



Írta:
CZÚNI LÁSZLÓ
TANÁCS ATTILA

KÉPI INFORMÁCIÓ MÉRÉSE

Egyetemi tananyag



2011

COPYRIGHT: © 2011–2016, Dr. Czúni László, Pannon Egyetem Műszaki Informatikai Kar
Villamosmérnöki és Információs Rendszerek Tanszék, Dr. Tanács Attila, Szegedi
Tudományegyetem Természettudományi és Informatikai Kar Képfeldolgozás és Számítógépes Grafika
Tanszék

LEKTORÁLTA: Dr. Szlávik Zoltán, Magyar Tudományos Akadémia Számítástechnikai és
Automatizálási Kutatóintézet

Creative Commons NonCommercial-NoDerivs 3.0 (CC BY-NC-ND 3.0)

A szerző nevének feltüntetése mellett nem kereskedelmi céllal szabadon másolható, terjeszthető,
megjelentethető és előadható, de nem módosítható.

TÁMOGATÁS:

Készült a TÁMOP-4.1.2-08/1/A-2009-0008 számú, „Tananyagfejlesztés mérnök informatikus,
programtervező informatikus és gazdaságinformatikus képzésekhez” című projekt keretében.



ISBN 978 963 279 494 5

KÉSZÜLT: a [Typotex Kiadó](#) gondozásában

FELELŐS VEZETŐ: Votisky Zsuzsa

AZ ELEKTRONIKUS KIADÁST ELŐKÉSZÍTETTE: Gerner József

KULCSSZAVAK:

képi információ, digitális képfeldolgozás, orvosi képfeldolgozás, optikai mérés, képelemzés, videó-
megfigyelés, távérzékelés, zajszűrés.

ÖSSZEFOGLALÁS:

A jegyzet megismerteti az olvasót a képi információ technikai jelentésének és mérésének alapelveivel,
alapvető módszereivel. A modern képalkotás és képfeldolgozás technikáját bemutatva megismerhető, milyen
módon érzékelik, tárolják és elemzik a különböző számítógépes alkalmazások a képeket. A jegyzetben
kiterünk az orvosi képalkotás és képelemzés, a képviszszakeresés, az ipari képelemzés, a távérzékelés és a
videó-alapú biztonsági rendszerek által alkalmazott módszerekre, az alapvető algoritmusokra is.

Tartalomjegyzék

1. Bevezetés	6
1.1. Kinek ajánljuk a könyvet?	6
1.2. A matematikai konvenciókról	8
2. A képbevitel jellemzése	9
2.1. A képfeldolgozás szintjei	9
2.2. Látás hullámokkal a gépi és biológiai rendszerekben	10
2.3. A kép projekciója	12
2.4. Az optika szerepe és jellemzése	14
2.4.1. Néhány speciális optika	17
2.4.2. Optikai hibák	17
2.5. CCD és CMOS képérzékelők	20
2.5.1. CCD és CMOS összehasonlítása	21
3. A kép felbontása	23
3.1. Digitalizálás	24
3.2. Reprezentáció pixel térben	24
3.3. Reprezentáció frekvenciatérben: felbontás hullámfüggvényekre	26
3.4. A mintavételezés korlátai, átméretezés	28
3.4.1. A kép kicsinyítése, nagyítása	30
4. A kép hibái és a kép szűrése	32
4.1. Elektronikus képi hibák	33
4.2. A képi minőség mérése: objektív és szubjektív megközelítések	34
4.3. A degradációs és szűrési folyamat	37
4.3.1. A kép elmosódása és az inverz művelet	37
4.3.2. Additív zaj és szűrése	40
4.3.3. Zaj szűrése a frekvenciatartományban	41
4.3.4. A kép elmosódása és additív zaj jelenléte	43
5. Képek visszakeresése	45
5.1. A digitális kép mint sokdimenziós adat	45
5.2. A szemantikai rés	46
5.3. Képi adatbázisok	47
5.3.1. A képi adatbázisrendszerek sajátosságai	47

5.3.2.	Dublin Core és más metainformációs rendszerek	48
5.4.	A JPEG2000 és az MPEG-7 visszakereshetősége	49
5.4.1.	JPEG2000	49
5.4.2.	MPEG-7	50
5.4.3.	Az MPEG-7 XM egy kísérleti kiértékelése	52
5.5.	A tartalmi elem kapcsolódásai	53
5.5.1.	Szemantika a szöveg és képi tartalom kapcsolatával	54
5.5.2.	"Bag of words"	56
6.	Orvosi képfeldolgozás	57
6.1.	Képalkotó technikák	59
6.1.1.	Röntgen, fluoroszkóp	59
6.1.2.	Rekonstrukció	60
6.1.3.	CT (Számítógépes tomográfia)	61
6.1.4.	MRI (Mágneses rezonancia képalkotás)	63
6.1.5.	Nukleáris medicina	63
6.1.6.	fMRI (funkcionális MR képalkotás)	64
6.2.	Digitális képek geometriája	64
6.2.1.	Mintavételezés	65
6.2.2.	Képi koordináta-rendszerek	66
6.2.3.	3D szeletorientációk	67
6.3.	Vizualizáció	68
6.4.	Orvosi információs rendszerek	72
6.4.1.	DICOM fájlformátum	73
7.	Orvosi algoritmusok	75
7.1.	Orvosi képregisztráció	75
7.1.1.	Regisztrációs algoritmusok fő komponensei	78
7.1.2.	A regisztrációs algoritmusok csoportosításai	81
7.1.3.	Klasszikus módszerek	81
7.2.	Orvosi képek szegmentálása	87
7.2.1.	Régió-alapú módszerek	88
7.2.2.	Él-alapú módszerek	90
7.2.3.	Energiaminimalizáló módszerek	91
7.2.4.	Modell-alapú megközelítés	92
8.	Távérzékelés, mikroszkópia	97
8.1.	Távérzékelés	97
8.1.1.	A távérzékelés alapfogalmai	97
8.1.2.	Képalkotás	99
8.1.3.	Műholdas programok	99
8.1.4.	Mezőgazdasági felhasználás	100
8.1.5.	Térképészeti alkalmazások	101
8.1.6.	Változáskövető alkalmazások	102
8.2.	Mikroszkópia	102

8.2.1.	Fénymikroszkópok	103
8.2.2.	Speciális vizsgálati módszerek	104
8.2.3.	Elektron- és pásztázószondás mikroszkópok	104
9.	Ipari képfeldolgozó rendszerek	106
9.1.	Az ipari alkalmazások komponensei	106
9.2.	Optika és kamera kiválasztása	107
9.2.1.	Telecentrikus optikák	108
9.2.2.	Ipari mikroszkópok	109
9.2.3.	Kamera és optika illesztése	109
9.3.	Megvilágítási technikák	111
9.3.1.	Felső megvilágítás	111
9.3.2.	Alsó megvilágítás	112
9.4.	Ipari számítógépek, intelligens ipari kamerák	113
9.5.	A méréseket, kiértékeléseket végző programok, algoritmusok	114
9.5.1.	Néhány ipari példa	115
10.	Biztonsági kamerarendszerek	118
10.1.	Alkalmazási területek	118
10.2.	A rendszerek összetevői és azok fejlődése	119
10.2.1.	Biztonsági kamerák	119
10.2.2.	Egyéb komponensek	122
10.3.	Biztonsági képek feldolgozása	123
10.3.1.	Kamera kalibráció	125
10.3.2.	Változás- és mozgásdetekció	125
10.3.3.	Optikai áramlás számítása	127
10.3.4.	Az elemzés magasabb szintű feladatai	131
A	Egy CT képszelet DICOM fejléce	132

1. fejezet

Bevezetés

1.1. Kinek ajánljuk a könyvet?

A képi információ a legtöbbünk számára a leggazdagabb információforrás a világ megismeréséhez. A képi információ, azaz az elektromágneses sugárzás szűk tartományának feldolgozása, elemzése, értelmezése lehetővé teszi az élőlények számára a kifinomult mozgást, a világ alkotóival való interakciót, közeli és távoli jelenségek megfigyelését és megismerését. Az emberi agy működésének több mint 50 százalékát közvetve vagy közvetlenül a vizuális feldolgozás foglalja le, a látásért felelős vizuális kortex az agy legnagyobb alrendszere. A látással kapcsolatos képességeinket a számítógépek korában egyre inkább elektronikus eszközökkel egészítjük ki, fejlesztjük tovább: digitális fényképezőgépek, távcsövek, mikroszkópok, videómegfigyelő rendszerek, a testbe beelátó orvosi képalkotó eszközök, képi adatbázisok állnak ma már rendelkezésünkre, hogy fokozzuk információszerző és feldolgozó képességeinket, így javítsuk az életminőséget.

A képi információt feldolgozó elektronikus rendszereknek az üzembe helyezése, használata és fejlesztése a szakemberektől folyamatos megújulást kíván, hiszen olyan nagyon gyorsan változó, fejlődő területről, eszközökről van szó, amelyek egyaránt használják az informatika, a szoftvertechnológia, az alkalmazott matematika, az elektronika és elektrotechnika legújabb módszereit és eszközeit. Magyarországon is egyre több olyan vállalkozás működik, ahol nemcsak használják, hanem fejlesztik is ezeket a technológiákat: digitális mikroszkópokat, ipari és biztonságtechnikai kamerákat, orvosi képalkotó műszereket és szoftvereket, világítástechnikai berendezéseket vagy akár gépjárművezetést támogató videórendszereket.

Tankönyvünk fő, általános célkitűzése a tipikus képalkotási és képfeldolgozási folyamat bemutatása: milyen fázisokon, átalakulásokon, elemzéseken megy át a képi információ, miután a fotonok eléri a kamerát, és amíg be nem fejeződik az információ feldolgozása. A munkánk terjedelme nem teszi lehetővé, hogy minden igényt kielégítően, teljeskörűen foglalkozzunk a témával, így dolgozatunk inkább bevezető jellegű, a terület áttekintését kívánja segíteni, az összefüggések megértését támogatni. A mélyebb, matematikailag igazolt tételeket sok esetben más könyvekből ([6],[48],[65],[68]) vagy magukon az egyetemi kurzusokon ismerhetjük meg. Fontosnak tartjuk, hogy ne csak az elmélet bemutatását tartsuk szem előtt, hanem a

gyakorlatban előforduló képalkotó és képfeldolgozó rendszerek működését úgy mutassuk be, hogy közben a megértéshez szükséges, legfontosabb elméleti alapokra is rálátásunk legyen. Mivel nem egy szűk szakterületet érint a tankönyv, ezért ennek megfelelően a hivatkozott irodalom is meglehetősen sokszínű. Ahol lehetséges volt, igyekeztünk magyar nyelvű hivatkozásokat használni, egyébként pedig az általunk legjelentősebbnek vagy éppen a legaktuálisabbnak tartott forrásokat tüntettük fel.

Könyvünben törekedtünk a fejezetek sorrendjét logikusan felépíteni, azt ajánljuk az olvasóknak, ennek megfelelően forgassák a könyvet. Természetesen az egyes fejezetek között előre és hátra utalásokat is tettünk az összefüggések jobb szemléltetése érdekében.

Tankönyvünk fejezetei a Pannon Egyetem (PE) és a Szegedi Tudományegyetem (SZTE) következő kurzusai esetében járulnak hozzá a tananyag kisebb–nagyobb részének elsajátításához:

- Képi információ mérése (PE: mérnök informatikus, MSc)
- Képfeldolgozás (PE: mérnök informatikus, BSc)
- Pixelgrafika (PE: informatika tanár, MA)
- Képfeldolgozás haladóknak (SZTE: programtervező informatikus, MSc)
- Képregisztráció (SZTE: programtervező informatikus, MSc)

Reméljük, hogy könyvünket nem csak egyetemi vagy főiskolai hallgatók használják majd tanulmányaik során, hanem a képalkotás, a fényképezés és képfeldolgozás iránt érdeklődők is hasznosnak találják munkánkat.

Veszprém és Szeged, 2011. augusztus 31.

Czúni László, PE, Műszaki Informatikai Kar
Tanács Attila, SZTE, Természettudományi és Informatikai Kar

1.2. A matematikai konvenciókról

Igyekeztünk a függvények, képek ill. transzformációk egységes jelölésére:

- vékony betű skalár értéket, vastag betű vektort jelöl
- nagybetűvel mátrixot, egy függvény Fourier transzformáltját vagy képpont térbeli megfelelőjének koordinátáit jelöltük
- a nagy, kalligrafikus F betű a Fourier transzformációt jelöli (\mathcal{F})
- a többváltozós értékeket oszlopvektorként használtuk
- a képfüggvény változójára kicsi indexeket (x, y, i, j, k, l, m, n) használtunk
- az időbeli változót t -vel jelöltük
- iterációk esetén az iterációs indexet zárójeles felső index-szel jelöltük

Mivel a digitális képek esetén diszkrét függvényekkel dolgozunk, ezért a folytonos esetek megfogalmazását a legtöbb esetben kerültük, csak a diszkrét megfogalmazást tárgyaltuk.

2. fejezet

A képbevitel jellemzése

2.1. A képfeldolgozás szintjei

Akár a komplex biológiai látórendszereket, akár az elektronikus képalkotó és képfeldolgozó rendszereket tekintjük, könnyen észrevehetjük, hogy az információ egy ún. képfeldolgozási láncban keresztül halad az érzékeléstől kezdve a magas szintű értelmezésig vagy a megjelenítésig. A láncolatban való áthaladás során a képi információt hordozó jelek folyamatos átalakuláson esnek át: miután a szenzorok detektálták a vizuális információt, az első lépések során nagymennyiségű nyers adat feldolgozását kell megoldani, majd a következő lépések során egyre absztraktabb, magasabb szintű információ processzálásáról, kinyeréséről van szó.

A jelek feldolgozása az emberi látórendszer esetében már a szemben elkezdődik: kimutatták, hogy a retina neurális rétegei képesek speciális képi tulajdonságok, mint pl. az élek detektálására (lásd laterális inhibíció [33]). Azt is megfigyelhetjük, hogy ezzel analóg módon egyes elektronikai szenzorokban is hasonló mechanizmusok működnek, pl. már a kilencvenes évek elején megjelentek olyan CMOS technológiájú látóchipek, amelyek tartalmaztak zajszűrésre vagy élkiemelésre alkalmas, pixel szinten integrált funkciókat ([51], [71]). Ezek a számítási lépések – az ún. *korai látás* (*early vision*) algoritmusai – jelentik a képi információ feldolgozásának *alacsony szintjét*.

A képfeldolgozás *középső szintjén* regionális információk feldolgozása történik, mint pl. az összefüggő régiók tulajdonság alapú kijelölése. Az azonos színű vagy mintázatú területeket össze lehet vonni, így homogén tulajdonságú képszegmensek jöhetnek létre.

A *legmagasabb* feldolgozási szinten a képen lévő foltok, mint objektumok azonosítása, felismerése, szemantikus értelmezése valósul meg. Az itt alkalmazott algoritmusok tágabb értelemben tekinthetők a mesterséges intelligencia eszközeinek is.

A feldolgozási láncban nagymennyiségű képi információ áramlik át, miközben számos szándékos vagy nem szándékos, hasznos vagy káros hatás éri a jeleket. Ezek jelentősen befolyásolják a képi információ minőségét, elemzésének módszereit, legvégül pedig értelmezését.

A képi információ mérése a feldolgozási szintek mindegyikén értelmezhető, így tehát könyvünkben ezt a láncolatot szeretnénk bemutatni, a képfeldolgozás néhány tipikus alkalmazási területének tárgyalásával.

2.2. Látás hullámokkal a gépi és biológiai rendszerekben

A képi információ feladatainak megértéséhez célszerű az optikai sugárzással, mint információ továbbító közeggel kapcsolatos néhány alapfogalom és alapkérdés tisztázása.

Bár hétköznapi értelemben a fény kifejezést a látható elektromágneses sugárzás jelenségének leírására használjuk, szigorúbb megfogalmazás szerint a fény az emberben kialakuló érzetet és észleletet jelöli. Ezt az optikai sugárzásnak (100nm-1mm) egy szűkebb, látható tartománya (380nm-780nm) tudja kiváltani. Ennek a szűkebb tartománynak az egyik végénél az ultraibolya tartomány (100-400nm), a másik végénél az infra tartomány (780-1mm) helyezkedik el. A láthatóságon kívül eső részeknek is fontos szerepe van a modern képalkotásban és képfeldolgozásban, ezen tartományokban való mérésekre speciális mérőeszközök, speciális kamerák állnak rendelkezésünkre (pl. infra kamerák, orvosi képalkotó berendezések, ipari röntgenek). Annak, aki az optikai ill. látható sugárzásról ill. annak humán érzékeléséről, észleléséről és számítógépes felhasználásáról szeretne többet megtudni, több magyar nyelvű könyv is a rendelkezésére áll, példaként említhetjük Schanda János: Színtervezés számítógépes felhasználás számára című egyetemi jegyzetét [72].



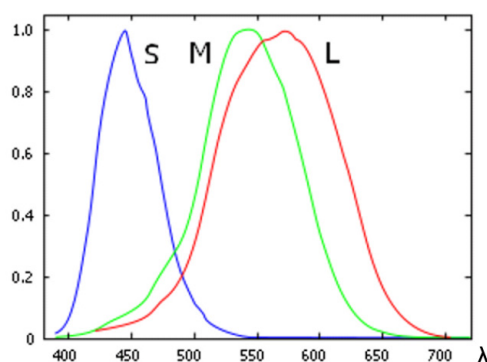
2.1. ábra. A vizuális információ feldolgozásának alapvető folyamatai az emberi látórendszerben és a gépi rendszerekben

Az elektromágneses sugárzáson kívül *részecskesugárzás* (alfa: kétszeres pozitív töltés, beta: elektronsugárzás, proton-, és neutronsugárzás) és *hangsugárzás* is használható képalkotásra. Mindhárom sugárzástípus leírható hullámként, alapvető tulajdonságuk a hullámhossz, ami lényegében meghatározza, hogy milyen mérettartományban tudunk segítségükkel méréseket végezni.

A látható sugárzás "feldolgozása" az emberi látórendszer esetén már a szemben elkezdődik, majd pedig az agy különböző területeinek bonyolult összjátéka alakítja ki a mindannyiunk által megszokott vizuális élményt. Az elektronikus és a biológiai látórendszerek között kis általánosítással könnyű párhuzamot felfedezni. A képalkotás és képfeldolgozás alapvető folyamatainak áttekintését a 2.1. ábra segíti.

Az elektronikus rendszerek sok esetben az emberi látás mechanizmusait igyekeznek lemásolni, képességeit reprodukálni, mivel az információ láncolat végén olyan képet szeretnénk megjeleníteni, ami a humán szemlélőben természetes észleletet kelt. Ez nyilvánvalóan nem minden esetben van így, gondoljunk például az orvosi képalkotásra, ahol az ember számára nem látható információt kell megjeleníteni, vagy a robot navigációra, ahol egy autonóm gépnek önállóan kell döntést hozni a vizuális információ alapján.

Az emberi szemben ún. pálcikák és csapok végzik az elektromágneses hullámok érzékelését: míg előbbiek gyenge fényviszonyok mellett és a szem periférikus területein is jelentős érzékenységgel bírnak, addig utóbbiak elsősorban nappali körülmények között, a szem központi részén érzékelik a látható sugárzást. A pálcikák nem képesek a színek megkülönböztetésére, viszont a csapokból 3 féle található, eltérő érzékenységgel: $L(\lambda)$, $M(\lambda)$ és $S(\lambda)$ (L : long - hosszú, M : medium - közepes, S : short - rövid hullámhossz tartományt jelöl). A 2.2 ábra a csapok relatív színeképi érzékenységét ábrázolja.



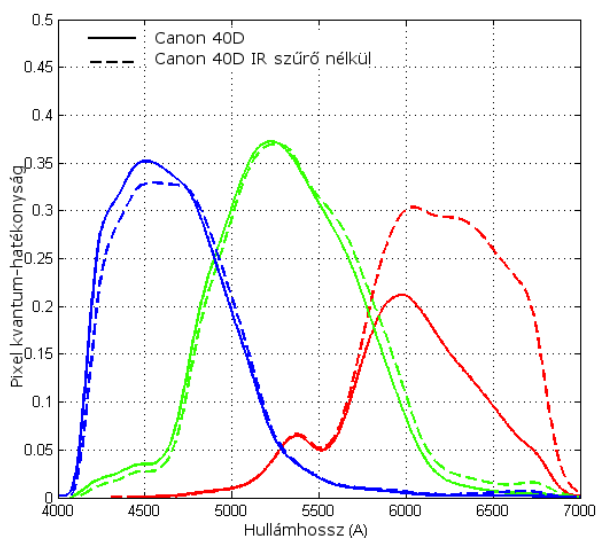
2.2. ábra. L , M és S csapok relatív színeképi érzékenysége

A színes kamerák színeképi érzékenysége az emberi szemhez nagyon hasonló módon képzelhető el: a különböző spektrális tartományok érzékelésére külön szenzorok felelősek. A kamerákból kiolvasható ún. R , G , B (R : red - vörös, G : green - zöld, B : blue - kék) értékek a három szenzor válaszának felelnek meg:

$$\begin{aligned} R &= \int L(\lambda)r(\lambda)d\lambda \\ G &= \int L(\lambda)g(\lambda)d\lambda \\ B &= \int L(\lambda)b(\lambda)d\lambda \end{aligned} \quad (2.1)$$

ahol $L(\lambda)$ a szenzorba beeső sugárzás teljesítménye adott hullámhosszon, $r(\lambda)$, $g(\lambda)$ és $b(\lambda)$ pedig a háromféle szenzor érzékenységét jellemző függvény. Természetesen a valós életben ennél az elvi modellnél valamivel bonyolultabb a konkrét R, G, B értékek meghatározása (pl. különböző szabványos színtranszformációk használata elkerülhetetlen). A képszenzorok érzékenységét ún. kvantum-hatékonysággal (*quantum efficiency*) jellemzik, ami megadja, hogy a szenzorra beérkező fotonok hány százaléka váltja ki elektron-lyuk párok kialakulását.

A 2.3. ábra a Canon 40D típusú kamera szerelt áramkör kvantum-hatékonyságát ábrázolja a gyárilag beépített IR szűrővel és annak eltávolítása után.



2.3. ábra. A Canon 40D érzékenysége gyári IR szűrővel és nélküle. A hullámhosszt sokszor nem SI mértékegységben, hanem angström-ben adják meg.

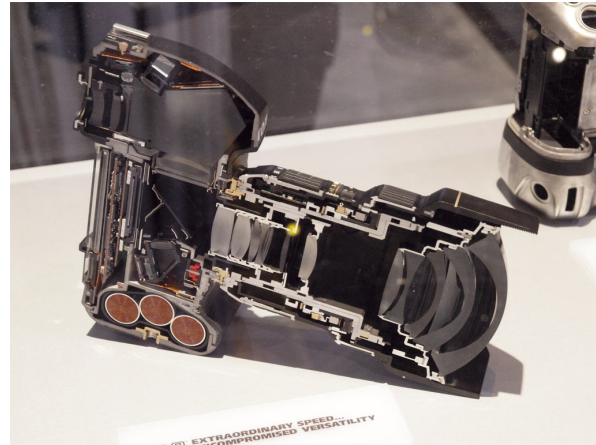
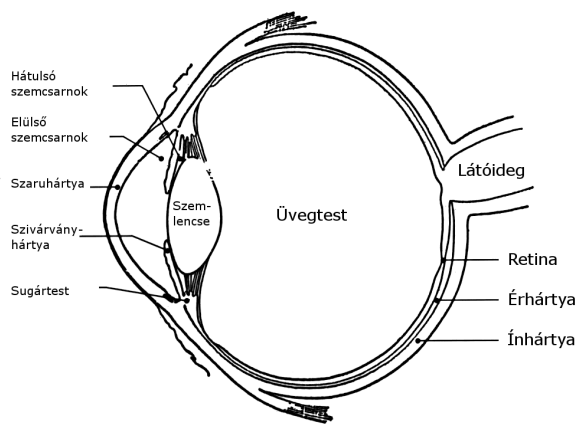
Gyakran merül fel kérdésként az emberi és a gépi látórendszerek képességeinek összehasonlítása. Érdemi összehasonlítást egy több száz oldalas tanulmány keretén belül lenne érdemes tenni, hiszen annak ki kéne terjednie az alacsony szintű jellemzőkre – mint pl. a kép térbeli és időbeli felbontása, az érzékelő spektrális érzékenységére – és a magas szintű, intelligenciát feltételező képességekre is. Nagyon tömören az összehasonlító mű konklúziója talán az lehetne, hogy az alacsony szintű jellemzők terén a gépi rendszerek már utolérték és meghaladták az ember képességeit, míg a bonyolultabb magas szintű feladatok elvégzésében még jó pár évig, évtizedig nincs veszélyeztetve az ember elsősege.

Végül érdekességképpen említjük meg, hogy 2011 óta Magyarországon is végeznek *retina implantátum* beültetéseket a Semmelweis Egyetem Szemészeti Klinikáján [56]. A beültetett chipnek köszönhetően a betegek látásában bekövetkezett javulás lehetővé teszi, hogy felismerjék a fényforrások irányát (ablak, lámpa), sötét alapon világos tárgyakat vegyenek észre.

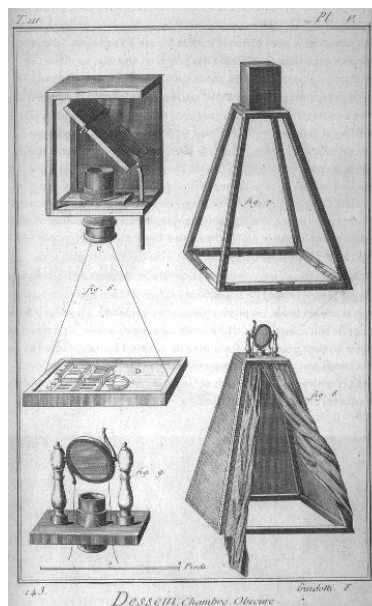
2.3. A kép projekciója

A magyar nyelvben a fényképezéshez, videózáshoz használatos képrögzítő eszköz általános elnevezése a kamera. Ennek a szónak latin és ógörög eredete is ismert, az előbbi esetén sötét kamrát, szobát (*camera obscura*), míg utóbbi esetében a *kamara* boltozatos termet, pincét, helységet jelent. Akár az emberi szem felépítését, akár egy klasszikus kamera felépítését vizsgáljuk, láthatjuk, hogy nagyon találó a kifejezés, mivel mindkét esetben egy sötét, zárt térről van szó, amin egy relatív kicsiny lyukon keresztül érkezik a fény (lásd 2.4. ábra). Természetesen az elv már több száz év óta ismert, a 2.5. ábra a fény leképezésének egy korai módját

ábrázolja a 18. századból.



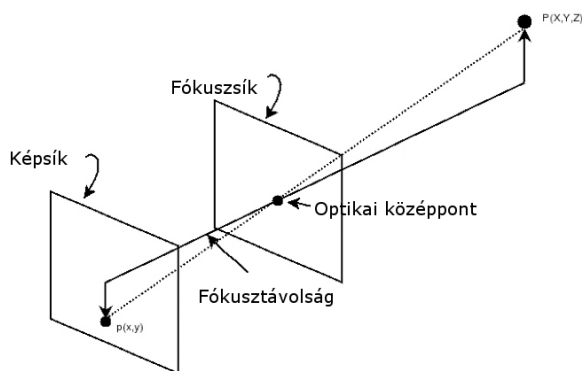
2.4. ábra. Az emberi szem és egy modern tükrorreflexes kamera keresztmetszeti képe



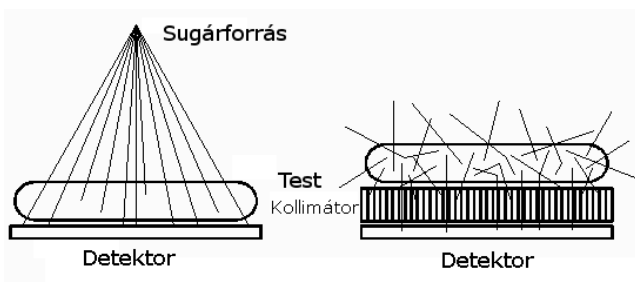
2.5. ábra. Camera obscura rajza egy francia enciklopédiából a 18. századból (*Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers*, Denis Diderot and Jean le Rond d'Alembert, 1751)

Tárgyalásunk során alapvetően kétféle képalkotási modellt fogunk használni. A klasszikus lyukkamera (pinhole) modell szerint a fény egy kis lyukon, a vetítés középpontján keresztül a kamera belső falán alakít ki fordított képet (2.6. ábra), míg a másik esetben a projekció középpontja és a fényérzékelő felület között helyezkedik el maga a vizsgált objektum, ill. a sugárforrás az objektumon belül van (lásd 2.7. ábra).

A kamerák képalkotásával kapcsolatosan bővebb információt a Kató Zoltán, Czúni László: Számítógépes látás c. egyetemi jegyzetben [48] található az olvasó.



2.6. ábra. Lyukkamera modellje, amely megfelel a perspektivikus leképezésnek.



2.7. ábra. Röntgen (bal) és gamma kamera (jobb) vázlatos képe.

2.4. Az optika szerepe és jellemzése

Az előzőekben láthattuk, hogy a képalkotás akár egy megfelelő méretű, kicsiny lyukkal rendelkező dobozzal is lehetséges, amelyben létrejön a perspektivikus leképezés. Kereskedelmi forgalomban kaphatók is igen egyszerű, optikát nélküli fényképezőgépek, de barkácsolással otthon is készíthető ilyen egyszerű készülék – a célnak akár egy konzerves- vagy cipősdoboz, de egy elsötétített hálószoza is megfelelhet (lásd 2.8. ábra).

A kicsiny nyílásból viszont következik, hogy megfelelően világos kép készítéséhez hosszú expozíciós idő szükséges, ami alkalmatlanná teszi a technikát a hétköznapi életben való normál fotózásra, hiszen a mozgó objektumok képe a felvételeken így elmosódik. Ennek a problémának a kiküszöbölésére különböző lencsét, ill. lencserendszereket építenek be a fényképezőgépekbe, kamerákba. Ezeknek a rendszereknek egyszerűsített modelljét láthatjuk a 2.9 ábrán. A gyűjtőlencse és az optikai tengely metszéspontjait (P_1 és P_2) főképnek nevezzük. A rajzon az X -szel jelölt objektum képe a lencse túloldalán x -ként áll elő. X egy kitüntetett pontjából három fénysugár van feltüntetve:

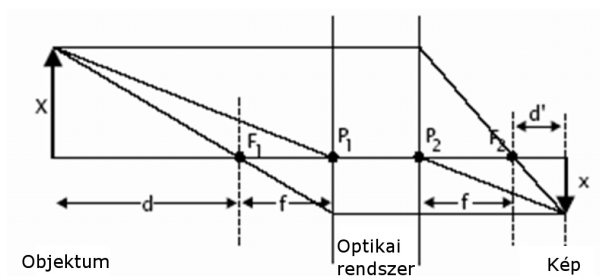
- I. az optikai tengellyel párhuzamosan induló, a lencse túloldalán F_2 fókuszpontra áthaladó sugár;
- II. az első fókusz (P_1) felé haladó, majd a második fókuszpontról (P_2) párhuzamosan továbbhaladó;



2.8. ábra. Hálószobából fényképezőgép: az elsötétített szoba ablakán egy mindössze 2cm átmérőjű lyukon érkezett be a fény. Fotó: Szász Péter – <http://bp.underground.hu>

III. F_1 fókuszpont felé haladó, majd a lencsét az optikai tengellyel párhuzamosan elhagyó.

Ahhoz, hogy éles képet kapjunk, a három fénysugárnak egy pontban kell találkozni. Ehhez d' -t megfelelően kell beállítani – d és f függvényében. Gyakorlatilag ez jelenti a kamera élesre állítását, amit köznyelven fókuszálásnak nevezünk. Az optika nagyításának változtatása a fókusz távolság változását jelenti.



2.9. ábra. Gyűjtőlencse modellje

Az *optikai alapegyenlet* a képtávolság, tárgy távolság és fókusz távolság kapcsolatát adja meg:

$$\frac{1}{f} = \frac{1}{d' + f} + \frac{1}{d + f} \quad (2.2)$$

Azaz a fókusz távolság (f) reciproka a kép- ($d' + f$) és tárgy távolság ($d + f$) reciprokanak összegével egyenlő. Mivel f és d' a kamera belső állapotát jellemzi, ebből következik, hogy amennyiben sikerült a képet élesre állítanunk, d – mint egyedüli ismeretlen – egyszerűen kiszámítható. A modern kamerákban ezt az elvet használják a tárgyak távolságának automatikus becslésére: a kamera elektronika addig állítja d' -t, míg az élességet kiértékelő függvény nem mutatja a maximumát. Természetesen a módszer pontossága függ az élességmérő eljárástól, amire hatással van magának az objektumnak a képe is (sima, kontraszt nélküli felületeken nem, vagy alig lehet a kép élességét megbecsülni).

Az optika fényerejét az ún. F értékkel lehet megadni:

$$F = \frac{f}{D}, \quad (2.3)$$

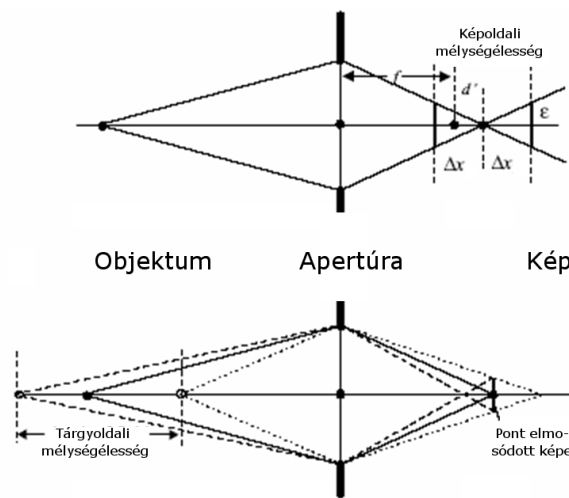
ahol D a nyílás (apertúra) átmérőjét adja meg. Azaz azonos fókusz távolság mellett kisebb F érték jelenti a nagyobb átmérőjű lencsét, így a világosabb képet.

A kamera laterális (oldalirányú) nagyítását a következő hányados fejezi ki:

$$m_l = \frac{x}{X} = \frac{f}{d} = \frac{d'}{f}. \quad (2.4)$$

Ezzel szemben az axiális (tengelyirányú) nagyítás nem konstans az optikai tengely mentén:

$$m_a \approx \frac{d'}{d} = \frac{f^2}{d^2} = m_l^2. \quad (2.5)$$



2.10. ábra. A kép- és tárgyoldali mélységélesség jelentése

Gyakran előfordul, hogy d' hibás megválasztásakor vagy a fényképezendő objektumok közeledése, távolodása miatt a kép életlen lesz. Természetesen ezeknek a hibáknak van egy bizonyos mértékű tűréshatára: amennyiben a képérzékelő síkjában még csak kevéssé tartanak szét az egy ponthoz tartozó fénysugarak, és nem egy szomszédos képérzékelő cellába csapódnak, úgy a hiba még elhanyagolható. Jelölje ε a széttartás sugarát a képérzékelő távolságában, ΔX pedig a képsík távolságát az ideális pozíciótól. Ekkor a kettő kapcsolata kifejezhető:

$$\Delta X = 2F \left(1 + \frac{d'}{f}\right) \varepsilon = 2F (1 + m_l) \varepsilon. \quad (2.6)$$

Azaz ha tudjuk, hogy mekkora ε -t tűr el a rendszerünk (ami alapvetően a képérzékelő lapka cellaméretétől függ) megkaphatjuk, hogy mekkora tartományon belül kapunk éles képet. Ezt a tartományt *képpoldali mélységélességnek* (*depth of focus*) nevezzük. Hasonlóan megadható a *tárgyoldali mélységélesség* (*depth of field*) is:

$$\Delta X \approx 2F \frac{1 + m_l}{m_l^2} \varepsilon \quad (2.7)$$

Mindkét fenti összefüggésből kiolvasható, hogy amennyiben növeljük egy optikai rendszer fényerejét (azaz csökkentjük az F -értéket) csökkenni fog a mélységelesség. Igen gyakori probléma, hogy a fenti beállítási okokból vagy az optikai rendszer egyéb hibájából a kép elmosódott, homályos lesz. Utószűréssel, a kép utólagos élesítésével valamennyire lehet korrigálni ezeket a hibákat, ennek matematikai modelljét a 4.3. fejezetben mutatjuk be.

2.4.1. Néhány speciális optika

A fenti megfontolások általában igazak a legtöbb kamerára, de természetesen sokféle speciális optikai rendszer van használatban, amelyeknek nem hagyományos módon működik a képalkotása. Néhány speciális esetet említünk meg itt röviden:

- *Telecentrikus képalkotás*: A módszer lényege, hogy a rekesz megfelelő pozicionálásával elérhető, hogy a leképezés perspektivikus hatása ne, vagy csak alig érvényesüljön. Ennek köszönhetően a távolabb lévő objektumok nem lesznek kisebbek a képen, így a telecentrikus optika mérési feladatokra igen jól használható. Ezen optikák fő alkalmazási területe az ipari méréstechnika, lásd a 9.2.1. fejezetet.
- *Nagylátószögű optikák*: A kamerák látószöge általában igen széles tartományban mozog az alkalmazási terület függvényében. Ha a látószög meghaladja a 60-70 fokot, akkor nagylátószögű optikáról beszélünk. Természetesen a látószög növekedésével az adott térszög alatt látott információ részletgazdagsága csökken változatlan szenzorfelbontás mellett, ugyanakkor az optika perspektivikus torzító hatása nő.
- *Halszem optikák*: Ezeknek az optikáknak a látószöge eléri a 180 fokot, sokszor kettőt egymásnak háttal állítva a teljes teret le tudják képezni. Működési elvük dioptrikus, azaz lencséből épülnek fel. A kép hagyományos perspektivikus megjelenítéséhez az eredeti kép geometriai transzformációja szükséges. (A 10.2. ábra mutatja egy halszem optika eredeti és perspektivikus leképezésre transzformált képét.)
- Egyéb speciális panoráma optikák: Hasonlóan a halszem optikákhoz léteznek más olyan optikai eszközök, amelyek nem csak fénytörést, hanem tükröződést is alkalmaznak¹ ahhoz, hogy a tér minél nagyobb részét a kamerába vetítsék (lásd a 2.11. ábrát). Természetesen itt is szükség van a kép geometriai transzformációjára ahhoz, hogy a megszokott négyyszög alakú, hagyományos perspektivikus torzítású képet kapjuk.

Utóbbi két kategória biztonságtechnikai alkalmazásáról a 10.2.1. fejezetben olvashatnak.

2.4.2. Optikai hibák

Általános értelemben optikai hibán értjük az optikai rendszer helytelen beállításából adódó hibákat ill. a lencse vagy lencserendszer tökéletlen kialakításából adódó hibákat. Előző csoportba tartozik a rossz fókuszálás miatti elmosódás: lásd képpoldali (2.6. egyenlet) ill. tárgyoldali

¹Amennyiben egy optikai rendszer egyaránt tartalmaz lencsét és tükröt, katadioptrikusnak nevezzük.



2.11. ábra. Viszonylag egyszerű eszközökkel (hagyományos kamera, tükör) is készíthető katadioptrikus panoráma kamera

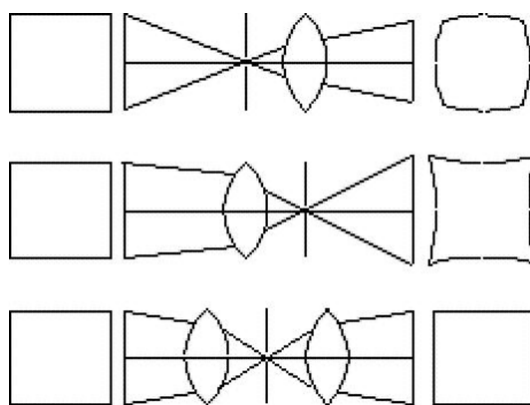
mélységelesség (2.7. egyenlet), utóbbi kategóriába pedig a következő típushibákat sorolhatjuk:

- Szférikus aberráció: A jelenség oka, hogy a lencse optikai tengelyénél, valamint a lencse szélein haladó fénysugarak fókusz távolsága eltérő, így egy adott pontból érkező fény a leképezés után foltszerűen terül el. Az optikai rendszerek gyártói a hiba javítására ún. aszférikus lencsetagokat használnak. A jelenség általában a képsarkok fokozott élettenségét okozza, fotózáskor ezt a rekesznyílás csökkentésével (rekeszeléssel) lehet bizonyos mértékig eliminálni.
- Kóma: A kóma a szférikus aberráció speciális fajtája, amikor a beeső fénysugarak a lencse optikai tengelyével nagy szöget zárnak be. Ekkor a leképezés során nem szabályos szóródási kört kapunk, hanem üstököszerű csóvát.
- Asztigmatizmus: Az optikai tengelytől távol lévő pontból induló fénysugarak közül a vízszintes síkban haladók a lencsén áthaladva nem azonos pontban fókuszálódnak, mint a függőleges síkban haladók, így eredményül egy pont helyett két rövid merőleges vonallá rajzolódnak. A két sík egyesülési pontjai között a tárgy pont képe ellipszis formájában jelenik meg. Az asztigmatizmus mértéke nagyobb beesési szög esetén jelentősebb. Napjainkban szinte minden összetett objektív tartalmaz olyan lencsetagokat, amelyek kiküszöbölik ezt a hibát.
- Képzőelhajlás: A képzőelhajlás oka, hogy a nagy kiterjedésű tárgysík pontjairól vetített éles kép a lencse görbületéhez hasonló gömbfelületen keletkezik, nem pedig síkban, azaz a képérzékelő síkjában nem keletkezik éles kép a tárgysík minden pontjáról. A jelenség mértéke függ a lencse alakjától: a kétszer domború lencsénél a legnagyobb, míg az ún. meniszkusz-lencsénél² a legkisebb. Kiküszöbölése megfelelő lencsetagok-

²Meniszkusz lencse: Olyan lencse, amelynek az egyik oldala konvex (domború), a másik konkáv (homorú).

kal lehetséges, az effektus mértéke rekeszeléssel csökkenthető.

- Fényelhajlás³: A fény útjába eső objektumok peremén a legszélső fénysugarak elhajlanak és gyenge szórt fénnel világítják meg az árnyékteret. Az objektívben lévő rekeszlamellák is fényelhajlást okoznak, ezért az objektív erős lerekeszelésénél mindez kontrasztcsökkenéshez és részletvesztéshez vezethet.
- Kromatikus aberráció: A különböző hullámhosszúságú sugarakra a lencsék törésmutatója eltérő: az ibolyaszínű sugarak törnek meg a legjobban, míg a vörösek a legkevésbé. Egy adott pontból a lencsére érkező fehér fénysugár összetevőire bomlik és különböző színű képei nem egy pontban, hanem egymás mellett jelennek meg. Kiküszöbölése alacsony diszperziójú lencsetagokkal történik. Fontos a lencsetagok precíz elhelyezése és összeillesztése, mert a hőtágulásból adódó elmozdulás élettenséghöz és más leképezési hibákhoz vezethet.
- Becsillanás (belső tükröződés): Erős fény hatására az objektív lencsetagjairól és egyéb belső alkatrészekről, vagy akár a képérzékelőről visszaverődő fénysugarak becsillanást okozhatnak, amely különböző foltok formájában jelenhet meg a képen, csökkenhet a kép kontrasztja és hamis elszíneződések jelenhetnek meg. A probléma csökkentésére a lencsék felületét tükröződést gátló bevonattal vonják be. Fényellenző használatával a jelenség jelentős mértékben mérsékelhető.

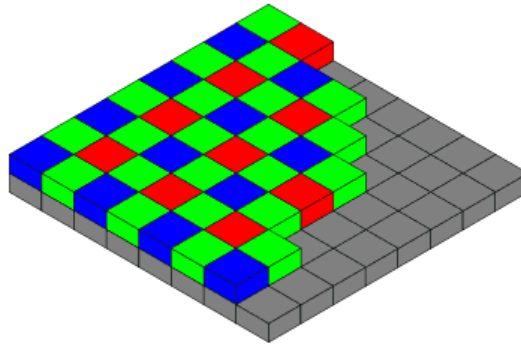


2.12. ábra. Geometriai torzítás a rekeszszerkezet függvényében: Első sor – hordó torzítás. Második sor – párna torzítás. Harmadik sor – torzításmentes kép.

