egy-egy átlagérték jellemmez. A következő ciklusban az első AC együtthetők szerinti módosítás történik meg minden blokkban; ez 4×4 pontos mintákat eredményez, s így haladunk az egyre finomabb felbontás felé. (Tipikus alkalmazási terület: képi adatbázisban kereséskor a durva képből egyre finomabb képet hozunk létre, a folyamat bármikor megszakítható - ha kiderül, hogy másik képre van szükség.)

42.2.4. Fraktál-tömörítés

Meg kell említenünk még egy másik, az eddigiektől gyökeresen eltérő elvű, igen jó hatásfokú képtömörítési eljárást is, amely a fraktálokon alapul és $1/40$ - $1/80$ tömörítési arányt eredményez, kismértékű látványromlás árán. (A fraktálokat Michael Barnsley, a számítógépes grafika egyik úttörő kutatója fedezte fel.)

Az eljárás lényege, hogy megkeresi a tónusos kép fraktálokkal történő előállításának szabályait és ezeket tárolja. Hátránya, hogy igen jelentős a tömörítés-irányú számítás-igénye; ezen speciális processzorok alkalmazásával segítenek.

Ellentétben a tömörítés viszonylagos lassúságával, a helyreállítás rendkívül gyors lehet; pl.: egy 320×200 képpontból álló 24 bit/képpont színfelbontású kép $386/25\text{MHz}$ PC-n kevesebb mint egy másodperc alatt készül el. További előny, hogy a helyreállítás során a kép tetszőlegesen nagyítható, s ekkor látszatra új részletek tűnnek elő, a globális mintázatok lokális ismétlésével.

ZÁRSZÓ

A már többször említett korlátozott terjedelem és az anyag bevezető jellege miatt nem állt módunkban foglalkozni a számítógépes képfeldolgozás harmadik, legmagasabb szintű fáziséval, a képfelismeréssel. Ez a terület nálunk egyelőre csak elméleti-kutatási jelentőséggel bír. Az ismertetett módszerek és eljárások végső soron arra irányulnak, hogy a jövőben mind "értelmesebb" gépeit (automatákat, robotokat) megtanítsuk látni. A felmerülő problémák súlyosak, és ez idő szerint csak egy-egy konkrét feladatra vannak többé-kevésbé kielégítő megoldások; az univerzális látórendszer kidolgozása még várat magára.

Mindenesetre a fejlődés íve - mondhatni - exponenciális. Napjainkra a képfeldolgozás bekerült minden hardware gyártó vagy software fejlesztő számítástechnikai cég programjába; különös hangsúlyt kapott az 5. generációs számítástechnikai távlati terv meghirdetése óta.

A szerzők remélik, hogy munkájukkal segítséget tudtak nyújtani a témával foglalkozóknak; egyszersmind felkeltik a figyelmet és az érdeklődést azokban is, akik számára a képfeldolgozás a jövőben hasznos lehet.

IRODALOMJEGYZÉK

I. részhez

1. AHMED, N. - RAO, K.R.: Orthogonal transforms for digital signal processing; Springer-Verlag (1975) Berlin.
2. BAHVALOV, N.SZ.: A gépi matematika numerikus módszerei; Tankönyvkiadó (1977) Budapest.
3. BRIGHAM, O.E.: The fast Fourier-transform; Prentice-Hall (1974) New Jersey.
4. DANIELSSON, P.E.: Incremental curve generation; IEEE-C-19, Nr.9 (1969) 783-793.
5. FREEMAN, H.: Computer processing of line-drawing images; Computer Survey, 6. (1974) 57-97.
6. GAAFAR, M.: Convexity verification, block-chords and digital straight lines; CGIP, 6. (1977) 361-370.
7. HAUSKA, H.: The Lulea image processing system (LIPS); Machine Processing of Remotely Sensed Data Symp. (1980)
8. HOBBS, L.C.: Computer graphics display hardware; IEEE Comp. Graphics and Applications, 1. (1981)
9. KOLMOGOROV, A.N. - FOMIN, : Bevezetés a valós függvénytanba; Műszaki Kiadó (1981) Budapest.
10. MAXWELL, P.C. -- BAKER, P.W.: The generation of polygons, representing circles, ellipses and hyperbolas; CGIP, (1979) 84-93.
11. PRESTON, K.: Image processing software. Progress in Pattern Recognition; North Holland (1981) Amsterdam.
12. RÉNYI, A.: Valószínűségyszámítás; Tankönyvkiadó (1968) Budapest.
13. RIESENFELD R.F.: Homogeneous coordinates and projective planes in computer graphics; IEEE Comp. Graphics and Applications, 1. (1981) 50-55.
14. ROGERS, D.F. - ADAMS, J.A.: Mathematical elements for computer graphics; Mcgraw-Hill Book Company (1976) New York.
15. ROSENFELD, A.: Digital straight line segments; IEEE-C-23 (1974) 1264-1269.
16. SEBESTYÉN, I.: The videodisc revolution, Working Paper IIASA (1981) Laxenburg.
17. TSOTSOS, J.K.: Knowledge of the visual process: content, form and use; IEEE Proc. 6th Int. Conf. Pattern Recognition (1982) 654-669.