- Képsarkok sötétedése (vignettálás): Az optikai tengelyhez képest ferdén haladó fénysugarak egy részét eltakarja az objektív foglalata, amely a képsarkok megvilágításának jelentős csökkenését okozza. Jelentősebb mértékben a nagy látószögű objektíveknel jelentkezik, rekeszeléssel csökkenthető, de a legmodernebb fényképezőgépek már szoftveresen is tudják korrigálni a vignettálást.
- Hordó és párnatorzítás (lásd 2.12): Geometriai torzítás során a tárgysíkban még egyenes vonalak a képsíkban nem egyenesként, hanem torzítva jelennek meg. A torzítás

³A fényelhajlás elméleti hátterét Fraunhofer diffrakció néven találhatjuk az irodalomban.

oka, hogy az objektív nagyítása nem állandó, hanem az optikai tengelytől távolodva változik. Ha a nagyítás mértéke a tengelytől távolodva nő, a torzítás párna alakú, ha pedig kisebb, akkor hordó alakú. A torzítás nagyban függ a rekeszszerkezet elhelyezésétől is. Ha a rekesz a lencse előtt van, akkor hordótorzítás, ha pedig a lencse mögött van, akkor párnatorzítás jelentkezik. A torzítás mértéke korrekciós lencsékkel csökkenthető. A gyakorlatban nagylátószög esetén általában hordó formájú a torzítás, míg teleobjektíveknél párna alakú. A geometriai hibák utólagos javítását a kép koordináta-transzformációjával lehet elvégezni. A különböző vetemítő (*warping*) módszerek leggyakrabban radiális vagy tangenciális transzformációval modellezik a kép torzítását és visszaállítását.



2.13. ábra. A Bayer szűrő mintázata. Mivel az emberi szem a zöld tartományban mutatja a legnagyobb érzékenységet, a zöld pixelek száma kétszerese a kéknek és pirosnak.

További olvasmánynak ajánljuk Ábrahám György: Optika című könyvét [6].

2.5. CCD és CMOS képérzékelők

A digitális kamerák alapvetően kétféle képérzékelő chip valamelyikét használják: CCD (Charge Coupled Device - töltéscsatolt eszköz) vagy CMOS (Complementary Metal Oxide Semiconductor - komplementer fém-oxid félvezető) szenzorok gyűjtik a kamerába jutó fotonokat, és alakítják át elektromos jellé a hullámokban terjedő információt. A képérzékelőkből 1 db ill. a drágább készülékekben akár 3 db is elhelyezkedhet. Előbbi esetben az ún. Bayer szűrő (lásd 2.13. ábra) gondoskodik arról, hogy az egyes pixelek R, G, B csatornáknak megfelelő értékeket adjanak, majd pedig interpolációs szűrőkkel (*demosaicing*) fognak minden pixelhez R, G, és B értéket számítani. Ettől a technikától egyedül az ún. Foveon érzékelő tér el, ahol a különböző színek érzékelése nem egymás melletti pozíciókban, hanem egymás alatti rétegekben történik (lásd [29] ill. 3.4 fejezet). 3 külön érzékelő esetén prizma segítségével juttatják el az optikai sugárzást az egyes R, G, B chipeknek.

2.5.1. CCD és CMOS összehasonlítása

Mindkét érzékelő fotodiódákból épül fel, ahol a beérkező fotonok hatására áram alakul ki. Minél több a becsapódó foton, annál nagyobb az áram, és jobban töltődnek a miniatűr kondenzátorok, amelyek az elektronikus képet hozzák létre. A két érzékelő azonban mind gyártási technológiában, mind működési mechanizmusában eltér.

A modern CCD alapjait – az analóg léptetőregisztert – George E. Smith és Willard S. Boyle 1969-ben fejlesztették ki a Bell Labsnál ⁴. Az analóg léptetőregiszter működésének lényege, hogy adott órajel hatására a bemeneti oldalon levő töltést mozgásra lehet bírni, azaz léptetni lehet a kimeneti oldal felé. Hamar világossá vált, hogy nem csak elektronikusan lehet feltölteni a regisztereket, hanem a fény fotonjaival is. 1970-re képesek voltak képet létrehozni az új eszközzel, s így megszületett a CCD. A CCD is MOS (Metal-Oxide-Semiconductor) technológiával készül, kereskedelmi forgalomba 1974-ben a Fairchild által gyártott, 100 x 100 pixeles érzékelő került először.

A CCD áramkörök esetében az érzékelők értékeit sorosan kell kiolvasni, egy-egy pixel megcímzésére, egyenkénti kiolvasására nincs lehetőség. A vezérlő áramkörök, az analóg-digitális átalakítók az érzékelő felületen kívül helyezkednek el. Ennélfogva csak a sorok végén lehet érzékelni a képpontok töltését, ahhoz, hogy az egész sor értékét megkapjuk, az egyes töltéseket pixelenkénti ugrással kell eljuttatni a sor végére.

A CMOS esetében ezzel szemben minden cella külön címezhető, a feszültségértékek egyével állnak elő és kiolvashatók, a digitális átalakítás itt is az érzékelő területen kívül történik. Mivel mindez nagyfokú integráltságot, egyenletes minőséget igényel, ezért eleinte csak a CCD-k voltak képesek jó minőségű képek előállítására. A szilícium technológia fejlődésével azonban sikerült a kezdeti problémákat megoldani, napjainkra a CMOS gyártás-technológiája valamivel egyszerűbb és olcsóbb is lett, mint a CCD áramköröké, ennélfogva az olcsó kameráknál CMOS érzékelő lapkát használnak előszeretettel. (Itt jegyezzük meg, hogy sok alacsony árú, CMOS-t alkalmazó kamerában sem elektronikus, sem mechanikus zárat ill. átmeneti tárolót sem használnak. Ennek következménye, hogy gyors mozgások esetében a tárgyak alakja torzul, hiszen kiolvasás közben a tárgy pozíciója megváltozik. Ez a jelenség tehát nem magának a CMOS technológiának a hátránya, de gyakran velejárója az olcsó kamerákban.)

2.1. táblázat. CCD és CMOS összehasonlítása

	CCD	CMOS
érzékenység	+	-
zaj	-	+
fogyasztás	-	+
sebesség	-	+
szűrők integrálása	-	+
gyártási költség	-	+

⁴A két feltaláló – Charles K. Kao-val, a fényvezető száloptika kutatójával megosztva – 2009-ben találmányukért Nobel díjat kapott.

A két technológia összehasonlítását a 2.1. táblázat tartalmazza. Bár egyik eszközről sem mondhatjuk általánosságban, hogy jobb minőségű képet készít a másiknál, a táblázat alapján úgy tűnik, a CMOS több perspektívát rejt magában. Egyedüli hátránya az érzékenység, amit a többi előnyös tulajdonsággal kompenzálni lehet, így összességében kiváló képminőséget ér el.

Napjainkban újabb technológiai fejlesztések folynak, amelyeknek két fő iránya körvonalazódik:

- ún. sCMOS hibrid architektúra, ahol CMOS kiolvasó áramköröket kapcsolnak CCD képalkotó szubsztráthoz;
- a CMOS érzékelők gyártási módszereit, nagyon kisméretű technológiáját alkalmazzák CCD struktúrájú érzékelők előállításához: az egyes poly-szilícium kapukat nagyon kicsi résekkel szeparálják el.

3. fejezet

A kép felbontása

A digitális kép diszkrét elemekből áll, a folytonosnak képzelt világ optikai leképezését véges számú ponttal reprezentálja, ahol a meghatározott geometriai struktúrában elhelyezkedő pontok számértékének ábrázolási pontossága is korlátozott.

Valójában az analóg fényképezésnél is vannak a felbontásnak korlátai. Nem csak az optika minősége (lásd lencse aberrációk), a fényelhajlás jelensége (Fraunhofer diffrakció), hanem az információ tárolására szolgáló film kristályszerkezete is meghatározza a rögzített kép információtartalmát, részletgazdagságát. A hagyományos analóg film érzékenység-növelésének alapvető eszköze, hogy az emulzióba nagyobb méretű fényérzékeny ezüst szemcséket helyeznek. Így gyengébb fényviszonyok között hiába érkezik kisebb valószínűséggel foton egységni idő alatt adott területre, a nagyobb szemcséket nagyobb valószínűséggel találja el egy-egy energiacsomag fotokémiai reakciót indukálva és kiváltva a film átlátszóságának csökkenését. A normál ISO 100-as helyett érzékenyebb film (pl. ISO 400) használatának következménye – a világosabb kép mellett – a kevésbé részletgazdag, szemcsés rajzolat. (Ehhez hasonló – bár ellentétes irányú – jelenség volt megfigyelhető a CCD képérzékelők felbontásának növekedése során is. A digitális fényképezőgépek tömeges elterjedésével párhuzamosan néhány év alatt megtöbbszöröződött a képet alkotó pixelek száma – tehát elvileg nőtt a képek részletgazdagsága – de a kisebb cellaméret miatt csökkent a pixelek érzékenysége, azaz romlott a jel-zaj viszony – egyébiránt azonos technológiai színvonalat feltételezve.)

A kép készítése, tárolása és feldolgozása során fontos, hogy mind a színminták számát, mind az ábrázolási pontosságot korlátok között tartsuk, de ezzel a korlátozással lényegi információt ne veszítsünk a jelből. Természetesen annak az eldöntése, hogy mi a lényegi információ nem triviális feladat. Sokszor a szakemberek is nehezen jutnak dűlőre az ügyben, hogy mi az a felbontás, ami egy adott alkalmazás esetén minőségileg már elfogadható, de ugyanakkor a rögzítés, tárolás és továbbítás költségeit tekintve még megengedhető. Matematikai értelemben egyszerűbb a helyzet: Shannon¹ mintavételi tétele [74] egyértelműen megmondja, hogy mi az a minimális mintavételi frekvencia, ami információvesztés nélkül lehetővé teszi sávkorlátolt jelek digitalizálását.

¹Claude Elwood Shannon (1916–2001): amerikai matematikus és villamormérnök, az információ elmélet megalapozója, a digitális áramkörtervezés úttörője.

3.1. Digitalizálás

Az analóg jelből digitális jellé való átalakítás két fő részre bontható:

- *Mintavételezés*: Az időben ill. térben folytonos reprezentációt bizonyos időpontokban ill. térbeli pozíciókban fogjuk mintavételezni. Általában egyenletes mintavételt alkalmazunk, de elvileg lehetséges, hogy a tartalomtól függően dinamikusan változtassuk meg a mintavétel sűrűségét.
- *Kvantálás*: A mintavételezés során előálló minták értékeit adott kvantálási szinteknek megfelelő értékekkel helyettesítjük. A kvantálási szintek távolsága nem feltétlenül egyenletes, de triviális esetektől eltekintve a kvantálás mindig információvesztéssel jár.

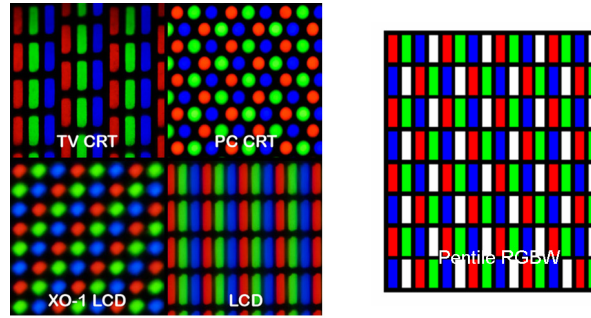
A digitalizált kép esetében beszélhetünk a kép felbontásáról és színmélységéről. Előbbi a mintavételezés sűrűségéből adódik, a képérzékelő lapka mintavételezi és elektronikus jelekké alakítja a felszínére érkező "folytonos" optikai sugárzást. (Mint korábban, a 2.5. fejezetben láttuk, sok esetben a mintavétel valójában kisebb felbontáson, azaz kisebb térfrekvencián történik, mint azt ahogy a kamerából kapott képből feltételeznénk, a kamerából normál módban kiolvasott kép felbontását interpolációval határozzák meg.) A színmélység megfelel a kvantálási szintek számának, de itt sem érdemes egyenlőséget tenni akár a színhűség és színmélység ill. színben való gazdagság és a színmélység között. Egy kép színmélységén azt értjük, hogy hány bitnyi információval kódoljuk a színcsatornák értékeit, de adott színmélység nem jelenti azt, hogy az adott kép ténylegesen minden színkódot használ, vagy hogy bár sokféle színkód elő is fordul a képen, de azok nagy része nem valamilyen zajhatásból adódik. (Félrevezető lehet, hogy néhány szkennert gyártója színcsatornánként 10 vagy 12 bites színmélységet ad meg, de a készülék A/D átalakítójának pontossága valójában csak 8 bites. Ez utóbbi természetesen nem derül ki a műszaki specifikációkból.)

(Orvosi képalkotó berendezésekre jellemző, hogy 14 vagy 16 biten szolgáltatnak adatokat, de a teljes dinamika tartománynak csak egy relatív szűk, 8-12 bites részében van zajtól eltérő lényegi információ.)

3.2. Reprezentáció pixel térben

A kép rögzítése ill. megjelenítése során pixelek ill. az azokat alkotó *szubpixelek* formájában áll rendelkezésünkre. Leggyakrabban szabályos rácpontokban, egyenletesen helyezkednek el az RGB komponensek, de léteznek olyan speciális struktúrák, ahol a képmegjelenítés bizonyos szempontjaira (pl. a kijelző effektív felbontása ill. fogyasztása) optimalizálva, ettől eltérő mintázatokat alkalmaznak. Szubpixel elrendezésekre mutat példát a 3.1. ábra, amelynek bal oldalán kétféle CRT (*Cathod Ray Tube*, katódsugárcső) és LCD (*Liquid Crystal Display*, folyadékkristályos kijelző) mintázatot láthatunk. Az ábra jobb oldali mintázata az ún. PenTile RGBW struktúra², amely az LCD-hez képest 33%-kal kevesebb subpixel tartalmaz, mégis ún. *effektív felbontása* azonos.

²PenTile RGBW: A Clairvoyante, Inc. által kifejlesztett, jelenleg a Samsung által birtokolt technológia.

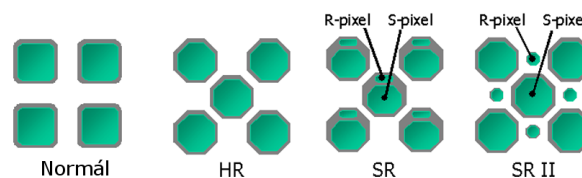


3.1. ábra. A bal oldalon hagyományos szubpixel elrendezések, jobb oldalon az ún. PenTile RGBW szubpixel-struktúra látható.

Mindeddig fizikai felbontásról volt szó, azaz a felbontás megfelelt a képet alkotó elemek számának. Ezzel szemben az *effektív felbontás*³ a szemlélő által észlelt felbontást veszi figyelembe: adott körülmények (látószög) között milyen részletgazdag képet tud egy megjelenítő produkálni [70].

Mint a 2.2. fejezetben láthattuk, a színes képalkotáshoz alapvetően három színcsatorna szükséges, de pl. a PenTile RGBW technológia már négy subpixelt használ a képek hatékony megjelenítéséhez. (Itt jegyezzük meg, hogy egyes képfájl formátumok a képszerkesztő, ill. képmegjelenítő alkalmazások támogatására egy ún. alfa-csatornát is bevezettek, amely a pixelek átlátszóságának mértékét kódolja.)

Természetesen az itt említettek mellett nem csak a képmegjelenítésben, hanem a képszenzorok között is jelennek meg újabb struktúrák. Legsikeresebb piaci jelenlétet a Fujifilm által kifejlesztett ún. Super CCD tudhat magáénak. A 3.2. ábrán láthatjuk, hogy a klasszikus elrendezésű és alakú változatnál több pixel helyezhető el egységnyi érzékelő felületen, ha nyolcszögletű pixelekből építkeznek. A Super CCD újabb változatánál nem csak a felbontást tudták növelni, hanem az érzékenységi dinamikát is, mégpedig különböző méretű fotodiódák beépítésével.

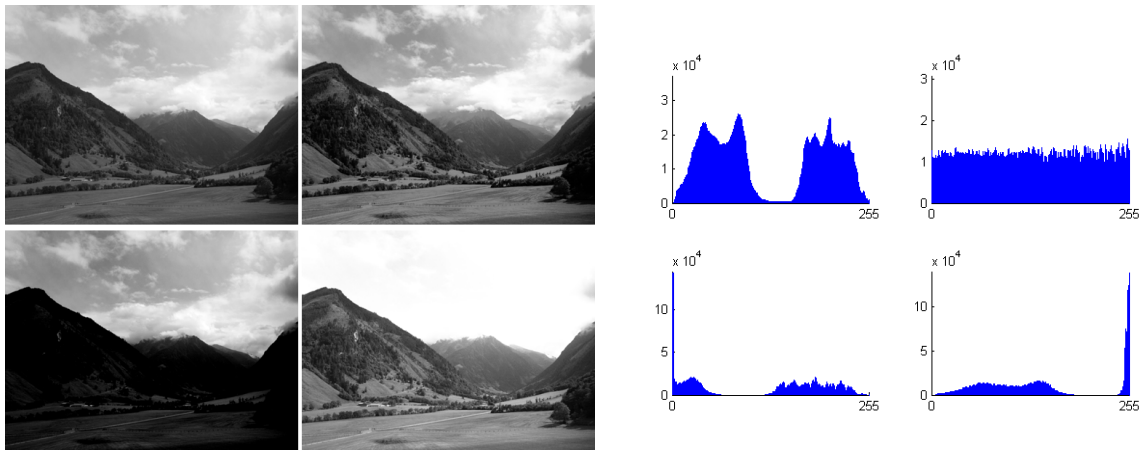


3.2. ábra. Fényérzékeny cellák normál elhelyezkedése és a Fujifilm által kifejlesztett Super CCD különböző generációi (HR: High Resolution, SR: Super Dynamic Range). Az SR és SR II. esetében a cél nem csak a felbontás, hanem a dinamika-tartomány növelése is volt.

A digitális kép képpontjaiból többféle statisztikai jellemző származtatható, mint például a színek gyakorisága (ennek grafikus megjelenése a kép hisztogramja), a színek várható

³Az *effektív felbontás* kifejezést használják annak a jelzésére is, hogy egy CCD vagy CMOS áramkör esetén valójában mennyi azon cellák száma, amelyek ténylegesen résztvesznek a fény érzékelésében, és nem egyéb jelfeldolgozó funkciót töltenek be, miközben felületük nem érzékeny a beeső fényre.

értéke, szórása, entrópiája. Természetesen ezeknél jóval összetettebb jellemzők és leírók is használatosak, gondoljunk például az MPEG 7 szabványra ([40]), amely a képek tartalom-alapú visszakereshetőségét támogatja tartalmi leírók szabványosításával (bővebben lásd az 5. fejezetet). A következő ábra egy szürkeskálás kép különböző változatait és azok hisztogramját mutatja. Az alulexponált kép hisztogramjából kiolvasható, hogy nagyon sok pont értéke 0, ami arra enged következtetni, hogy a kép rögzítése során információvesztés történt, mivel igen sok képpont értéke a legkisebb ábrázolható kóddal egyenértékű. A túlexponált kép esetén hasonló történt, de ott a fehér 255-ös kódnál látható csúcs jelzi, hogy a felhők rajzolata már nem látható. A hisztogramkiegyenlítés [68] segítségével a hisztogram közel vízszintessé válik és a kontraszt jelentősen megnő.



3.3. ábra. Kép és különböző változatai a megfelelő hisztogramokkal. BF: normál expozíció; JF: a kép kiegyenlített hisztogrammal; BA: alulexponált kép; JA: túlexponált kép

3.3. Reprezentáció frekvenciatérben: felbontás hullámfüggvényekre

A képeken alkalmazhatunk különféle transzformációkat, és tárolásukat, feldolgozásukat elvégezhetjük a transzformációk után, ezekben az ún. transzformált terekben. Ilyen esetekben nem közvetlenül a pixelek értékeit tároljuk, vagy alakítjuk tovább, hanem a transzformáció után megkapott ún. transzformációs együtthatókat. Amennyiben a képet meg szeretnénk jeleníteni, el kell természetesen végezni az inverz transzformációt, ami az együtthatókból a pixelek kiszámítását végzi. Könyvünk további részében (pl. 4. fejezet) látni fogjuk, hogy igen sok előnye van annak, hogy nem a pixeltérben, hanem pl. a frekvenciatérben dolgozzuk fel a képi adatokat.

A transzformációk jelentős része az ún. unitér transzformációk csoportjába tartozik. Az unitér transzformációk olyan lineáris, invertálható transzformációk véges dimenziós térben, ahol a transzformációs kernel orthogonális és igazak a következő egymással ekvivalens megállapítások (U transzformációra megfogalmazva):

- U unitér transzformáció;

- $UU^*T = I$ (I az ún. identikus mátrix), azaz U inverze komplex konjugáltjának transzponáltja;
- $\langle f|g \rangle = \langle Uf|Ug \rangle$, ahol f és g képfüggvények a véges dimenziós tér elemei, $\langle \cdot | \cdot \rangle$ pedig a skaláris szorzatot jelöli;
- U oszlopai és sorai ortonormált bázist alkotnak.

A harmadik pontban megadott ún. normatartó tulajdonság igen fontos, ez ugyanis annyit tesz, hogy két függvény skaláris szorzattal való összehasonlítását elvégezhetjük csakúgy a pixeltartományban, mint a transzformált tartományban (ill. ha $f = g$, akkor f energiájának számítását is végezhetjük a transzformált térben). Képek összehasonlításáról bővebben a 4.2. fejezetben olvashatunk.

Az unitér transzformációk családjába igen sok transzformáció tartozik, mint pl. a Fourier, cosinus, Hadamard, Haar, Hartley. Könyvünkben csak az elsőt fogjuk bemutatni, mivel tradícionálisan ennek a transzformációnak a legjelentősebb a felhasználása a képek feldolgozásában és tömörítésében.

A transzformációknak általában folytonos verziója is ismert, de mivel esetünkben a képek diszkrét függvényeire alkalmazzuk őket, ezért csak a diszkrét esetet tárgyaljuk. Az egyszerűség kedvéért a képfüggvényünk kétváltozós (j, k indexek jelölik a sorokat és oszlopokat) és skalár értékű.

A mérnöki alkalmazásokban talán az egyik legjelentősebb unitér transzformáció a Fourier transzformáció. Egy 2D-s f jelű képfüggvény diszkrét Fourier transzformáltja:

$$F(u, v) = \frac{1}{N} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} f(j, k) \exp\left\{\frac{-2\pi i}{N}(uj + vk)\right\}, \quad (3.1)$$

ahol $i = \sqrt{-1}$, és N a kép vízszintes és függőleges mérete. Míg j, k egy adott pixel sor és oszlop pozícióját rögzíti, u és v jelenti a függőleges és vízszintes frekvenciákat. Azaz a Fourier transzformáció adott u és v frekvenciákhoz egy komplex számot rendel, ami megmondja, hogy az f függvényben milyen súllyal szerepel az adott frekvenciájú komplex hullámfüggvény (bázisfüggvény). $F(u, v)$ értékeit frekvencia együtthatóknak nevezzük. A fenti egyenlettel egyenértékű a következő felírás:

$$F(u, v) = \frac{1}{N} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} f(j, k) \left\{ \cos\left(\frac{2\pi}{N}(uj + vk)\right) - i \sin\left(\frac{2\pi}{N}(uj + vk)\right) \right\}, \quad (3.2)$$

azaz az exponenciális bázisfüggvény megadható szinusz és koszinusz periódikus függvények segítségével is.

A fenti egyenletekből látható, hogy maga a transzformáció valójában egy korrelációnak felel meg, ahol a diszkrét képfüggvényt összekorreláljuk az egyes frekvenciák által specifikált bázisfüggvényekkel: minél nagyobb a hasonlóság az adott frekvenciájú bázisfüggvényekkel, annál nagyobb lesz az adott együttható értéke.

Az inverz transzformáció alakja nagyon hasonló:

$$f(j, k) = \frac{1}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u, v) \exp\left\{\frac{2\pi i}{N}(uj + vk)\right\}. \quad (3.3)$$

A Fourier transzformáció egyes tulajdonságai (linearitás, skálázás, eltolási tulajdonság, konvolúciós tulajdonság, Parseval egyenlőség, autokorrelációs elmélet, differenciál tulajdonság) nagymértékben hozzájárulnak széles körű felhasználhatóságához, de ezek kifejtése, magyarázata terjedelmi okokból nem került bele jegyzetünkbe [1]. Egyedül a konvolúciós tulajdonságot ismertetjük röviden, miszerint:

$$f * g = \mathcal{F}^{-1}\{F \cdot G\}, \quad (3.4)$$

ahol \mathcal{F}^{-1} jelöli az inverz Fourier transzformációt, F és G felel meg f és g függvények Fourier transzformáltjainak. A konvolúció matematikai művelete:

$$(f * k)(i, j) = \sum_m \sum_n f(m, n) \times k(i - m, j - n), \quad (3.5)$$

ahol a két szumma csupán a konvolúciós kernel méretének megfelelő területen végzi az összegzést. Mivel a konvolúció a képfeldolgozás egyik leggyakrabban használt matematikai művelete (bővebben a 4.3. fejezetében olvashatunk alkalmazásáról) ill. a Fourier transzformáció a jelfeldolgozásban az egyik leggyakrabban használt unitér transzformáció, könnyen belátható, hogy a 3.4 egyenlet által kimondott átjárhatóság igen sok esetben szerepet ad mind a Fourier transzformációnak, mind a konvolúciónak.

A Fourier transzformáció rendelkezik gyors implementációval (Fast Fourier Transform – FFT [24]), a legtöbb mérnöki alkalmazásban az FFT-t alkalmazzák, aminek a komplexitása $O(N^2)$ helyett csupán $O(N \log N)$.

3.4. A mintavételezés korlátai, átméretezés

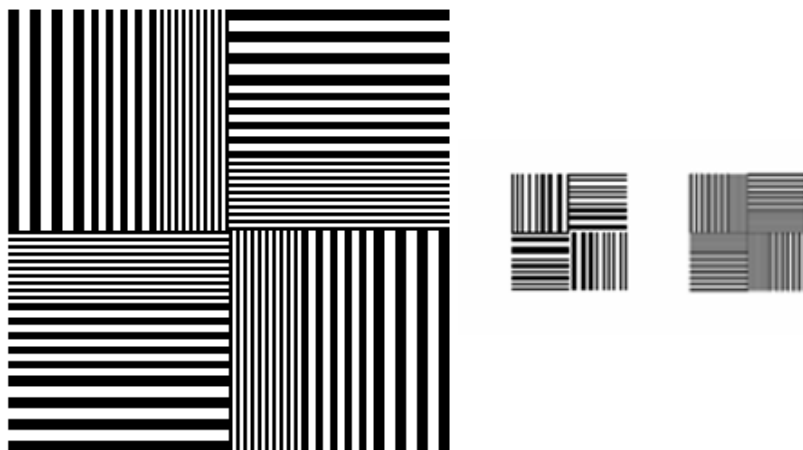
A fejezetünk elején említettük, hogy egy analóg jel mintavételezésekor van egy elméleti korlát, ami meghatározza, hogy mi az a mintavételi frekvencia, ami feltétlenül szükséges ahhoz, hogy a diszkrét jelekből az analóg verzió később hibamentesen visszaállítható legyen, azaz a mintavétel ne okozzon információvesztést. Shannon mintavételi tétele kimondja: ahhoz, hogy a sávkorlátolt analóg jel visszaállítható legyen, a mintavételi frekvencia a jel maximális frekvenciájának kétszeresét el kell, hogy érje [74]. Ezt a frekvencia kritériumot Nyquist⁴ kritériumnak nevezik. A tétel bizonyítását nem mutatjuk be dolgozatunkban, az több forrásból elérhető az érdeklődők számára.

A gyakorlati életben mind a túlmintavételezést (indokolatlanul nagy mintavételi frekvencia megválasztása), mind az alulmintavételezést érdemes elkerülni. Míg az előbbi esetén feleslegesen sok mintát tárolunk, addig az utóbbi esetén zavaró, ill. félrevezető képi hibák jelentkeznek. Ezt a fajta hibát spektrumátfedési hibának, aliasing hibának, ill. Moiré hatásnak is nevezik.

A spektrumátfedési hiba nem csak akkor fordulhat elő, ha analóg jelek digitalizálásáról van szó, hanem akkor is, ha egy adott felbontású digitális jelet akarunk kicsinyíteni egyszerűen bizonyos pixelek elhagyásával. Erre a 3.4. ábra mutat példát, ahol a középső képen a vastag

⁴Harry Nyquist (1889. február 7. - 1976. április 4.) svéd származású amerikai villamosmérnök, aki műszaki kutatásait az AT&T kutató részlegénél (később Bell Telephone Laboratories) folytatta és legfontosabb eredményei az információ-elmélet területén születtek.

és vékony vonalak sorrendisége tűnik felcserélődni az alulmintavételezés miatt. A simított esetben kicsi homályosabb a kép, de a hiba nem jelentkezik.



3.4. ábra. Alulmintavételezési hiba kép kicsinyítésekor. Bal oldal: eredeti kép. Középen: ötödére kicsinyített kép az oszlopok és sorok elhagyásával. Jobbra: kicsinyítés előszűrés alkalmazásával.

Amennyiben az alulmintavételezést el akarjuk kerülni, akkor előszűrést kell alkalmazni, ami egy simító jellegű szűrőt jelent a magas frekvenciás komponensek szűrésére. Egy képfeldolgozó program esetében ez a kép egyszerű konvolúciós előszűrését jelenti (ill. ezzel ekvivalens aluláteresztő szűrő alkalmazását a frekvenciatérben), míg a képérzékelő szenzorok esetén azok felületére ültetett – lencseként funkcionáló – optikai réteget.

Az érzékelő gyártók igyekeznek minél nagyobb felbontású szenzorokat előállítani csökkentve az alulmintavételezés esélyét. A Foveon nem egyedül a pixelek számának növelésében, hanem azok speciális elhelyezésében látja a probléma megoldását [29]: az RGB színcsatornákért felelős pixelek egymás alatt helyezkednek el.

A mintavétellel előállított diszkrét adatsorozatunkból interpoláció segítségével kaphatjuk vissza az eredeti jelsorozatot. Belátható, hogy az ideális, veszteség nélküli visszaállítás *sinc* (azaz $\frac{\sin(x)}{x}$) függvénnyel való interpolációval lehetséges. Mivel a *sinc* interpolációs függvény végtelen kiterjedésű, ezért ennek korlátozott méretű verzióját alkalmazzák a gyakorlatban Lanczos szűrő⁵ néven. A Lanczos szűrőnél is egyszerűbb közelítő megoldás valósítható meg ún. spline vagy bicubic interpolációval.

⁵Cornelius Lanczos (Székesfehérvár 1893. február 2. - Budapest 1974. június 25.) szakmai munkájának nagy részét az USA-ban és az írországi Dublinban végezte, kimagasló eredményeket ért el elsősorban az alkalmazott matematika területén. Közvetlenül az 1974-es ELTE-n tett látogatása után hunyt el Magyarországon.

3.4.1. A kép kicsinyítése, nagyítása

A különböző képfeldolgozó eljárások során gyakran van szükség a képek kicsinyítésére, nagyítására. Bár ezekben az esetekben nem analóg, hanem diszkrét adatokat dolgozunk fel, a mintavétellel és visszaállítással kapcsolatos elméletek alapvetően érvényesek maradnak.

Vegyük példának a kép sorozatos kicsinyítését és ezáltal egy kép-piramis kiszámítását. Általában a piramis minden szintje fele akkora felbontással (negyed akkora területtel) rendelkezik, mint a közvetlenül alatta lévő szint. A piramist elsősorban olyan esetekben használják, amikor:

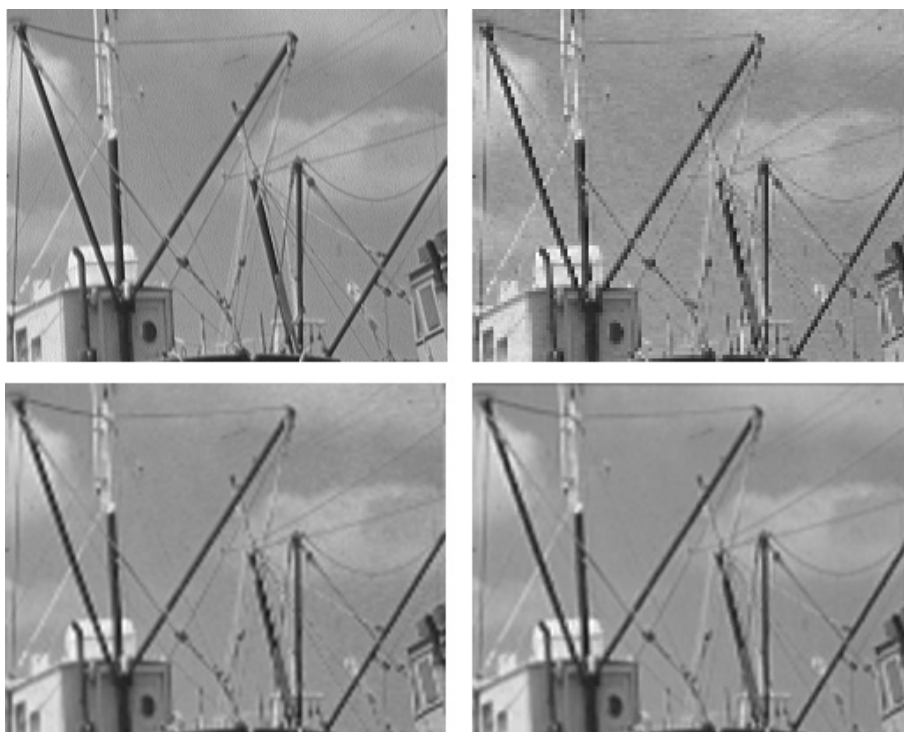
- egy feladat gyors közelítő megoldása kisebb felbontáson is elérhető;
- a keresett objektum vagy képi tulajdonság méretét előre nem tudjuk, ezért többféle nagyításban is keresni kell;
- valamilyen operátornak kicsi a hatóköre, a kép kicsinyítésével implicit kiterjeszhető a hatókör (pl. optikai áramlás számítása a blokkegyezés módszerrel (lásd 10.3.3 fejezet)).

A piramis szintjeinek kiszámítása során minden újabb szint alulmintavételezést jelent. Ezért Shannon tételéhez igazodva először a nagyobb felbontású képet simítani kell, majd el lehet hagyni minden 2. pixelsort és pixeloszlopot. Ha egy kép nagyítására van szükség, tekinthetjük úgy, hogy a kiindulási állapot egy korábbi mintavétel eredménye, ezért *sinc* interpoláció fogja a legjobb nagyítást eredményezni.

A kép szuperfelbontása

Szuperfelbontásról akkor beszélhetünk, ha a kiindulási kép nagyítása során olyan finom részletek is előállnak, amit a legjobb (*sinc* interpoláció) interpoláció sem tenne lehetővé. A szuperfelbontást előállító algoritmusok két kategóriába sorolhatók:

- Több képkockán alapuló módszerek: az adott objektumról kisebb elmozdulások során készült több felvétel áll rendelkezésre. Az elmozdulás pontos megbecslése után tudunk becsléseket adni a finom részletekre.
- Egy képkockán alapuló módszerek: ezek a módszerek megpróbálják kitalálni, hogyan nézhetett ki az eredeti nagyfelbontású verzió. Ehhez vagy az adott kép különböző részleteit, vagy hasonló képekből korábban épített képrészlet adatbázisokat használnak mintául. Utóbbi módszereket példa alapú szuperfelbontás (lásd 3.5. ábra) vagy kép-halucináció néven említi a szakirodalom.



3.5. ábra. Kép kétszeres nagyítása kétszeres kicsinyítése után. BF: eredeti kép; JF: Legközelebbi szomszéd módszer; BA: Bicubic interpoláció; JA: példa alapú szuperfelbontás [20]

4. fejezet

A kép hibái és a kép szűrése

Egy digitális kép a rögzítésének pillanatától kezdve a megjelenítéséig igen sokféle hatáson, átalakításon, torzításon megy keresztül. Ezeknek egy része nem szándékos és a kép minőségére káros hatással van, a másik része pedig ezeket a hatásokat kívánja csökkenteni ill. a kép kezelését szándékozik hatékonyabbá tenni, vagy pedig a hibák kiszűrését célozza meg.

A *képek szűrése* egy gyakran használt általános kifejezés, ami annyit jelent, hogy a kép valamilyen szándékos átalakításon, transzformáción esik át. Pl. a népszerű képmanipulátor, képszerkesztő programok (pl. a GIMP [26]) is ún. szűrőket alkalmaznak a képek átalakítására, valamilyen effektus létrehozására.

A képfeldolgozó eljárásokat igen sokféleképpen lehet csoportosítani, könyvünk legelején egy alapvető, három szintű hierarchiát ismertettünk (alacsony-, közép- és magas szintű feldolgozás), a különböző kép-transzformációkat tartalmazó függvénykönyvtárak (pl. OpenCV[64], Matlab[58]) általában más szempontok alapján végzik a csoportosítást, mint pl. :

- egy, kettő (esetleg több) képet felhasználó algoritmusok ;
- pixel szintű operátorok, amelyek a pixelek módosítását egymástól függetlenül végzik (pl. hisztogram transzformációk) ;
- lineáris vagy nemlineáris szűrők ;
- Fourier transzformációt használó szűrők ;
- morfológiai operátorok ;
- geometriai (koordináta) transzformációk ;
- stb.

Természetesen ezek a besorolások nem kizárólagosak, azaz lehetnek olyan eljárások, amelyek több csoportba is tartoznak. Ebben a fejezetben olyan szűrőkről lesz szó, amelyek a képek minőségét próbálják javítani bizonyos képi hibák korrekciójával, mérséklésével. Meg kell jegyeznünk, hogy munkánkban csak néhány, nagyon alapvető módszert tudunk bemutatni területi okok miatt.

Ahhoz, hogy viszonylag hatékony működést érjünk el, ismerni kell a hibák alapvető tulajdonságait.

Elsőként áttekintjük az elektronikus forrásból származó torzulásokat, majd megnézzük, hogyan tudjuk a kép minőségét mérni, később modellezzük a képalkotási folyamatot, majd végül röviden áttekintünk néhány alapvető rekonstrukciós módszert.

4.1. Elektronikus képi hibák

Egy digitális kép igen sokféle hibától szenvedhet. Az optikai hibák alapvető típusait a 2.4.2. fejezetben tárgyaltuk. A következőkben az elektromos jelek előállítása és továbbítása során szerzett zajok és torzulások hatásait tekintjük át.

- A kép expozíciója során hibás beállításból származó hibák: alulexponált vagy túlexponált kép, színegyensúly hibák, elmosódás.
- Elkenés (*smear*): Erős fény esetén függőleges fehér csík jelenik meg a CCD képén. Mivel a pixelek kiolvasása oszloponként történik, ezért az erős fény miatt felhalmozódott elektronok torzítják a szomszédos területek értékeit is.
- Blooming: Erős fény hatására előfordulhat, hogy a szenzor már nem tud több töltést felhalmozni, túlcsoordulás történik, a töltés a szomszédos cellákba áramlik.
- Erősítési hiba: A képérzékelő lapkákon felhalmozódó töltések jeleit erősíteni kell, eközben a jelek bizonyos mértékben torzulnak.
- Kvantálási hiba: Mivel a jelek érzékelése során A/D átalakítás történik, ennek az átalakításnak a hibája a kvantálási hiba. Sok esetben a jel kódolása, feldolgozása során változik a kvantálási szintek száma, a hiba ekkor is felléphet.
- Aliasing hiba: Más néven spektrumátfedési hiba vagy Moiré hatás. Az analóg jel mintavételezése vagy a digitális kép kicsinyítése során alakulhat ki. Lásd 3.4 fejezet.
- Termikus zaj: Melegedés hatására az elektronok heves, véletlen mozgásba kezdenek, ami a képérzékelőkben véletlenszerű töltéshalmozáshoz vezet.
- Foton zaj: A fotonok időben nem teljesen egyenletesen érkeznek a képérzékelőbe, egyfajta – statisztikailag jellemezhető – ingadozás szerint hol kevesebb, hol több – egységnyi idő alatt. Ennek az a következménye, hogy homogén felületek sem lesznek tökéletesen homogének a képen, különösen akkor, ha igen rövid volt az expozíciós idő. (Ezt a jelenséget valójában nem is tekinthetjük a képérzékelő hibájának.)
- Forró pixel (*hotpixel*): Az érzékelő pixel hibája, amikor a beérkező fotonok számával nem arányos bizonyos pixelek kiolvasott értéke, egyes hibás pixelek akkor is nagy értékkel bírnak, ha alig érkezett foton az érzékelőbe.
- Halott pixel (*dead pixel*): A hibás érzékelő cellák nem reagálnak a fotonok becsapódására, mindig 0 értéket adnak.

- Beégett pixel (*stuck pixel*): A hibás pixelek mindig ugyanazt az értéket, általában fehér szint jeleznek a beérkező fotonoktól függetlenül.
- Interlész (*interlace*) hiba: Valójában nem hibáról van szó, az effektus a váltott soros képrögzítési módszer természetes velejárója. Leginkább a gyors, vízszintes mozgást végző kontrasztos képrészeknél jelentkezik fésűhöz hasonló mintázat, javítására többféle szűrőt dolgoztak ki.
- Tömörítési hibák: A veszteséges képtömörítési eljárások többféle káros vizuális hatást gyakorolnak a tömörített képre. Ilyenek a moszkító hatás, a blokkosodás, és az alulkvantáltság. A hibák a tömörítés során alkalmazott szándékos információvesztés következményei.
- Átviteli adatvesztés: Akár analóg kódolásról, akár digitális átvitelről van szó, előfordulhat adatvesztés az átviteli csatorna hibájából. Esetenként valamilyen hibás érték hozzáadódhat az eredeti jelhez (additív zaj), máskor a jel értéke torzulhat (multiplikatív zaj), vagy akár az eredeti jel teljesen el is veszhet. Digitális átvitel esetén pl. hiányozhatnak a képből egyes képterületek (blokkok).



4.1. ábra. Digitális fényképezőgép által, hosszú expozíció során keltett termikus zaj. A kép jobb szélén jól látszik az erősítő által okozott melegedés hatása.

Mielőtt megnéznénk, miként hat a képre a leggyakoribb két hiba (az elmosódás és az additív zaj) tekintsük át, miként mérhetjük a képek minőségét.

4.2. A képi minőség mérése: objektív és szubjektív megközelítések

Ha egy kép minőségére gondolunk, akkor az vonatkozhat a kép tartalmára (esztétika) ill. a kép technikai jellemzőire (pl. felbontás, kontraszt, világosság). Természetesen szélsőséges esetben az utóbbi is befolyással van az esztétikai élményre, de normális esetben alig van hatással a kép tartalmáról alkotott véleményünkre.

Műszaki szempontból fontos, hogy a minőséget mérni tudjuk, és ne csak szubjektív benyomások, hanem *objektív*, jól definiált, mérhető információnk legyen róla. Amikor két jelet hasonlítunk egymáshoz, használhatjuk a klasszikus, egy dimenziós jelfeldolgozásban is használt

módszereket: a kép, mint függvény minőségét úgy határozzuk meg, hogy milyen mértékben hasonlít egy kiindulási, torzítatlan, ideális változatára.

Két függvény (f és g) hasonlóságát azok korrelációjával (keresztkorrelációval) lehet legegyszerűbben jellemezni. Magát a korrelációt skaláris szorzattal tudjuk felírni:

$$\langle f|g \rangle = \sum_{i=0}^M \sum_{j=0}^N f(i, j) \cdot g(i, j), \quad (4.1)$$

ahol a két kép $M \times N$ -es méretű. Minél nagyobb a korrelációs szorzatösszeg, annál nagyobb a hasonlóság a két függvény között. A gyakorlati alkalmazásokban a keresztkorreláció különböző normalizált változatait használják, hogy a mérés invariáns legyen a jel bizonyos – megengedett – torzulásaira.

A másik igen gyakran használt, egyszerű módszer két kép átlagos négyzetes különbségét (MSE , *Mean Square Error*) meghatározni:

$$MSE(f, g) = \frac{\sum_{i=0}^M \sum_{j=0}^N (f(i, j) - g(i, j))^2}{NM}. \quad (4.2)$$

Minél kisebb ez a különbség, a két vizsgált kép annál inkább hasonló egymáshoz. Az MSE -ből könnyen származtatható az ún. jel-zaj viszony (SNR , *Signal-to-Noise Ratio*):

$$SNR(f, g) = 20 \times \log_{10} \frac{RMS(f)}{RMSE(f, g)}, \quad (4.3)$$

ahol $RMSE$ MSE gyöke, RMS pedig a jel átlagos négyzetes értékének a gyöke:

$$RMS(f) = \sqrt{\frac{\sum_{i=0}^M \sum_{j=0}^N f(i, j)^2}{NM}}. \quad (4.4)$$

A képfeldolgozásban igen gyakran a csúcs jel-zaj viszonyt ($PSNR$, *Peak Signal-to-Noise Ratio*) használják, ami 8 bites pixelértékek esetén a következő:

$$PSNR(f, g) = 20 \times \log_{10} \frac{255}{RMSE(f, g)}. \quad (4.5)$$

Természetesen könnyű belátni, hogy előfordulhatnak olyan eltérések a képek között, amelyek az emberi látórendszer számára alig észrevehetőek, ellenben a fenti metrikák szerint nagy eltérést tapasztalunk. (Példának okáért toljunk el egy oszloppal egy nagyfrekvenciás mintázatot tartalmazó képet. Ha elég nagy a kép felbontása, 1 pixelnyi elmozdulást nem fogunk szemmel észrevenni, de a két kép között igen nagy számszerű eltérés adódhat a fenti mutatók valamelyikét alkalmazva.)

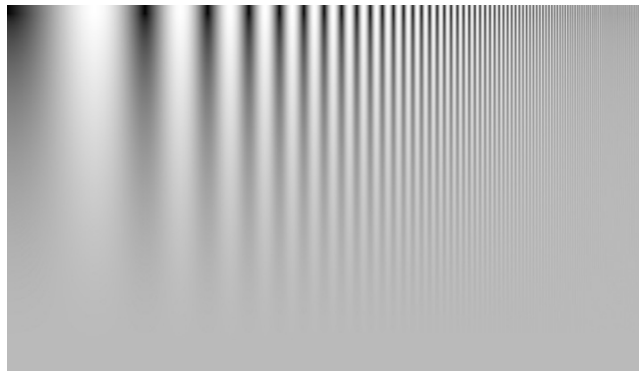
A kutatók már korán rájöttek arra, hogy az emberi látás komplex mechanizmusok összességként működik, és a vizuális minták, különbségek, képi hibák észlelését több nemlineáris hatás befolyásolja. Az egyik ilyen jól ismert jelenség *Weber*¹ törvénye néven ismert, miszerint

¹A német pszichológus Ernst Heinrich Weber 1834-ben súlyokra fogalmazta meg megfigyelését, később tanítványa Gustav Theodor Fechner általánosította a törvényt.

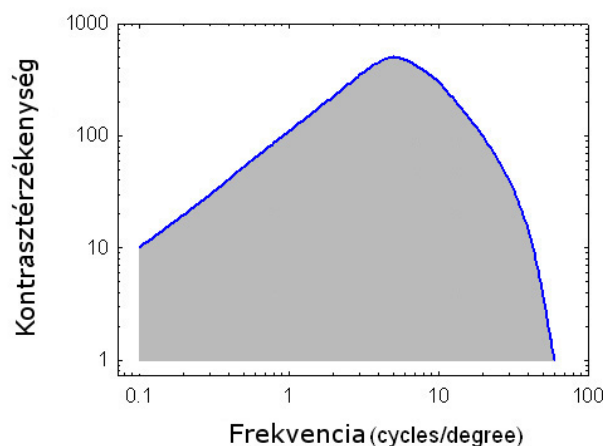
a látórendszer által éppen észlelhető intenzitásküszöb (ΔI) és a háttérintenzitás (I) hányadosa (a Weber hányados) viszonylag széles tartományban konstans:

$$K = \frac{\Delta I}{I}. \quad (4.6)$$

Ugyanakkor azt is korán megfigyelték, hogy az intenzitás változás észlelése a térfrekvencia függvénye is (lásd 4.2 ábra). Azaz felállítható egy kontrasztérzékenységi függvény (4.3 ábra), amely megmutatja, hogy milyen térfrekvenciák mellett vagyunk a legérzékenyebbek az intenzitás változására, a kontrasztra. Ez a függvény időben nem mindig állandó, a látórendszer képes adaptálódni bizonyos mintázatokhoz, megváltoztatva a kontrasztérzékenységi függvény alakját.



4.2. ábra. Frekvencia-kontraszt tesztkép a kontraszt-érzékenység vizsgálatához. A frekvencia balról jobbra, a kontraszt lentől felfele nő. Megállapítható az a frekvencia, ahol a legkisebb kontraszt mellett már észleljük a hullámfüggvény változását.



4.3. ábra. A kontrasztérzékenységi függvény

A képi hibák érzékelésére jelentős hatást gyakorol az ún. *maszkolás* jelensége: bizonyos térfrekvenciák képesek más frekvenciák észlelését elnyomni, ennél fogva ugyanaz a várható

értékű additív zaj a kép bizonyos részein alig vehető észre, míg más részein pedig nagyon zavaró. A hatás illusztrációját szolgálja a 4.4 ábra.

A látórendszer itt említett és egyéb, dolgozatunkban nem tárgyalt nemlineáris jelenségeinek feltérképezése és megértése olyan módszerek kidolgozását segíti, amelyek képesek az emberek által észlelt képi hibák automatikus objektív mérésére. Ugyanakkor fontos szerepük van a modern veszteséges képtömörítő eljárások kifejlesztésében, amelyek a pszichovizuális redundancia kiaknázásával érnek el nagy tömörítési arányokat minimális látszólagos torzulás mellett. Jó példa utóbbira a népszerű JPG tömörítési eljárás alapját szolgáló DCT együtthatók kvantálási mátrixainak optimalizálása a DCTune technológiával [22].

Az emberi látással jól korreláló, objektív mérőmódszerek kidolgozására nemzetközi kutatócsoport jött létre több intézmény tagjainak részvételével Video Quality Experts Group néven [82].

Ha humán kísérletek, tesztek alapján szándékozunk egy kép minőségét jellemezni, akkor *szubjektív* minőségről beszélünk. Ha ilyen, az emberi látórendszert felhasználó *méréseket* használunk, szintén van mód az eljárás sztandardizálására: az ITU-R BT.500 mérési szabvány [9] adja meg egy szubjektív összehasonlító kísérlet kereteit.

4.3. A degradációs és szűrési folyamat

A képképzési láncon áthaladó információ először optikai torzulásokat szenved, majd pedig a digitalizálás során ill. után különböző elektronikus zajhatások érik. Ebből a folyamatból kettő, nagyon gyakran előforduló jelenséggel (ill. ezek együttes hatásaival) foglalkozunk a továbbiakban:

- az optika elmosó, a kép finom részleteit simító hatásával,
- az additív zaj következményeivel,
- e két torzító jelenség együttes hatásával.

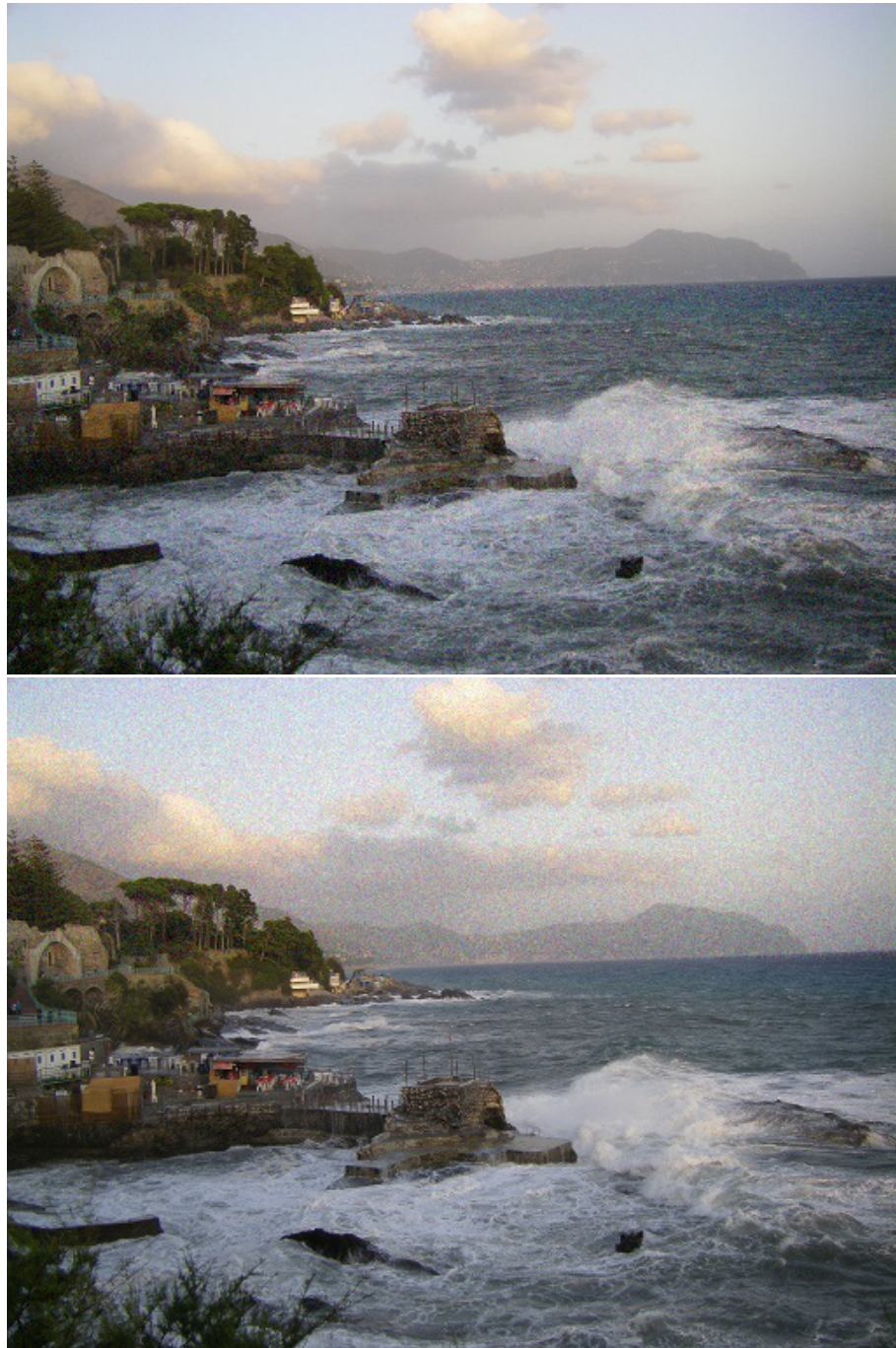
4.3.1. A kép elmosódása és az inverz művelet

Az optika elmosó hatása eredhet az optikai rendszer helytelen beállításából, szférikus aberrációból vagy akár a Fraunhofer diffrakcióból is. Az elmosódás közelítő matematikai modellje a torzítatlan kép (f) és a torzítást leíró függvény (k) konvolúciójával adható meg:

$$g(i, j) = (f * k)(i, j) = \sum_m \sum_n f(m, n)k(i - m, j - n), \quad (4.7)$$

ahol a két szumma csupán a konvolúciós kernel méretének megfelelő területen végzi az összegzést, ugyanis azon kívüli területeken a kernel értéke mindenhol 0.

A k függvényt pontszóródási függvénynek nevezik (*PSF, Point Spread Function*), szemléletes elnevezése azt takarja, hogy ha pontosan egy piciny pontot képezünk le az optikával,



4.4. ábra. A fenti és a lenti kép azonos mértékű zajjal van terhelve, de más-más helyen (a fenti a horizont alatt, a lenti a horizont felett). A zaj észlelt hatása eltérő a két kép esetén, a kép alampintázata maszkolja a zaj hatását.

akkor az elmosó hatás következményeként a pontból érkező sugarak a képérzékelőn egy kicsiny korong szerű foltként szóródnak szét. A legegyszerűbb simító konvolúciós szűrő az ún. doboz szűrő (*box filter*):

$$k_B = \frac{1}{25} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \quad (4.8)$$

A valóságos esetekhez közelebb áll egy olyan konvolúciós kernel, ahol egy adott pont elmosódásához a tőle távolabbi pixelek kevésbé járulnak hozzá. Ilyen a Gauss függvény, aminek egy diszkrét közelítése pl. egy 5x5-ös mátrixon a következő:

$$k_G = \frac{1}{80} \begin{bmatrix} 0 & 1 & 2 & 1 & 0 \\ 1 & 4 & 8 & 4 & 1 \\ 2 & 8 & 16 & 8 & 2 \\ 1 & 4 & 8 & 4 & 1 \\ 0 & 1 & 2 & 1 & 0 \end{bmatrix}. \quad (4.9)$$

A PSF Fourier transzformáltját optikai átviteli függvénynek, OTF-nek (*Optical Transfer Function*) nevezzük. A konvolúció korábban ismertetett (lásd 3.4) tulajdonsága miatt:

$$g(i, j) = f * k = \mathcal{F}^{-1}(F \cdot K), \quad (4.10)$$

ahol $K = \mathcal{F}(k)$ az optikai átviteli függvény, $F = \mathcal{F}(f)$.

Az elmosódás szűrését, azaz f visszaállítását g megfigyelésből ún. inverz művelettel, dekonvolúcióval kaphatjuk meg. Ez a frekvencia térben a

$$H = \frac{1}{K}, \quad (4.11)$$

szűrővel való szorzással valósítható meg, azaz:

$$f = \mathcal{F}^{-1}(G \cdot H), \quad (4.12)$$

ahol $G = \mathcal{F}(g)$. Utóbbi egyenlet valójában pontosan nem kivitelezhető a legtöbb esetben. Ennek az az oka, hogy K függvény tipikusan a magasabb frekvenciatartományban gyengíti, vágja le a jelet, itt közel 0 az értéke. Ennek a reciproka elvileg végtelen erősítést jelentene, aminek elkerülése érdekében beállítanak egy határfrekvenciát, ami felett H értéke már nem nő, azaz nem végeznek tetszőlegesen nagy erősítést.

A pixeltérben iteratív módszerekkel oldják meg az inverz, rekonstrukciós problémát. Az egyik leginkább elterjedt módszer a Lucy-Richardson (LR) algoritmus, ami f -re a következő iterációs formulát használja:

$$f^{(t+1)} = f^{(t)} \left(\frac{g}{\mathbf{c}^{(t)}} * k \right), \quad (4.13)$$

ahol t a felső indexben az iterációs számláló és

$$\mathbf{c}^{(t)} = f^{(t)} * k. \quad (4.14)$$

Az LR megoldás feltételezi, hogy f Poisson eloszlású, és kísérletileg igazolható, hogy ha a megoldás konvergál, akkor a legvalószínűbb megoldáshoz konvergál [75].

A f visszaállításának feltétele – függetlenül attól, hogy a pixeltérben vagy a frekvencia-térben dolgozunk – k , azaz a degradációs szűrő ismerete. Ha ez nem teljesül, akkor ún. vak dekonvolúciót (*blind deconvolution*) kell alkalmazni, ami a rekonstrukció során iteratív módon becslést tesz k -ra. Természetesen utóbbi módszerek kisebb sikerrel tudják megoldani a problémát.

Az LR módszernél egyszerűbb, gyorsabb, de sokkal durvább közelítő megoldást kaphatunk egy kép élesítésére egyszerű konvolúciós szűrővel is. Ennek az egyszerű rekonstrukciós szűrőnek az előállításához vegyük a degradált g függvény 2. deriváltját és vonjuk ki magából a jelből.

A g függvény második deriváltját az ún. Laplace operátorral állíthatjuk elő (3x3-mas esetben):

$$L = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \quad (4.15)$$

Ebből számítható a rekonstrukciós szűrő és adható durva becslés f -re:

$$f \approx g - (g * L) = g * k_r, \quad (4.16)$$

ahol

$$k_r = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}. \quad (4.17)$$

4.3.2. Additív zaj és szűrése

Az additív zaj jelenlétének többféle oka lehet, a 4.1. fejezet számos lehetséges esetet sorolt fel. Tekintsük f_z zajjal terhelt képet:

$$f_z = f + z, \quad (4.18)$$

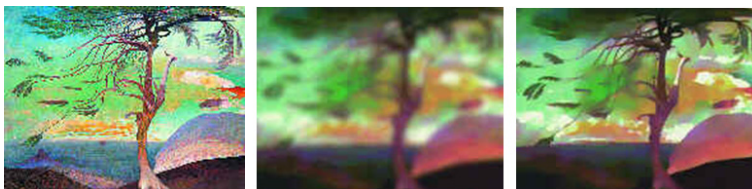
ahol f a zajmentes kép, $z \in \mathcal{N}(0, \sigma^2)$ zaj jelenti a degradációt. A feladat tehát f előállítása f_z és z statisztikai jellemzőinek ismeretében.

Lineáris szűrés a pixeltérben, a kép simítása

Mivel z várható értéke 0, ez azt jelenti, kiátlagolva z -t, 0-t fogunk kapni. Tehát egy átlagoló konvolúciós szűrőt alkalmazva f_z -re a hiba kiejthető:

$$f_z * k_B = f * k_B + z * k_B = f * k_B. \quad (4.19)$$

Sajnos a módszer nyilvánvaló hátránya, hogy bár a zaj kiszűrhető, de a kép magas frekvenciás komponensei csökkenni fognak, a kép homályosabbá válik. A probléma megoldására vannak azonban viszonylag egyszerű módszerek (ún. anizotróp ill. nemlineáris diffúziós szűrők), amelyek a kép simítását² csak olyan helyeken végzik el, ahol nincsen jelentős éltartalom [67]. (Belátható, hogy a Gauss függvénnyel való konvolúciós szűrés a hődiffúziós differenciálegyenlet megoldását jelenti. Ezért is használják a kép Gauss függvénnyel való simítására, elmosására gyakran a diffúzió, hődiffúzió kifejezést.) A 4.5 ábra egy kép lineáris elmosását és nemlineáris elmosását illusztrálja.



4.5. ábra. Bemeneti kép, lineáris diffúziós ill. nemlineáris diffúziós szűrése

Nemlineáris szűrés a pixeltérben

Ha a zaj előfordulásának valószínűsége kicsi, de amplitúdója nagy (és nem normális eloszlást követ) a fenti megközelítés nem vezet sikerre. Impulzus zaj esetében nagy szélsőértékű intenzitásértékek jelentkeznek a képen, ezért só-bors zajnak is nevezik. A kép simítása esetén ezek a kiugró értékek a környező pixelek értékeit is eltorzítanak, ezért ilyen esetben (többek között) ún. outlier vagy médián szűrőket lehet alkalmazni:

- Médián szűrő : a szűrő a vizsgált pont adott környezetében lévő képpontok értékeit nagyság szerint sorba rakja, majd pedig a sorrend szerinti középső értékkel helyettesíti a vizsgált pixelt.
- Outlier szűrő : a szűrő a vizsgált pont adott környezetében lévő képpontok átlagát veszi és megvizsgálja, hogy ettől az átlagtól abszolút értelemben mekkora a pont eltérése. Ha az eltérés túl nagy, a képpontot az átlagértékkel helyettesítjük.

A két szűrő hatása hasonló, a 4.6 ábra só-bors zajjal terhelt kép simítását és médián szűrésének eredményét illusztrálja. Jól látható, hogy a simító szűrő belemossa a környező területekbe az impluzus zajt, miközben a képünk kicsit homályos lesz. A médián szűrő nagy eséllyel kicseréli a zajos pixeleket a környezetéből származó, nem szélsőséges értékre, kismértékben a hasznos részletek is áldozatul esnek.

4.3.3. Zaj szűrése a frekvenciatartományban

Az eddig tárgyalt módszerek a különböző zajokat a pixeltartományban próbálták eltüntetni. Bár a 4.12. egyenlet a frekvenciatérben fogalmazta meg a magas frekvenciás komponensek

²Az angol *blur* kifejezést magyarul simításnak, elmosásnak, elkenésnek egyaránt szokták mondani.



4.6. ábra. Bal: kép impulzus zajjal terhelve; Jobb fent: a zajos kép lineáris diffúziós változata (3x3-mas doboz szűrő); Jobb lent: a zajos kép médián szűrt változata.

erősítését, ott nem kimondottan zajok szűrése, hanem a magas frekvenciás komponensek erősítése volt a célunk.

Amennyiben a képeken nemkívánt alacsony- vagy magas frekvenciás komponensek jelennek meg, a Fourier térben elvégezhetjük az alacsony vagy magas frekvenciás komponensek szűrését: előbbit *felüláteresztő*, utóbbit *aluláteresztő* szűrőnek nevezzük. Amennyiben egy meghatározott frekvenciasávban jelentkezik a zaj, egy viszonylag szűk sávra is korlátozhatjuk a szűrésünket ún. *sávszűrőt* definiálva.

A szűrő (legyen az alul-, felüláteresztő vagy sávszűrő) alkalmazhat 0-val való szorzást a frekvenciatérben, azonban ilyen esetekben mellékhatásként a jelentősebb élek mentén szellemképszerű hullámok jelennek meg a képen. Ha a szűrő karakterisztikája nem tartalmaz hirtelen 0-1 átmenetet, hanem a két érték között folytonos átmenet jellemzi, elkerülhető ez a nemkívánt hatás. Ilyen szűrő az ún. Butterworth szűrő [11], aluláteresztő esetben:

$$B(u, v) = \frac{1}{1 + \left(\frac{(u^2+v^2)^{1/2}}{C}\right)^{2n}}, \quad (4.20)$$

ahol u és v jelenti a sor és oszlop szerinti frekvenciákat, n a szűrő fokát definiálja, a C vágási frekvencia esetén pedig 50%-os csillapítást alkalmazunk.

A 4.7. kép periódikus zaj szűrésére mutat példát.



4.7. ábra. Fent: Periódikus zajt és periódikus mintát tartalmazó kép és nagyított részlete sávszűrés előtt ill. után. Lent: Fourier tartomány valós és képzetes része a kerítés és a zaj jellemző csúcaival megjelölve.

4.3.4. A kép elmosódása és additív zaj jelenléte

Az eddigi modellekben vagy a kép elmosódásával, vagy additív zaj meglétével számoltunk. Azonban a valós életben sokszor mindkét jelenség egyszerre jelentkezik:

$$g = f * k + z, \quad (4.21)$$

azaz a kép konvolúcióját (pl. lencse elmosó hatása) additív zaj (képezékelő chip termikus zaja) követi. Ebben az esetben inverz szűrőt (lásd 4.11. egyenlet) alkalmazva sajnos a zaj erősítését is elvégeznénk:

$$F_r = F + \frac{Z}{K}. \quad (4.22)$$

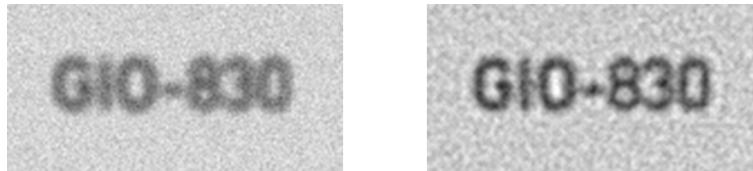
Mivel a zaj tipikusan túlnyomórészt magas frekvenciás komponensekkel jellemezhető, ezért az inverz szűrő magas frekvenciát erősítő hatása kimondottan káros. Ha célként tűzzük ki, hogy:

$$E(F - F_r)^2 = 0, \quad (4.23)$$

azaz a rekonstrukciós hiba négyzetének várható értékét 0-nak szeretnénk elérni, az ideális megoldást a Wiener szűrő adja meg [85]:

$$H_W = \frac{K^* C_f}{|K|^2 + \frac{C_n}{C_f}}, \quad (4.24)$$

ahol C_n a zaj függvény, C_f pedig az eredeti jel kovarianciájának Fourier transzformáltja, a $*$ pedig a komplex konjugáltat jelöli. Az így meghatározott szűrő esetén bár szükségünk van a zajt jellemző kovarianciára, de úgy leszünk képesek a kép élesítésére, hogy a magas frekvenciás zajok erősítését is el tudjuk közben kerülni. A 4.8 ábrán egy homályos, de egyben erős zajjal terhelt képet látunk. A szűrő eredményén láthatjuk, hogy a betűk élesebbek lettek, miközben a zaj nem rontja jelentősen a láthatóságot.



4.8. ábra. Homályos és erős zajjal terhelt kép és annak Wiener szűrővel javított változata

5. fejezet

Képek visszakeresése

Az egyre elérhetőbbé váló digitális fényképezési és videózási technológia elterjedésével egyre több kép halmozódik fel kultúránkban, és ezzel párhuzamosan nő a vizuális információ szerepe. Egy 2003-as felmérés szerint [36] több mint hétszázézer óra mozi és tévéfilm készült eddig a világban, de valójában ennek a többszöröse lehet a fényképek és egyéb elektronikus vagy papírképek száma, különösen, ha figyelembe vesszük a háztartásokban, hobbi szinten keletkező vizuális információt. Egyes becslések szerint a legnépszerűbb videómegosztó internetes portálon több mint 140 millió videó volt letölthető 2010-ben, és ez a szám percről percre rohamosan növekszik.

Az Interneten oly népszerű Google képkeresőjével is több százmillió kép közül kereshetünk, böngészhetünk, de hamar rájövünk arra, hogy sok esetben a megadott keresőszó alatt gyökeresen más képi tartalmat találunk, mint azt elvárnánk. A Google (jelen dokumentum írásának idejében) ugyanis alapvetően a szövegkörnyezet alapján rendezzi, "értelmezi" a tartalmat, és lássuk be, egyéni vizuális fantáziánk szavakkal néha nehezen leírható, hát hogy is várhatnánk el, hogy néhány hasonló szó alapján mindig a megfelelő képre asszociáljunk. Különösen nehéz garantálni a sikeres keresést, ha a szöveg nem feltétlenül a kép illusztrációját szolgálja, hanem éppen annak egy ellentétéről ír. Ugyanakkor az otthoni felhasználásban napjainkban tipikusnak mondható néhány Gigabyte-os memóriakártyára több ezer digitális fénykép rögzíthető, és így már akár a hétköznapi életben is egyre nagyobb feladat a vizuális információ rendszerezése, visszakeresése, bizonyos esetekben annak megértése vagy összefüggések felfedezése.

5.1. A digitális kép mint sokdimenziós adat

Mint korábban, a 3. fejezetben tárgyaltuk a digitális kép képpontokból, ún. pixelekből áll hasonlóan, mint egy fotópapír vagy filmnegatív kristályszemcséi, de ezek a pixelek alapvetően szabályos rácsszerkezetben helyezkednek el egymás mellett. A számítógép monitorját távolról szemlélve egy sima, folytonos képet látunk, bár az valójában millió és millió apró alkotóelemből (piros, zöld és kék ún. "szub-pixelből") áll össze. Színes képeknél egy képpont tipikusan 3 komponensből (csatornából) áll, amik a szemünkben és agyunkban színi érzetet ill. észleletet generálnak (természetesen mindenkinél kicsit másképpen).

Ha a képpontok közül egyik sincs megkülönböztetve a többitől, és a kép előállítása alapvető-

en azok egymás utáni (általában balról jobbra, fentről lefele haladó) kirajzolásával történik, akkor *raszteres képről* beszélünk. Ilyen képek készülnek a digitális fényképezővel vagy ha egy fényképet lapszkennelrel beolvasunk például JPG vagy BMP fájlformátumban. *Vektoros képek* esetén viszont a kép objektumokból (pontokból, vonalakból, poligonokból) áll össze, ezeket az objektumokat egyenként tudjuk manipulálni. Ha egy vektoros képet nagyítunk, az nem fog durva felbontásúnak, pixelesnek tűnni, mivel a vonalak, poligonok értékét az adott pixelekhez újra tudjuk számolni. Természetesen egy vektoros kép is olvasható raszteres módban, de tipikusan pont az a lényege a vektoros adathalmaznak, hogy az egyes objektumokhoz valamilyen olyan jelentés kapcsolódik, ami alapján az adatok tárolása, a kézi vagy gépi értelmezés, elemzés vagy manipuláció könnyebben elvégezhető. (Napjainkban gyakran használt vektoros fájlformátum a Microsoft Windowsban elterjedt WMF vagy a térképészetben kedvelt .shp kiterjesztésű Shape fájl.)

A raszteres képek önmagukban tehát a gépek számára nem sok támpontot adnak az értelmezésre, a tartalom jellemzése érdekében azokat valamilyen képelemző programmal fel kell dolgozni. A képek azonban óriási adatmennyiséget jelentenek: pl. egy 3 csatornás 1000×1000 méretű kép (1 MP) értelmezhető egy 1 millió dimenziójú vektornak, ahol természetesen az egyes, térben szomszédos (darabonként 3×8 bites) komponensek között erős összefüggések lehetnek. Amennyiben filmek visszakereshetőségéről beszélünk, akkor másodpercenként minimum 24 képkocka vizsgálatára kell gondolnunk, természetesen itt is erős lehet az időbeli kohézió. A képi tartalmat általánosan úgy jellemezhetjük, hogy a homogén, sima területeken nagy hasonlóság fedezhető fel, míg az élek, kontúrok környékén az élre merőleges irányban erősen eltérő színű pontokat látunk. Természetesen ennél jóval bonyolultabb összefüggéseket kell felfedeznünk a képi adatbázisok elemzése közben ahhoz, hogy hatékonyan tudjuk leírni a képi tartalmat, és azt visszakereshetővé tegyük.

5.2. A szemantikai rés

A számítógépes programok közvetlen vagy közvetett felhasználói egy kommunikációs forradalomnak voltak tanúi a XX. század végén, XXI. század elején. A nyomtatott írás, majd pedig a számítógépes adatrögzítés révén olyan mértékű írott (vagy fényképezett, elektronikusan rögzített) információ állt elő (és az Internet révén vált elérhetővé), ami tömegesen csak számítógépes módszerekkel dolgozható fel, érthető meg, kereshető vissza, ez pedig az ún. "szemantikai rés" (*semantic gap*) áttörését teszi szükségessé.

Egy kép keresése vagy visszakeresése során az információ után kutató felhasználó magas absztrakciós szinten fogalmazza meg kéréseit, elvárásait, pl.:

- Olyan képet keresek, amin a naplemente látható.
- A megadott képen lévő virághoz hasonló képet keresek.
- Egy futballmérkőzés fontos eseményeit (pl. gólok, büntetőrúgások) szeretném megnézni.

Ezzel szemben a számítógép csupán pixelek sokaságát, videók esetében képkockák sorozatát, esetleg primitív geometriai objektumokat lát. A két eltérő elvonatkoztatási szintnek a

közelítéséhez a képek elemzése, feldolgozása, de ugyanakkor sokszor a felhasználó gondolkodásának megértése is szükséges.

Egy digitális kép értelmezése alapvetően három absztrakciós szinten lehetséges:

- A legalacsonyabb szinten képi primitívekről beszélünk. Ilyen egy képpont színe, a kép mikrostruktúrája (textúrája), egy folt vagy alakzat megléte.
- Ennél magasabb absztrakciót jelentenek a képen látható tárgyak, objektumok.
- A legelvontabb szinten pedig a képeken megjelenő események, ill. az általuk kiváltott hangulati elemek, érzelmek fogalmazódnak meg.

A szemantikai rés – azaz az alacsony szintű és magas szintű értelmezés közti különbség – áttöréséhez tehát olyan intelligenciára van szükség, amely képes az alacsony szintű komponensek elemzésével magasabb szinten megfogalmazott kérdések megválaszolására. A különböző felhasználói attitűdök, az eltérő vizuális világok és asszociációk, a más és más verbális megfogalmazás viszont megnehezíti a probléma megoldását még a viszonylag jól definiált speciális alkalmazások körében is, szükség van tehát adaptációra, a mesterséges intelligencia alkalmazó képességére. Esetünkben kitűnő példa erre a fontossági visszacsatolás [17] vagy a hosszú távú tanulás a képkereső rendszerekben.

5.3. Képi adatbázisok

Míg a szöveges vagy más rövid karakterekkel operáló adatbázisokban viszonylag könnyen tudunk keresni, addig a nagyméretű képi adatbázisokban sokkal nehezebb a felhasználók számára fontos információ kinyerése, visszakeresése, összehasonlítása. Egyrészt a gépi intelligencia számára nehéz a kép szemantikai értelmezése, másrészt meglehetősen számításigényes a több ezermilliárd képpont feldolgozása. Ahhoz, hogy sikert tudjunk elérni, a képeket metainformációval kell ellátni, illetve a képek strukturális vagy globális tulajdonságait az adatbázisba való bekerüléskor kivonatolni és tárolni kell.

5.3.1. A képi adatbázisrendszerek sajátosságai

Egy képi adatbázis alapvetően abban tér el egy hagyományos szöveg alapú adatbázistól, hogy egy-egy rekord igen nagymennyiségű adatot jelent, ahhoz viszont, hogy gyors és lehetőleg magas szintű lekérdezéseket tudjuk végrehajtani, tárolni és indexelni kell a feldolgozáshoz szükséges járulékos információkat.

Alapvetően kétféle ilyen járulékos információt ismerünk a jelenlegi rendszerekben: a *metaadatokat* és az *indexstruktúrákat* (itt jegyezzük meg, hogy a két fogalom mögött álló tartalom nem áll messze egymástól, egyes rendszerekben össze is mosódhat a kettő jelentése). Az indexstruktúrák egyfajta kivonatok vagy segéd adatok a rekordokhoz, vagy azok csoportjaihoz rendelve. Segítségükkel az adatok jobban rendszerezhetők, nyilvántarthatók, visszakereshetők. A lekérdezések során ugyanis nem szükséges egy rekord tartalmának (pl. magának

a képnek) vizsgálata, elégséges az index bejegyzéseket feldolgozni. Az indexstruktúrák szerkezete pedig úgy van kialakítva, hogy az adott alkalmazásban minél inkább elősegítse a gyors vagy helytakarékos (kevésbé memóriaiigényes) működést.

A metaadatok plusz, járulékos információk az információkról. Képi adatbázisoknál ilyen lehet egy kép neve, sorszám, keletkezési dátuma, vagy akár a képen látható objektumok vagy események leírása. Kitűnő példa erre az amerikai kosárlabda játékok adatbázisa, ahol – eleinte operátorok manuális módon, napjainkban pedig automatikus módszerekkel – rögzítik a játék pontos menetét, a labdakezelést, eseményeket, és ez alapján könnyen lehet visszakeresni egy adott szituációt, lehet készíteni statisztikákat a csapatokról vagy játékosokról [78].

Szükségünk van tehát objektumok, alakzatok, speciális képi tulajdonságok, mozgásfajták, események gépi felismerésére, szabványos leírására, kódolására, elemzésére, összehasonlítására, visszakeresésére. Megfelelő sorrendben a következő lépésekből áll egy tartalom-alapú képviszakereső adatbázisrendszer működése:

- I. Kép (videó) felvétele az adatbázisba;
- II. Metaadatok bevitele;
- III. Képi tulajdonságok kinyerése;
- IV. Képi tulajdonságok leírása, indexelése;
- V. Ezen tulajdonságok alapján magas szintű információ kinyerése, azok indexelése;
- VI. Felhasználói kérdések kezelése (GUI);
- VII. A lekérdezés és a rekordok összehasonlítása;
- VIII. Eredmény (eredménylista) megjelenítése;
- IX. Rövid és hosszú távú tanulás felhasználói visszacsatolás által.

Az, hogy milyen tulajdonságokat választunk az indexeléshez, és hogyan írjuk le ezeket a tulajdonságokat, majd pedig milyen további elemzések szükségesek a hatékony működéshez, az adott alkalmazás határozza meg, de valójában ezen a területen napjainkban igen jelentős kutatási aktivitás figyelhető meg.

5.3.2. Dublin Core és más metainformációs rendszerek

A vizuális tartalom szemantikai feldolgozását legjobban humán erőforrások alkalmazásával tudjuk elérni, azaz el kell látni a rekordokat olyan szöveges információval, ami szerint a későbbiekben az adatokat rendezni, összehasonlítani, visszakeresni szeretnénk. Ez a terület elsősorban a könyvtárosok világában fejlődött, így kidolgoztak különböző

- osztályozási rendszereket,
- kötött név- és tárgyszóregisztereket,

- katalógusokat, tezauruszokat.

A modern adatbázis-rendszerek elterjedésével természetesen lépést tartanak a könyvtári rendszerek is. Jelentős törekvés a digitálisan tárolt információk rendszerezéséhez az MDC Open Information Model vagy az SMPTE által kidolgozott UMID (Unique Material Identifier) azonosítók rendszere. A Making of America II. (MOA II) Testbed Project figyelemre méltó folytatása a kilencvenes évek első felében megkezdett digitális könyvtárakkal és más egyéb intézményekkel foglalkozó kezdeményezéseknek. Ebben a rendszerben leíró, strukturális (az objektum belső szerkezetét leíró) és adminisztratív metaadatokat különböztettek meg.

Az Internet térhódításával újabb feladatokat kellett megoldani: a térben erősen szétszóródott és sokféle digitális információt egységes kezelői és leíró felülettel kellett ellátni. A *Dublin Core* kezdeményezés az internetes forrásfeltárási munka megkönnyítését, az Interneten található források leírásának egységesítését, a hozzáférés és az egységes értelmezhetőség szabványosítását tűzte ki célul. 15 leíróban állapodtak meg – azaz definiáltak egy minimális konszenzust – a tartalom jellemzésére, ezek a következők: cím, alkotó, tárgyszavas leírás, leírás, kiadó, hozzájáruló, dátum, típus, formátum, egyedi azonosító, forrás, nyelv, forrás hivatkozás, tér és idő paraméterek, szerzői jogok. A Dublin Core sémának az a legfontosabb szerepe, hogy az egymással konkuráló szabványok és módszerek között átjárást biztosítson, amennyiben integrálják más sémákba, vagy hivatkoznak rá.

5.4. A JPEG2000 és az MPEG-7 visszakereshetősége

Nem kétséges, hogy a digitális képek hatékony megosztásához szükséges azok szabványos formátumú kódolása. Ugyanezt tudjuk elmondani a képek tartalmi leírásáról is, így az egyedi tartalomleíró megoldások mellett létrejöttek különböző szabványos rendszerek. Ezek közül a két legjelentősebbet mutatjuk be röviden.

5.4.1. JPEG2000

Függetlenül a könyvtári metaadat rendszerektől a JPEG2000¹ [46] tömörítési eljárás kidolgozásakor létrehozták annak metaadatokat tároló formátumát a JP2-t ill. annak egy kiterjesztését, kiegészítését, a JPX formátumot. A képet leíró tulajdonságokat ún. "dobozokban" tudjuk tárolni a JPX formátum szerint. A következő doboz típusokat definiálták:

- Asszociációs;
- ROI Leíró (ROI: Region of Interest - a kép azon kijelölt része, amit feldolgozunk);
- XML;
- MPEG-7 Bináris;
- Szabad felhasználású doboz.

¹JPEG: Joint Photographics Pictures Group 2000

Az Asszociációs "szuperdoboz" (olyan doboz, ami egyéb dobozokat is tartalmazhat) lehetőséget biztosít adatok egymáshoz rendeléséhez dobozok vagy azok tartalma között, így a ROI Leíró doboz és az Asszociációs doboz segítségével a kép egyes részeihez metaadat rendelhető. A XML doboz a nevéhez híven XML adatokat tartalmaz, míg az MPEG-7 Bináris doboz BiM formátumú adatokat tárol (a BiM az MPEG-7 bináris fájlformátumát jelöli). A Szabad felhasználású doboz típus segítségével pedig tetszőleges formátumú adattárolást oldhatunk meg. A JPEG2000 metaadatai négy csoportba sorolhatók:

- A kép készítésével kapcsolatos adatok (mint pl. a fényképezőgép típusa, képszerkesztő szoftver).
- A tartalmat leíró információk (szöveges adatok: ki, mikor, hol és mit fényképezett le, ill. a vizuális képi tartalmat jellemző adatok).
- A metaadat előzményeivel kapcsolatos metaadat leírja a képen eddig végzett átalakításokat: mi történt a képpel addig, amíg a jelenlegi állapotába került. Megadása történhet a kép metaadataiba ágyazásával vagy referenciával a kép előző változatára. Több képből összeállított kép esetében ez hierarchikus szerkezetű is lehet.
- Szellemi tulajdonnal és szerzői jogokkal kapcsolatos információk.

A szabvány XML séma szintaxis szerint adja meg a metaadatokat, az adatoknak jól formált XML-ben kell lenniük, érvényességüket ún DTD (Document Type Definitions) definiálja.

5.4.2. MPEG-7

Az MPEG-7 [40] általánosan megfogalmazott célja, hogy szabványos eszközt biztosítson a hang, kép, mozgókép és multimédia anyagok tartalmi, magas szintű (szemantikai) leírásához, és átjárást tegyen lehetővé a tartalomleíró eszközök között. A szabványt "multimédia-tartalom leíró interfésznek" is nevezik, és bár már meglévő szabványokra építkeznek (pl. MPEG-2-es mozgásvektorok, vagy az MPEG-4-es alak leírók felhasználása) lényegében a legteljesebb olyan eszköz, aminek lényege, hogy szabványosan kódolja:

- az alacsony szintű tulajdonságokat,
- a felhasználó számára könnyen érthető magas szintű jellemzőket,
- ill. a tartalomból nem származtatható, de azzal kapcsolatos információkat (metaadatok).

A szabvány a következő részekből épül fel:

- I. Rendszer: az adatfolyam kezelésével foglalkozik;
- II. DDL ("Description Definition Language"): a leíró eszközökhöz adja meg a szintaxist;
- III. Vizuális: képi információt leíró eszközök;
- IV. Audio: audio információt leíró eszközök;

- V. MDS ("Multimedia Descriptor Scheme"): általános és multimédia tartalmat leíró eszközök;
- VI. Referencia szoftver: a szabvány eszközeihez készített kísérleti szoftver (XM: "experimentation model" - kísérleti modell);
- VII. Megfelelőségi teszt;
- VIII. MPEG-7 leírók kivonása és használata;
- IX. Profilok;
- X. Séma definíció.

A szabvány hatásköre a tartalom *leírásának módjára* terjed ki, nem feladata definiálni azt, hogy mit, miért, és hogyan jellemezzünk, illetve a felhasználás, pl. képvisszakeresés módjával sem foglalkozik. A multimédia anyagok jellemzésére az XML jelölőnyelvet használja, kiterjesztve azt a hang, illetve képi információk jellemzéséhez szükséges elemekkel, definíciókkal (például mátrixok, vektorok).

Mint a felsorolásból látszik, az MPEG-7 alapvetően külön foglalkozik a vizuális, audio és multimédia leírókkal, most nagyon röviden csak a Vizuális rész (a szabvány 3-ik része) elemeit soroljuk fel, mivel itt vannak definiálva azok a tulajdonságok, amelyek a képek vizuális jellegét leginkább leírják és így alapul szolgálnak az MPEG-7 felett működő adatbázis alkalmazásoknak. Ezen rész elemei hét fő csoportba sorolhatók:

- I. Alaptípusok: Rács elrendezés, Idősor (reguláris, irreguláris), 2D-3D nézet, Síkbeli koordináták (lokális, integrált), Időbeli interpoláció
- II. Szín leíró: Színtér, Szín kvantálás, Domináns szín, Skálázható színek, Színeloszlás, Szín-struktúra, GoF / GoP Szín
- III. Textúra leíró: Homogén textúrák, Él-hisztogram, Textúra tallózás
- IV. Alakzat leíró: Régió alakja, Kontúr alakzat, Háromdimenziós alakzat
- V. Mozgás: Kamera mozgása, Mozgás trajektória, Parametrikus mozgás, Mozgás tevékenység
- VI. Lokalizáció: Régió Lokátor, Térbeli-időbeli lokátor
- VII. Egyéb: Arcfelismerés

Ahhoz, hogy jobban megértsük egy visszakereső rendszer működésének logikáját, a továbbiakban három – kifejlesztésének idejében, megközelítésében, módszerében – különböző visszakereső mechanizmust mutatunk be.

5.4.3. Az MPEG-7 XM egy kísérleti kiértékelése

Az MPEG-7 tervezését alapvetően az motiválta, hogy a multimédia információ leírását hatékony szabványok által tudjuk elvégezni, és ezáltal a tartalom széles alkalmazási körben váljon összehasonlíthatóvá és kereshetővé. Az utóbbi célok eléréséhez viszont nem elegendő a szabványban megadott leírók generálása, azokat hatékonyan fel is kell tudni dolgozni, a célnak megfelelően értelmezni. Tehát bár a szabványleíró eszközök nagy halmazát hozta létre, viszont nem tudjuk, hogy egy adott cél, egy adott alkalmazás esetén milyen eszköz lesz számunkra a legalkalmasabb, és az sem könnyű kérdés, hogyan kell a szabványban definiált eszközöket hatékonyan felhasználni.