18. VINCZE, I.: Matematikai statisztika; Tankönyvkiadó (1974) Budapest.
19. WADE, B.T.: A virtual memory system for picture processing; CACM, 5. (1984)
20. WIE, V.P. et al.: A transportable execution system for use with remote sensing applications software; in [7]

II. részhez

21. ANDREWS, H.C.: Computer techniques in image processing; Academic Press (1970), New York.
22. ANDREWS, H.C.: Digital image restoration: a survey; IEEE Computer, Vol. 7, Nr. 5 (1974) 36-45.
23. BOUTLER, J.F.: Interactive digital image restoration and enhancement; CGIP, 11. (1979)
24. BRACCINI, C. - MARINO, G.: Fast geometrical manipulations of digital images; CGIP, 13. (1980) 127-141.
25. BURT, P.J.: Fast filter transforms for image processing; CGIP, 16. (1981) 20-51.
26. DETTORI, G.: An on-line algorithm for polygonal approximation of digitized plane curves; in [17] 739-741.
27. DORST, L.: A local contrast enhancement filter; in [17] 604-606.
28. GEUEN, W. - WIDZGOWSKI, E.: A simple edge detection algorithm on the basis of visual contour perception; in [17] 843-845.
29. GONZALEZ, R.C. - WINTZ, P.: Digital image processing; Addison-Wesley (1977).
30. HUECKEL, M.: An operator which locates edges in digital pictures; JACM, 18, Nr. 1 (1971) 113-125.
31. INGRAM, W.L. - HOOKER, H.L.: The selection of approximating functions for tabulated numerical data; Technical Report, X-64658, G.C.Marshall Space Flight Center (1972) Alabama.
32. JULESZ, B.: Textons, the elements of texture perception and their interactions; Nature, Vol. 290 (12. March, 1981), 91-97.
33. KUNDU, A. et al.: Application of two-dimensional generalized mean filtering for removal of impulse noises from images; IEEE Trans.Acoustic Speech and Signal Processing, 3. (1984)
34. MATSUYAMA, T. et al.: A structured analysis of natural textures by Fourier transformation; in [17] 289-292.
35. NEVATIA, R.: A color edge detector and its use in scene segmentation; IEEE-SMC-11 (1977).
36. PERNY, D. et al.: Perspective mapping of planning textures; Computer Graphics, Vol. 16, Nr.1. 72-89.
37. PIETIKÄINEN, M. et al.: Texture classification using averages of local pattern matches; in [17] 301-303.

38. PRATT, W.K.: Digital image processing; Wiley Interscience (1977) New York.
39. PRATT, W.K.: Generalized Wiener filtering computation techniques; IEEE-C-21 (July, 1972) 636-641.
40. ROBINSON, E.S.: Edge detection by compass gradient masks; CGIP, 6, Nr. 5 (1977)
41. ROSENFELD, A. - KAK, A.C.: Digital picture processing I-II.; Academic Press, (1982) New York.
42. ROSENFELD, A. - THURSTON, M.: Edge and curve detection for visual scene analysis; IEEE-C-20, Nr. 5 (1971) 562-569.
43. WANG, D.C. - VAGNUCCI, A.H.: Gradient inverse weighted smoothing scheme and the evaluation of its performance; CGIP, 15. (1981) 167-181.
44. WEIMAN, F.R.: Highly parallel digitized geometric transformations without matrix multiplication; Proc.Int.Conf. Parallel Processing (1976), 1-10.

III. részhez

45. ARATÓ, M.: Fejezetek a matematikai statisztikából számítógépes alkalmazásokkal II., SZÁMKI Közlemények 22 (1979) Budapest.
46. CSETVERIKOV, D.: 2-dimenziós alakfelismerési módszerek; Munkaanyag, MTA-SZTAKI (1980).
47. CSETVERIKOV, D.: Threshold-averaged characteristic length as a measure of texture separateness; Working paper MTA-SZTAKI (1981).
48. DUDA, R.O. - HART, P.E.: Pattern classification and scene analysis; Wiley-Interscience (1973), New York.
49. FOLEY, D.H. - SAMMON, J.W.: An optimal set of discriminant vectors; IEEE-C-24, Nr. 3 (1975).
50. FRITZ, J.: Az alakfelismerés statisztikus módszerei, MTA-MKI kiadvány (1974) Budapest.
51. FU, K.S. - YU, T.S.: Statistical pattern classification using contextual information; Research Studies Press (1980) Chichester.
52. FUKUNAGA, K.: Introduction to statistical pattern recognition; Academic Press (1972) New York.
53. HARALICK, R.M.: Statistical and structural approaches to texture; Proc. IEEE Vol. 67, Nr. 5 (1979).
54. HUMMEL, R.A. - ZUCKER, S.W.: On the foundations of relaxation labelling process; IEEE-PAMI-5, Nr. 3 (1983).
55. KITTLER, J. - DEVIJVER, P.: Pattern recognition: a statistical approach; Prentice Hall (1982) London.
56. KOONTZ, W.L.G. - FUKUNAGA, K.: A nonlinear feature extraction algorithm using distance transformation; IEEE-C-21, Nr. 1 (1972).

57. LANDGREBE, D.A.: The development of a spectral-spatial classifier for earth observational data; Pattern Recognition, Vol. 12 (1980).
58. MERICKEL, M.B. et al.: A spatial processing algorithm to reduce the effects of mixed pixels and increase the separability between classes; Pattern Recognition, Vol. 17, Nr. 5 (1984).
59. PELEG, S.: A new probabilistic relaxation scheme; IEEE-PAMI-2, (1980)
60. ROSENFELD, A. et al.: Scene labelling by relaxation operations; IEEE-SMC-6, (1976)
61. SWAIN, P.H. et al.: Contextual classification of multi-spectral data; Pattern Recognition, Vol. 13 (1981).
62. WHARTON, S.: A contextual classification method for recognising land use patterns; Pattern Recognition, Vol. 15 (1982).

IV. részhez

63. ABRAMSON, N.: Information theory and coding; McGraw-Hill, New-York, 1963.
64. CHEN, W. PRATT, W.K.: Scene adaptive coder; IEEE Trans. on Comp. (March, 1984)
65. GONZALEZ, R.C., - WOODS, R.E.: Digital image processing: Addison-Wesley Publishing Company, Inc.; 1992.
66. EasyTech/Codec, Owner's Manual; AGI AutoGraph International A/S, Denmark, 1990
67. Initial draft for adaptive discrete cosine transform technique for still picture data compression standard; ISO document (1988) ISO/JTC1/SC2/WG8 N800.
68. HUFFMAN, D.A.: A method for the construction of minimum redundancy codes; Proc. IRE-40/10 (1952) pp. 1098-1101.
69. JAIN, A.: Image data compression: a review; Proc. IEEE - 69/3 (1981) pp. 349-389.
70. KUNT, M. - IKONOMOPOULOS, A. - KOCHER, M.: Second-generation image-coding techniques; Proc. IEEE-73/4 (1985) pp. 549-574
71. LANGDON, G.G.: An introduction to arithmetic coding, IBM J.Res. Develop.-28, (1984) pp. 135-149.
72. LANGDON, G.G. - RISSANEN, J.: A simple general binary source code; IEEE Trans. IT-28/5 (1982) pp. 800-803.
73. MITCHELL, J.L. - PENNEBAKER, W.B.: Optimal hardware and software arithmetic coding procedures for the Q-coder; IBM-Res, Report RC-12658, 1987.
74. SEABROOK, G.R.: Arithmetic coding - an alternative strategy for video coding; 3-rd Int.Conf. on Image Processing (IEEE-ICIP) (1989) pp. 613-617.