Nagyméretű általános témájú képi adatbázisoknál tipikus feladat az, amikor a képeket téma szerint osztályokba kell sorolni. Ojala és társai a cikkükben leírtak szerint [63] csupán a képet leíró jellemzők összehasonlításával próbálták meghatározni azt, hogy a lekérdezést jelentő példaképet megadva mely leíró lesz a legjobb összehasonlítási alap, ha ugyanabba az osztályba tartozó képeket szeretnénk az adatbázisból visszakapni. Kísérletükben 4db különböző, de színnel kapcsolatos MPEG-7 leírót vizsgáltak; a szabványos leírók mellett (*Domináns szín, Skálázható szín, Színeloszlás, Szín-struktúra*) az összehasonlításba belevették a Huang által korábban javasolt ún. HSV Autokorrelogram módszert is [42].

A cikkben szereplő leírókat az MPEG-7 XM Reference Software version 5.3 alapján implementálták, a GoF/GoP Szín szintén színekkel kapcsolatos leírót viszont nem vizsgálták, mivel az több kocka együttesére van definiálva, az adatbázisban viszont különálló képek voltak. Ahhoz, hogy megértsük a kísérlet eredményét, nagyon tömören ismertetjük a kísérletben szereplő leírókat:

- A Színeloszlás (Color Layout) leírót a kép 64 lokális domináns színéből generáljuk (a képterület 8x8-as felosztásának megfelelően).
- A Szín-struktúra (Color Structure) - egy hisztogramhoz hasonlóan - a globális színi jellemzőket valamint a lokális színeket vektorban tárolja, amiket az ún. L_1 normával² hasonlítunk össze.
- A Domináns szín (Dominant Color) leíró a kép tetszőleges régiójának a domináns színeit tudja jellemezni. Az ún. Általánosított Lloyd Algoritmussal kell elvégezni a színek kvantálását a CIE LUV térben, így a kép egyes régióihoz csak néhány fontos színt kell rendelni. A képek összehasonlításánál figyelembe vesszük a régiók térbeli elhelyezkedését is.
- A Skálázható szín (Scalable Color) egy egyenletesen kvantált 256 elemű hisztogram. A hisztogram értékeit nemlinárisan kvantáljuk a Haar transzformáció segítségével. A leíró összehasonlítását a Haar térben L_1 normával vagy Hamming távolsággal illetve a hisztogram térben L_1 normával végezhetjük - a hivatkozott cikkben ez utóbbit használták.

² L_1 norma jelentése $\|x\|_1 := \sum_{i=1}^n |x_i|$

A korrelogrammok megadják, hogy egy adott távolságban milyen valószínűséggel találhatók adott színpárok. Az autokorrelogramm annyiban speciális eset, mivel a színpár két eleme azonos - Huang szerint az autokorrelogram hatékonysága nem, viszont a számítási igénye jelentősen lecsökken a korrelogrammhoz képest képi keresések esetén. A cikkben HSV autokorrelogrammot teszteltek négyféle távolságértékkel (1,3,5,7), miközben a HSV értékeket kvantálták: 12 színességi, 3 szaturáció és 3 intenzitás értékkel, tehát csupán 108 különböző szín maradt a képeken. A 4×108 dimenziós autokorrelogrammot L_1 normával hasonlították össze. A kísérleti adatbázisban szereplő 2445 tesztképet 8 szemantikus kategóriába sorolták humán felhasználók, ez adta a referenciát a különböző módszerek kiértékeléséhez. 1623 kép nem tartozott egy kategóriához sem - hasonlóan a mindennapi gyakorlathoz, amikor is egyes képek vagy nem sorolhatók semmilyen ésszerű kategóriába, vagy csak az adott esetben nem létezik megfelelő osztály. A szubjektív besoroláskor a szemantikai jelentés játszott döntő szerepet, nem pedig a színek elhelyezkedése a képen. Mindez egyszerűen azt jelenti, hogy egy-egy adott kategóriába néha meglehetősen eltérő szerkezetű és színvilágú képek kerültek. Néhány esetben egy kép több kategóriába is tartozhatott.



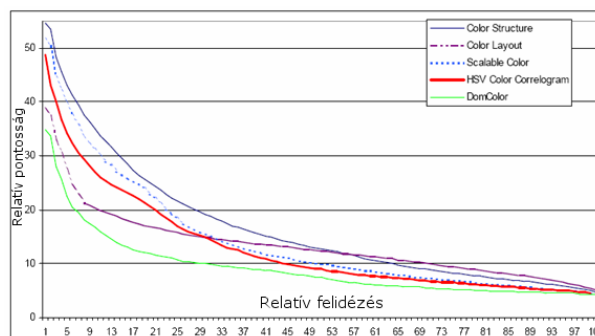
5.1. ábra. Néhány példa az épületek kategóriából [63].

Az idézett cikk szerint a leírók képességeinek vizsgálatához nem alkalmaztak tanuló mechanizmust, hanem egyszerűen a leíróknál említett normákat használták a találatok sorba rendezéséhez. A kiértékeléshez a pontosságot (*precision*) (a helyes kategória aránya az összes visszaadott találatban) és az ún. felidézés (*recall*) értéket (helyes visszanyert kategóriájú képek aránya az adott kategória összes elemszámához képest) vizsgálták. Az összes, előzőleg osztályozott 822 képet külön-külön megadták lekérdezésekhez, majd pedig az első találatot (ami maga a lekérdezés volt) elhagyták és az eredményeket a 822 képre kiátlagolták. A következő grafikonból (5.2. ábra) jól látható, hogy a Szín-struktúra teljesített a legjobban a kísérlet során.

5.5. A tartalmi elem kapcsolódásai

A tartalom alapú képviszakeresést támogató rendszerek fejlesztése már a kilencvenes években elkezdődött. A felhasználás sokrétű volta miatt többféle típust lehet megkülönböztetni [76]:

- kategória alapú: A képeket osztályokba soroljuk és az adott osztályba tartozó képeket kell megtalálni.



5.2. ábra. A visszakeresés pontossága a felidezés függvényében egyszerű szabványos képi leírók alkalmazásával néhány száz kép esetén [63].

- célzott keresés: Pontos elképzelésünk van a keresendő kép tartalmáról, pl. egy konkrét kép másolatait szeretnénk megkeresni.
- asszociatív: A keresés elején nincs pontos elképzelés a keresés tárgyáról, interaktív folyamat során alakul ki a kép a felhasználóban, hogy pontosan mire is kíváncsi. Fontos lehet a felhasználói értékelések folyamatos visszacsatolása a keresési motor irányába.

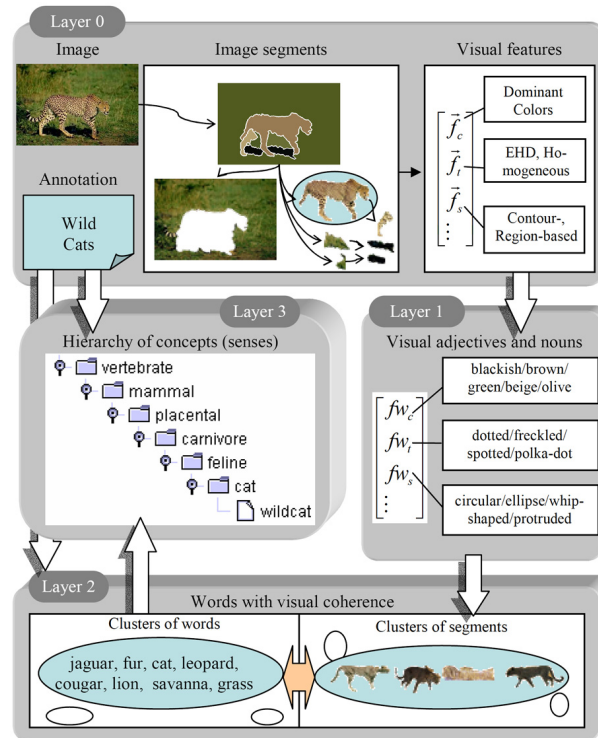
A különböző megközelítések során fontos a hasonlósági fogalom pontos megfogalmazása és olyan módszerek kidolgozása, amelyek jól tudnak illeszkedni az emberi észlelés és gondolkodás logikájához az adott keresési szituációban. Napjainkra igen sokféle megközelítés fejlődött ki, ezek közül kettő módszert ismertetünk igen tömören. Először egy olyan módszert mutatunk be, ahol a szemantikai modellt a szavak és a képi elemek kapcsolata, illetve a szavak jelentését és kapcsolatát leíró adatbázisok határozzák meg. A másik technika pedig annak az illusztrációját mutatja be, hogy miként lehet olyan képi jellemzőket automatikusan kinyerni, amelyek megfelelően invariánsak lesznek a különböző képi transzformációkkal szemben.

5.5.1. Szemantika a szöveg és képi tartalom kapcsolatával

Érdekes és ígéretes törekvés a szöveges és vizuális információ összekapcsolása a képi adatbázisok indexelésében. Az ilyen módszerek alacsony szintű vizuális tulajdonságokat emelnek ki, majd pedig összekapcsolják azokat szöveges megjegyzésekkel manuális bevétel, vagy tanítási módszerek segítségével. Jó példája ennek a megközelítésnek a [50]-ban ismertetett módszer, ahol hierarchikus struktúrában kötik össze a képi részleteket a szavakkal. Az adatbázisban szereplő képeket először nemlináris, inhomogén diffúziós modell segítségével szegmentálják. Míg a hagyományos diffúziós módszerek nem képesek a textúrázott (erős mikrostruktúrával rendelkező terület) elkenésére és így a kép szegmentálására, addig az általuk javasolt diffúziós módszer úgy lett kialakítva, hogy az erős textúrával rendelkező területeket is szegmentálni tudják.

Ezek után a képek 4-6 jelentősebb területét választják ki - a kiválasztás a mérettől, geometriai tulajdonságtól, elhelyezkedéstől ill. valamilyen egyedi tulajdonságtól függ. Majd ezeknek a kiválasztott régióknak a leírását generálják az MPEG-7 leírók segítségével (Skálázható szín,

Domináns szín, Él-hisztogram, Homogén textúrák). A momentumot leíró invariáns jellemzőket, kontúr és régió leírókat is használnak az alak leírásához. Ezen kívül a szegmensek területét, koordinátáit, projekcióit és a fő tengelyek hosszát is kiszámítják, ill. a fő szegmensek hierarchikus struktúráját is feljegyzik.



5.3. ábra. Szemantika és képi tartalom kapcsolatának kialakítása [50] szerint

A feldolgozásnak ezek után három fő szintjét különböztethetjük meg:

- I. A képi szegmenseket és a szavakat is csoportokba rendezik hasonlóság alapján, majd pedig a csoportok között kapcsolatokat definiálnak. Ezek a kapcsolatok vizuális jellemzőkre vonatkoznak, pl. egy elipszist ábrázoló képrészlet a kör, elipszis, szavakhoz lesz rendelve.
- II. A második szinten tartalmi koherenciával rendelkező szó és szegmens klaszter kapcsolatokat definiálnak. Itt az absztrakt szavakat kiszűrik szótári adatbázisok segítségével.
- III. A legmagasabb szinten fogalmi hierarchiákat generálnak a WordNet³ segítségével [87].

A szintek működését a 5.3. ábra illusztrálja. Az így kialakított rendszer célzott és kategória alapú keresésre is alkalmas, a lekérdezéshez lehet használni kulcsszavakat, kifejezéseket, példa képeket vagy képrészleteket.

³A WordNet az angol nyelv szótári adatbázisa a szavak csoportosításával és kapcsolatok leírásával a szövegelemző eljárások támogatására.

5.5.2. "Bag of words"

Ez a módszer a nyelvi visszakereső rendszerek analógiájára olyan képi elemeket keres és rendszerez, amelyek egyfajta vizuális szótárt hoznak létre [21]. A rendszer tanításakor képi jellemzőket keresünk (pl. élek, sarkok), majd ezeket olyan leírókkal tároljuk, amelyek lehetőleg invariánsak lesznek a kép nagyítására, elforgatására. A vizuális szótár a leírók csoportosítása során jön létre. A képeket, a rajtuk szereplő vizuális szavak előfordulásának gyakorisága modellez, legegyszerűbben egy hisztogrammal lehet mindezt szemléltetni, és a keresés során felhasználni. Mivel a képi elemek egymáshoz való (térbeli) viszonyát a "zsák" alapvetően nem tartalmazza, ezért az alap módszer a strukturális, geometriai jellemzők figyelembe vételét nem tudja megoldani.

6. fejezet

Orvosi képfeldolgozás

Az 1800-as évek végére megnyílt az út az emberiség egyik több ezer éves álmának megvalósulása felé. Wilhelm Conrad Röntgen vákuumcsövekkel végzett fizikai kísérletei közben egy új sugárzást fedezett fel, amely a legtöbb tárgyon áthalad, haladása során a tárgyak anyagától függő mértékben gyengül, és a fényérzékeny fotólemezen erősségének megfelelően nyomot hagy¹. Így lehetővé vált az élő szervezet – elsősorban a csontozat – vizsgálata külső beavatkozás nélkül. Ez a korábban ismert, vagyis az elektromágneses spektrum látható tartományába eső sugárzások segítségével nem volt megvalósítható. A sugárzásnak az ismeretlent jelölő X nevet adta, amit később Röntgen-sugárzásnak neveztek át. A hír gyorsan bejárta a világot és hamarosan megjelentek az első képalkotó berendezések a kórházakban. Röntgen a felfedezésért 1901-ben kapta meg a fizikai Nobel-díjat – ez volt egyben az első fizikai Nobel-díj.

A Röntgen-készülék 2D vetületi képet ad eredményül, ez önmagában térbeli lokalizációra nehézkesen használható, mivel csak az objektumok irányát kapjuk meg. Egymáshoz képest ismert helyzetű képek készítésével (szemből és oldalról) viszont mélységi információ is nyerhető, amit az I. világháborúban már rutinszerűen alkalmaztak bombarepeszek eltávolításánál.

A (folytonos) függvények vetületekből történő helyreállításának, az ún. *rekonstrukció* elméleti hátterét már 1917-ben kidolgozta Johann Radon, viszont még hosszú évtizedekig nem állt rendelkezésre technológia a szükséges számítások elvégzésére. Így a valódi 3D képalkotás megjelenésére az 1970-es évek elejéig kellett várni. 1963-64-ben Allan McLeod Cormack² publikálta a Röntgen-alapú CT képalkotás elméleti hátterét, amit figyelembe véve 1971-ben Sir Godfrey Hounsfield³ készítette el az első működő prototípust Angliában. A berendezést EMI-szkennernek is nevezték, mivel a fejlesztéshez szükséges anyagi forrásokat az akkor a Beatles együttesel együtt dolgozó lemezkiadó cég biztosította. Cormack és Hounsfield 1979-ben megosztva kapta a fiziológiai és orvostudományi Nobel-díjat a számítógépes tomográfia kidolgozásáért.

Az 1970-es években egy újabb, az atommagok rezonanciás tulajdonságain alapuló 3D képalkotó berendezés is készült. Ez a mágneses rezonancia képalkotás (MRI – Magnetic Reso-

¹Wilhelm Conrad Röntgen (1845–1923) Nobel-díjas német fizikus, a később róla elnevezett sugárzás felfedezője.

²Allan McLeod Cormack (1924-1998) dél-afrikai születésű Nobel-díjas amerikai fizikus, a CT képalkotás elméleti hátterének kidolgozója.

³Sir Godfrey Hounsfield (1919-2004) Nobel-díjas angol villamosmérnök, az első CT berendezés készítője.

nance Imaging), amit eleinte NMR-nek hívtak (Nuclear Magnetic Resonance), de a nukleáris jelzőt a hozzá kapcsolódó negatív felhang miatt elhagyták, mivel nincs köze a radioaktivitáshoz. A technika alapjait Paul Lauterbur⁴ és Sir Peter Mansfield⁵ dolgozták ki 1973-ban. Az első, élő szervezetről készült kép 1974-ben, míg emberről 1977-ben készült. A felfedezők 2003-ban kapták meg a fiziológiai és orvosi Nobel-díjat. A díjátadást botrány övezte, Raymond Vahan Damadian⁶ magát tartja az MRI technika kidolgozójának, de díjazásban nem részesült.

Az 1930-as években kutatók sikeresen állítottak elő mesterséges radioaktív anyagokat, aminek eredményeképpen 1946-ban radiofarmakonok (radioizotópok) szervezetbe juttatásával már sikeresen kezeltek betegeket. Ennek sugárzását a testen kívül elhelyezkedő érzékelők mérni tudják, ami képpé alakítható. Ez az alapja a nukleáris medicinában alkalmazott képalakító eljárásoknak, a 2D-s Gamma- vagy Anger-kamerának (Hal O. Anger⁷, 1950-es évek), és a több irányból történő mérés eredményéből tomografikus rekonstrukcióval előálló 3D-s SPECT-nek (Single Photon Emission Tomography; 1960-as évek, David Kuhl⁸ és Roy Edwards) és a PET-nek (Positron Emission Tomography; 1975, Michael E. Phelps⁹). A technikák az alkalmazott izotópok, és ennek megfelelően a sugárzás típusában különböznek. Gyengébb térbeli felbontóképességük ellenére nagy hasznuk abban rejlik, hogy segítségükkel a szervezet funkcionális, biomechanikai folyamatai feltérképezhetők, követhetők.

Az utóbbi évtizedekben a képalakító technikák továbbfejlődtek, gyorsabbá, pontosabbá váltak. 1991-ben jelentek meg az MRI képalakítás olyan válfajai, amelyek a vér oxigénszintjének a mérésével funkcionális jellemzők kinyerését is lehetővé tették. Az ezredforduló környékén jelentek meg a különféle hibrid kamerák (PET/CT, SPECT/CT) amelyek a funkcionális és anatómiai képalakításokat ötvözik egymással. Számos más fizikai jelenség felhasználásával készíthetők még képek. Képként ábrázolható tetszőleges olyan mérés eredménye, amit 2D vagy 3D rácson tudunk végezni. Az 1940-es évek végétől használt ultrahang technika esetében például a különböző hangvezetési képességgel rendelkező szövetek határáról visszaverődő hanghullámok tulajdonságait detektálják és ábrázolják 2D képként. Hőkamerák segítségével a bőrfelszínt pásztázzák, a mért eredményeket képként ábrázolva, pl. gyulladásszerű területeket keresve. Látható fény tartományban is készülnek képek hagyományos fotó vagy videó technikával, pl. a bőr felszínéről, vagy endoszkópos kamerákkal az ér- vagy bélrendszerből.

A következő részben a fontosabb képalakító technikák közül néhányuk sematikus működési elvét mutatjuk be.

⁴Paul Christian Lauterbur (1929–2007) Nobel-díjas amerikai kémikus, az MR képalakítás egyik kidolgozója.

⁵Sir Peter Mansfield (1933–) Nobel-díjas brit fizikus, az MR képalakítás egyik kidolgozója.

⁶Raymond Vahan Damadian (1936–) örmény származású amerikai orvos és feltaláló, az MR technika első publikálója.

⁷Hal Oscar Anger (1920–2005) amerikai villamosmérnök és biofizikus, a 2D-s Gamma-kamera készítője.

⁸David Edmund Kuhl (1929–) amerikai tudós, a SPECT képalakítás egyik atyja.

⁹Michael Edward Phelps (1939–) amerikai biofizikus, a PET képalakítás egyik kidolgozója.

6.1. Képpalkotó technikák

*Anatómiai*nak nevezzük azokat a képeket, amelyek a beteg szerveinek, szöveteinek megjelenítését biztosítják. A *funkcionális képeken* ezzel szemben elsősorban az látható, hogy milyen a vizsgált terület működése (például az oxigénfelhasználása, cukorfelvétele, vérátfolyása). A funkcionális képpalkotó berendezések térbeli felbontóképessége rendszerint gyengébb, mint az anatómiaiaké, de olyan információkat biztosítanak, amelyek más módon nem elérhetők.

Képek készíthetők úgy egy tárgyról, hogy egy külső forrásból származó sugárzás (például röntgen vagy neutron) áthalad rajta, a tárgy belsejében különböző mértékben elnyelődik, a tárgyból kilépve a mögötte elhelyezett érzékelő pedig az átjutott sugárzás intenzitását méri. Ezek a *transzmissziós* technikák, ilyen például a röntgenkép is. Ha a radioaktív anyagot a szervezetbe juttatják és a bomlás során keletkező sugárzást mérik a tárgyon kívül, akkor *emissziós* technikáról beszélünk (ilyen például a gammakamera).

A képek igen sokáig fotótechnikai úton készültek, tanulmányozásuk megvilágított háttér előtt átvilágítással történt. A mai kórházak többségében található orvosi képparchiváló és -továbbító (ún. *PACS - Picture Archiving and Communication System*) rendszer, amely az elkészült képeket digitális formában küldi és tárolja [37], a megjelenítés számítógépek segítségével történik.

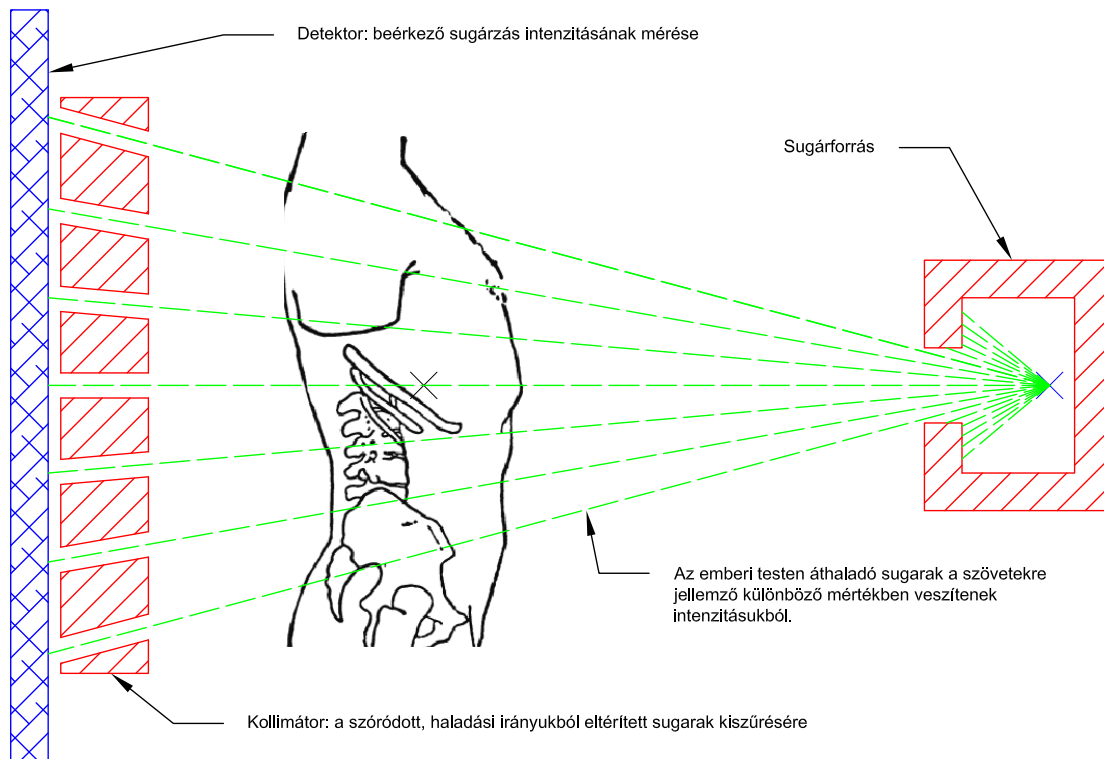
6.1.1. Röntgen, fluoroszkóp

A módszer alapja a katódsugárcsőben a magas feszültség (50–120 kV) által felgyorsított elektronok fémes anyaggal történő ütköztetésekor keletkező Röntgen-sugárzás felhasználása. A pontszerű forrás által kibocsátott és az emberi testen áthaladó Röntgen-sugarak kilépési intenzitása függ az emberi szövetek abszorpciójának a röntgensugarak menti integráljától. A testen áthaladó jelek filmre (hagyományos Röntgen-készülék) vagy képerősítőkre (fluoroszkóp) kerülnek (6.1. ábra).

A röntgen képpalkotási technika elsősorban a csontszerkezet ábrázolására használatos, de szemléltetheti még a kontrasztanyagok, mint a pl. jódot helyzetváltoztatását is (pl. vérerekben). Általánosan használják különféle szűrővizsgálatok nagy tömegű és gyors elvégzésekor is (pl. TBC szűrés, mammográfia), valamint számítógéppel támogatott műtévtvégrehajtás esetében, a műtét közben az elvégzett beavatkozások ellenőrzésére.

A Röntgen-berendezések elterjedtek és relatíve olcsók (összehasonlítva a 3D képpalkotó berendezésekkel), de mivel a képek a beteg 3-dimenziós testéről készült vetületek, így egy képpont intenzitása egy, a betegen áthaladó egyenes irányában található összes szövet együttes tulajdonságát jellemzi. Különböző nézőpontból készült képek esetén más irányok mentén készül a vetület, így a képpontok intenzitásértékei nehezen vethetők össze. Ebből adódóan a 2D képek elsősorban diagnosztikai célokat szolgálnak. Ha segítségükkel a diagnózis nem állítható fel egyértelműen, szükség lehet valamilyen 3-dimenziós, sokszor jóval időigényesebb és költségesebb vizsgálat pótlólagos elvégzésére.

A Röntgen-készülékek használata az ionizáló sugarak miatt egészségkárosító lehet. A sugárzás képes kapcsolatba lépni a sejtek DNS-ével, azokban mutációkat okozni, ami több év alatt akár daganatos megbetegedésekhez is vezethet. A technika korai alkalmazói körében ez igen gyakori halálok volt. A legnagyobb sugárterhelésnek a berendezéseket rendszeresen



6.1. ábra. 2D röntgen képalkotás sematikus váza.

használó személyzet, orvosok vannak kitéve. A veszélyek felismerésével ezt sikerült visszaszorítani. Az ólom a sugárzást elnyeli, így ólom tartalmú védőruhát, védőfalat alkalmaznak.

6.1.2. Rekonstrukció

Radon¹⁰ 1917-ben megmutatta, hogy (folytonos) 2D függvények egyértelműen visszaállíthatók, amennyiben minden lehetséges, vagyis végtelen sok síkbeli irányban ismerjük a függvényértékek vonalmenti integrálját:

$$R_f(m, b) = \int_{-\infty}^{\infty} f(x, m \cdot x + b) dx = p(\theta, s)$$

ahol θ jelenti az egyenes normálvektorának irányszögét, s pedig a normálvektor irányában az origótól mért távolságát. Adott irányhoz tartozó R_f értékek összességét *vetületnek* nevezzük.

Az eredeti függvényt a vetületek megfelelő irányokba történő „visszavetítésével” kapjuk meg. A vetületek és az eredeti függvény között a vetület-szelet tétel teremti meg a kapcsolatot, miszerint egy 1D vetület Fourier transzformáltja és az eredeti 2D függvény Fourier transzformáltjának a vetület irányába eső értékei megegyeznek egymással. A vetületek Fourier transzformáltjaiból fel tudjuk építeni a kép Fourier transzformáltját, majd az inverz Fourier

¹⁰Johann Karl August Radon (1887–1956) osztrák matematikus, a rekonstrukciós technika matematikai hátterének kidolgozója.

transzformáció alkalmazásával helyreállíthatjuk a függvényt. Belátható, hogy

$$f(x, y) = \int_0^\pi \left(\int_{-\infty}^{\infty} P(w, \theta) \cdot e^{-2 \cdot \pi \cdot i \cdot w \cdot s |w|} dw \right) d\theta$$

vagyis a vetületek $P(w, \theta)$ Fourier transzformáltján visszavetítéskor végre kell hajtani egy $|w|$ (lejtő) szűrést. Ez a Fourier térben a magasabb frekvenciás komponenseket „felerősíti”. A módszert az alkotóelemei alapján *szűrt visszavetítésnek* nevezzük.

Az orvosi gyakorlatban képalkotáskor csak véges számú, ráadásul diszkrét (nem folytonos) vetületi információ áll rendelkezésünkre. Ezek összességét *szinogramnak* nevezzük, ami képként ábrázolható: a kép egy oszlopa egy vetületi irány értékeinek felel meg. A vetületi adatok hibával terhelt, zajos mérési eredmények. A zaj általában magas frekvenciás komponensként jelentkezik, és ahogyan láttuk, a visszavetítéskor alkalmazott szűrő ezeket erősíti fel. A gyakorlatban ezért inkább más, a magas frekvenciákat kevésbé kiemelő, pl. Shepp-Logan szűrőt alkalmaznak. A rekonstrukció minőségét nagymértékben befolyásolja a mérés pontossága és a vetületi irányok száma. Néhány rekonstrukciós eredményt a 6.2. ábrán láthatunk.

A fenti képletek azt feltételezik, hogy a vetítési irányok egymással párhuzamosak. A gyakorlatban más konfiguráció is elképzelhető, pl. a CT esetén pontszerű forrásból kiinduló, legyezőnyaláb alakú a leképezés. A módszer ilyen esetekre is kidolgozható. Szintén megoldható 2D vetületi képek alapján a 3D képek rekonstrukciója is.

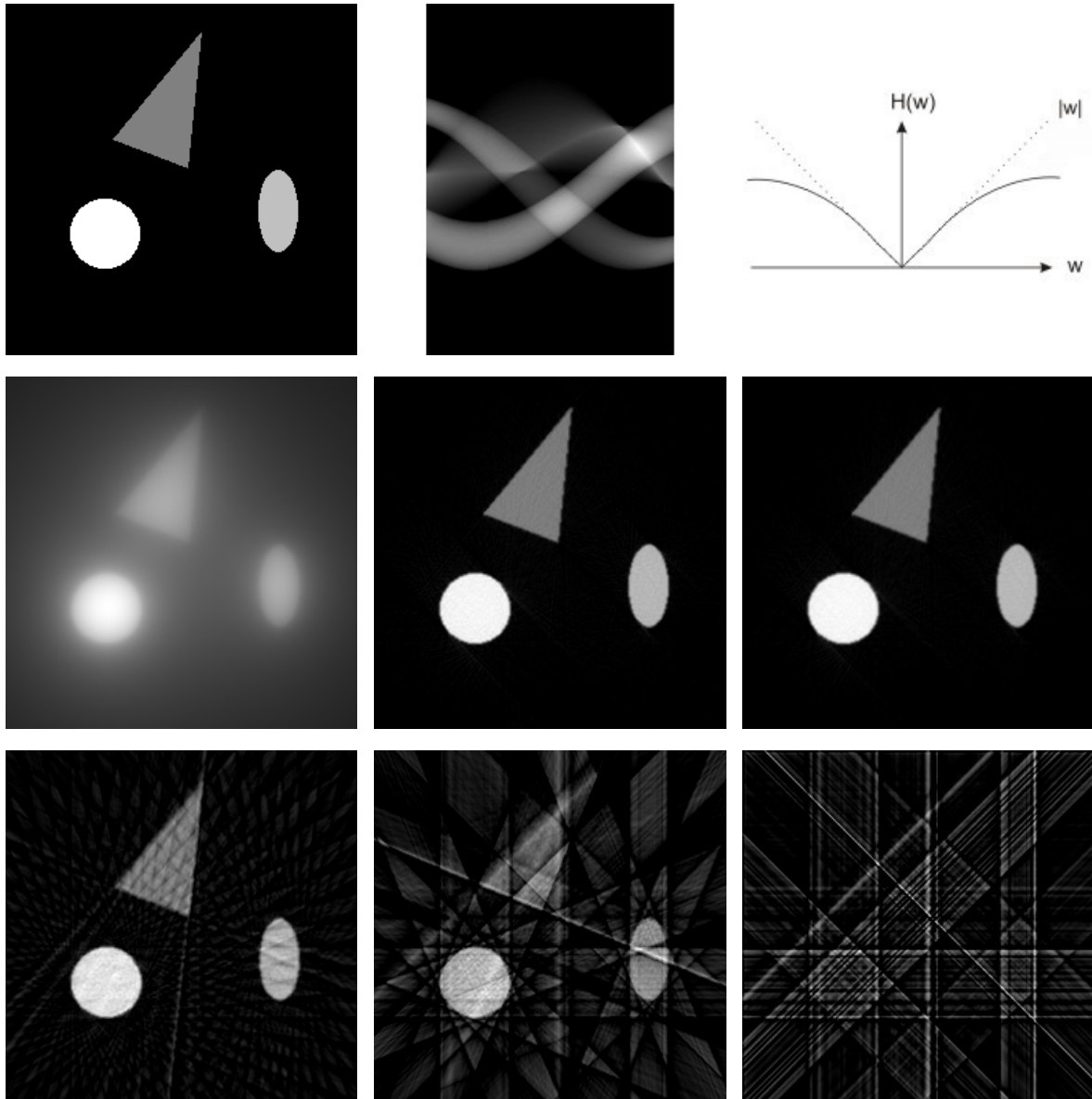
A szűrt visszavetítés az egyik leggyakrabban alkalmazott rekonstrukciós technika a CT, PET és SPECT képalkotáskor, de számos más rekonstrukciós megközelítés is létezik. Ezeket itt nem tárgyaljuk.

6.1.3. CT (Számítógépes tomográfia)

A CT (*Computed Tomography*) fizikai elve megegyezik a 2D röntgenével. A hagyományos CT berendezések a detektorgyűrű egy körülfordulásával egy 2D szelet rekonstrukcióját teszik lehetővé, majd a szelettávolságnak megfelelő mértékben az asztalt továbbmozgatják. A forrásnak és a detektornak a beteg hosszanti tengelye körüli forgatásával kapott projekciókból transzaxiális párhuzamos szekciók sorozata állítható elő tomografikus rekonstrukcióval (6.1.2. alfejezet). Nagy hátránya ennek a technikának a lassú képalkotás, mivel 1 szelet elkészítése kb. 3 másodpercebe telik. A teljes mellkas képének előállítása közben emiatt légzés is történik, ami az egyes lágy szövetek jelentős elmozdulását is maga után vonja. Az 1989-ben megjelent *spirál CT*-k a detektor körülfordatását és az asztal mozgatását egyszerre végzik, ami gyorsabb képalkotás biztosít. A rekonstrukció utólagosan tetszőleges szelettávolsággal előállítható a mért adatokból, viszont az eredmény zajosabb, mint a hagyományos technikával készült. További gyorsítás érhető el *multidetektoros* berendezésekkel, ami több, pl. 4 érzékelő egymás melletti elhelyezését jelenti.

Mivel a képalkotás korrekt közelítését adja a szövetsűrűségeknek, így a CT képek megkülönböztetik a csont- és a lágy szöveteket. Emellett a CT képek intenzitásértékei normalizálhatók; a μ_X elnyelődési tényezővel rendelkező X anyag ún. Hounsfield értéke (HU) megkapható az alábbi képlettel:

$$HU = \frac{\mu_X - \mu_v}{\mu_v - \mu_l} \cdot 1000,$$



6.2. ábra. Rekonstrukció szűrt visszavetítéssel. Felső sor: eredeti kép, a 180 vetületi irány alapján készült szinogram, valamint a lejtő (pontozott vonal) és a Shepp-Logan (folytonos vonal) szűrők képe a frekvenciatérben. Középső sor: rekonstrukció szűrés nélkül, lejtő és Shepp-Logan szűréssel, 180 különböző vetületi irány felhasználásával. Alsó sor: rekonstrukció 36, 8 és 4 vetületi irány felhasználásával, Shepp-Logan szűréssel. A vetületek számának csökkenésével jelentősen romlik a rekonstrukció minősége.

ahol μ_v és μ_l jelentik a víz és a levegő elnyelődési tényezőit. A CT képalkotás a csontszerkezet ábrázolásában különösen jó, a lágy szövet rész viszont csak egy szűk, ráadásul szervenként átfedő intervallumban jelenik meg.

A CT-k 70-es évekbeli megjelenése inspirálta a számítógéppel segített műtéttervezést. Geometriailag korrekt képet ad. Elterjedt és gyors technika, károsító hatása ugyanaz, mint a röntgennek. Térbeli felbontása nagyon jó: axiális síkon 1 mm alatti, egy szelet vastagsága

néhány mm. Általában 512×512 képpont méretű síkmetszeteket biztosít.

6.1.4. MRI (Mágneses rezonancia képalkotás)

Az MRI (*Magnetic Resonance Imaging*) a szövetek tulajdonságairól nyújt információt. Először a testet egy konstans intenzitású mágneses mezőbe helyezik. Ennek hatására a szövetekben található, spinnel vagy páratlan számú protonnal vagy neutronnal rendelkező atommagok (döntő többségben hidrogénatomok) – kis mágnesként viselkedve – a mágneses erővonalaknak megfelelően egyensúlyi helyzetbe kerülnek. Az MRI a szövetek viselkedését vizsgálja, amikor azokat gyenge mágneses mező perturbációjának teszik ki: a rádiófrekvenciás pulzusok a spinek elmozdulásait indukálják, amelyek aztán rádióhullámot gerjesztve térnek vissza az egyensúlyi állapotba. Egy kiszemelt perturbáció-sorozat használatával információ nyerhető ki az atomok sűrűségéről és a fizikai-kémiai környezetről (T1 és T2 válaszüvek, protonszűrűség¹¹). A képalkotás során rendszerint a kapott jelek számos komponensét használják fel. A mágneses mező inhomogenitása miatt a kép nemlineáris torzítást szenved. Ez a torzítás kis-mértékű, de idegsebészeti alkalmazásokban, ahol a precíz mérések fontosak, gondot okozhat.

Az MRI elterjedt, bár lassú (egy felvétel 15-45 perc is lehet) és drága technika, egészségkárosító hatása nem ismert. Az erős mágneses mező miatt a vizsgálat nem minden betegen végezhető el, például a ferromágneses implantátumok vagy szívritmusszabályzók kizáró tényezők. A szűk berendezésben töltött hosszú idő klausztrófóbiás reakciókat is kiválthat.

Térbeli felbontása nagyon jó: transzaxiális síkon 1 mm körüli, egy szelet vastagsága néhány mm. Általában 256×256 -os síkmetszeteket biztosít.

6.1.5. Nukleáris medicina

A nukleáris medicinában (*Nuclear Medicine* vagy *Scintigraphy*) a radioaktív izotópot tartalmazó anyagot a szervezetbe juttatják (például a vérkeringésbe fecskendezik vagy a beteg lenyeli), ami eloszlik a szervezetben. Bizonyos idő után a radioaktív elemek kibocsátását vizsgálják külső vonal- vagy síkdetektorokon.

Gamma- vagy Anger-kamera

Elve a SPECT-tel egyezik meg, csak itt egy 2D kép készül, ami radioaktív elemsűrűség egy ortogonális projekciójának felel meg.

SPECT (Single Photon Emission Tomography)

SPECT esetén olyan izotópot használnak, amely a bomlása során gamma sugárzást bocsát ki. Ezek az izotópok olcsón előállíthatók, és nagy a felezési idejük. Az egyik legfontosabb ilyen izotóp a technécium-99m, amelynek kb. 6 órás felezési ideje jóval nagyobb, mint más izotópoké, és a bomlás után nagyon alacsony sugárterhelést okozó normál technéciummá alakul, ami ráadásul pár nap alatt teljesen kiürül a szervezetből.

¹¹ A T1- és T2-súlyozások a zsír és a víztartalmú szövetek szétválasztására használatos. T1 esetén a víz sötét, a zsírszövet világos intenzitástartományban ábrázolódik. T2 esetén a víz világos, a zsír sötét. Protonszűrűség kép a szabad és kötött víz elkülönítésére használható.

A gamma kamera beteg körüli forgatása 2D projekciók egy halmazát biztosítja, melyeket tomografikus algoritmusokkal dolgoznak fel. Agyi vérátfolyás vizsgálatára, perfúziós mérésekre használják. Elterjedt technika. Térbeli felbontása gyenge: axiális síkon néhány mm körüli, egy szelet vastagsága akár 6-10 mm is lehet. A síkmetszetek általában 64×64 vagy 128×128 méretűek.

PET (Positron Emission Tomography)

PET esetén olyan izotópot használnak, amely a bomlása során pozitront bocsát ki. Minden kibocsátott pozitron összeütközik egy, a környezetében lévő elektronnal és két, ellenkező irányban haladó fotont idéz elő. A beteg köré helyezett detektorok gyűrűje érzékeli ezeket a fotonpárokat, így a társított esemény egy egyenes vonalon van két megfelelő detektor között. Tomografikus rekonstrukcióval állítják elő a 3D-s képet.

Elsősorban anyagcsere folyamatok megfigyelésére és daganatok korai detektálására használatos. A leggyakoribb alkalmazott izotóp a 2-fluorodeoxy-D-glükóz (2FDG), amely működését tekintve a cukorral egyenértékű. Ez a daganatos szövetekben mutat dúsulást olyan stádiumban is, amikor anatómiai képen még nem látható elváltozás.

Kevésbe elterjedt technika: az izotópok előállításához elektrongyorsító (ciklotron) szükséges, felezési idejük nagyon kicsi (pár perc), így szinte csak az előállítás helyén használhatók fel. Magyarországon sokáig csak Debrecenben volt PET centrum. 2005 óta Budapesten is üzemel két PET/CT hibrid berendezés, és időközben Kecskeméten is nyílt ilyen centrum. Térbeli felbontása gyenge: axiális síkon néhány mm körüli, egy szelet vastagsága 4-8 mm is lehet. A síkmetszetek általában 128×128 vagy 256×256 képpontból állnak.

6.1.6. fMRI (funkcionális MR képalkotás)

1990-től a nukleáris medicina mellett egy másfajta funkcionális képalkotó eljárás is megjelent, amely az MR képalkotásra épül. Az fMRI az idegi aktivitás hatására bekövetkező agyi változásokat, általában a vér oxigénellátottságának változásait detektálja az ún. *BOLD effektus*¹² felhasználásával. Az agyról először egy normál MR kép készül, amelyet később az anatómiai részek pontosabb meghatározására használnak fel. Ezután alacsony térbeli felbontással (kb. 4 mm-es szeletszintű felbontás), de gyorsan, pár másodpercenként készülnek képek 10-15 percen keresztül. A vizsgálat ideje alatt a beteget időnként bizonyos ingereknek teszik ki, például képet mutatnak neki, kezével mozgást végeztenek, hangokat hallgattatnak vele. A nyugalmi helyzet és az aktív szakasz közötti különbségek feltérképezésével megkapható, hol került felhasználásra több oxigén, ami arra utal(hat), hogy az inger hatására az agy mely részei voltak aktívak. Képfúzióval ezeket az aktív területeket az MR képre lehet vetíteni.

6.2. Digitális képek geometriája

A képi információ számítógépes méréséhez, feldolgozásához a képeket digitális formában várjuk. A legtöbb mai képalkotó berendezés eleget tesz ennek az elvárásnak. Megvizsgáljuk,

¹²Vér oxigén-szint függőség (*Blood-oxygen-level dependence*).

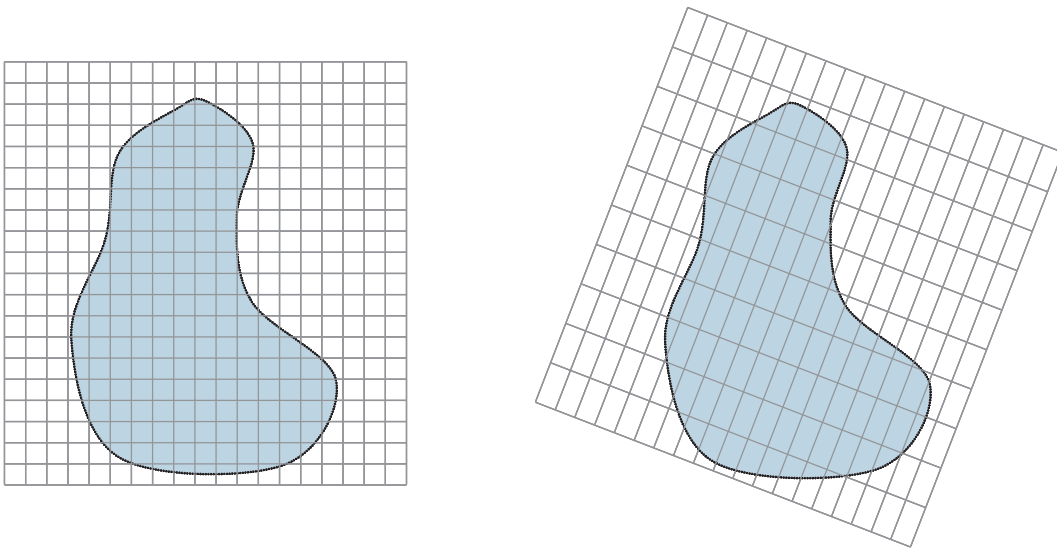
hogy a különféle mérésekből hogyan készül digitális kép és mi ezek geometriai reprezentációja.

6.2.1. Mintavételezés

A jegyzet további részeiben [30] alapján a következő jelöléseket használjuk. Jelölje X a vizsgált tárgyat, A a róla készült 2D vagy 3D digitális képet, Ω_A pedig A képtartományát

$$A : x_A \in \Omega_A \mapsto A(x_A),$$

ahol $A(x_A)$ az x_A térbeli pozícióhoz tartozó intenzitásértékeket jelöli. Az intenzitásértékek valamilyen mérhető anyagjellemző értéket reprezentálnak X térbeli pozícióiban. Ilyen például a röntgensugarak elnyelődési tényezője CT képek esetén, protonok állapotváltozásai a mágneses mező tulajdonságainak változásakor MR képek esetén, radioaktív jelzőanyagok szerkezetbeli eloszlása SPECT és PET képalkotáskor, illetve a látható fény intenzitása fényképezéskor.



6.3. ábra. Ugyanazon tárgyról készült két különböző térbeli felbontású 2D digitális kép mintavételezési rácsai. A képpontok helye a rácspontokban található.

Mivel a digitális képek az objektum véges számú pontban történő mintavételezésével készülnek, így az Ω_A tartományt a következő módon definiálhatjuk:

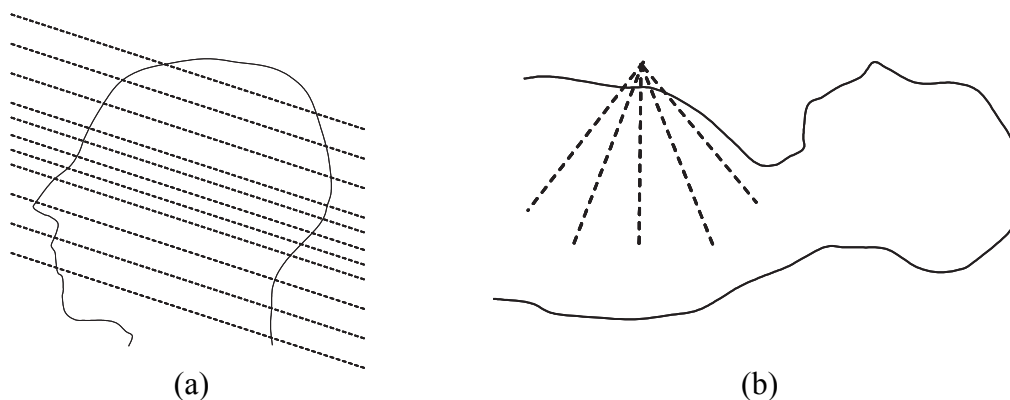
$$\Omega_A := \tilde{\Omega}_A \cap \Delta,$$

ahol $\tilde{\Omega}_A$ egy folytonos korlátos halmaz, valamint Δ egy végtelen diszkrét mintavételezési rács, amely a $\delta = (\delta^x, \delta^y, \delta^z)$ anizotropikus mintavételezési lépésközökkel jellemezhető. Ez a rács képenként más és más lehet (6.3. ábra). A rácspontokat és a hozzájuk tartozó mintavételezett értékeket képpontnak nevezzük. A mintavételezés történhet pontosan az adott pontban, de igen gyakran annak egy szűk környezete is szerepet játszik az intenzitásérték meghatározásában, emiatt a rács definíciójával duális *mozaik* definíció is széles körben használatos.

Ebben az esetben megkülönböztetjük a *szelettávolság* és *szeletvastagság* fogalmakat. Előbbi a mintavételezési rács Z-irányi lépésközét jelenti (δ^z), míg az utóbbi annak a térrésznek a vastagságát, amelyből a mérési információk származnak. Optimális esetben ez a két távolság megegyezik, de előfordulnak a szelettávolságnál kisebb és akár nagyobb szeletvastagságok is. Különösen ez utóbbi kezelése lehet problémás, mivel így egy térrész információtartalma több szeleten is megjelenik.

A képpontokat 2D kép esetén *pixelnek*, 3D esetén *térfogatelemnek* vagy *voxelnek* is nevezzük. Ezen képpontok, pixelek vagy térfogatelemek összesége alkotja a *digitális képet*. A képpontokhoz rendelt mérési érték az *intenzitásérték*. Ezek tárolása történhet előjeles egész számként (8-16 biten), vagy lebegőpontos számként is. *Idősorozatról* beszélünk akkor, ha egyező méretű 2D vagy 3D képek egy sorozatát készíti el a berendezés.

A 3D képek reprezentációjakor a legegyszerűbben kezelhető eset az, amikor a mintavételezési rács $\delta = (\delta^x, \delta^y, \delta^z)$ lépésközű a képkalkotó berendezés fő tengelyeivel egyező irányok mentén. Általában teljesül, hogy $\delta^x = \delta^y$ (ezek alkotják a 2D képszeleteket), a δ^z pedig gyakran nagyobb lépésközű, mint a másik kettő. Bizonyos CT vizsgálatok esetében előfordul, hogy a Z-tengely menti mintavételezés lépésköze nem konstans, a vizsgálat szempontjából fontosabb területeken sűrűbb. Szintén a CT vizsgálatokra jellemző, hogy a képkalkotás síkját megdönthetik, így a tengelyek nem lesznek egymásra merőlegesek. Ezekre mutat példát a 6.4. (a) ábra.



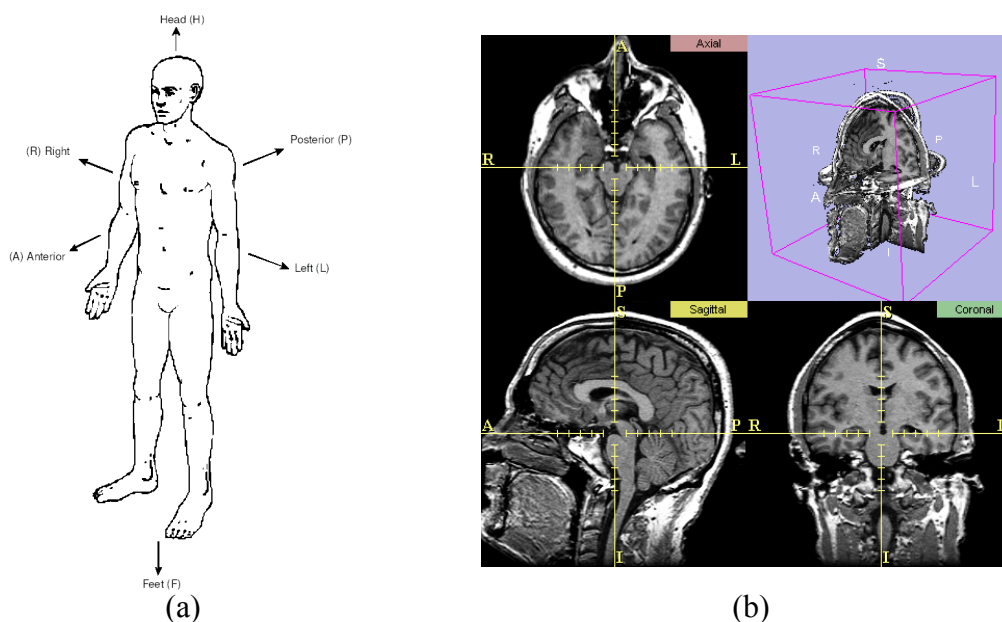
6.4. ábra. Képszeletek elhelyezkedése. Döntött képkalkotó sík és változó lépésközű Z-tengely menti mintavételezés CT képkalkotáskor (a). Egymással nem párhuzamos képszeletek ultrahang képkalkotáskor (b).

A legáltalánosabb esetben a képszeletek nem párhuzamosak egymással. Ekkor minden képszelethez hozzárendelnek egy térbeli pozíciót (ez lesz például a képszelet bal alsó sarka), valamint két irányvektort, amely a mintavételezés irányát adja. A δ^x és δ^y lépésközök rendszerint ekkor is megegyeznek. A 6.4. (b) ábra egy ilyen helyzetet ábrázol.

6.2.2. Képi koordináta-rendszerek

A képpontok helyének leírása többféle koordináta-rendszer szerint is lehetséges. Amennyiben a kép 2D vagy 3D mátrixként reprezentálható, a képpontok legegyszerűbben a sor, oszlop, képszelet hármassal érhetők el. Ezt nevezzük IJK koordináta rendszernek. Előnye, hogy

könnyen megfeleltethető a programozási nyelvek tömb-fogalmának, így a tényleges megvalósításkor rendszerint más koordináta-rendszerekről is erre térnek át. Hátránya, hogy általában nem izotropikus, az irányonként gyakran különböző mintavételezési lépésközök miatt Euklideszi távolság mérésére alkalmatlan, megjelenítéskor torzított képet ad. Ezt hidalja át a világ (WLD) koordináta-rendszer, amely már egy milliméter egységű rendszer, ami például a képkalkító berendezés koordináta-rendszerével egyezik meg. A mintavételezés lépésközének ismerete, általános esetben a képszeletek sarokpontja és a két mintavételezési irány ismerete elegendő az IJK–WLD átváltáshoz. Orvosi képek esetében rendkívül fontos, hogy a képek felhasználásakor egyértelmű legyen, mi a beteg orientációja, melyik például a bal és jobb oldala. Az orvosi képmegjelenítő szoftverek emiatt a beteg-központú, ún. RAS koordináta-rendszert használják. A tengelyek itt a balról-jobbra (*Right*), hátulról-előre (*Anterior*), alulról-felfelé (*Superior*) irányokat jelentik. A WLD koordináta-rendszerhez hasonlóan ez is milliméter egységű. Az orvosi képek küldésére és tárolására létrehozott DICOM szabvány a beteg fekvési helyzetét és a képszeletek irányát is tartalmazza, amiből az IJK–RAS koordináta-rendszer váltás meghatározható [2].



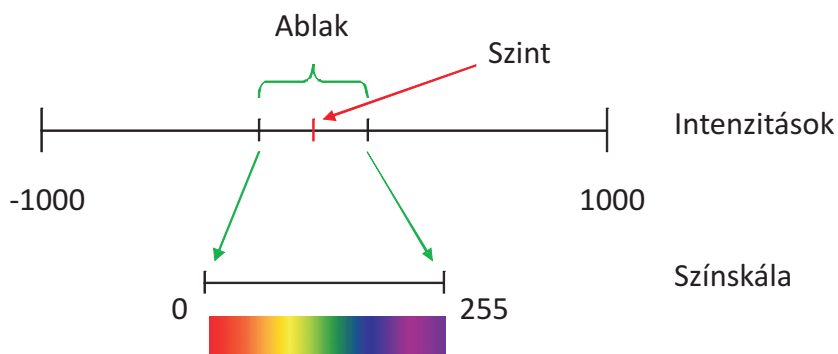
6.5. ábra. Beteg-központú RAS koordináta rendszer (a). Transzaxiális, sagittális és koronális szeletorientációk agyvizsgálat esetén (b).

6.2.3. 3D szeletorientációk

A RAS koordináta-rendszert figyelembe véve három fő szeletorientációt definiálhatunk. A *transzaxiális* a beteg hosszanti tengelyére, a *sagittális* a bal-jobb tengelyre, a *koronális* a hátulról-előre tengelyre merőleges képsíkokat jelenti. Ezekről a fő irányoktól eltérő térbeli irányokra merőleges képsíkok esetén *ferde* szeletekről beszélünk. A RAS koordináta-rendszert és a szeletorientációkat a 6.5. ábrán szemléltetjük.

6.3. Vizualizáció

A berendezések különféle fizikai tulajdonságok (pl. röntgensugárzás elnyelődése, emissziós radiokatív sugárzás erőssége, hangterjedés és visszaverődés) mérésének eredményeit használják fel, tehát legtöbbször nem szín-jellegű információt hordoznak. Ilyen esetekben a képernyőn való megjelenítéshez szükséges az intenzitástartomány egy általunk kiválasztott részének színskálára történő leképezése. A leggyakrabban használatos a szürkeárnyalatos színskála, de elsősorban funkcionális információt tartalmazó képeknél a színkódolás (például szivárvány színskála¹³, vagy a forró fém skála¹⁴ alkalmazása) is közkedvelt (6.6. ábra).



6.6. ábra. Intenzitástartomány leképezése szivárvány színskálára ablakolással.

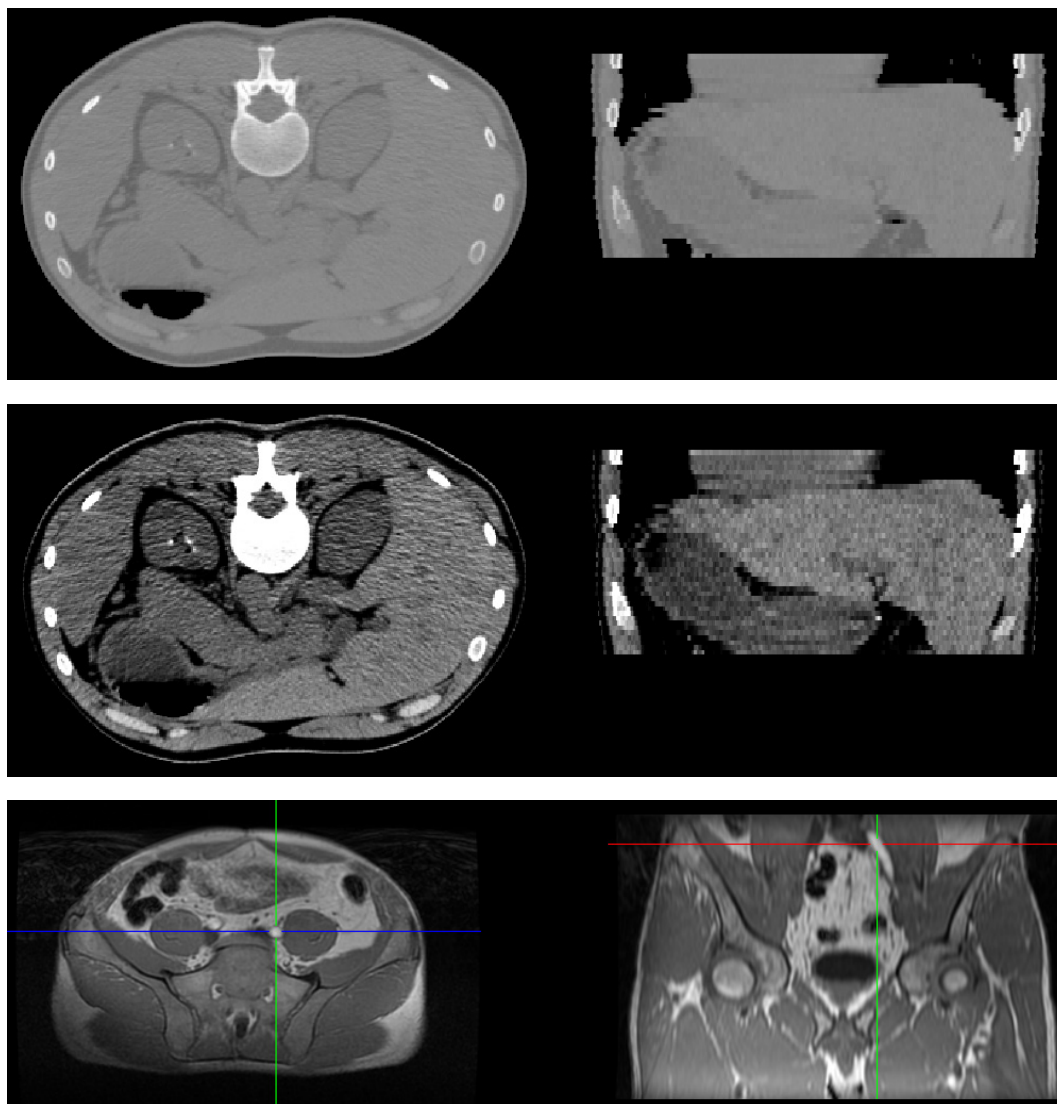
Az orvosi képek intenzitástartománya sokszor igen nagy, több ezer értéket is felvehet. A megjelenítéshez felhasznált színskálák ezzel szemben rendszerint jóval kevesebb, pl. 256 különböző szint definiálnak. Emiatt a színskálára történő leképezés információvesztéssel jár. Fontos, hogy ez a leképezés csak a megjelenítéshez kerüljön kiszámításra, a képi mérések elvégzéséhez az eredeti intenzitástartományt használjuk fel. Az intenzitástartomány vágásával, ún. *ablakolással* lehetőségünk van egy szűkebb résztartomány kontrasztosabb megjelenítésére, pl. a CT képből az igen szűk látgy szöveteket megjelenítő tartományt kiemelhetjük. Ezt a tartományt megadhatjuk a legkisebb és legnagyobb intenzitásértékével is, de az orvosi szoftverek esetében a *szint* és az *ablak* paraméterek használatosak. Az ablak paraméter a kívánt intenzitásintervallum szélességét, míg a szint paraméter az ablak középső értékét adja meg (6.6. és 6.7. ábrák).

A 3D képek térbeli megjelenítésére többféle klasszikus módszer is rendelkezésünkre áll. A natív képszeletek mellett tetszőleges más szeletorientáció előállítható mesterségesen a 3D mátrix megfelelő bejárásával, ezek együttes megjelenítése az ún. *többszeletes rekonstrukció* (6.5. (b) ábra). (A rekonstrukció kifejezés itt nem a vetületi adatokból történő helyreállítást, hanem a rendelkezésre álló képi adatok strukturált megjelenítését jelenti).

A *felszín megjelenítés* esetén egy előzetes szegmentáló lépésre van szükségünk, majd a szegmentált területek határát 3D poligonhálavá kell alakítanunk. Ezek a poligonok a 3D grafika szokásos eszköztárával megjeleníthetők (szervenként különböző színek, mintázatok, átlátszóság), akár a többszeletes rekonstrukcióval együtt is (6.8. ábra). A modellek térben körüljárhatók.

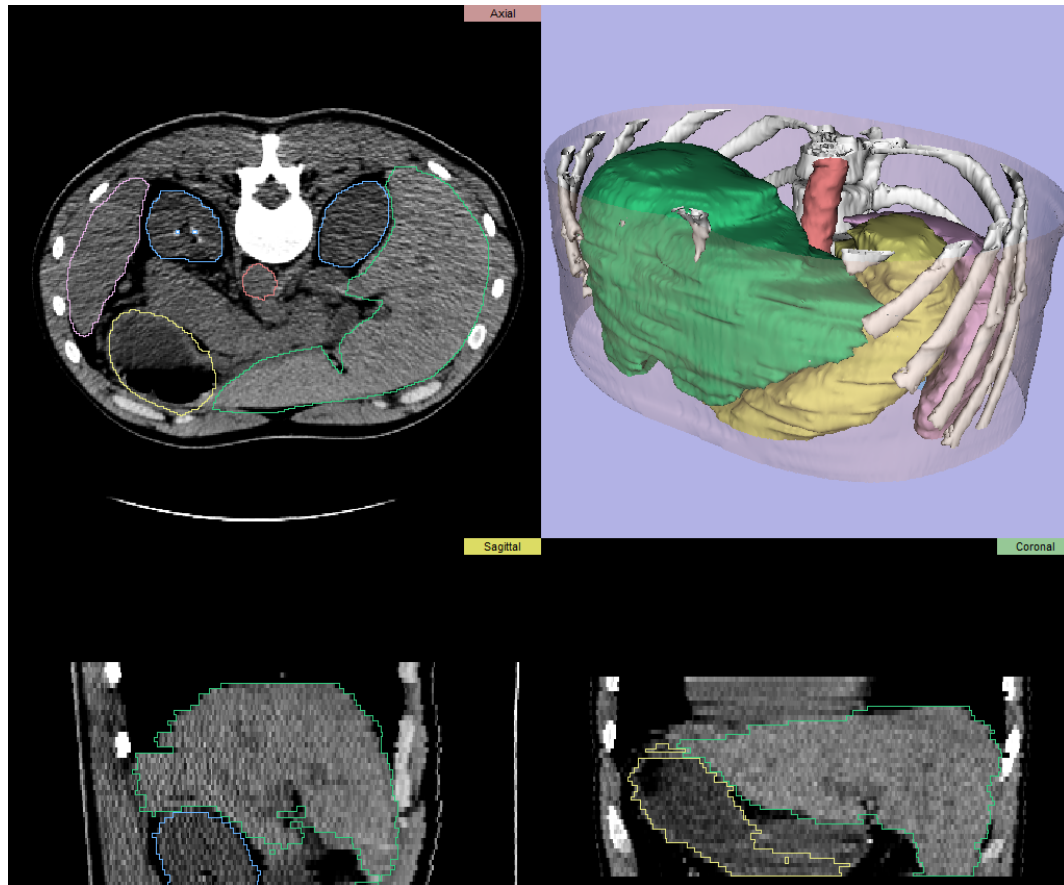
¹³A színek a látható fény spektrumának (a szivárvány színeinek) felelnek meg.

¹⁴Az egyre forróbb hőmérsékletű izzó fém színváltozásai a sötétvöröstől a narancssárgán át a fehérig.



6.7. ábra. Alhasi CT és MR felvételek. Felső sor: contra ablakozott CT vizsgálat transzaxiális és koronális szeletei (ablak: 1500, szint: 0). Középső sor: ugyanazon szeletek lágy szövetre történő ablakozással (ablak: 300, szint: 50). Alsó sor: medence-környéki MR vizsgálat transzaxiális és koronális szeletei.

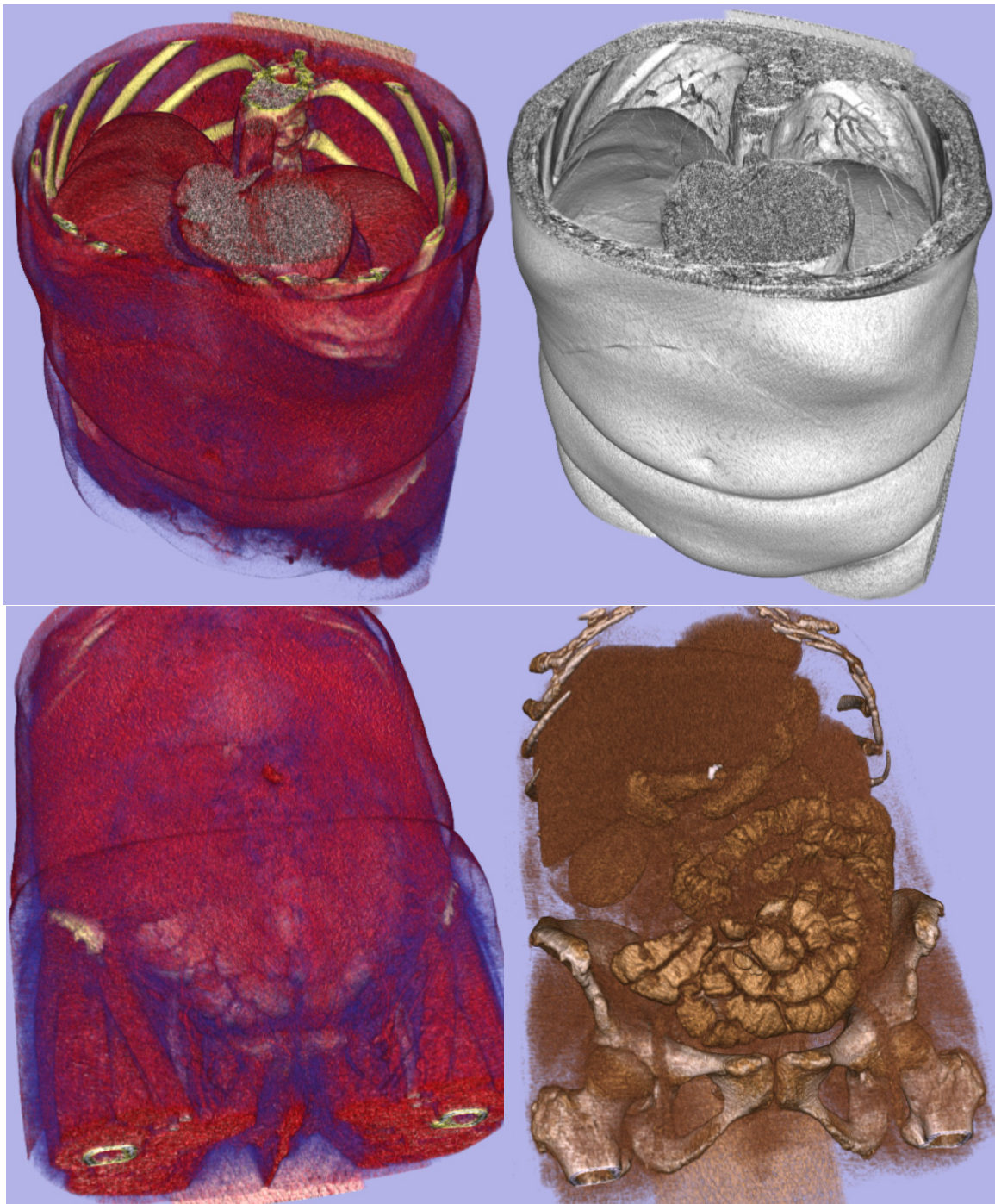
A *térfogat megjelenítő* módszerek geometriai adatok kinyerése nélkül képesek a képpont adatokat közvetlenül megjeleníteni a képernyőn. Sugárkibocsátó technikával jó minőségű, nagy felbontású képeket kaphatunk. A módszer elve, hogy a 3D képen keresztül sugarakat bocsátanak, amelyek interakcióba lépnek az útjukba kerülő intenzitásértékekkel. Ezek különböző mértékben elnyelik a sugár egyes komponenseit, majd az objektumon átjutó sugarak adják a 2D vetületi képet. A módszer kritikus része az interakció definiálása. Ez leggyakrabban transzfer függvények segítségével történik, így minden intenzitásértékhez könnyen és gyorsan hozzárendelhetünk szín és átlátszósági paramétereket. Hasonló módon járhatunk el a gradiens értékek esetén is, ekkor a szomszédos képpontok intenzitáskülönbségei alapján történik a szín és átlátszóság hozzárendelése, ez az objektumok határvonalának kiemelésére



6.8. ábra. Felszín megjelenítés. A CT szeleteken az egyes szervek körvonalai láthatók. A szegmentálás manuálisan történt.

használható. A gyakorlatban használatosak még az ún. MIP technikák is, amikor egy sugár útjába jutó intenzitásértékek minimumát vagy maximumát határozzák meg.

Az általános vizualizációs megközelítések mellett alkalmazás specifikus algoritmusokkal is találkozhatunk. Csőszerű objektumok esetében (pl. bélrendszer, légutak, artériák) a rendelkezésre álló szegmentálást felhasználva virtuális endoszkópiát, kolonoszkópiát, disszekciót (boncolást) végezhetünk a szegmentált objektum középvonala menti kameramozgatással.



6.9. ábra. Egy alhasi CT vizsgálat térfogat vizualizációja többféle transzfer függvény alkalmazásával, kétféle nézőpontból.

6.4. Orvosi információs rendszerek

A kórházakban, egészségügyi intézményekben naponta óriási mennyiségű információ keletkezik, többek között vizsgálati eredmények, laboreredmények, zárójelentések, műtéti protokollok, adminisztrációs adatok formájában. A hatékony működéshez kulcsfontosságú, hogy ezek az információk rendezetten tárolásra kerüljenek és gyorsan visszakereshetők legyenek. Ez a *kórházi információs rendszerek (HIS – Hospital Information System)* alapfeladata. A számítógépes rendszerek nagy segítséget nyújtanak a legtöbb részfeladat megvalósításában. Általánosan elmondható, hogy a kórházi rendszerek annyira komplexek, hogy nehéz mindenre kiterjedő, egységes megközelítést kidolgozni.

Igen fontos részfeladat az orvosi képek tárolása és visszakeresése. Manapság a legtöbb kórház már rendelkezik *képtároló- és továbbító*, ún. *PACS (Picture Archiving and Communication System)* rendszerrel. A PACS rendszerekkel szembeni fő elvárások az alábbiak:

- A képkötő berendezések digitális formában állítsák elő a vizsgálatokat, azok formátuma szabványos legyen.
- A vizsgálatok biztonságos hálózati kapcsolaton keresztül automatikusan kerüljenek be egy adatbázisba.
- A képi adatbázisból az orvos el tudja érni a szükséges vizsgálatokat kliens szoftveren vagy akár a weben keresztül.
- A vizsgálatokhoz rendelt új információk (pl. szegmentálás, leletezés) kerüljenek be az adatbázisba.
- Rendkívül fontos a hozzáférési jogosultságok kezelése, a vizsgálatokhoz tartozó privát adatok védelme érdekében.
- Gondoskodni kell az adatok megfelelő mentéséről az adatvesztések elkerülése érdekében.

A vizsgálatok esetén a képi adatok (geometria és intenzitásértékek) mellett számos járulékos információt is célszerű tárolni. Ilyen például a vizsgálat és a képsorozat egyedi azonosítói; a beteg neve, azonosítója; a képet készítő radiológus neve, a kórház megnevezése; a képkötő berendezés paraméterei; a felvétel készítésének körülményei, dátuma, időpontja; a megjelenítéshez alkalmazandó színskála; szegmentálás eredménye, leletezés.

A képi információ általában tömörítetlen formában kerül tárolásra. Lehetőség van tömörítő eljárások használatára is, viszont kritikus fontosságú, hogy csak veszteségmentes eljárást szabad alkalmazni. A veszteséges tömörítés jóval nagyobb fokú méretcsökkenést lenne képes biztosítani, ami szemmel látható változást nem nagyon okozna, viszont a képi információ mérésekor ezek az eltérések komoly gondokat okozhatnak.

Látható, hogy az igen szerteágazó képkötő technikák és a sokféle járulékos információ miatt nem egyszerű egy egységes szabványt kidolgozni. A képkötő berendezések gyártói eleinte saját, zárt formátumokat használtak, amelyek között nem volt egyszerű az átjárás biztosítása. Az igény hamar megmutatkozott az egységes szabvány iránt. 1983-ban bizottság

alakult erre a célra, aminek eredményeképpen 1985-ben jelent meg az első változat ACR-NEMA 1.0 néven. Ezt a második főverzió 1988-ban követte, a harmadik pedig 1993-ban. Ez a 3.0 verzió már a DICOM (Digital Imaging and Communications in Medicine) elnevezést kapta. A DICOM formátum ISO-szabvány, az 1990-es években a PACS rendszerek de facto szabványává vált. A fejlődése ma is folyamatos, az újonnan érkező igényeknek megfelelően.

6.4.1. DICOM fájlformátum

A DICOM magában foglalja a képi tárolás fájlformátum definícióját, valamint a képek küldéséhez és fogadásához szükséges TCP/IP-alapú hálózati részt is. Itt most a fájlformátumának egy gyors áttekintését végezzük el. A leírás nem teljes, részletesebb információk a DICOM szabványban találhatóak [2].

A DICOM reprezentációban minden pácienshez több *vizsgálat*, vizsgálatonként több *sorozat*, sorozatonként több képszelet is tartozhat. A képszeletek különálló fájlokban, de azonos könyvtárban kerülnek tárolásra. Az összetartozó adatok leválogatására egyedi páciens, vizsgálat és sorozat azonosítók szolgálnak, amelyet minden fájl kötelezően tartalmaz.

Egy DICOM fájl *csoportok* sorozatából áll. A csoportok egymáshoz logikailag kapcsolódó információkat, *elemeket* tartalmaznak. A csoportok és elemek egyedi azonosítására 2 bájt hosszúságú számok szolgálnak, ezeket 16-os számrendszerben írják le.

6.1. táblázat. Néhány fontosabb DICOM csoport

Azonosító kód	Leírás
0x0002	Fájl meta elemek
0x0008	Vizsgálat információk
0x0010	Páciens információk
0x0018	Képkészítés körülményei
0x0028	Kép információk
0x7FE0	Kép intenzitásértékei

A fájlban az adatok bináris formában kerülnek tárolásra. Minden csoport első (0x0000 kódú) eleme a csoport hosszát tárolja bájtban. Így a beolvasáskor lehetőségünk nyílik átugrani azokat a csoportokat, amelyek tartalma érdektelen a számunkra.

Az elemek különféle reprezentációjú adatokat (egyedi azonosító, sztring, decimális szám, binárisan kódolt lebegőpontos szám, stb.) tartalmazhatnak. Az egyes csoport-elem párokra a DICOM szabvány egyértelműen definiálja a reprezentáció típusát, de explicit érték-reprezentáció esetén maga a DICOM fájl is leírja ezeket 2 bájtos VR (Value Representation) azonosítókkal.

A mellékelt példa egy CT felvétel egy képszeletének legfontosabb dekódolt fejléc információit tartalmazza. A teljes lista a függelékben megtekinthető. Figyeljük meg, hogy a páciensre vonatkozó privát elemek anonimizálásra kerültek.

```
(0010,0010) PN Patient's Name (16): ??????????????????
(0010,0020) LO Patient ID (8): ??????????
(0018,0050) DS Slice Thickness (8): 3.000000
(0018,1100) DS Reconstruction Diameter (14): 480.0000000000
```

(0020,000d) UI Study Instance UID (48):
1.2.840.113619.2.1.1.322986825.674.993564792.356
(0020,000e) UI Series Instance UID (48):
1.2.840.113619.2.1.1.322986825.674.993564792.358
(0020,0013) IS Image Number (2): 45
(0020,0032) DS Image Position (Patient) (36):
-234.500000\ -240.000000\104.000000
(0020,1041) DS Slice Location (14): 104.0000000000
(0028,0010) US Rows (2): 512 (0028,0011) US Columns (2): 512
(0028,0030) DS Pixel Spacing (26): 0.9375000000\0.9375000000
(0028,0100) US Bits Allocated (2): 16
(0028,0101) US Bits Stored (2): 16
(0028,0102) US High Bit (2): 15
(0028,0103) US Pixel Representation (2): 1
(0028,1052) DS Rescale Intercept (6): -1024
(0028,1053) DS Rescale Slope (2): 1
(7fe0,0000) UL Group 7FE0 Length (4): 524296
(7fe0,0010) OX Pixel Data
(524288): Data starts at position 1382

7. fejezet

Orvosi algoritmusok

Az orvosi képek leggyakoribb szerepe a diagnózis megállapításában van. Az orvos megjeleníti a képet és áttanulmányozza azt, majd a tapasztalata alapján döntést hoz. Bonyolultabb, nem egyértelmű esetekben további vizsgálatokra, más képalkotó berendezés használatára is szükség lehet.

A képek megjelenítése mellett azokon további műveletek, pl. kvantitatív mérések végezhetők, több kép esetén felmerülhet az igény a közös térben, egymásra vetítve történő megjelenítésre, valamint szükséges lehet a képpontok osztályozására aszerint, hogy melyik szervhez tartoznak. A felsorolt feladatok összetett műveleteket, algoritmusokat igényelnek. A továbbiakban két fontos területet ismertetünk részletesebben, az orvosi képek regisztrációját és szegmentációját. Mindkét területhez alapos és átfogó összefoglaló munkák születtek [3, 30, 77].

7.1. Orvosi képregisztráció

A képfeldolgozás számos problémájának megoldásakor szükség van olyan módszerre, ami különböző nézőpontból, különböző időpontban, vagy különböző képalkotó berendezésekkel készült képeket egymással fedésbe hoz. Az előző fejezetben tárgyaltaknak megfelelően jelölje X a vizsgált tárgyat, valamint A és B jelölje a 2D vagy 3D képeket, amelyeket ugyanazzal, vagy más képalkotó berendezéssel készítettünk X -ről. A képek általában más látómezővel rendelkeznek, így az Ω_A és Ω_B képtartományok különbözők lesznek:

$$\begin{aligned} A & : x_A \in \Omega_A \mapsto A(x_A), \\ B & : x_B \in \Omega_B \mapsto B(x_B), \end{aligned}$$

ahol $A(x_A)$ és $B(x_B)$ az x_A illetve x_B térbeli pozícióhoz tartozó intenzitásértékeket jelöli. Az intenzitásértékek valamilyen mérhető anyagjellemző értéket reprezentálnak X térbeli pozícióiban. Ilyen például a röntgensugarak elnyelődési tényezője CT képek esetén, protonok állapotváltozásai a mágneses mező tulajdonságainak változásakor MR képek esetén, radioaktív jelzőanyagok szervezetbeli eloszlása SPECT és PET képalkotáskor, illetve a látható fény intenzitása fényképkészítéskor. Mivel A és B képek ugyanazt a tárgyat ábrázolják, így az A és B térbeli pozíciói között kapcsolatot van: a tárgy egy tetszőleges $x \in X$ pontja az A képen x_A helyen, a B képen x_B helyen jelenik meg. A regisztráció célja annak a geometriai transzformációnak a megkeresése, amely x_B pozíciókat a megfelelő x_A -beli pozíciókba képezi le

a számunkra érdekes képtérben, amely jelentheti például az A kép teljes képterét vagy akár a két képtér metszetét, átfedő részét is. Ez az átfedő rész az A és B képektől, valamint a T transzformációtól függ:

$$\Omega_{A,B}^T = \{x_A \in \Omega_A \mid T^{-1}(x_A) \in \Omega_B\}.$$

Tetszőleges T transzformáció esetén a diszkrét Ω_A és Ω_B képterek metszete üres lehet, amennyiben a rácspontok nem kerülnek pontosan egymásra. Ennek áthidalására a B kép intenzitásértékeit újra kell mintavételeznünk Ω_A pontjaiban. A legegyszerűbb újramintavételezési módszer az Ω_B legközelebbi rácspontjához tartozó értéknek a felhasználása. Lineáris vagy akár összetettebb interpolációs módszerek szintén használhatók (pl. köbös konvolúció, B-Spline). A továbbiakban jelölje \mathcal{T} azt a transzformációt, amely a térbeli pozíciót és a hozzá rendelt intenzitásértéket együttesen képezi le, valamint legyen B^T az Ω_A rácson újramintavételezett B kép.

Regisztráció kapcsán három egymásra épülő feladatot különböztethetünk meg.

- *Képregisztrációnak (image registration)* nevezzük a *bázis- vagy referenciakép* és az *illesztendő kép* közötti legjobb fedést biztosító geometriai transzformáció meghatározását.
- A *képilllesztés (image matching)* feladata a regisztráció során megtalált transzformáció alkalmazása az illesztendő képre, vagyis az *illesztett kép* előállítása.
- A *képfúzió (image fusion)* alkalmazásával a referencia- és az illesztett képből egy új képet állítunk elő, amelyen a képek eltérései, illetve az egymást kiegészítő képtartalmak együttesen vizsgálhatók.

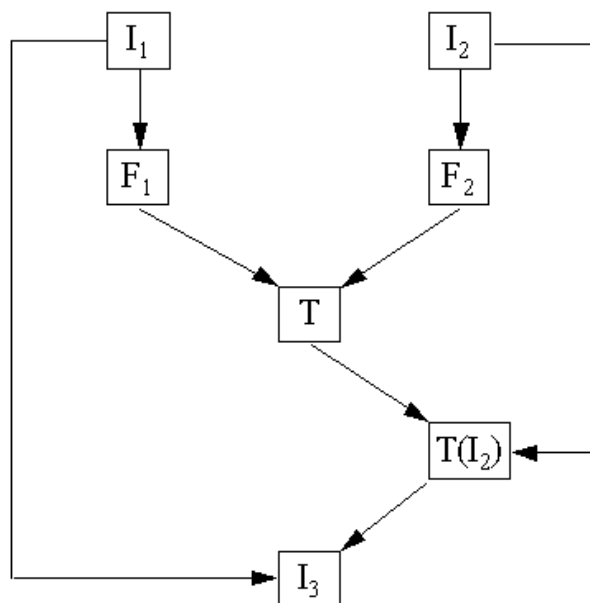
Mint látható, az illesztés feltételezi a regisztrációt, a képfúzió pedig a regisztráció és az illesztés előzetes végrehajtását.

Regisztrációs probléma nem csak képi adatok között merülhet fel. Amennyiben olyan tárgyról készül a felvétel, amelynek a pontos geometriai modelljét is ismerjük, a kép és a modell között is kapcsolatot tudunk teremteni. Orvosi területen a műtétségrehajtás igényel rendszerint ilyen megközelítést, például a műtét előtt a betegről készült kép és a műtőeszköz modelljének együttes megjelenítése céljából. Ezek alapján megfogalmazhatjuk a regisztráció általános definícióját.

- A regisztráció feladata a különböző forrásokból származó képi és/vagy geometriai információtartalmak közötti geometriai kapcsolat megteremtése.

Az 7.1. ábra a regisztrációt, az illesztést és a fúziót egyaránt tartalmazó eljárás fő lépéseit mutatja be. Nem feltétlenül van szükségünk minden esetben mindhárom lépés végrehajtására. Bizonyos feladatoknál elegendő lehet csak a legjobb illesztést biztosító transzformáció meghatározása, vagy ezen túl az illesztés végrehajtása. Megjegyezzük, hogy a regisztrációs lépés gyakran nem közvetlenül a bemeneti adatokkal, hanem egy előfeldolgozó lépés keretében előállított képi vagy geometriai jellemzők halmazával dolgozik. A jellemzőkkel és kivonásukkal a következő alfejezetben foglalkozunk részletesebben.

A képek közötti különbségek három típusát különböztetjük meg.



7.1. ábra. Egy regisztrációs eljárás fő lépései. Az I_1 referencia- és az I_2 illesztendő információtartalmakból kinyerjük az F_1 és F_2 jellemzőket. Ezek felhasználásával kiszámítjuk a legjobb illesztést biztosító T transzformációt (regisztráció). A T geometriai transzformációt alkalmazzuk I_2 -re (illesztés). Az illesztett adatokból egy új I_3 adatot állítunk elő, ahol a közöttük található eltérések, illetve az egymást kiegészítő információtartalmak vizsgálhatók (füzió).

- A legegyszerűbb esetben a képek között *csak térbeli eltérés van*, maga az objektum változatlan. Ilyen esetekben a geometriai eltérés általában jól számítható.
- A második típus esetén a *felvételek készítésének körülményei változnak*, például a képalkotó berendezés pozíciófüggő nemlineáris torzítást okoz a kép intenzitásértékein (ugyanazon szövet a kép különböző részein más intenzitásértékkel jelenik meg), a vizsgált objektum más állapotban van a két vizsgálat alatt (például normál és terheléses vizsgálatok, kontrasztanyag használata) vagy más-más képalkotó berendezéssel készültek a képek. Az ebből fakadó különbségeket általában nehéz modellezni.
- A harmadik típus a legbonyolultabb, ekkor ugyanis maga a *vizsgált objektum változik meg* (például daganat eltávolítása előtt és után készültek a felvételek, vagy agyi struktúrák időbeli változásait figyeljük). Ebben az esetben a regisztráció célja ezen változások detektálása, megfigyelése, így ezeket a különbségeket nem szabad eltüntetni a regisztráció során.

A regisztrációs, képfúziós technikák egyik fontos alkalmazási területe az orvosi képfeldolgozás. A különböző időpontokban készített felvételek illesztésével lehetőség van pl. egy daganat méretváltozásának megfigyelésére, a különböző képalkotó berendezések képeinek fúziójával pedig pl. az anatómiai és a funkcionális képek együttes megjelenítése pontosabb diagnosztikai információt biztosít, illetve segíti a műtét és a terápia tervezését. Műtétvégrehajtás közben a betegről készült kép és a műtőeszköz modelljének együttes megjelenítésével

az eljárás vezérelhető illetve ellenőrizhető. Több alapos áttekintő cikk is megjelent erről a területről [35, 54, 57, 59].

Regisztrációs módszerek fontosak még a légi- és műholdfelvételek feldolgozásakor (geológia, urbanisztika, célazonosítás), illetve a számítógépes látás egyes feladataiban (mélységi információ számítása sztereoképekből, képszegmentálás, mozgáselemzés) is [8]. A továbbiakban orvosi képregisztrációhoz használható módszereket ismertetünk, de közülük több sikeresen alkalmazható más területeken is.

7.1.1. Regisztrációs algoritmusok fő komponensei

Bár a regisztrációs problémák igen változatosak lehetnek és sokszor egyedi megközelítést igényelnek, a legtöbb módszer jellemezhető a következő négy komponens megválasztásával [8]:

- A *keresés tere* az illesztéshez használt geometriai transzformáció típusát határozza meg. A keresési tér dimenziója a keresett transzformáció szabad paramétereinek számával egyezik meg, ebben a térben kell az optimális pontot megkeresni.
- Meg kell határoznunk, hogy a megfelelő illesztést biztosító transzformáció kereséséhez milyen jellemzőket használunk fel, vagyis meg kell adni a jellemzők terét (*feature space*). A jellemzők alkalmas megválasztásával a feldolgozandó adatmennyiség radikálisan csökkenthető, így felgyorsítható a keresés. Ehhez természetesen szükséges az, hogy a kinyert jellemzők relevánsak legyenek, vagyis jól jellemezzék az eredeti adatot.
- A *hasonlósági mérték* olyan függvény, amely tetszőleges transzformációs paraméterek esetén (vagyis a keresés terének bármely pontjára) megadja, hogy az adott transzformáció „milyen jól” illeszti a bemeneti adatokat. A regisztráció feladata ezen függvény (globális) maximum- vagy minimumhelyének meghatározása.
- A *keresési stratégia* a hasonlósági mérték globális optimumának meghatározására alkalmazott módszert jelenti.

Keresési tér

A képek és/vagy geometriai információtartalmak közötti geometriai kapcsolatot *geometriai transzformáció* segítségével adhatjuk meg. A keresési tér dimenzióit a geometriai transzformáció szabad paraméterei alkotják. Ebben a térben kell az optimális pontot (vagyis transzformációt) megkeresni, ami a legjobb illeszkedést biztosítja.

A geometriai transzformáció egy $T : D \rightarrow R$ leképezés, ahol $D \subseteq \mathbb{R}^m$, $R \subseteq \mathbb{R}^n$ és $m, n \geq 1$. Az $m > n$ eseteket vetítő transzformációknak nevezzük, vagyis a transzformáció hatására dimenzióvesztés következik be. Ilyen például a 3D tájról készült 2D fotó készítése is. Ha $m=n$, akkor normál transzformációról beszélünk¹.

A T geometriai transzformáció elvileg tetszőleges lehet, a gyakorlatban azonban célszerű a képek/objektumok közötti geometriai különbözőségnek legjobban megfelelő *megszorításokat, kényszerfeltételeket* bevezetnünk. Néhány példa:

¹A transzformáció ilyen esetben is okozhat dimenzióvesztést, amennyiben elfajuló.

- T legyen invertálható.
- T legyen diffeomorfizmus: invertálható, differenciálható, és az inverze is differenciálható. „Sima”, vagyis nem okoz „szakadást” és „gyűrődést”.
- T lineáris: előáll a koordináták lineáris kombinációjából + eltolásból.
- T őrizze meg a pontok közötti távolságokat (merev-test transzformáció).

Egy geometriai transzformációt megadhatunk *parametrikus* formában, vagyis az egyes pontokhoz rendelt új koordinátát egy véges, általában a pontok számánál jóval kevesebb számú paraméterrel rendelkező képlettel adjuk meg (például transzformációs mátrixszal vagy véges számú alapfüggvény súlyozott összegeként). ϕ megszorításait ilyenkor a képlet implicit módon tartalmazza. A *nem-parametrikus* esetben minden egyes ponthoz egyenként adjuk meg képének koordinátáit eltolási vektorokkal, így egy ún. *elmozdulási mezőt* definiálunk. Ekkor rendszerint kényszerfeltételeket vezetünk be az elmozdulási mező elemeire, amely általában valamilyen fizikai modellre épül. Ezt a lépést *regularizációnak* nevezzük. Ez utóbbi megadási mód csak véges méretű diszkrét digitális képek esetében használható.

Képjellemzők tere

15–20 évvel ezelőtt a tárolókapacitás szűkössége és az akkori processzorteljesítmény miatt különösen fontos volt, hogy a nagy méretű képek ne közvetlenül kerüljenek illesztésre, hanem a radikális méretcsökkenéssel járó releváns jellemzők kivonása segítségével. Manapság a bemeneti adatokat közvetlenül felhasználó módszerek is igen népszerűek.

A képi jellemzők lehetnek *külsők* (*extrinsic*) vagy *belső* (*intrinsic*). A külső jellemzők mesterségesen kerülnek a képre, kifejezetten a regisztráció elősegítésére, míg a belső képjellemzőket magából a képtartalomból származtatjuk.

A regisztráció megkönnyítése céljából a mesterségesen a képhez adott külső képjellemzők a képeken rendszerint jól láthatók, akár automatikusan meghatározható a helyzetük. Általában nagy pontosságot igénylő, például számítógéppel támogatott idegsebészeti beavatkozásokhoz, illetve anatómiai és funkcionális képek illesztésekor használják. Ez utóbbi esetben a funkcionális képen gyakran nehéz jól azonosítható anatómiai pontokat találni. A legnagyobb pontosságot a csontba rögzített markerek adják, azonban ezek nagyon kényelmetlenek a beteg számára. A bőrhez, fogakhoz rögzített markerek nem okoznak ilyen problémát, viszont könnyen elmozdulhatnak, rontva ezzel az illesztés pontosságát.

A külső képjellemzőkre épülő algoritmusok legnagyobb problémája az, hogy nem retrospektívek, vagyis csak azokat a képeket lehet így illeszteni, amelyek a markerek eltávolítása előtt, tehát egy rövid időintervallumban készültek. Pár hónap múlva egy ellenőrző vizsgálatkor nem lehet pontosan ugyanúgy visszahelyezni a markereket, így más módszer szükséges a képek összehasonlításához.

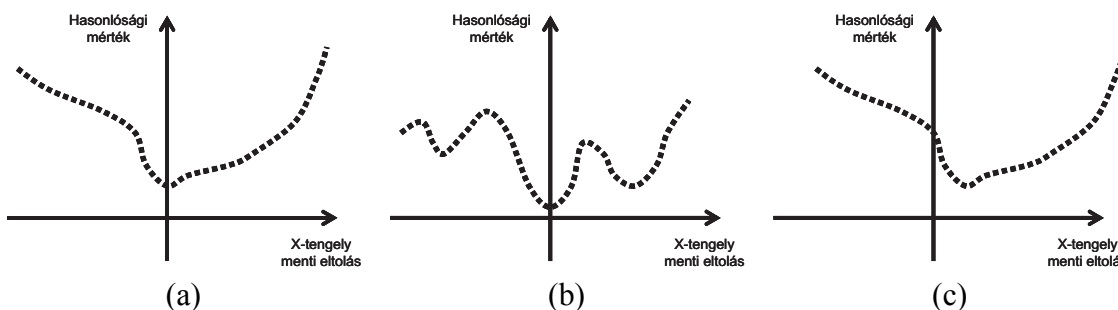
Egy érdekes megoldás ilyen esetekre a fejrögzítő használata a képalkotáskor. A beteg felfekszik az asztalra és első alkalommal egy lágy, de gyorsan szilárduló műanyag hálót helyeznek a fejére, amelynek a széle az asztalhoz mereven rögzíthető. Pár perc után, a megfelelő szilárdság elérésekor készül el a felvétel. A következő alkalmakkor úgy fektetik a beteget az

asztalra, hogy a merev háló pontosan illeszkedjen a fejhez és rögzítse azt. Ez a módszer ugyanazon képalkotó berendezéssel készült képek esetén alkalmazható, és mivel a fej a maszkhoz képest kis elmozdulásokra képes lehet, illesztési pontatlanság felléphet. Nincs viszont szükség regisztrációs algoritmusra, az így készült képek közvetlenül egymásra vetíthetők.

A belső képjellemzőket a beteg anatómiáját, szerveinek működését megjelenítő képtartalomtól származtatjuk. Ilyen belső képjellemzők lehetnek például jól azonosítható anatómiai pontok, határvonalak, felszínek, objektumok, vagy akár a kép intenzitásai is közvetlenül felhasználhatók.

Hasonlósági mérték

A hasonlósági mértékekkel szemben támasztott legfontosabb kritérium az, hogy a globális optimumát ott (vagy attól még elfogadható távolságban) érje el, ahol a két adat valóban legjobban fedi egymást. A keresés szempontjából hasznos, ha csak egy globális optima létezik, lokális optimumok pedig nem fordulnak elő. Ez utóbbi valós orvosi képek esetében nem igazán valószínű, a lokális optimumok elkerülése ekkor a keresési stratégia (az optimalizáló módszer) feladata lesz. A 7.2. ábra három képzeletbeli 1-dimenziós hasonlósági mértéket mutat be.



7.2. ábra. Fiktív 1D hasonlósági mértékek értékei a paraméter (X-tengely menti eltolás) változásának függvényében. Egy ideális (a), egy valós életben általában előforduló (b), és egy használhatatlan (c) hasonlósági mérték. Feltételezzük, hogy a valós illeszkedés az origóban a legjobb.

Az (a) mérték lenne a legkönnyebben használható, de a gyakorlatban csak speciális feladatok esetén sikerül ilyen típusút megadni. A (b) mérték már jobban közelíti a valós eseteket. Rendelkezik lokális optimumhelyekkel, de a globális optima egyszeres és azt a megfelelő helyen veszi fel. A (c) mérték optimumának meghatározása könnyű, viszont azt nem a valós illeszkedésnek megfelelő pont környezetében veszi fel, így felhasználásával hibás eredményt kapunk.

Kinyert geometriai képjellemzők esetén (pontok, határvonalak, felszínek) azok valamilyen távolságfogalmon alapuló illesztése a leggyakoribb. Ilyen távolságdefiníció lehet például az euklideszi legkisebb négyzetes eltérés, a Hausdorff, vagy a Chamfer távolság. A képpontok intenzitásértékén alapuló mértékek is széles körben használatosak. Ezek az egymással fedésbe kerülő képpontpárok hasonlóságát mérik. A 7.1.3. alfejezetben ismertetünk néhány gyakran használt ilyen mértéket.

Keresési stratégia

A függvények optimalizálása a regisztrációnál jóval nagyobb tudományterület, ennek összefoglalását itt nem kíséreljük meg. Általában egy jó tulajdonságokkal rendelkező, a feladat számára megfelelő klasszikus eljárást választunk.

Kinyert geometriai jellemzők esetén az optimális transzformáció gyakran egy direkt módszerrel közvetlenül megkapható. Bonyolultabb esetekben pl. iterációs módszerek, dinamikus programozási technikák, illetve ún. „durvától a finomig” többfelbontású kereső módszerek alkalmazhatók.

7.1.2. A regisztrációs algoritmusok csoportosításai

A regisztrációs problémák megoldásai nagyon szerteágazók, többféle szempont szerint csoportosíthatjuk ezeket az algoritmusokat [57]. A négy fő komponens megválasztása mellett további jellegzetességeket emelünk ki a most következő részben.

Vizsgálhatjuk például azt, hogy az algoritmus igényel-e felhasználói beavatkozást, ha igen, akkor milyen jellegűt. A *manuális módszerek* esetében általában egy képmegjelenítő szoftver áll rendelkezésre, az optimális transzformációt a felhasználó határozhatja meg a transzformáció paramétereinek közvetlen elérésével és változtatásával. *Interaktívnak* vagy *félautomatikusnak* nevezünk egy módszert, ha a felhasználó inicializálja az algoritmust képjellemezők kivonásával, például egymásnak megfeleltethető pontok, kontúrok vagy felszínek kijelölésével és ezek optimális illesztését határozza meg a módszer. Az *automatikus módszerek* nem igényelnek felhasználói beavatkozást.

A manuális módszer alkalmazása elég nehézkes és időigényes, különösen 3-dimenziós képek esetén. A félautomatikus módszerek megbízhatóak és gyorsak, orvosi képek esetében viszont a képjellemezők kivonása tapasztalt, szakértő radiológust igényel, a kijelölés pedig akár 10–30 percet is igénybe vehet. Az automatikus módszerek a felhasználó szempontjából a legegyszerűbben használhatók, viszont minden esetben szükség van a regisztráció eredményének vizuális ellenőrzésére, mivel csak a képi információtartalom alapján a módszerek gyakran nem képesek a nyilvánvalóan rossz eredmények kiszűrésére sem. Nagy képméret és/vagy bonyolult transzformációtípus esetében a futásidő nagyon nagy is lehet (akár több óra, vagy több nap is).

A képek forrásai alapján a regisztrációs probléma *egymodalitásos (unimodal)*, ha az illesztendő képek ugyanabból a képképző berendezésből származnak és ugyanarról betegről készültek. Különböző időpontokban, például műtét előtt és után készült képek, illetve normál és terheléses, kontrasztanyagot vizsgálatok összehasonlítására használják. *Többmodalitásos (multimodal)* probléma esetén a különböző berendezésekből származó, egymást kiegészítő képtartalmak egyesítése a cél. Ezek mellett megkülönböztetünk egy beteghez tartozó (*intra-patient*), valamint különböző beteg közötti (*interpatient*) regisztrációt. Ez utóbbit gyakran ún. *standardizált atlasz* készítésére használják.

7.1.3. Klasszikus módszerek

A regisztrációs probléma megoldására számos megközelítés született az elmúlt évtizedekben. Ezek egy része általánosan felhasználható szélesebb körben is, könnyen a saját igényeinknek

megfelelőre szabhatjuk őket. Ezek közül a klasszikus megközelítések közül mutatunk be néhányat a most következő részben.

Ponthalmazok illesztése

A pontok általánosan használt képjellemzők valós orvosi regisztrációs problémákban. Egy általános pont-alapú módszer a következő lépésekből áll: először kijelöljük a pontokat a képeken, majd meghatározzuk, hogy az illesztendő képen kijelölt pontokhoz a báziskép melyik kijelölt pontja tartozik, végül az egymásnak megfelelőített pontpárok felhasználásával kiszámítjuk a legjobb illesztést biztosító transzformációt:

$$D_P(T) = \sum_{i=1}^K \|x_i - T(y_i)\|^2.$$

A pontok meghatározása lehet manuális, félautomatikus vagy teljesen automatikus. A *manuális módszer* általánosan használható regisztrációs feladatok megoldására, ugyanis ekkor a pontok kijelölése és párosítása a felhasználó feladata. A regisztrációs algoritmusnak nem kell a kép intenzitásaival dolgozni, azokat értelmezni, így szinte tetszőleges képi adat esetén használható. A kijelölt pontok száma általában kevés, 4–20 között mozog. Hátránya, hogy orvosi képek esetén a munka szakértő radiológust igényel, időigényes, valamint a képpontok nem mindig jelölhetőek ki elegendő pontossággal.

Hartkens és Rohr javított, félautomatikus módszert javasol ezen problémák kezelésére [31]. A kiválasztott pont egy adott, például $7 \times 7 \times 7$ méretű környezetében 3-dimenziós sarokpont detektáló algoritmus segítségével tovább finomítják a kijelölést. Az adott térrészben természetesen akár több esélyes pont is előfordulhat, a legvalószínűbb párosítás megtalálása is az algoritmus feladata.

A felhasználó számára a legkényelmesebb, ha a pontok kijelölése sem igényel semmilyen beavatkozást, vagyis a módszer teljesen *automatikus*. A sarokpont detektáló algoritmusok ekkor a teljes képen végigfutnak, különböző, akár nagy számú ponthalmazokat adva eredményül. Előfordulhatnak olyan pontok is, amelyeket nem lehet párosítani. A pontpárok meghatározása ekkor igen összetett feladat lehet. Goshtasby invariancián, klaszterezésen és képkoherencián alapuló algoritmusokat foglal össze a párosításra [27].

A automatikus kijelölés esetén kulcsfontosságú szerepet tölt be a sarokpontokat detektáló algoritmus. Ennek a képek között feltételezett geometriai deformációval szemben invariánsnak kell lennie, valamint robusztusnak kell lennie a képeken található zajterheléssel szemben. Orvosi képek esetében Hartkens és Rohr a cikkükben összehasonlított 9 sarokpont detektáló algoritmus közül az elsőrendű parciális deriváltakra épülőket találta a legmegfelelőbbnek.

Amennyiben az automatikusan detektált pontok nem sarokpontok, hanem például kontúrok vagy felszínek, de különálló pontok halmazaként reprezentáljuk őket, akkor lehetőségünk van kontúr- illetve felszínillesztő algoritmusok használatára is.

Kontúr- és felszínillesztés

Különösen az 1980-as években illetve az 1990-es évek első felében örvendtek nagy népszerűségnek a kontúr- illetve felszínillesztő algoritmusok. Az akkori számítógépek – a maiak telje-

sítményéhez viszonyítva – kis memóriamérete és gyenge számítási kapacitása miatt szükség volt a képi adat nagymértékű csökkentésére. A pontpárok kijelölésével összemérhető időigényű, egymásnak megfeleltethető kontúrok illetve felszínek félautomatikus vagy automatikus meghatározása hatékony és elfogadható pontosságú módszerek kidolgozását tette lehetővé.

A kontúrok és felszínek reprezentációja többféle lehet. Megadhatjuk például pontfelhőkkel, kontúr mentén elhelyezkedő pontokkal és az őket összekötő egyenes szakaszokkal vagy spline-okkal, a felszínen elhelyezkedő pontok által kifeszített hálóval, de akár geometria modellel is, amennyiben azt pontosan ismerjük. Nehézséget jelent, hogy a kontúr/felzínpontok közötti párosítás nem ismert, ezt a megoldás keresésének iterációs lépéseiben becsülni kell. Az általános felszínillesztő hasonlósági mérték az alábbi:

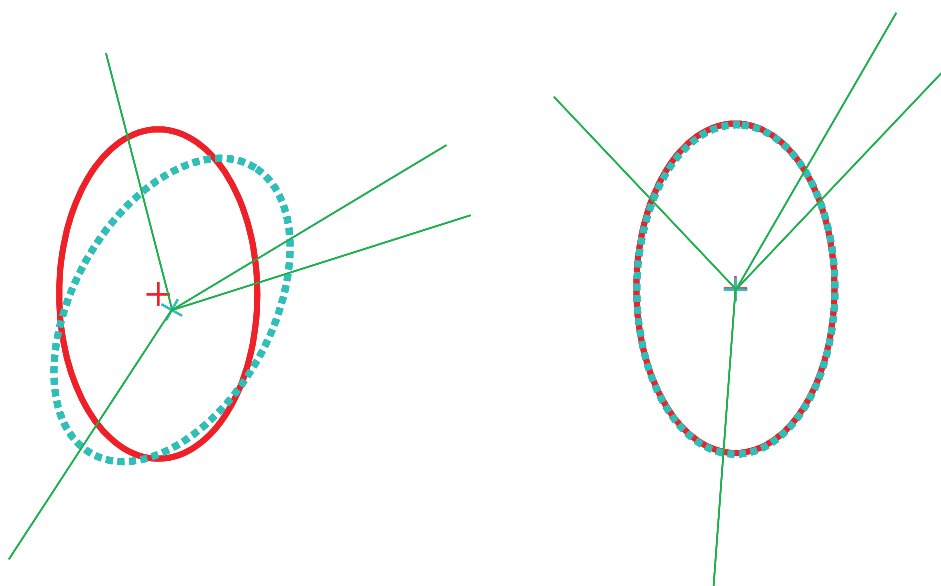
$$D_S(T) = \sqrt{\sum_{i=1}^K \|x_i - P(T(Y), x_i)\|^2}.$$

Itt a $P(.,.)$ függvény feladata a párosítás becslése, megvalósítása az adatreprezentációtól függ. Lehet például az x_i ponthoz legközelebbi pont a transzformált Y pontfelhőben, vagy ha az Y háromszöghálót jelent, akkor az x_i -hez legközelebbi pont a háromszögháló felszínén.

Ezen módszerek legnagyobb problémája az egymásnak meg nem feleltethető felszínarabokból fakad. Az eltérő részek „elhúzzhatják” a felszíneket egymástól. Ilyen eltérések adódhatnak szegmentálási hibákból, de okozhatja az objektum megváltozása is a két képalkotás között (pl. szerv természetes deformációja vagy műtét). Egy lehetséges megoldás az lehet, ha tudunk becslést adni az eltérés mértékére, akkor a hasonlósági mérték kiértékelésekor az egymástól legtávolabb kerülő pontpárok adott százalékát kiszűrjük.

Az alábbiakban három klasszikus felszínillesztő algoritmust mutatunk be.

„Kalapot a fejre (*Head-Hat*)” algoritmus



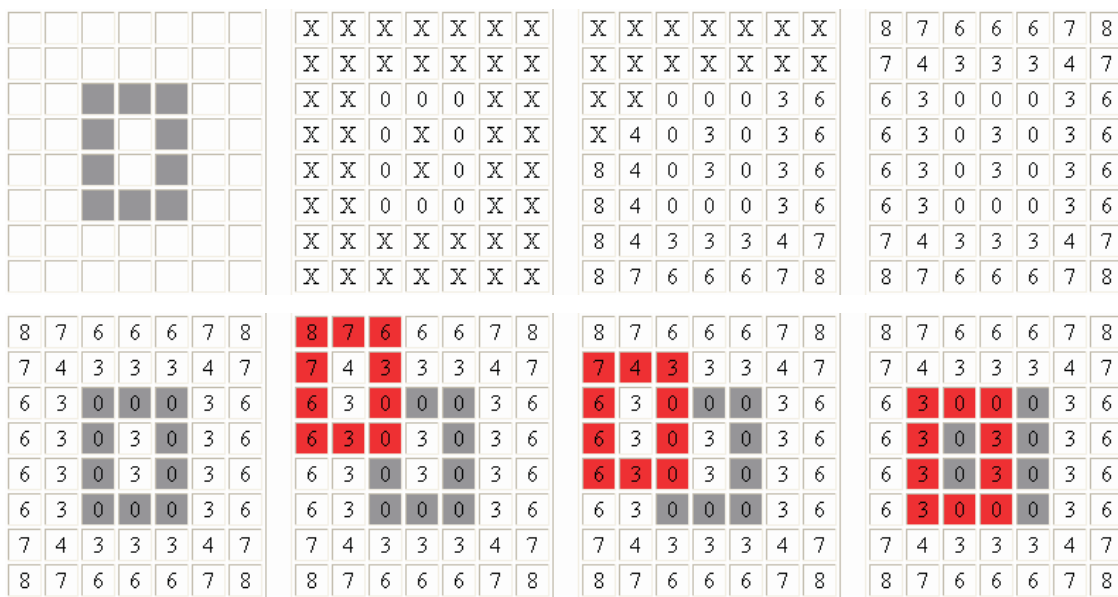
7.3. ábra. „Kalapot a fejre” algoritmus. A kiinduló állapot (balra) és az optimális illesztés (jobbra).

Pelizzari dolgozta ki MR-PET agyfelvételek merev illesztésére 1989-ben [66]. Mindkét képből ki kell nyerni ugyanazon felszínt, a nagyobb felbontású képet (az MR-t) zárt kontúrok sorozataként, a gyengébbet (PET) pontok sorozataként reprezentálja. Első lépésként a két halmaz súlypontja kerül meghatározásra, a kiindulási transzformáció az ezeket egymásba juttató eltolás lesz. A súlypontból a pontokon keresztül húzott félegyenesek elmetszik valamelyik zárt kontúr, ezen pontpárok négyzetes távolságösszege adja a felszínnek távolságát. A hasonlósági mérték optimumának meghatározására a Powell-módszert használták [69].

A módszer nehézkesen használható. Az egymásnak megfeleltethető felszínek kinyerése félautomatikus módszerrel akár 20–30 percet is igénybe vehet. Ezen előkészítés után a transzformáció keresése pár másodpercig tart. Csak zárt, gömbszerű felszínek illesztésére használható, viszont lényeges, hogy az alakzat ne legyen forgásszimmetrikus.

Távolságtérkép-alapú illesztés

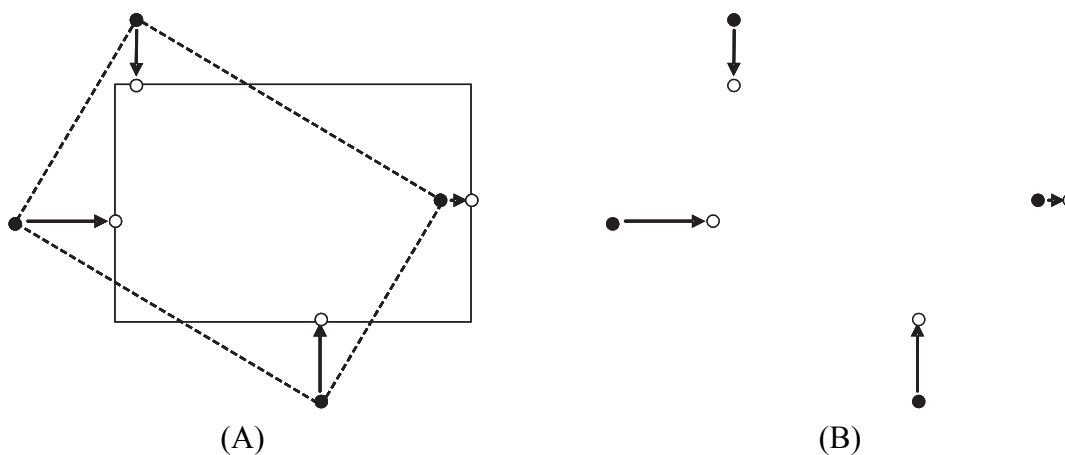
Az előző módszernél minden egyes transzformáció esetén újra kell számolni a pontok és a felszín távolságát. A távolságtérképen alapuló módszer esetében ez a számítás egyszer, egy előkészítő lépés keretében kerül végrehajtásra: minden egyes képponthez hozzárendelődik a legközelebbi felszínponttól mért valamilyen mérték szerinti távolsága Euklideszi távolság helyett célszerű diszkrét távolságokat (például a pontok közötti négy- illetve nyolc-összefüggő utak hossza, Chamfer-távolság) választani. A Chamfer-távolság alkalmazása egyrészt azért előnyös, mert egész aritmetikát használ és alkalmazásával az Euklideszi távolság jól közelíthető, másrészt számításához rendelkezésre áll egy gyors, tetszőleges dimenzióban lineáris idejű algoritmus [5]. A kontúrok közötti távolság ekkor például az illesztendő kontúr aktuális pontjaihoz tartozó távolságvértékek négyzetes összegeként áll elő (7.4. ábra). A hasonlósági mérték optimalizálásához tetszőleges numerikus módszert lehet használni.



7.4. ábra. Illesztés távolságtérkép használatával. Távolságtérkép számítása a szürke színnel ábrázolt kontúrhoz előre- és visszairányuló pásztázással (felső sor), valamint egy illesztendő piros színű kontúr mozgása (alsó sor). A kontúrok távolság rendre 288, 191 és 54.

Iteratív legközelebbi pont módszer

Az egyik legnépszerűbb felszínillesztő technika a számítógépes látás feladataiban és az orvosi képek illesztésekor az iteratív legközelebbi pont (*ICP – Iterative Closest Point*) módszer. Besl és McKay publikálta 1992-ben [4], azóta számos módosított változata jelent meg. Egy objektum pontosan ismert geometriájú modellje és a róla készült mérési eredmény illesztésére dolgozták ki, de kis módosítással orvosi képek illesztésére is használható (ekkor ugyanis a modell is a mérési eredményből származik). Legfontosabb jellemzője, hogy a modell geometriáját többféle formában megadhatjuk (ponthalmazzal, háromszöghálóval, parametrikus felszínnel, stb.), a mérési eredményből kinyert felszínt pedig ponthalmazként ábrázoljuk. A módszer iteratív, egy iterációs lépés két lépésből áll. Először a mérési eredmény pontjaihoz megkeressük a modell felszínén található legközelebbi pontot, majd az így kapott egymásnak megfeleltett pontpárok felhasználásával, a pontillesztő módszereknél ismertett módon megkeressük a legjobb illesztést biztosító transzformációt (7.5. ábra). Az iteráció addig tart, míg a kívánt pontosságot, vagy a maximális iterációs lépésszámot el nem érjük.



7.5. ábra. Az ICP algoritmus egy iterációs lépése. A ponthalmazként megadott illesztendő kontúrhoz legközelebbi pontok keresése a modellen (A), majd az így előálló egymásnak megfeleltetett pontpárok illesztése (B).

Ahhoz, hogy a keresés kisebb eséllyel akadjon el lokális optimumokban, előfeldolgozó lépésként érdemes a kezdeti transzformációra valamilyen módon becslést adni. Ilyen előfeldolgozó lépés lehet például a súlypontokat egymásba juttató eltolás alkalmazása, főkomponens analízis által meghatározott tengelyek egymásba forgatása, vagy egyszerűen egy véletlenszerű kiindulási transzformáció választása. Ez utóbbi esetben érdemes többször is futtatni az algoritmust más-más kiindulási transzformációkkal és a legkisebb hibát okozó eredményt elfogadni.

Intenzitások hasonlóságán alapuló módszerek

Ezek a módszerek csak képek illesztésére használhatók, geometriai adatok nem kezelhetők velük. Az előző két megközelítéssel szemben nem igénylik jellemzők kinyerését az adatokból, akár változtatás nélkül képesek a képek intenzitásértékeivel dolgozni. Felhasználásuk az

utóbbi bő 15 évben terjedt el, a megfelelően nagy tárolási és számítási kapacitással rendelkező olcsó személyi számítógépek megjelenésével. Ezen módszerek kulcsa a *hasonlósági mérték*, amely az egymással fedésbe kerülő intenzitásértékek hasonlóságát jellemzi.

A hasonlósági mérték optimumát valamilyen ismert optimalizáló módszerrel keresik (lejtő módszer, Powell-módszer, stb.). A regisztráció sebességének növelése és a lokális optimumokban való elakadás esélyének csökkentése érdekében a képek többfelbontású (például Gauss) piramis reprezentációját is gyakran alkalmazzák [10]. Emellett szükség lehet a képek intenzitástartományának transzformációjára is. Bizonyos mértékek esetén az intenzitástartomány csökkentése szükséges például lineárisan (12 bitesről 8 vagy akár 6 bites méretre) vagy ablakozásos technikával (a tartomány egy részének kivágásával és lineáris transzformációjával).

A legegyszerűbb ilyen mérték az intenzitáskülönbségek abszolút (*SAD – Sum of Absolute Differences*) vagy négyzetes (*SSD – Sum of Squared Differences*) különbsége:

$$SSD = \frac{1}{N} \sum_{x_A \in \Omega_A} \left(A(x_A) - B^T(x_A) \right)^2,$$

$$SAD = \frac{1}{N} \sum_{x_A \in \Omega_A} |A(x_A) - B^T(x_A)|.$$

Ezek a mértékek akkor optimálisak, ha a két kép csak legfeljebb normális eloszlású zajban különbözik egymástól.

A kereszt korreláció (*C - Cross Correlation*) és a korrelációs együttható (*CC - Correlation Coefficient*), ami az előbbi normalizált változata, a képek intenzitásértékei közötti $I_2 = a \cdot I_1 + b$ lineáris kapcsolatot is magában tudja foglalni:

$$C = \frac{1}{N} \sum_{x_A \in \Omega_{A,B}^T} A(x_A) \cdot B^T(x_A),$$

$$CC = \frac{\sum_{x_A \in \Omega_A} (A(x_A) - \bar{A}) \cdot (B^T(x_A) - \bar{B})}{\sqrt{\sum_{x_A \in \Omega_A} (A(x_A) - \bar{A})^2 \cdot \sum_{x_A \in \Omega_A} (B^T(x_A) - \bar{B})^2}},$$

ahol \bar{A} és \bar{B} a képek átlagintenzitását jelölik.

Az előző mértékek csak olyan képek esetén használhatók, amelyek ugyanazon képalkotó berendezésből származnak. Sokan próbálkoztak multimodális, vagyis más-más berendezésből származó képek illesztésére jól használható hasonlósági mértéket adni. A Woods által bevezetett „particionált intenzitás uniformitás” mérték (*PIU – Partitioned Image Uniformity*) volt az első, amelyet sikeresen tudtak MR-PET illesztésre használni, sőt erre a feladatra még ma is a legjobbak között tartják számon [86]:

$$PIU = \sum_a \frac{n_a}{N} \cdot \frac{\sigma(a)}{\mu(a)}$$

$$n_a = \sum_{\Omega_a} 1, \quad \mu(a) = \frac{1}{n_a} \cdot \sum_{x_A \in \Omega_a} B^T(x_A), \quad \sigma(a) = \sum_{x_A \in \Omega_a} \left(B^T(x_A) - \mu(a) \right)^2.$$

Az alapfeltételezés az, hogy egyfajta szövettípushoz egy jól meghatározható, természetesen képalkotó berendezésenként különböző intenzitásérték tartozik mindkét képen. A mérték azt vizsgálja, hogy az egyik kép egy adott intenzitásértékével milyen intenzitásértékek kerülnek párba a másik képről, ezen intenzitásértékek uniformitását méri és ezeket összegzi. A mérték nem szimmetrikus: a képek felcserélésével más eredményt kaphatunk.

Az igazi áttörést az intenzitások együttes előfordulási mátrixán alapuló mértékek hozták. A Wells és Viola, valamint Collignon és munkatársai által javasolt kölcsönös információtartalom (MI – *Mutual Information*), valamint a Studholme és munkatársai által javasolt normalizált kölcsönös információtartalom (NMI – *Normalized Mutual Information*) jól használható MR-CT és MR-PET illesztési problémák megoldására is [14, 80, 83, 84]. A mértékek számításához szükséges a képek entrópiájának és együttes entrópiájának meghatározása:

$$\begin{aligned} H(A) &= - \sum_a p_A(a) \cdot \log p_A(a) , \\ H(B) &= - \sum_b p_B(b) \cdot \log p_B(b) , \\ H(A, B^T) &= - \sum_a \sum_b p_{AB^T}(a, b) \cdot \log p_{AB^T}(a, b) . \end{aligned}$$

p az egyes intenzitásértékek és intenzitásérték-párok relatív előfordulási valószínűségeit jelöli. Ezek alapján MI és NMI számítható az alábbi módon:

$$\begin{aligned} MI(A, B^T) &= H(A) + H(B) - H(A, B^T) , \\ NMI(A, B^T) &= \frac{H(A) + H(B)}{H(A, B^T)} . \end{aligned}$$

Számos más, a képpontok együttes előfordulási mátrixán alapuló mértéket találhatunk Bro-Nielsen doktori disszertációjában [7].

7.2. Orvosi képek szegmentálása

Az elkészült orvosi képek megjelenítésével már sok diagnosztikai feladat ellátható. Ha a képeken kvantitatív méréseket, biomechanikai számításokat is végeznénk vagy összetettebb megjelenítést szeretnénk elérni, általában nem elegendők önmagukban az intenzitásértékek, szükségünk van azok osztályozására is.

Szegmentálás alatt a kép particionálását, részekre osztását értjük, jól meghatározott kritériumok alapján. Szegmentálás feladata lehet például az egyes szervekhez tartozó képpontrégiók meghatározása, betegségre specifikus területek detektálása.

Szegmentálás szükséges például besugárzástervezés esetén. A daganatos szövetek kezelésének egy lehetséges módja azok ionizáló sugárzással történő roncsolása. A sugárzás a környező szöveteket is éri, ezért rendszerint több irányból végeznek célzást, amelyek középpontjában a kérdéses terület található. A sugárzás hatása ott így összegződik. A terápia tervezésekor különösen fontos a célterület megfelelő detektálása, valamint azon szervek helyének meghatározása, amelyeket lehetőség szerint minél kisebb sugárzás kell érjen. Ilyen védendő területek pl. a szív, a gerincvelő, az artériák.

Orvosi képek szegmentálása igen összetett feladat. Különösen 3D képek esetén nagy mennyiségű információ áll rendelkezésre, a szervek közötti határok pedig sokszor nehezen detektálhatók. A szakértő által végzett manuális szegmentálás eredménye általában a legjobb, viszont rendkívül időigényes, vizsgálatonként akár több órahosszába is telhet. Kulcsfontosságú tehát az orvos munkájának számítógépes segítése. A módszereknél megfigyelhető a minél nagyobb automatizmusra való törekvés. Bizonyos feladatok esetén (pl. agyvizsgálatokon a szürke- és fehérállomány szétválasztása, vagy csontozat detektálása CT képeken) ez el is érhető. A legtöbb módszer azonban interaktív, az orvostól kisebb-nagyobb beavatkozást igényel. Szükség lehet például a módszer inicializálására pontok, ívek, térfogatrészek gyors kijelölésével; részeredmény ismeretében szegmentálási paraméterek változtatására, a módszer ismételt végrehajtására; illetve sok esetben a szegmentációs eredmény gyors és hatékony manuális módosíthatóságára, szerkeszthetőségére. Ez utóbbi funkcióra még az egyébként teljesen automatikus megközelítések esetén is szükség lehet, a szegmentálások hibára hajlamossága miatt.

A szegmentáló algoritmusok két fő csoportja a *régió-alapú* és az *él-alapú*. Az első egy homogenitási kritériumnak megfelelő, szomszédos képpontokat sorol egy osztályba. Ilyen kritérium lehet a hasonló intenzitásérték, textúrázottság. Az él-alapú módszerek az egyes régiók közötti határvonalakat, nagy intenzitáskülönbségeket keresik. Az összetettebb módszerek energiaminimalizáción alapulnak: lehetnek él-alapúak, amikor is az élek keresése mellett a határvonal alakjára adhatunk megszorításokat („sima” legyen), de akár a határvonal által közrezárt terület intenzitásértékeit is magukban foglalhatják.

7.2.1. Régió-alapú módszerek

A legegyszerűbb régió-alapú módszer a *(globális) küszöbölés*, amely egy bináris képet ad eredményül:

$$g(i, j) = \begin{cases} 1 & \text{ha } f(i, j) \geq T, \\ 0 & \text{egyébként,} \end{cases}$$

vagyis az f képet előtér és háttér régiókba osztályozza. Jól meghatározott képalkotási protokoll esetén a T küszöbérték konstans lehet, vagy mivel rendkívül gyors műveletről van szó, akár interaktívan állítható és az eredmény rögtön megtekinthető. A módszer alkalmazási lehetőségei szűkösek, de alkalmas lehet pl. csontozat szegmentálásra röntgen vagy CT képeken. Figyelni kell viszont arra, hogy bélrendszerbe vagy a vérkeringésbe juttatott, nagy intenzitású kontrasztanyag is szegmentálásra kerül, amit detektálni kell. Ez különösen akkor nehéz feladat, ha a különböző szervekhez tartozó részek összeérnek egymással. Amennyiben a képen található objektumok intenzitástartományáról előzetesen ismerjük, hogy egy sötét és egy világos régióban találhatók, akkor a T értékét automatikusan is meghatározhatjuk *optimális küszöbölő* technikákkal. Olyan küszöbértéket kell keresni, amely az intenzitásértékek varianciáját maximalizálja az objektum és a háttér között. A módszer több küszöbérték detektálására is kiterjeszhető. *Lokális küszöbölés* esetén T értéke a kép egyes részein megváltozik, a képpont lokális környezetének megfelelően.

A *régiónövelő* módszerek esetében a szegmentálás a felhasználó által manuálisan kiválasztott vagy automatikus pontkereső eljárás által biztosított 1 vagy több pontból, ún. *magból* indul. Ezen pontok szomszédságát vizsgálva a *homogenitási kritériumnak* megfelelő pontok

hozzáadódnak a szegmentálási eredményhez. A módszer akkor fejeződik be, ha nem lehet több pontot hozzáadni a régióhoz. A homogenitási kritérium lehet rögzített, vagy a pontok hozzávétele után bizonyos paraméterei változhatnak is. Homogenitási kritérium lehet például a vizsgálat alatt álló pixel intenzitásértékének és a régió átlagintenzitásának eltérése: ha ez pl. a régió intenzitásértékeinek szórása kétszeresénél kisebb, akkor hozzáadódik a régióhoz, egyébként nem. A megközelítés hátránya, hogy a szegmentálás eredménye erősen függ a homogenitási kritériumtól, más pontokból elindított szegmentálás nem garantált, hogy egyező eredményt ad, valamint hasonló intenzitású szomszédos régiók esetén könnyen „átfolyás” alakulhat ki. Jól használható, ha a szomszédos régiók intenzitástartományja jelentősen eltér egymástól. További előnye, hogy összefüggő komponenseket ad eredményül, valamint könnyen implementálható. Hasonló ötleten alapulnak a *régiőfelosztó* technikák: itt egy nagy régióból indulunk ki, és ebből kerülnek eltávolításra a homogenitási kritériumnak nem megfelelő részek. A régiónövelő és -felosztó technikák kombinálhatók is egymással.

A *vízválasztó* szegmentálás az alapötletét a topográfiából meríti. Hegyes vidéken a le hulló csapadék a hegygerinc két oldalán más-más vízgyűjtőkbe vezetődik el. A vízgyűjtő területeket ún. vízválasztó vonalak választják el egymástól, a módszer ezeket detektálja. A módszert szemléletesen úgy képzelhetjük el, hogy az intenzitásértékeknek domborzatot feleltetünk meg: a sötét intenzitások az alacsony, a világos intenzitások a magasabb felszín-pontokat jelentik. A legalacsonyabb intenzitásokban „kilyukasztjuk” a domborzatmodellt és vízbe mártjuk. A lyukakon befolyó víz feltölti a medencéket. Ahol két, egymástól független medence található ott gátat emelünk, vagyis nem engedünk összefolyásokat. A domborzat teljes feltöltésével a gátak összessége adja a régiók határát. A módszer rendszerint erősen túlszegmentált eredményt, vagyis túlságosan sok régiót ad eredményül. Ez csökkenthető élkép készítésével és simítás alkalmazásával (a domborzat kisebb egyenetlenségeinek eltüntetésére), de így szükséges lehet a szomszédos régiók összeolvasztását lehetővé tenni utófeldolgozó lépésként.

Az előző módszerek határozott döntést hoznak minden egyes vizsgált képpont esetén arra vonatkozóan, hogy egy régióhoz tartoznak-e vagy sem. A képpontok közötti *fuzzy összekötés* fogalmának bevezetésével lehetőség nyílik annak jellemzésére, hogy egyes képpontok *milyen mértékben tartoznak össze*. Jelölje $\mu(c, d)$ két képpont között a *fuzzy szomszédsági relációt* n -dimenzióban:

$$\mu(c, d) = \begin{cases} \frac{1}{1+k_1\sqrt{\sum_{i=1}^n (c_i-d_i)^2}} & \text{ha } \sum_{i=1}^n |c_i-d_i| \leq n, \\ 0 & \text{egyébként,} \end{cases}$$

ahol k_1 nemnegatív konstans. $\mu(c, d)$ a $[0,1]$ intervallumból kap a távolság növekedésével csökkenő értéket, ami akkor különbözik nullától, ha a koordináták különbségösszege nem nagyobb n -nél, vagyis a képpontok egymás „közelében” helyezkednek el. A fuzzy kapcsolatban álló képpontok között $\psi \in [0,1]$ jelöli az összetartozásuk mértékét, amit *fuzzy affinitásnak* nevezünk. ψ definiálása alkalmazásfüggő, és kulcsfontosságú a sikeres szegmentálás eléréséhez. Megalkotásánál figyelembe vételre kerül a képpontok fuzzy szomszédságának mértéke, valamint az intenzitásértékek és egyéb, képi információból kinyert jellemzőik (pl. él információk).

Két tetszőleges, akár a kép átellenes pontjain található c és d képpont között $\pi = \langle c^{(0)}, \dots, c^{(N)} \rangle$

út vezet, ahol $c=c^{(0)}$ és $d=c^{(N)}$. Ezen az úton minden szomszéd között számítható $\Psi(c^{(n)}, c^{(n+1)})$ affinitás. Egy π út affinitását az úton található párok affinitásainak minimális értéke (a „leggyengébb láncszem”) jellemzi:

$$\Psi'(\pi) = \min_{0 \leq n \leq N-1} \Psi(c^{(n)}, c^{(n+1)}) .$$

Ez alapján a c és d képpontok közötti *fuzzy összeköttetést*

$$\mu_{\Psi}(c, d) = \max_{\pi \in M} \Psi'(\pi)$$

alakban kapjuk, ami a pontok között vezető összes lehetséges út halmazából (M) a maximális affinitás értékűt jelenti. Ez a keresés dinamikus programozással hatékonyan megoldható.

Az ún. *fuzzy összeköttetés alapú szegmentáláshoz* ki kell jelölnünk egy kiindulópontot, amely biztosan a szegmentálandó objektumhoz tartozik, majd minden más képponthez meghatározásra kerül a fuzzy összeköttetés értéke. Így egy *fuzzy összeköttetési térképet* kapunk, aminek az értékei tehát azt jellemzik, hogy az egyes képpontok milyen mértékben tartoznak össze a kiindulási képponttal. A szegmentálás végeredményét ennek a térképnek a küszöbölésével kapjuk meg, rendszerint a küszöbérték interaktív kiválasztásával. Számos további változatát kidolgozták a fuzzy szegmentálásnak, ezeket itt nem tárgyaljuk.

7.2.2. Él-alapú módszerek

Képen ott érzékelünk éleket, objektumok között határokat, ahol a képpontok közötti intenzitáskülönbség elegendően nagy. Így él információt az $f(x, y)$ képfüggvényből a $|G| = \sqrt{G_x^2 + G_y^2}$ gradienseknek, vagyis az elsőrendű deriváltjának vizsgálatával kaphatjuk meg. A gradiens iránya (a „legnagyobb csökkenés” iránya) $\tan^{-1} \frac{G_y}{G_x}$ alakban áll elő, ahol G_x és G_y az x - és y -irányú gradiensek. Az él iránya erre a gradiensre merőleges. Élkeresésre a másodrendű derivált közelítése is felhasználható: ahol az elsőrendű derivált szélsőértéket vesz fel, ott a másodrendű derivált nulla és előjelet vált, ezeket a váltásokat kell detektálni. A másodrendű deriváltat *Laplace operátornak* hívjuk, jelölése $\nabla^2 f(x, y)$.

Diszkrét digitális képek esetében a gradienst és a Laplace operátort intenzitáskülönbségek számításával közelíteni tudjuk. Rendszerint konvolúciós maszkokat (pl. Roberts, Prewitt, Sobel, Frei-Chen; Laplace) alkalmazhatunk, majd az élkép küszöbölésével kaphatjuk meg az élék helyeit. Ezek a módszerek gyorsak, nagy problémájuk viszont, hogy az objektumok körüli határvonal általában nem folytonos, így az éldarabok összekötése, és így a régiók elkülönítése utófeldolgozó lépést is megkövetel. A egyszerű módszer az él *hiszterézis*, az élképen a képpontok lokális környezetének vizsgálata. Amennyiben egy képpont szomszédságában vannak elegendő mértékben egyező nagyságú és irányú gradiensek, akkor az adott pontot hozzávesszük az élhez. Egy élpont gradiense hiába nagy, ha az él megszakad, azt el kell nyomni: ha az él iránya egy pontban jelentősen eltér a szomszédságában található élekétől, akkor kijelöljük törlésre. Összetettebb detektorok alkalmazásával kevesebb, relevánsabb él információhoz juthatunk. Az egyik legnépszerűbb ilyen detektor az élelnyomást, hiszterézist és jellemző szintézist is magában foglaló *Canny éldetektor*. További nagy problémája az

élkereső módszereknek a zajérzékenység. A zaj hatását egy előzetes simító lépéssel csökkenteni lehet. Marr és Hildreth javasolta először a Gauss simítás alkalmazását a Laplace operátor előtt, amit így együttesen LoG (*Laplacian of Gaussian*) operátornak neveznek.

Élkereső módszerek gráf algoritmusokkal is kombinálhatók. A gráf csúcspontjainak a képpontok kerülnek megfeleltetésre, a 4-szomszédjukkal élek kötik össze őket. Az élek a gradiens értékeknek megfelelően kapnak súlyt. Két pont között így lehetővé válik a legkisebb költségű út megkeresése, ami az objektum határát adja. Az ún. *livewire* szegmentálás ezt az elvet használja fel interaktív formában. Az orvos az objektum határa mentén kontúrpontokat helyez el. A módszer a legutóbbi pont és a mozgó szátkereszt között keres minimális költségű utat. Mivel a keresés gyors, az eredmény valós időben meg is jeleníthető. A szegmentálás eredménye így folyamatosan követhető, a szátkereszt pozícionálásával befolyásolható.

Élkereső módszerek közé sorolhatók még a deformábilis kontúrok, ún. *snake*-ek is, ezeket a következő pontban tárgyaljuk.

7.2.3. Energiainimalizáló módszerek

Az előző alfejezetben az objektumok határvonalát egymástól függetlenül detektált élpontok utólagos feldolgozásával próbáltuk megtalálni. Egy másik lehetőséget biztosít a kontúrok parametrikus megadási módja, és alakjának *a priori* és képi információk alapján történő deformációja. A deformációhoz a kontúr alakjára (folytonos, „sima”) és a szétválasztott területek intenzitásértékeire (pl. legyenek homogének) vonatkozó energiatagok kerülnek definiálásra, és ezek egyensúlyi állapota adja a szegmentálás eredményét.

A klasszikus *aktív kontúr modell* vagy *snake* egy ilyen energiainimalizáló spline, amit az 1980-as évek második felében dolgozott ki Kass, Witkin és Terzopoulos. Az energiatag a kontúr alakjából kinyert *belső erőből*, valamint a képi információból származó *külső erőből* áll elő. Maga a kontúr $\mathbf{v}(s) = [x(s), y(s)]$ parametrikus alakban adott, ahol $x(s)$ és $y(s)$ jelentik a kontúr x és y koordinátáit $s \in [0,1]$ mellett. A minimalizálandó energiatag az alábbi alakban írható fel:

$$E_{\text{snake}} = \int_0^1 \left(E_{\text{belső}}(\mathbf{v}(s)) + \left(E_{\text{kép}}(\mathbf{v}(s)) + E_{\text{megszorítás}}(\mathbf{v}(s)) \right) \right) ds ,$$

ahol $E_{\text{külső}} = E_{\text{kép}} + E_{\text{megszorítás}}$. $E_{\text{megszorítás}}$ a képi információból származón túli egyéb, például a felhasználó által adott külső megszorító erőket jelenti. $\mathbf{v}(s)$ közelítésére célszerű spline-t választani. A belső erő ekkor

$$E_{\text{belső}} = \alpha(s) \left| \frac{d\mathbf{v}}{ds} \right|^2 + \beta(s) \left| \frac{d^2\mathbf{v}}{ds^2} \right|^2$$

alakban kapható meg, ahol $\alpha(s)$ a kontúr elasztikusságát, míg $\beta(s)$ a merevségét határozza meg. A képi információból származó külső erő több tag súlyozott összegére bontható:

$$E_{\text{kép}} = w_1 \cdot E_{\text{vonala}} + w_2 \cdot E_{\text{él}} + w_3 \cdot E_{\text{végpont}} .$$

E_{vonala} szabályozhatja, hogy sötét vagy világos intenzitású vonalakra húzódjon a kontúr, $E_{\text{él}}$ a magas gradiensű pontok felé vonzza, $E_{\text{végpont}}$ segítségével pedig a képen található sarokpontok, vonal végpontok fejthetnek ki vonzó erőt. Az optimális $\mathbf{v}(s)$ az ún. Euler-Lagrange

egyenlet megoldásával kapható meg:

$$\frac{d}{ds}E_{\mathbf{v}_s} - E_{\mathbf{v}} = 0,$$

ahol $E_{\mathbf{v}_s}$ jelöli E dv/ds szerinti, $E_{\mathbf{v}}$ pedig a \mathbf{v} szerinti parciális deriváltját. A hatékony megoldáshoz célszerű egy optimumhoz közeli kontúrt választani és iteratív keresést végezni. A megfelelő inicializálás nagyon fontos az aktív kontúr algoritmus esetében, a megfelelő súly paraméterek meghatározása is nehézséget okozhat, valamint az optimum megkeresése numerikusan instabillá válhat.

A numerikus instabilitás elkerülésére bevezethető egy külső, ún. *felfújó erő*, amely a kontúr folyamatos „kifelé” mozgását végzi. Az optimumhoz közeli inicializálás problémájára Xu és Prince javasolt megoldást, az ún. *gradiens vektor folyam* (GVF) bevezetésével. A GVF mező erői a kép homogén területein az élek felé mutatnak, így távolabbi kezdőpozícióból is biztosítható a konvergencia.

A kontúr parametrikus megadási módja helyett a kontúr helyzetét parciális differenciálegyenlet megoldásával is megkaphatjuk. Ezek az ún. *level-set* módszerek. A kontúr időbeli mozgását egy sebességfüggvény vezérli, a mozgás élek mentén áll meg. Képes kezelni a kontúr topológiaváltásait is. Chan és Vese javasolt egy olyan kiterjesztést, ahol a kontúr által szétválasztott régiók intenzitásértékeinek homogenitása is szerepet játszik a megállási feltételben [12].

7.2.4. Modell-alapú megközelítés

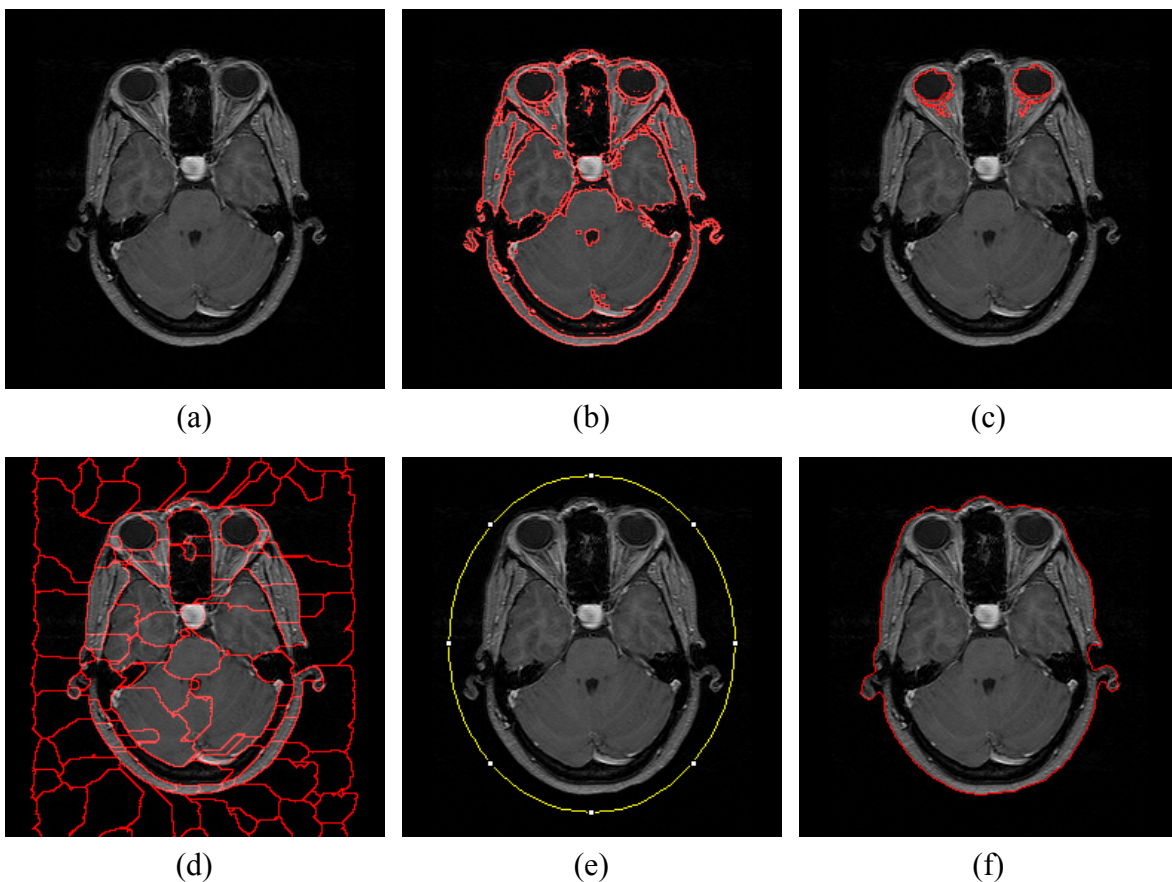
Az emberek anatómiai felépítése, a szervek alakja, elhelyezkedése *globálisan* nézve nagy hasonlóságot mutatnak. Alak és méret szerint viszont nagy *lokális* különbségek is előfordulhatnak. Szegmentáló módszerek számára nagy segítséget jelent, ha a különbségekre vonatkozóan *statisztikai információ* rendelkezésre áll.

Statisztikai információ reprezentációnak egyik gyakran használt módja a Cootes által javasolt ún. *pont-eloszlási modell* (PDM – *Point Distribution Model*) [15]. Ennek fő lépései az alábbiak:

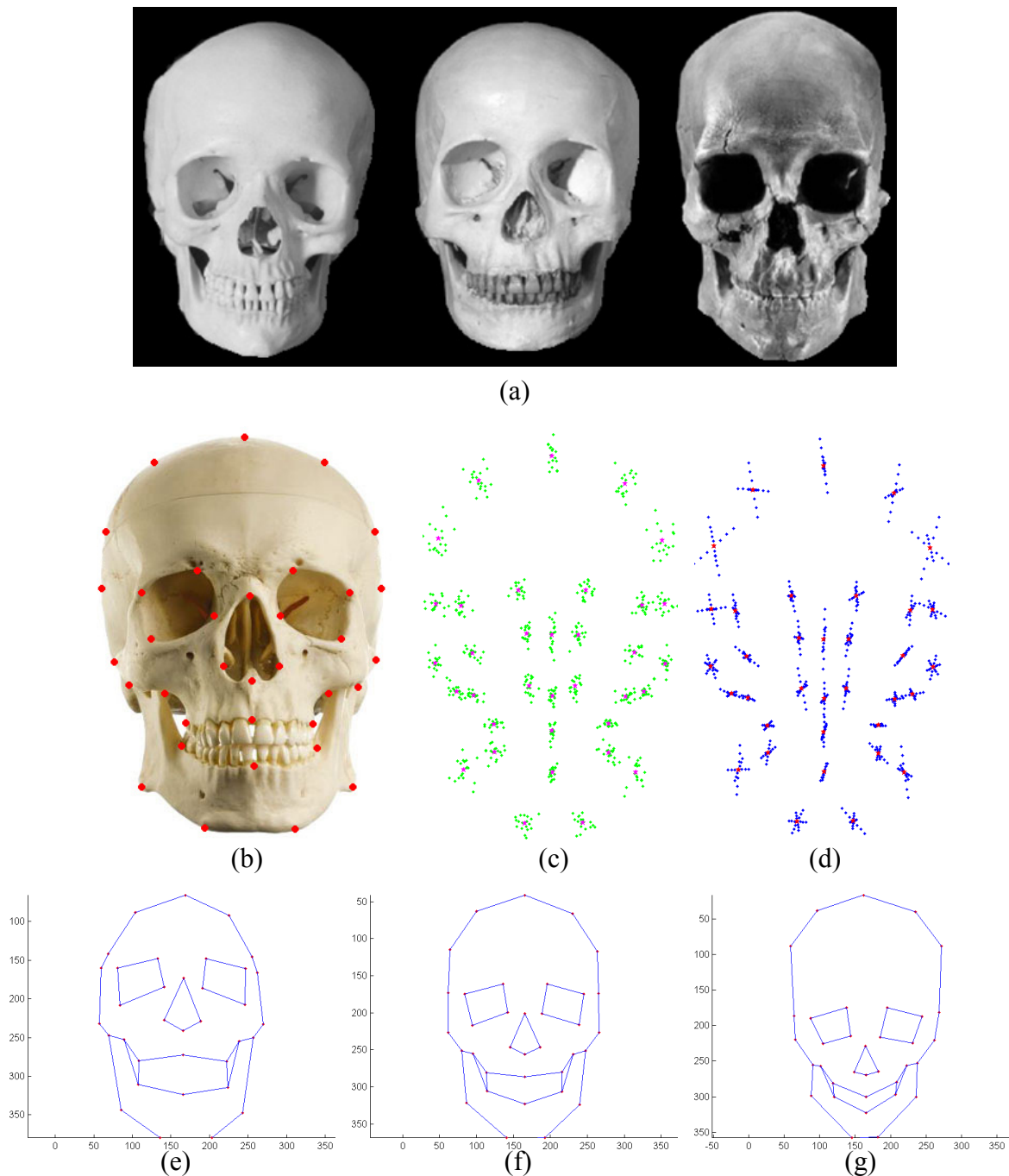
1. Gyűjtünk össze ugyanarról a szervről képeket minél nagyobb számban (több tízes vagy százas nagyságrendben, jelölje M). Fontos, hogy ezek jól reprezentálják a szerv populációban előforduló alak és méret változatosságát.
2. Jelöljük ki minden képen egymásnak megfeleltethető N darab pontot. Ez történhet manuálisan vagy automatikusan is. A pontokat összekötő vonalak (2D-ben) illetve fel-színhálók (3D-ben) közelítsék a szerv körvonalát (D jelölje a dimenziót).
3. A ponthalmazokat transzformáljuk egy közös referencia térbe, pl. pontpár-illesztő módszerrel, eltolás, elforgatás és skálázás figyelembe vételével. Ettől a lépéstől azt várjuk, hogy az egymásnak megfeleltethető pontok egymás „közelébe” kerülnek. A megfeleltethető pontok „szóródása” fogja az alakra vonatkozó statisztika alapját képezni.
 - Minden $1 \leq n \leq N$ anatómiai pontra számítsuk ki az M darab pont átlagát. Az átlagokat vonjuk ki az egyes koordináta-értékekből, és ezekből képezzünk $D \cdot N \times D \cdot N$ méretű kovarianca-mátrixot.

- Számítsuk ki a kovariancia-mátrix sajátértékeit és sajátvektorait. A sajátvektorokat a sajátértékeik szerint rendezzünk csökkenő sorrendbe.
- Új alakzatot úgy tudunk képezni, hogy a pontátlagokhoz hozzáadjuk az első t ($1 \leq t \leq \leq D \cdot N$) darab sajátvektort tetszőleges súlyozással. t értékét célszerű rögzíteni.

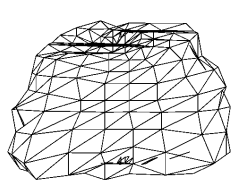
Nagy előnye a megközelítésnek, hogy a súlyok állításával olyan alakzatokat is megkapunk, amelyek a betanító halmazban nem szerepeltek. Szegmentálásra a legegyszerűbben úgy használható a modell, hogy vizsgáljuk a súlyok állításával keletkező kontúrok és a képen detektált élek egybeesését. Egy másik lehetőség, hogy energiaminimalizáló módszernél külső erőként alkalmazzuk az aktuálisan vizsgált kontúr és a hozzá illeszthető legközelebbi statisztikai modell távolságát.



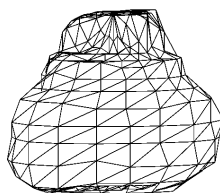
7.6. ábra. Szegmentáló algoritmusok 2D képszeleten. Eredeti kép (a); Otsu-küszöbölés eredeti képre vetített körvonalai (b); szemgolyók közepéből indított régiónövelő szegmentálás körvonalai (c); vízválasztó szegmentálás partícióinak határa (d); aktív kontúr modell inicializálása ellipszissel (e) és az optimális végeredmény (f).



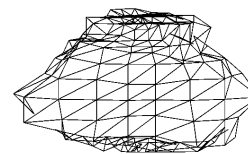
7.7. ábra. Statisztikai modell készítése 2D frontális koponya képekhez. Három frontális koponya fotó a 18 képből álló adatbázisból (a). Minden fotón ki kell jelölni az egymásnak megfelelő anatómiai pontokat (b). A 35 pont szóródása (zöld pontok) az átlagukhoz (vörös pontok) képest (c). A három legnagyobb sajátértékkel rendelkező irány szerinti pont koordináták (kék pontok) a súlyok módosításával. (e–g) A deformált körvonal a legnagyobb változási irányban alkalmazott 3 különböző súly értékkel. (A képekért köszönetet mondunk Tasi Tamás Sámuelnek.)



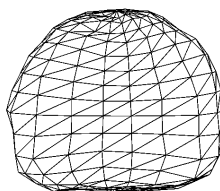
(a)



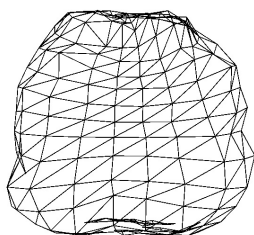
(b)



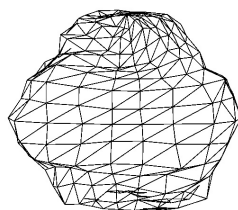
(c)



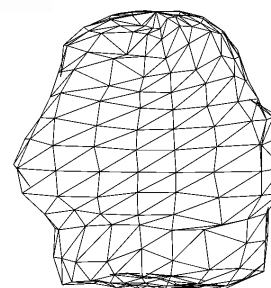
(d)



(e)

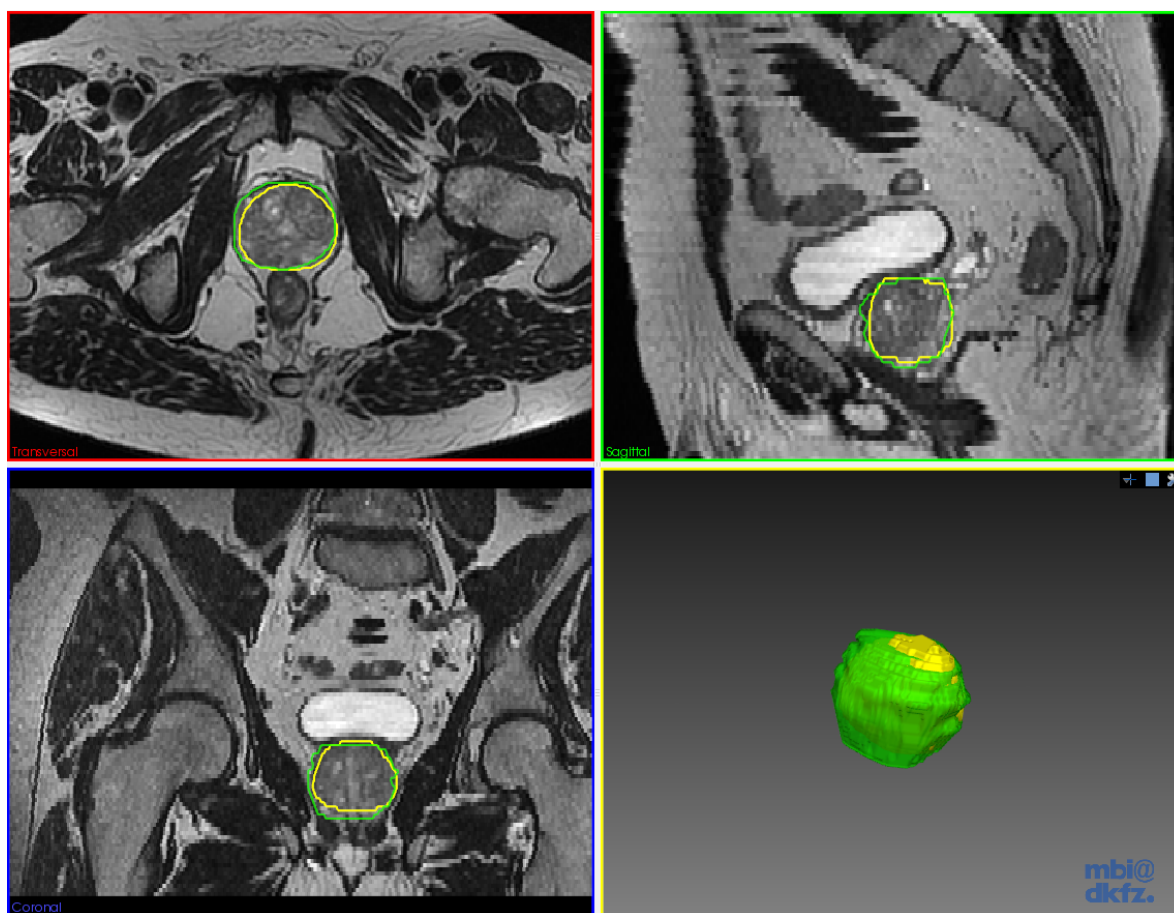


(f)



(g)

7.8. ábra. Statisztikai modell készítése 3D prosztatára szegmentáláshoz. Három, orvos által szegmentált prostata felszínhálója az adatbázisból (a–c). Az átlag prostata (d). Új alakzatok generálása a 8 legnagyobb sajátértékkel rendelkező irány szerinti súlyok módosításával (e–g). (Az eredmények közzlése a GE Healthcare szegedi irodájának engedélyével történt. Köszönetet mondunk Fidrich Mártának, Nyiri Gergelynek, Osztróluczki Andrásnak, Nagy Viktornak és Bara Norbertnek.)



7.9. ábra. prosztatata szegmentálás alakmodell felhasználásával. A képen az orvos által szegmentált prosztatata körvonala zöld színnel, a szegmentáló módszer eredményének körvonala sárga színnel látható. (Az eredmények közzlése a GE Healthcare szegedi irodájának engedélyével történt. Köszönetet mondunk Fidrich Mártának, Nyiri Gergelynek, Osztroluczki Andrásnak, Nagy Viktornak és Bara Norbertnek.)

8. fejezet

Távérzékelés, mikroszkópia

8.1. Távérzékelés

A távérzékelés során objektumokról, tárgyokról úgy gyűjtünk adatokat, hogy a mérőberendezés nincs közvetlen kapcsolatban a megfigyelés tárgyával. Az adatok gyűjtése mellett azok feldolgozása is a terület része. A fenti definícióba a hagyományos fotózás és az orvosi képalkotás is belesznek, de a gyakorlatban távérzékelés alatt a földfelszínről nagyobb távolságból készült adatok készítését és feldolgozását értjük. A szakterületről több alapos magyar nyelvű áttekintő munka is született [44, 53].

8.1.1. A távérzékelés alapfogalmai

Az adatok gyűjtése történhet felszínközeli méréssel, repülőgépekről és műholdakról. Az első légifelvételek az 1850-es években készültek, az akkoriban feltalált fényképezési eljárások és a repüléstechnika (léggömbök) együttes alkalmazásával, néhány száz méteres magasságokból. A repülőgépek megjelenése nagy lökést adott a légifotózásnak, az 1910-es években már földtani kutatások céljából is készültek fotók. A domborzat tanulmányozását a sztereo-képalkotó technikák segítették, vagyis az emberi látás működési elvét felhasználva a repülőgép két különböző pontján elhelyezett kamerák más nézőpontból biztosították egy-egy terület képét. Az 1940-es években a rakéatechnika felhasználásával már több száz kilométeres magasságba sikerült eljutni. 1959. augusztus 7-én készült az első fényképfelvétel a Földről mesterséges műholdról, majd az 1960-as évektől kezdődően sorozatban indultak műholdprogramok meteorológiai, földtani kutatások segítésére (TIROS, NIMBUS, Kozmosz, Landsat, SPOT). A képalkotó technológiák fejlődésével egyre jobb felbontású és nagyobb mennyiségű adat érkezik, amit a meteorológia és a földtan mellett a mezőgazdaságban (termésbecslés), környezet- és természetvédelemben, vízgazdálkodásban és a térképészetben is eredményesen felhasználnak állapotfelmérésre és állapotváltozások nyomon követésére.

A távérzékelés során leggyakrabban az elektromágneses hullámokat használják fel. Az elektromágneses tartományt több részre oszthatjuk fel a sugárzás hullámhossza alapján. Az egyik legfontosabb tartomány a $0,4-0,7 \mu\text{m}$ közötti ún. látható fény tartomány, ami a teljes spektrum nagyon kis szeletét teszi ki. A látható tartománynál kisebb hullámhosszú sugárzások az UV, majd a röntgen és gamma sugárzások. Ezek nagyrészt a légkör kiszűri. A másik

irányban 0,7-1,3 μm között közeli-infravörös, 1,3-3 μm között középső infravörös, míg 3-100 μm között hőtartományról beszélünk. A látható tartománnyal együttesen ezeket *optikai hullámhossz-tartománynak* nevezzük. A közeli-infravörös sáv általában a Nap sugárzásából származik, amit a növényzet visszaver, a víz pedig elnyel. Az infravörös hőtartományt a földfelszín által kibocsátott hősugarak alkotják, mérésével a felszín hőtérképe készíthető el. 1 mm és 30 cm között a mikrohullámok, azon túl pedig a rádióhullámok helyezkednek el. A mikrohullámok a felhőtakarón jól áthatolnak, bizonyos mértékig a felszín alá is bejutnak. A felszínről visszavert mikrohullámok (radar) segítségével a domborzat magasságértékei mérhetők. Mérések végezhetők még sok más módon (pl. lézerrel). A gyakorlatban az adatok mérése szűrők alkalmazásával több elektromágneses hullámsávban egyidejűleg történik, így a vizsgált terület más-más jellemzői detektálhatók a segítségükkel, ami a képelemzési, osztályozási feladatot jelentősen segíti.

Megkülönböztetünk *passzív* és *aktív* távérzékelő rendszereket. A passzív esetben a berendezés a természetes sugárzást érzékeli. Az aktív rendszerek a sugárzást maguk bocsátják ki, és annak a tárgyról való visszatükröződését mérik.

A legfontosabb sugárforrás a Nap, de minden 0 Kelvin foknál melegebb anyag bocsát ki elektromágneses sugárzást. A sugárzás energiája a hullámhossz növekedésével csökken, vagyis például a földfelszín által kisugárzott mikrohullámú sugárzás nehezebben érzékelhető, mint a rövidebb hullámhosszú infravörös hőtartomány energiája. A sugárzás jelentős utat tesz meg az atmoszférában, azon egyszer (földfelszín által kibocsátott sugárzás) vagy kétszer (aktív rendszerek vagy a Nap sugárzása) áthalad, és kapcsolatba lép azzal. A hatása függ a sugárzás által megtett út hosszától, a sugárzás energiájától, az atmoszféra összetételétől (páratartalom, szennyezettség), valamint a hullámhossztól.

A két legfontosabb atmoszférikus hatás a *szóródás* és az *elnyelődés*. A szóródásnak több típusa van. Jellemző rájuk, hogy a különböző hullámhosszúságú sugárzásokat más mértékben térítik el, ennek a látható fény tartományban is érzékelhető a hatása (pl. kék színű égbolt, vörös-narancs színű naplemente). A szóródás gyengíti a felvétel minőségét, kontrasztcsökkenést okoz. A szóródással szemben az elnyelődés valódi energiaveszteséggel jár. A légkör egyes összetevői egyes hullámhossz-sávokban teljesen elnyelik a sugárzást, míg az ún. *atmoszférikus ablakokban* átengedik azt. Távérzékelő rendszerek tervezésénél az elnyelődési tartományokat mindenképpen figyelembe kell venni.

A földfelszín anyaga a rá érkező sugárzást egyes részeit elnyeli vagy továbbítja, más részét visszaveri. A visszaverődés típusa és aránya anyagonként, valamint azok állapotától függően különbözik, függ a sugárzás hullámhosszától és a domborzati viszonyoktól is. *Spektrális reflektancia* értékét (p_λ) az adott hullámhosszon visszavert (E_r) és a beeső teljes sugárzás (E_b) hányadosából kapjuk:

$$p_\lambda = \frac{E_r(\lambda)}{E_b(\lambda)} \cdot 100\% .$$

Az egyes hullámhosszokra meghatározott értékeket grafikonon ábrázolva kapjuk az ún. *spektrális visszaverődési (reflektancia) görbét*. Ez a görbe nem csak az objektum spektrális tulajdonságait jellemzi, hanem meghatározza azokat a hullámhossz-tartományokat, amelyekben a kérdéses objektumok más objektumoktól elválaszthatóan részletesen vizsgálható.

8.1.2. Képkalkotás

A detektorokat rendszerint repülőgépeken vagy műholdakon helyezik el. A *pánkromatikus szenzorok* szélesebb sáv tartományt ölelnek fel, viszont jobb térbeli felbontóképességgel rendelkeznek. Az ún. *multispektrális detektorok* tartalmaznak *spektrométert*, amely a beérkező sugárzást szűrők segítségével sávokra bontja és a lencserendszere azokat a megfelelő detektorra juttatja. A többcsatornás spektrométerek a 0,4-15,0 μm -es tartományt akár több tíz vagy száz egyforma szélességű csatornára bontják és ezzel közel folytonos spektrumot alkotnak a felszínről.

Az adatok rögzítése történhet hagyományos filmes fotózási technikával, videósugárzással is, de manapság a digitális képkalkotás, tárolás, adatküldés és feldolgozás célravezetőbb megoldás. A műholdakhoz földi vevőállomás rendszer is kiépítésre került, amelyek a jeleket fogadni tudják. A két leggyakoribb távérzékelő rendszer a *kamera* és a *pásztázó* rendszer.

A kamera rendszer a vizsgált felszín teljes területéről egyidejűleg készíti a képet, amelyek rendszerint részben átfedik egymást. Az érzékelési tartomány általában 0,3-0,9 μm közé esik. A pásztázó rendszerek a detektorral párhuzamos irányban, egy vékony sávban vizsgálják a területet. A *keresztsávós pásztázás* során egy, a repülés irányával megegyező tengely körül motor által meghatározott szög tartományban forgatott tükör gyűjti össze a sugárzást és juttatja a detektorra. A repülés magasságát és sebességét figyelembe véve a forgásnak elgondően gyorsnak kell lennie, hogy ne maradjon ki terület a képkalkotás során. Ez egy-egy területdarabra igen rövid időt jelent, ami csökkenti a beérkező energia mennyiségét, így nehezebb a jelek megfelelő minőségű detektálása. A forgatás miatt a kép egyes pontjain reprezentált felszínelemek detektortól mért távolsága változó, így a kép széle felé haladva csökken a térbeli felbontóképesség. A *köríves pásztázó rendszerek* esetén a tükör egy függőleges tengely körül forog, így egy ív alakú területet vizsgál. Az egy képponthez tartozó felszín területe megegyezik és a pásztázás is gyorsabb, mint a keresztsávós esetben, viszont a képek feldolgozása előtt geometriai korrekcióra van szükség. A *sávmenti pásztázó technika* sorban vagy mátrix alakban több detektort is tartalmaz, így megnövekszik az egy területre eső felvételezési idő, ami növeli a térbeli és a spektrális felbontást. Ez utóbbi akár 0,1 μm is lehet.

8.1.3. Műholdas programok

Az utóbbi fél évszázadban számos kormányzat, ügynökség, szervezet és cég indított műholdas programokat erőforráskutatás, meteorológia, térképészeti célokkal. A teljesség igénye nélkül mutatunk be ezek közül néhányat.

Az egyik legrégebben futó műholdas erőforráskutató program a NASA és az Amerikai Belügyminisztérium által 1966-ban indított *Landsat*¹ program. Az első műhold pályára állítása 1972-ben történt meg. Jelenleg (2011-ben) a Landsat 5 és 7 működnek. A Landsat holdak sarkokhoz közeli, Napszinkron pályán keringenek, 920 km (Landsat 1-3) illetve 700 km átlagmagasságban (Landsat 4-7), 103 illetve 99 perces keringési idővel. Mozgásuk során fokozatosan nyugatra haladnak, így egy hold 18 illetve 16 napos visszatérési idővel rendelkezik (vagyis ilyen időközönként tudja pontosan ugyanazt a felszínrészt vizsgálni). A Landsat 1-3 holdak RBV kamerát (videókép visszasugárzó) és multispektrális kamerát (MSS) hordoztak.

¹<http://landsat.usgs.gov/>

A Landsat 4-5 holdak az MSS mellett egy fejlettebb érzékelőt, ún. tematikus leképezőt (TM – Thematic Mapper) tartalmaztak az RBV helyett. A Landsat 7 detektora ETM+ (Enhanced Thematic Mapper Plus) névre hallgat. Az MSS detektor sávmenti pásztázó elven működik, a látható tartomány és a közeli infravörös tartomány 2-2 intervallumában, 4 csoportban elhelyezett 6-6 vonaldetektorral. Az 1-3 csatorna 0-127 közötti egész típusú mérési eredményeket produkál, a 4-es csatorna 0-63 közöttit. A térbeli felbontása kb. 80 méter (aminél kisebb objektumok már 1 képponton belülre esnek). A TM szenzor már 7 sávban mér, 30 méteres felbontással, 0-255 közötti értékeket szolgáltatva. Az ETM+ ezeket egészíti ki egy 15 méteres felbontású pánkromatikus szenzorral a 0,5-0,9 μm tartományban. Egy teljes kép elkészítése 25 másodpercet igényel, ami alatt a Föld keleti irányban elmozdul, így a szélességi koordinátától függő mértékben az elkészült képet geometriailag korrigálni kell.

A Landsathoz hasonló célt tölt be a Francia Űrügynökség által indított *SPOT* program², amelynek első holdja 1986-ban állt pályára. Jelenleg a SPOT 4 és 5 holdak működnek. Ezek is poláris, napszinkron pályán mozognak 830 km magasságban, nyugatra tolódva, 26 napos visszatérési idővel. A képalkotó berendezése elnevezése HVR (High Resolution Visible), amely pánkromatikus módban 10 méteres, multispektrális módban (zöld, vörös és közeli-infravörös tartományokban) 20 méteres felbontásra képes. A SPOT holdak két kamerával rendelkeznek, amelyek a függőleges irányhoz képes 27 fokkal elforgathatók, így lehetőség nyílik egyrészt sztereo képpárokat készíteni, másrészt egy-egy felszín terület a visszatérési időnél rövidebb időközönként is fotózható.

A technológia fejlődésével a térbeli felbontóképesség tovább nőtt. Az 1999-ben felbocsátott IKONOS műhold a 0,45-0,90 μm pánkromatikus tartományban már 80 cm-es, míg a kék, zöld, vörös és közeli infravörös multispektrális tartományokban 4 méteres felbontással rendelkezik és 11-bites mérési eredményeket ad. 2008-ban követte a GeoEye-1 műhold, ami már 41 cm, illetve 1,65 m felbontással készít képeket, a 2013-ra tervezett GeoEye-2 ezt 25 cm-re csökkenti, de az Egyesült Államok kormányzata a fél méternél jobb felbontásokhoz nem ad hozzáfértést. A három műhold együttesen képes lesz a földfelszín szinte bármely pontjáról naponta új, 1 méter alatti felbontással képet szolgáltatni. Érdekesség, hogy a GeoEye-1 műholdban pénzügyi befektetőként a Google cég is szerepet játszik, a műhold adatait térképészeti céljaira használja fel.

Meteorológiai alkalmazásokhoz már az 1960-as években pályára állítottak műholdakat. Az első sikeres kísérlet az amerikai TIROS-1 volt 1960-ban, ami 78 napig működött. Poláris pályán keringve kb. 700 km-es magasságból 10 másodpercenként készített és sugárzott televíziós felvételeket a légkör állapotáról. Poláris pálya mellett ún. geostacionárius pályán (egyenlítő felett, mindig azonos pozícióban látszó) is található meteorológiai műholdak. Ilyen pl. az amerikai GOES vagy az európai EUMETSAT. 36000 km-es magasságból a földfelszín kb. negyedéről készítenek kilométer nagyságrendű felbontásban képeket. Ezek a holdak alkalmasak pl. a felhő- és jégborítottság, időjárási frontok, ózontartalom érzékelésére.

8.1.4. Mezőgazdasági felhasználás

A felszín multispektrális vizsgálata a reflektanciagörbék alkalmazásával lehetőséget biztosíthat a különböző fajtákhoz tartozó növényzet osztályozására, betegségük észlelésére, a talaj

²<http://www.spotimage.com/>

típusának megállapítására. A növények levelében található klorofil erősen elnyeli a kék és vörös tartományt, a zöldet pedig erősen visszaveri. Az egészséges növényzet ezért általában zöld színben látható. A barnás, sárgás színek betegség jelei lehetnek. A 0,7-1,3 μm közötti közeli infravörös tartományban a visszaverődés jelentősen megnő, a 40-50%-ot is eléri, amely elsősorban a levélzet belső szerkezeti sajátosságaiból adódik. Mivel a belső szerkezet fajtánként erősen eltérő lehet, így ennek a tartománynak a vizsgálatával még akkor is különválaszthatók lehetnek növényfajok, ha a látható tartományban a színük nem válik el. Ismert fajta esetében a beteg populációt detektálhatjuk, amennyiben az a belső szerkezetre van hatással. Az 1,3 μm -esnél nagyobb hullámhosszú infra tartományban a visszaverődés mértékéből a levélzet nedvességtartalmára következtethetünk – az arány közöttük nagyjából fordított. A talaj visszaverő képességét többek között a nedvességtartalma, a szerkezete, a szerves anyag tartalma és a felszín egyenetlenségének bonyolult kapcsolati rendszere befolyásolja. Nyílt vízfelületek az 1,4 és 2,7 μm -es ún. *víznyelési sávok* segítségével detektálhatók.

8.1.5. Térképészeti alkalmazások

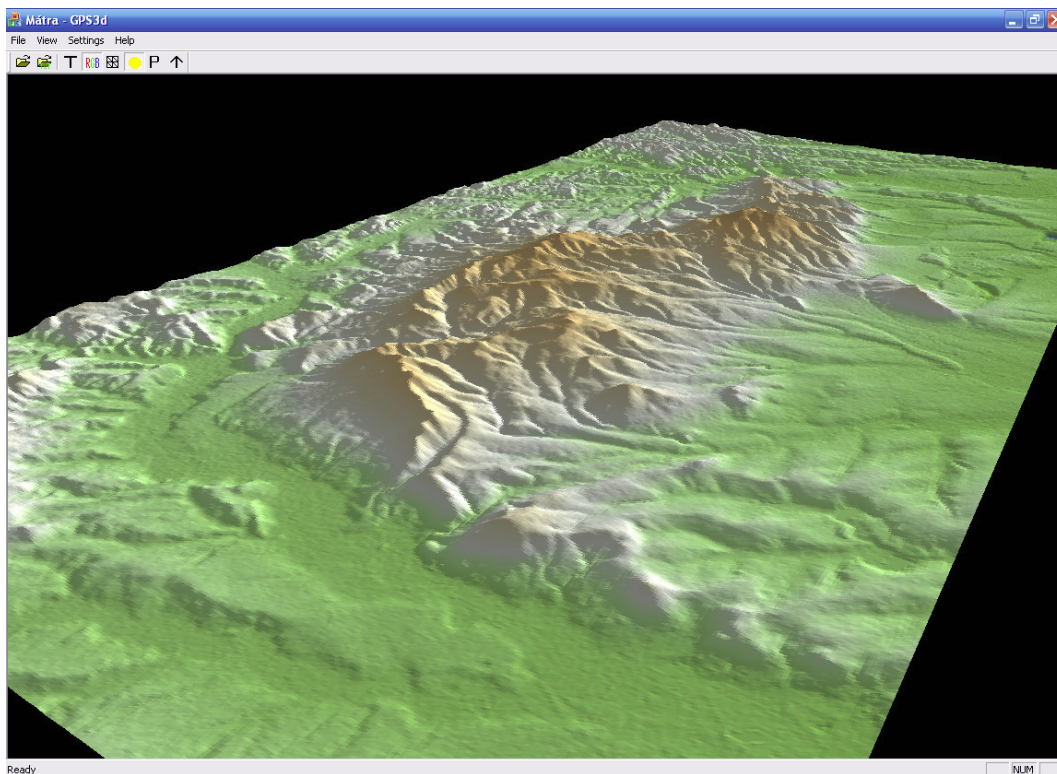
A légi- és űrfelvételek felhasználása nagy befolyással van a térképészetre is. Segítségével gyorsan és egyszerűen gyűjthető adat nagy kiterjedésű, akár nehezen megközelíthető helyszínekről is; a sztereoképek készítésével pedig geometriailag pontos, ún. ortofotók készíthetők rektifikáló eljárással, illetve a domborzat magasságáról nyerhetünk ki információt.

Mikrohullám (radar) segítségével a Föld domborzati viszonyai jól feltérképezhetők. A NASA SRTM (Shuttle Radar Topography Mission) programjának a célja a Föld felszínének mintegy 80%-át lefedő digitális domborzati térkép elkészítése volt³. 2000. február 11-én bocsátották fel az Endeavour űrrepülőgépet, fedélzetén a méréshez szükséges berendezésekkel. A mérés néhány centiméter hullámhosszúságú radarjelekkel történő radarinterferometrián alapult. Az interferenciát két, egymástól fix 60 méteres távolságban elhelyezett vevő biztosította: a nyitott űrrepülőgépből egy 60 méteres rögzített tartószerkezet nyúlt ki, amelynek végén volt a másik érzékelő. Az adatokat az űrrepülőgép fedélzetén rögzítették. Az űrrepülőgépek pályaelhajlása miatt az SRTM-program keretében feltérképezhető terület a 60 fokos északi, illetve az 57 fokos déli szélességi körök közötti régióra terjedt ki.

A mérési folyamat 11 napig tartott. Az ezt követő adatfeldolgozás munkaigényét jól jellemzi, hogy az első feldolgozási fázis csak 2004 közepére fejeződött be, a pontosított, javított második változat 2006-ban készült el. Gondot okozott a hiányzó adatok pótlása (pl. nagy magasságkülönbségű helyekről nem érkezett mérhető adat), valamint a vízfelszínen megjelenő egyenetlenségek korrigálása.

A magassági adatok digitális formában bárki számára elérhetők az interneten. Az adatok 1 foktrapéz területű darabokra vannak felosztva. Az Amerikai Egyesült Államok területén 1 szögperces (kb. 30 méteres), míg a világ többi részén 3 szögperces a beosztás (kb. 90 méter). Az SRTM magassági adatait használja fel domborzat ábrázolásra pl. a Google Föld programja, valamint több navigációs szoftver is (8.1. ábra).

³<http://www2.jpl.nasa.gov/srtm/>



8.1. ábra. A Mátra 3D domborzati modellje SRTM adatok alapján. A felszínháló színezése a magassági adatok értéke szerint egy színskála alkalmazásával történt. A képernyőkép Perényi Tamás diplomamunkájából származik.

8.1.6. Változáskövető alkalmazások

Ugyanarról a területről különböző időpontokban, több hónapos, akár éves eltéréssel készült képek összehasonlító elemzése nagy segítséget nyújt a természetes és mesterséges felszíni elemek változásainak követésében. Megállapítható például az urbanizáció (egy terület beépítésének) folyamata, az esőerdők pusztításának üteme, katasztrófák esetén a változás mértéke. Ezekhez a feladatokhoz jellemzően nagy térbeli felbontású képekre van szükség – nagy területet lefedő vizsgálatok esetén ez hatalmas adatmennyiséget jelent. Képfeldolgozási módszerek segítségével a munkát hatékonyabbá tehetjük. Mivel légifelvételek esetén általában nem biztosítható a képek pontos geometriai illeszkedése, ezért képregisztrációs módszereket célszerű alkalmazni. Az osztályozás előtt szükséges a képen található elemek szegmentálása, felismerése. Ezen műveletek többsége jól automatizálható, a felhasználónak csak a problémás esetekben kell közbeavatkoznia.

8.2. Mikroszkópia

Az emberi szem felbontóképessége – bár egyénekenként nagy változatosságot mutat – korlátozott, a tárgyak apró összetevőit, illetve a nagy távolságban elhelyezkedő tárgyak részleteit nem tudjuk érzékelni. 25 cm távolságból kb. 1 tized mm méretű pontok különböztethetők meg,

amely méret a távolsággal arányosan nő. Azokat az eszközöket, amelyek valamilyen fizikai elv felhasználásával a vizsgált tárgyak jelentősen nagyított képét állítják elő *mikroszkóp*oknak hívjuk. A mikroszkópok többféle fajtája létezik.

8.2.1. Fénymikroszkópok

Történetileg az ún. *fénymikroszkóp alakult* ki először az 1600-as évek elején, ami egy optikai és mechanikai alkotórészekből álló összetett nagyítórendszer. A nagyítás eléréséhez fénytörő lencsákat alkalmaznak, az elérhető nagyítás mértéke kb. 1500-szoros. Működésének alapelve az, hogy egy külső fényforrás látható tartományba eső elektromágneses sugárzását a vizsgáló – rendszerint vékony – tárgyon keresztülvezetik, majd egy lencserendszeren keresztülhaladva a képe felnagyítódik. A fénymikroszkóp d feloldóképességét a felhasznált elektromágneses sugárzás (fény) hullámhossza korlátozza, amit az Abbé-féle képlet alapján számíthatunk⁴:

$$d = \frac{\lambda}{n \cdot \sin \alpha},$$

ahol λ a hullámhossz, n a tárgy és az objektív lencséje közötti közeg törésmutatója, α pedig az objektív nyílásszögének a fele. A feloldóképesség szabályozására legalkalmasabb így az n paraméter, vagyis a kitöltő közeg anyagának megfelelő megválasztása. A levegőben a fénysugarak megtörnek és így kisebb arányban jutnak a lencsébe. Nagyobb sűrűségű anyagot (desztillált vizet, olajat) választva n növekszik, aminek hatására a feloldóképesség értéke (a legkisebb méretű pont, ami még különállóként jelenik meg) csökken.

A fénymikroszkóp mechanikai alkatrészei a megvilágító rendszer, a talp, az oszlop, a tárgyasztal, a tubustartó, a tubus, valamint a makro- és mikrocsavar [41, 49]. A talp U alakú fémlap, amely csavarokkal a vizsgálóasztalhoz rögzíthető. Az oszlop a további mechanikai alkatrészeket, a megvilágítás irányát szabályzó tükröt, a fénysugarakat a tárgyra összpontosító kondenzort, a vizsgálandó tárgyat tartó tárgylemezt tartalmazza. Az oszlop és a tubus között az íves alakú tubustartó helyezkedik el. A tubus általában egy megtört cső, amelyben prizma segítségével törlik meg a fény útját. Felső végéhez a szemlencse vagy okulár, az alsó végéhez a tárgylencsék csatlakoznak. Egy mikroszkóp rendszerint több (3-5) különböző tárgylencsét is tartalmazhat, ekkor egy ún. revolverfoglat biztosítja ezek elforgatással történő kiválaszthatóságát. A makro- és mikrocsavarral a tárgy és a tárgylencse távolságát állíthatjuk be, durvább, illetve finomabb léptékkel.

Az optikai nagyítórendszer két fő részből, a tárgylencséből és a szemlencséből áll. A tárgylencse kis gyújtótávolságú, több tagból álló összetett lencserendszer, amely a gyújtópontján kívül eső tárgyról fordított állású, valós nagyított képet ad. Minél kisebb a frontlencse átmérője (így a fókusz-távolsága), annál nagyobb a nagyítóképessége, vagyis a nagy nagyítás esetén a tárgynak csak igen kis részéről kapunk képet. A szemlencse a tubus felső részében található, általában két lencse (egy alsó gyűjtő és egy felső nagyító) alkotja, és a tárgylencse képét nagyítja tovább. Eredményül fordított állású, látszólagos képet kapunk. A tárgylencsével maximálisan 100-szoros, a szemlencsével további 5-25-szörös nagyítás érhető el. A tényleges

⁴Ernst Karl Abbe (1840–1905) a modern optikai egyik alapítója, számos eszköz fejlesztője, a Carl Zeiss AG egyik tulajdonosa volt.

nagyítás a kettő szorzatából áll elő, viszont a gyakorlatban 1500-szoros nagyítás felett már nem növekszik a kép részletgazdagsága, csak a tárgy pontjai válnak nagyobbá.

8.2.2. Speciális vizsgálati módszerek

A fénymikroszkópok feloldóképességét, mint láttuk, az alkalmazott fény hullámhossza korlátozza. Speciális vizsgálati módszerekkel viszont lehetőség adódik a kontraszt fokozására.

A sejtszerkezet részeinek eltérő törésmutatóját használja ki az ún. *fáziskontraszt mikroszkópia*. A kontraszt növelése festés nélkül elérhető, így az alacsony kontraszttal rendelkező élő minták jól vizsgálhatók a feloldóképesség jelentős csökkenése nélkül. A minta optikailag sűrűbb részén áthaladó fény fáziskésést szenved a ritkább részekben áthaladókhöz képest, majd az objektívben ezek interferenciáját használják fel a kontrasztos kép elkészítéséhez. Feltalálója, Frits Zernike 1953-ban kapott Nobel-díjat⁵.

A *fluoreszcens mikroszkópia* azt az elvet használja ki, hogy bizonyos anyagok kis hullámhosszú sugárzással (UV fénnel) történő megvilágítása esetén nagyobb hullámhosszú, látható fényt bocsátanak ki. Néhány mintatípus önmagában is rendelkezik ilyen fluoreszcens tulajdonsággal, az esetek többségében azonban festéssel, fluoreszcens fehérjékkel, molekulákkal kell a mintát ellátni. A festés alkalmazásakor az élő minta elpusztul, viszont a fehérjék segítségével az élő sejt vizsgálható, mivel a genetikailag módosított fehérje funkcióját a fluoreszcens rész nem befolyásolja. Különböző anyagok használatával más-más részletek emelhetők ki. Megvilágításra higanygőzlámpát vagy fémelektrodokkal működő ívlámpát használnak, amely fényéből a látható fényt teljesen kiszűrjük.

A *pásztázó lézer konfokális mikroszkópia* esetén egy fókuszált lézersugarat irányítanak a minta egy pontjára, majd a szóródott fényt, vagy fluoreszcencia esetén a kibocsátott fényt tükör segítségével egy másik pontban összegyűjtik és mérik. Az esetlegesen szóródott sugarakat két diafragmával⁶ szűrjük, amelyek azonos távolságra helyezkednek el az objektívtől. Ezen technika legnagyobb előnye, hogy a tárgyfelszíne helyett képes annak belsejében egy kiválasztott síkot vizsgálni. A síkszelet képét pásztázással, sorfolytonos bejárással kapjuk meg. Egy vizsgálat során több sík leképezésével 3D információt nyerhetünk a mintáról. Feloldóképessége nagyjából a hagyományos fénymikroszkóppal egyezik meg.

8.2.3. Elektron- és pásztázószondás mikroszkópok

Az *elektronmikroszkópok* esetén a vizsgálathoz elektronsugarakat használnak, amelynek a hullámhossza 5 nagyságrenddel kisebb, mint a látható fényé, így jóval nagyobb mértékű nagyítás érhető el (kb. 2 millió-szoros). A transzmissziós elektronmikroszkóp esetén – a fénymikroszkópokhoz hasonlóan – a mintán áthaladó sugárzást mérik. A pásztázó elektronmikroszkóp a minta felszíne és a rá bocsátott elektronsugár kölcsönhatását méri.

Pásztázószondás mikroszkópok esetében egy fókuszált sugárzást kibocsátó szondát vezetnek végig a tárgy fölött sorról sorra, és mérik a fellépő kölcsönhatás erősségét. A vizsgált kölcsönhatás alapján további csoportosítások végezhetők, a két leggyakoribb típus az atom-

⁵Frits Zernike (1888–1966) holland fizikus.

⁶Átlátszatlan, fekete lemez optikai eszközökben, amely a lencsére eső szélső sugarak visszatartására szolgál.

erő mikroszkóp (AFM) és az alagútelektron-mikroszkóp (STM). Mivel a mozgás precizitása atomi méretű, így a nagyítás mértéke is ennek megfelelő lehet.

Elektronmikroszkópok és pásztázószondás mikroszkópok esetében a mért értékek a magas hullámhossz miatt nem szín-jellegű információk, így képként álszínezéssel jeleníthetők meg.

9. fejezet

Ipari képfeldolgozó rendszerek

Az ipari képfeldolgozó rendszerek célja, hogy szigorú elvárásoknak megfelelően biztosítsák a gyártás során elvárt műszaki paramétereket a termékek vizuális jellemzőinek folyamatos monitorozásával. Lehet szó bizonyos alkatrészek formai, színbeli minőségének ellenőrzéséről, részkomponensek meglétéről vagy pozíciójáról, feliratok, jelzések detekciójáról – sok esetben fontos elvárás, hogy mostoha körülmények között, gyakran napi 24 órában kell megfelelni a technológiai elvárásoknak. Amennyiben az ellenőrző rendszerben üzemzavar fordul elő, a lehető legrövidebb időn belül el kell hárítani a problémát, ellenkező esetben a gyártás leállása magas költségeket vonhat maga után. A gépi ellenőrzés alkalmazásával:

- a humán munkaerő megkímélhető a túlzottan monoton munkafolyamatoktól,
- a véletlen, figyelmetlenségekből adódó hibák száma minimalizálható,
- szubjektív értékelési szempontok helyett objektív mérések végezhetők,
- az adatgyűjtés automatizálható és a termelés-irányítási rendszerbe integrálható.

9.1. Az ipari alkalmazások komponensei

Egy optikai minőségellenőrzést végző rendszer a következő főbb komponensekből épül fel: kamera (vagy kamerák), a kamerához illesztett lencse, állványzat, megvilágítás, számítógép, mérést és adatgyűjtést végző szoftver(ek). A mérési módszer megtervezésekor figyelembe kell vennünk minden olyan információt, amit a termékről tudunk és hatással lehet a mérés kimenetére. Amennyiben üzemi körülmények között kell a méréseket végezni, vegyük figyelembe, hogy a környezeti hatásoktól megfelelően védeni kell a mérőberendezést. Így a kamerát, annak optikáját, a számítógépet ill. az egyéb elektronikai berendezéseket védeni kell a fröccsenésektől, portól és egyéb szennyeződésektől.

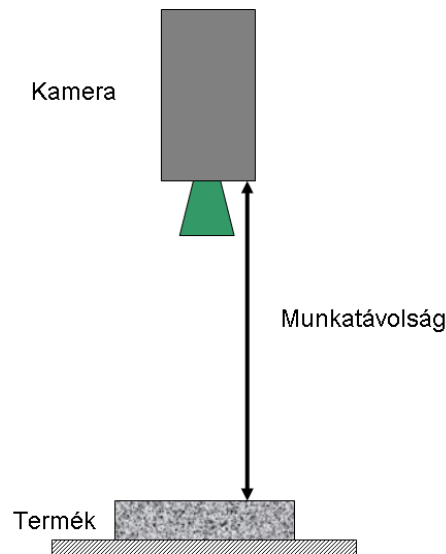
Mivel sok képfeldolgozó algoritmus érzékeny a megvilágítás változására, célszerű állandó megvilágítási körülményeket biztosítani a mérésekhez, kizárva a napsugárzás vagy egyéb fényforrások zavaró hatását.

Az objektívek rázkódás hatására elállíthatódnak, célszerű csavarral rögzíthető objektíveket

alkalmazni. A megfelelő teherbírású és stabilitású állványzat szerepet játszhat a vibráció kiküszöbölésében, csökkentve a mechanikai sérülések előfordulásának esélyét és hozzájárulva a stabil képminőséghez.

9.2. Optika és kamera kiválasztása

Általában elmondható, hogy a kamera látómezőjébe a vizsgálandó tárgynak kb. 10% ráhagyással kell beleférnie, gondolva a tárgy esetleges elmozdulására, elfordulására, a kamera-objektum távolság megváltozására. A munkatávolság (*working distance*) – a vizsgálandó tárgy és a kamera távolsága (nem beleértve az optikát ill. közgyűrűt) – lehetőleg ne haladja meg az 50cm-t, ezzel biztosítva a megvilágítás egyenletességét és zavartalanságát, ill. elkerülve túlzott teljesítményigényét.



9.1. ábra. A munkatávolság a kamera és a tárgy távolságát jelenti, általában nem haladja meg az 50cm-t.

A fizikai mérések szempontjából a kép fizikai felbontását a vizsgált tárgy legkisebb, a képen még megkülönböztethető részlete határozza meg. A kérdés tehát most az, hogy a vizsgált tárgy adott méretű (pl. 1mm-es), az ellenőrzés számára fontos részletei a képen hány pixelen jelennek meg. A felbontás függ:

- az érzékelő pixelfelbontásától,
- a munkatávolságtól,
- a látómező méretétől (ezt meghatározza a lencse nagyítása, látószöge),
- a lencse torzításától (lásd szférikus aberráció és geometriai torzítás a 2.4.2. fejezetben),
- az optika pontos beállításától (lásd a 2.6. és 2.7. egyenleteket),

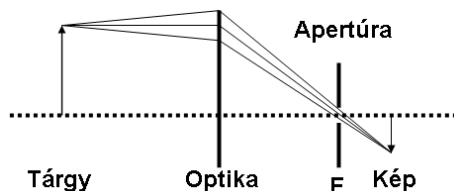
de valójában a képek elemzését végző algoritmusnak is jelentős szerepe van, pl. interpoláció segítségével lehetséges a pixelnél nagyobb pontossággal méréseket végezni (lásd 3.4. fejezet).

Ipari alkalmazásoknál a szükséges mérési pontosság meghatározza, hogy normál lencsére, macro objektívra, esetleg ipari mikroszkópra van szükség. Amennyiben az objektum mélységében is kell pontos méréseket végezni, célszerű ún. telecentrikus lencsét alkalmazni.

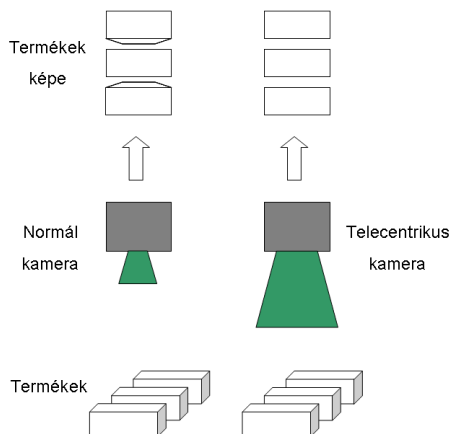
9.2.1. Telecentrikus optikák

Optikai méréseknél gyakran okoz problémát, hogy a távolabbi objektumok kisebbnek látszódnak és a vizsgált tárgy bizonyos részei takarásba kerülnek a közelebbi részek által. A perspektivikus vetítés ezen problémáit képesek kiküszöbölni az ortografikus projekciót alkalmazó telecentrikus lencsék. Az apertúra fókuszpontban való elhelyezésével elérhető, hogy a lencse úgy viselkedjen, mintha fókusz távolsága a végtelenben lenne és a távolságtól ne függjön a nagyítása (9.2. ábra). (Meggjegyezzük, hogy itt is csak egy bizonyos távolságra lévő tárgyak lesznek élesek a képen.)

A telecentrikus képalkotás hátránya, hogy nagyméretű lencsét kell használni és a normál lencsékhez képest drága az előállításuk, beszerzésük. A 9.3. ábra egy hagyományos és egy telecentrikus kamera által készített kép különbségét illusztrálja.



9.2. ábra. A fénysugarak útját a fókuszpontban erősen lekorlátozva elérhető, hogy csak az optikai tengellyel közel párhuzamos sugarak vegyenek részt a képalkotásban.



9.3. ábra. Telecentrikus lencsék alkalmazásával a leképezés után a távolabbi objektumok mérete a képernyőn megegyezik a közelebbi tárgyakéval.

9.2.2. Ipari mikroszkópok

A mikroszkópok ipari felhasználása igen jelentős, a nyomtatott áramkörök gyártásában, a fémipari megmunkálásban, az anyagtechnológiában sok esetben több tízszeres vagy akár százszoros nagyításra is szükség van. A mikroszkópok elvi felépítéséről és működéséről az előző fejezetben olvashattunk bővebben.

Kialakításukat tekintve a fő különbség a biológiai és az ipari mikroszkópok között az, hogy az ipari mikroszkópok túlnyomó esetében a megvilágítás nem alulról, hanem felülről történik, illetve a mikroszkópot sok esetben egy jól mozgatható (de stabilan rögzíthető) karon helyezik el, hogy a vizsgálandó tárgyakhoz könnyen hozzáférjen. A biológiai mikroszkópokhoz hasonlóan léteznek kamerával gyárilag felszerelt ill. hagyományos okkularéval ellátott ipari mikroszkópok is. Utóbbi esetben, ha nincs külön csatlakozási pont kamerák részére, az okkularé helyére tudunk kamerát (esetenként tubus adapterrel) elhelyezni.

A 9.4. ábrán kamerával, megvilágító gyűrűvel és LCD panellel egybeépített ipari mikroszkópot látunk, a flexibilis kar lehetővé teszi a mikroszkóp könnyű pozicionálását.

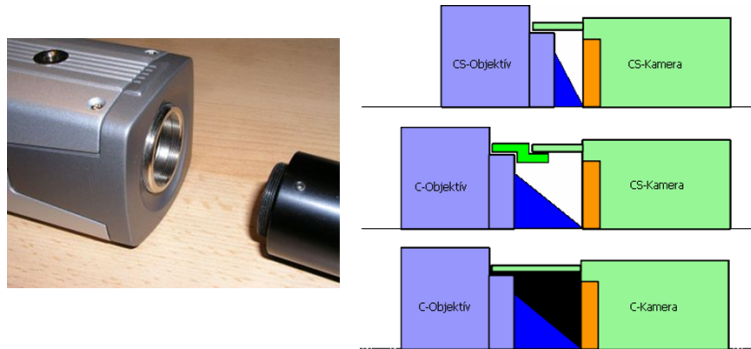


9.4. ábra. LCD panellel felszerelt ipari mikroszkóp. Az állítható kar könnyű hozzáférést tesz lehetővé kiterjedt objektumokhoz.

9.2.3. Kamera és optika illesztése

A kamera és optika illesztésére különböző szabványok alakultak ki. A kamera és optika gyártók több tucat csatoló típust fejlesztettek ki fényképezőgépek és filmes kamerák részére, de szerencsére az ipari alkalmazásokban alapvetően kétféle típus terjedt el, a C mount és a CS mount:

- A C (Cine) mount csatoló esetében az optika az "apa", a kamera az anya szerepét tölti be. A menet colonként 32-t emelkedik, a képszenzor síkja és lencse 17.526 mm-re van egymástól. Egy C mount objektív egy CS mount kamerával közgyűrű segítségével használható.
- A CS mount csupán a lencse és szenzor távolságában tér el a C mounttól. Ez a távolság CS mount esetében 12.52 mm. CS mount objektív csak CS mount kamerával használható.



9.5. ábra. C mount kamera és objektív, ill. C és CS mount lencsék és kamerák lehetséges kombinációi

Vizsgáljuk meg, hogy egy adott feladat esetén milyen nagyítású (azaz fókusztávolságú) lencsét és mekkora felbontású kamerát kell alkalmazni. A képzelt feladat szerint egy kb. 25 cm-es négyzet alakú, nagyjából sík tárgy méretét kell félmilliméteres pontossággal meghatározni. A feladat megoldására háttérvilágítást alkalmazunk. A következő gondolatmeneten kell végighaladnunk:

- I. Néhány százalékos ráhagyással számolva a látómező (lm) legyen 30 cm-es élhosszúságú.
- II. A munkatávolság (mt) legyen 50 cm-es, így szükség esetén a munkaasztal és az optika között kényelmesen elférünk.
- III. Valójában a fénysugarak a tárgy szélén – annak a kialakításától függően – valószínűleg szóródni fognak, ill. elképzelhető, hogy az optika is torzít valamelyest a képen. Ennélfogva nem fogjuk a tárgy határvonalát feltétlenül élesnek látni, így a határvonal detekció során elképzelhető, hogy interpolációt kell alkalmaznunk. Ha 0.5 mm a mérés elvárt pontossága és $lm = 30\text{cm}$, akkor legyen a szenzor minimálisan $30 \times 10 \times 2 \times 2$ pixelméretű, azaz legalább 1200 pixeles legyen a kisebbik dimenziója.
- IV. A kamerabeszállítótól kapott adatok szerint a felbontásban megfelelő szürkeskálás képszenzor mérete (szm) 6,6 mm ($2/3''$ a szenzor magassága), felbontása ebben az irányban 1500 pixel és C mount a kamera csatolófelülete.
- V. A gyűjtőoptika kép- és tárgytávolságának összege C mount esetén:

$$t + k = mt + 17,526. \quad (9.1)$$

- VI. Hasonló háromszögek alapján k kifejezhető:

$$k = t \frac{szm}{lm}. \quad (9.2)$$

- VII. A fentiek alapján a képtávolság meghatározható:

$$t = \frac{mt + 17,526}{\frac{szm}{lm} + 1} = 506,39. \quad (9.3)$$

VIII. A t -ből immár 9.2 segítségével megkapható, hogy $k=11,14$, majd pedig az optika alapegyenletéből következik, hogy

$$f = \frac{1}{\frac{1}{t} + \frac{1}{k}} = 10,9. \quad (9.4)$$

IX. A megfelelő optika kiválasztásánál a fókusz távolságon kívül ügyelni kell arra, hogy a lencse felbontása megfeleljen a 2 megapixeles szenzor-felbontásnak, illetve érdemes egy próbát tenni, hogy a lencse-kamera párosítás igényli-e közgyűrű alkalmazását.

9.3. Megvilágítási technikák

A megvilágítás igen nagyban befolyásolja a képalkotást és ezáltal azt, hogy a termék milyen tulajdonságait milyen minőségben tudjuk mérni. A vizsgálandó tárgy megvilágítása sokféle lehet:

- megkülönböztetünk felülről (előről) vagy alulról (hátról) történő megvilágítást,
- a fénysugarak lehetnek párhuzamosak, vagy diffúzak,
- a beesési szög lehet kicsi vagy nagy, ennél fogva (ill. a kamera pozíciójától függően) a látómező lehet világos (*bright field*) vagy sötét (*dark field*).

Természetesen a fenti esetek kombinációit is használhatjuk a legjobb eredmény elérése érdekében. Röviden tekintsük át, hogy mit is jelentenek ezek a beállítások és milyen hatást tudunk elérni velük.

9.3.1. Felső megvilágítás

Amennyiben felülről világítjuk meg a vizsgálandó tárgyat, a lámpa és tárgy helyzete szempontjából alapvetően kétféle eset lehetséges:

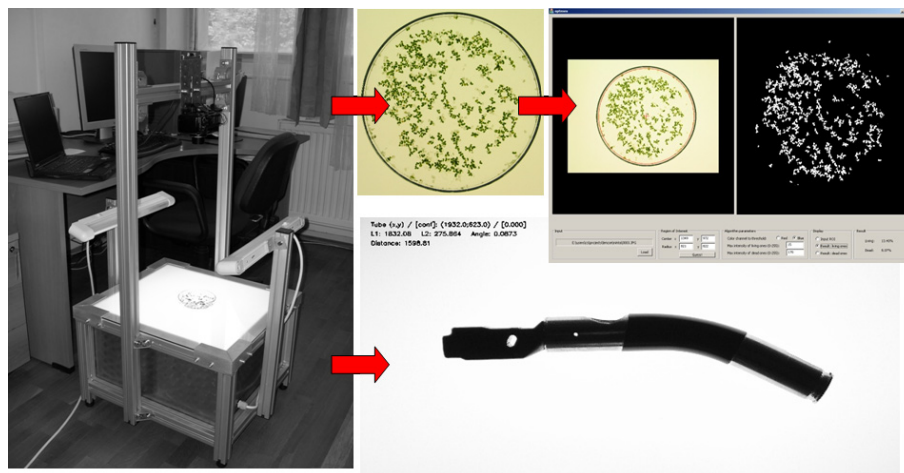
- *Sötét látómező*: Ha a kamera optikai tengelye közel merőlegesen áll a tárgy felületére, de a fényforrás iránya kicsi szöget zár be azzal, akkor a fény energiájának nagy része nem jut a kamerába és így sötét képet kapunk. Ha azonban a felületen egyenetlenségek, pl. kitüremkedő szennyeződések vagy forrasztások vannak, azok a képen világosnak fognak látszódni.
- *Világos látómező*: Ha a fényforrásból visszaverődő fénysugarak jelentős része a kamerába tükröződik, akkor a látómező világos lesz. Csak azokon a helyeken lesz sötét a kép, ahol a fény iránya eltérítődött valamilyen felületi hiba, pl. karcolás vagy szennyeződés miatt.

A megvilágítás fénysugarai haladhatnak *párhuzamosan* vagy *szóródva*, *diffúz* fényt alkotva. A diffúz fény könnyen behatol a repedésekbe, mélyedésekbe, és csökkenti azok detektálhatóságát, de egyúttal a festett mintázatok, feliratok olvashatóságát javítja.

9.3.2. Alsó megvilágítás

Alsó megvilágítást több esetben célszerű használni:

- a vizsgálandó objektum átlátszó, és a belsejében lévő részek az átvilágítás által jól láthatóvá válnak,
- a vizsgálandó objektumon lyukak vannak, azok meglétét, méretét szeretnénk ellenőrizni,
- az objektum körvonalán, sziluettjén szeretnénk méréseket végezni.



9.6. ábra. Felső-oldalsó megvilágítással, diffúz hátsó megvilágítással és kamera állvánnyal ellátott labor munkaasztal. Békalencse és fém alkatrész hátulról megvilágított képe.

A hátulról jövő fény lehet diffúz, vagy *kollimátor* használatával a fénysugarak párhuzamosíthatók. Utóbbi esetben jóval egyszerűbb az alakra vonatkozó pontos méréseket végezni a képfeldolgozó algoritmusoknak, mivel a fénysugarak nem szóródnak a vizsgálandó objektum szélén (lásd 9.7. ábra).

Néhány tipikus megoldást a 9.1. táblázat tartalmaz.



9.7. ábra. Kondenzátor képe normál háttérmegvilágítás és kollimátor használatával

9.1. táblázat. Néhány tipikus feladathoz a megfelelő megvilágítás kiválasztása

Termék/feladat	Megvilágítás típusa
íves felületű termékek	diffúz megvilágítás
közel sima, de megbillenthető, fényes felületek	diffúz megvilágítás
karcok detekciója közel sima felületeken	koaxiális vagy sötét látómezős megvilágítás
termék meglétének detekciója	alsó megvilágítás vagy sötét látómezős megvilágítás
átlátszó termék vizsgálata	alsó megvilágítás, esetleg kollimátorral
feliratok detekciója	diffúz megvilágítás

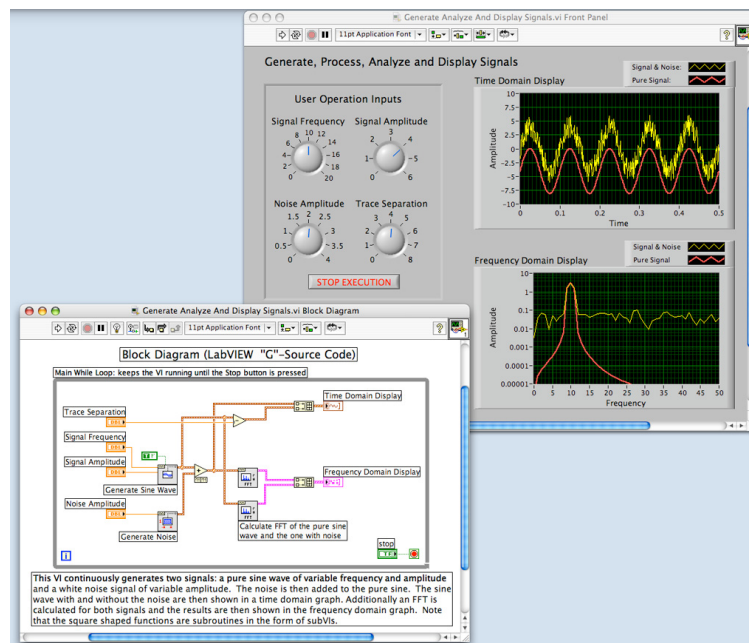
9.4. Ipari számítógépek, intelligens ipari kamerák

A minőségellenőrzés során elvégzendő mérési/számítási feladat lehet viszonylag egyszerű és lehet nagyon komplex is. Ennek megfelelően a felhasznált számítógép lehet egy kisméretű, de csak egyszerűbb műveletekre képes beágyazott rendszer vagy egy különálló, nagy teljesítményű munkaállomás is. Az ipari gyakorlatban a következő számítási platformokat különböztethetjük meg:

- PC alapú rendszerek: Az irodai felhasználásban alkalmazott olcsó, de meghibásodás esetén könnyen és gyorsan pótolható PC architektúrák megfelelő körülmények között ipari környezetben is használhatóak.
- Ipari PC-k: A különböző ipari PC típusok közül az egyik legelterjedtebb a PC/104-es, illetve ennek különböző busszal épített változatai (a PCI-104-es PCI busszal, a PCI/104-Express PCI és PCI Express busszal, a PCIe/104 pedig PCI Express busszal szerelt). A kisméretű (90.17 mm × 95.89 mm) lapra integrálták a leggyakoribb perifériás interfészeket, több alaplapi modul egymás tetejére építhető. A PC/104-es ipari PC anyagfelhasználását, csatlakozó felületeinek anyagminőségét, mechanikai és elektromos tulajdonságait szabványban rögzítették.
- Intelligens kamerák: Több gyártó készít ipari környezetben használható intelligens kamerákat, mint pl. a Sony vagy az Omron. A Sony XCI-100-as sorozata 1 GHz VIA Pentium processzorral, legalább 512MB RAM-mal, Gigabit Ethernet és USB 2.0 csatlakozóval van ellátva. Az Omron cég ipari feladatokra menüből tanítható kamerákat, többprocesszoros operációs rendszert és 3D-s szenzorokat is gyárt.

9.5. A méréseket, kiértékeléseket végző programok, algoritmusok

A különböző ipari algoritmusokat egyedileg fejlesztett keretrendszerbe, vagy valamilyen általános platformra implementálhatjuk. Az egyik leginkább elterjedt általános keretrendszer a National Instruments által fejlesztett LabVIEW (Laboratory Virtual Instrumentation Engineering Workbench), amely vizuális programozási nyelvvel rendelkezik (ennek neve G) [52]. A LabVIEW adatfolyam programozási paradigmát használ, a grafikus interfészek a program alapvető részei. Ún. virtuális eszközök (VI-k, Virtual Instruments) képviselik a programokat, szubrutinokat, a VI-k blokk diagramból, front-panelből és konnektor panelből állnak (lásd 9.8).



9.8. ábra. A LabVIEW G forráskódja és egy front-panelje

A különböző minőségellenőrzési alkalmazások során gyakran előforduló mérési feladatok sokrétűek lehetnek, mint pl.:

- képi jellemzők geometriai mérése: távolság- és területmérés,
- objektumok számának meghatározása,
- feliratok megléte, értelmezése, optikai karakterfelismerés,
- sérülések, karcok, szakadások, egyenetlenségek detekciója,
- formák, alkatrészek meglétének ellenőrzése,
- stb.

Ezeket a feladatokat a digitális képfeldolgozás és képelemzés alapvető módszereiből (hisztogram műveletek, éldetekció, morfológia, szegmentálás, mintaegyezés, stb.) felépített algoritmusokkal végezzük el, ezért tárgyalásuktól jelen esetben eltekintünk, fejezetünkben később egy-egy példát mutatunk be.

Megjegyezzük, hogy amennyiben a mérési pontosság igényli, kalibrált kamerát kell alkalmazni, így korrigálhatók a leképezés és az optika geometriai hibái (a kamera kalibráció említésre kerül a 10.3.1 fejezetben is, az elméletét részletesen a [48] segítségével is megismerhetjük). Kalibrált rendszer esetében a képfeldolgozás egyik első lépése a korrekciós algoritmus alkalmazása, ami a kép geometriai transzformálásával lehetővé teszi, hogy a képen alkalmazott mérések pontosan megfeleljenek a fizikai valóságnak.

A minőségellenőrzési alkalmazások kimenetét egy osztályozási feladatnak lehet tekinteni, ahol a termékeket a *hibás* vagy a minőségileg *megfelelt* osztályba kell sorolni. Természetesen a képelemző algoritmusok is hibázhatnak: elmulaszthatják egy hibának a detekcióját (hibás negatív döntés, másnéven elsőfajú hiba), illetve olyan esetben is hibát jelezhetnek, amikor az valójában nem fordul elő (hibás pozitív döntés, másodfajú hiba). Előfordul, hogy nincs lehetőség minden egyes termék minőségét leellenőrizni, ilyen esetben a mintavétel módszerét kell alkalmazni, és bizonyos megbízhatósággal tudunk majd megállapításokat tenni a hiba arányáról.

A különböző alkalmazások esetén a képelemzést végző rendszereknek különféle elvárásoknak kell megfelelni, mint például:

- valós-idejű feldolgozás, ami bizonyos időkritikus esetben valós-idejű operációsrendszer igényel, ahol garantálva van az azonnali beavatkozás lehetősége;
- egyéb perifériák és folyamatok vezérlése, kamerák szinkronizált használata;
- adatgyűjtés, statisztikák készítése és elemzése, kapcsolódás vállalatirányítási rendszerekhez;
- felhasználók, operátorok hozzáféréseinek szabályozása, adatok hitelesítése.

9.5.1. Néhány ipari példa

A következőkben két egyszerűbb ipari feladat megoldását fogjuk tömören ismertetni. Mindkét esetben objektumok számolását kell elvégezni. Az első esetben a problémát a csomagolás eltérő átlátszósága jelenti, míg a második feladatban nem csak számlálást, hanem különböző alakú objektumok felismerését is el kell végezni, s mivel a különböző objektumok mérete eltérő, ezért az első módszer nem alkalmazható.

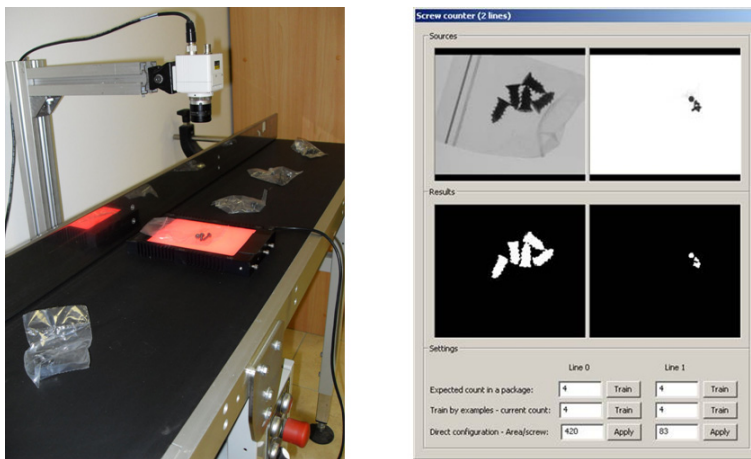
Csavarszámlálás

Gyakori feladat az ipari minőségellenőrzésben, hogy adott alkatrészek számát kell meghatározni - tipikusan háttérvilágítás alkalmazásával. Megfelelő optikai kalibráció esetén ez egyszerű területméréssel kivitelezhető, természetesen ehhez az szükséges, hogy a képpontok színe alapján az objektumok képe jól detektálható legyen. Ez alapvetően a hisztogram küszö-

bölése által lehetséges. Azonban sok esetben nem biztosíthatóak az állandó fényviszonyok, ill. a munkaterületen is lehetnek zavaró hatások, mint pl. az, hogy a termékek különböző átlátszóságú csomagolásban vannak. Az ezek kiküszöbölésére alkalmas adaptív módszer főbb lépései:

- I. élkép meghatározása (pl. Prewitt vagy Sobel konvolúciós operátorral);
- II. vágási küszöb meghatározása ún. együttes-hisztogram számításával:
 - az együttes-hisztogram megmutatja, hogy adott szürkeségi kódú pixelhez milyen átlagos élerősség tartozik;
 - a hisztogram balról első csúcsának megfelelő szürkeségi értéke jelenti a vágási küszöböt;
- III. vágás (binarizálás);
- IV. zajok, túl kicsi bináris foltok eltüntetése (bináris morfológiai szűrés méret alapján);
- V. fekete pontok számlálása.

Ha a rendszert megfelelően kalibráltuk, akkor a fekete pontok számából következtethetünk a csavarok számára.



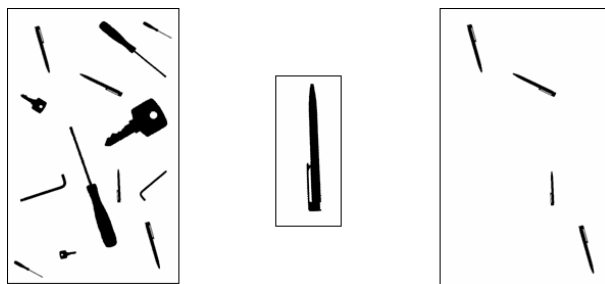
9.9. ábra. Átlátszó csomagokban lévő csavarok számlálása adaptív küszöböléssel

Algoritmus objektumok számlálására és felismerésére

Az összefüggő területek vagy komponensek (*connected components*) számlálására, azonosítására több elvi módszer is létezik. Egy ilyen módszer a rekurzív bejárás módszere, amely során a terület egy kijelölt (tetszőleges) kezdőpontjából indulva a szomszédos tárgy-képpontok vizsgálatával és feljegyzésével haladunk mindaddig, amíg a felderítendő terület minden pontját be nem járjuk. A bejárás során a komponensről térképet (olyan kép, amely csak az éppen vizsgált összefüggő területet tartalmazza) készíthetünk és különböző statisztikákat vezethetünk, mint pl. a befoglaló téglalap vagy a terület (képpontok száma).

Az alaki vizsgálat esetén ügyelni kell arra, hogy az alaki jellemzés ne függjön a komponens elhelyezkedésétől (elfordulásától) vagy a nagyságától. Utóbbira akkor lehet szükség, ha a rendszernek különböző nagyítású optikákkal, vagy különböző kamera-tárgy távolságokkal is utánállítás nélkül kell működnie. Ilyen alaki jellemzést érhetünk el a Hu momentumok mérésével. A Hu momentumok a komponens legfeljebb harmadrendű centrális momentumaiból, zárt képlettel számolható hét érték, amelyek az eltolásra, elforgatásra és skálázásra is invariánsak, ezen felül a tükrözés is jól azonosítható a használatukkal, mert ez esetben a hét érték egyike fog előjelet váltani. Osztályozás céljából a hét értéket vektorként kell összehasonlítani valamilyen vektornorma (pl. maximális abszolút eltérés) használatával. Az eljárás lépései:

- I. Binarizálás (használható az előző pontban leírt módszer)
- II. Bináris objektumok bejárása és felcímkézése rekurzív algoritmussal
- III. A példa objektum és a felcímkézett objektum Hu momentumának számítása
- IV. Hu momentumok összehasonlítása.



9.10. ábra. Bináris objektumok felismerése: bemeneti kép, felismerendő objektum, a felismerés eredménye

10. fejezet

Biztonsági kamerarendszerek

Fejezetünkben áttekintjük, hogy napjainkban mi a videó alapú távfelügyeleti rendszerek szerepe, az elmúlt néhány évtized során milyen fejlődés volt megfigyelhető, és mik az újabb várható fejlesztési irányok. Bemutatjuk a biztonsági kamerarendszerekben történő képfeldolgozás különböző szintjeit, alapvető eszközeit, algoritmusait.

10.1. Alkalmazási területek

A biztonságtechnikában különböző szenzorok szavatolhatják a védendő területek biztonságát: pl. infra mozgásérzékelők, tűz- és füstjelzők, nyitásérzékelők, légnyomásváltozás detektorok, mikrofonok, kamerák; az integrált rendszerekben rendszerint ezek kombinációját alkalmazzák. A videó alapú módszerek – bár áruk nem számít a legalacsonyabbnak – azzal tudnak igazán versenyképesek lenni, hogy a kép és esetleg hang segítségével sokrétű információt tudnak szolgáltatni, és így nagymértékben tudnak hozzájárulni az események felismeréséhez és rekonstrukciójához. Egyedülálló módon segítségükkel lehetséges – akár egyidejűleg – mozgások detekciója, személyek azonosítása, tűz vagy füst detekciója, speciális események felismerése. A kép- és videófeldolgozás, elemzés, mintafelismerés eszközeivel bonyolult, összetett folyamatok lokalizációja is lehetséges.

A megfigyelő kamerarendszerek felhasználási területe egyre széleskörűbb: eleinte leginkább azokon a helyeken alkalmazták őket, ahol fontos volt a biztonsági szint magasán tartása (katonai, állami, kiemelt vagyónvédelmi alkalmazások), napjainkra azonban a katonai, kormányzati és vállalati alkalmazások mellett egyre gyakoribbak – elsősorban a költségek csökkenésének következményeként – a közösségi ill. otthoni (civil) alkalmazások. Felsőrendűen lássunk néhány jelentősebb alkalmazási területet:

- Kereskedelem (üzletek, bankok) és közösségi szervezetek (hivatalok, kórházak, iskolák): személyek, futó személyek, tömeg detekciója, személyek számlálása, bejárt útvonalak felfedezése, sorban álló emberek detekciója, tárgyak eltűnésének vagy elvesztésének detekciója.
- Ipar: folyamatok elemzése, rendhagyó események jelzése, minőség monitorozása, objektumok számlálása, tiltott területekre való behatolás.

- Közlekedés: sávokban haladó gépjárművek számlálása, sebesség becslése, tiltott területeken való mozgás jelzése, rendszámfelismerés, rendhagyó események jelzése.

A videó-felügyeleti rendszerek szerepének felértékelődése több okra vezethető vissza, mint pl.:

- alacsonyabb áron elérhető technológia,
- növekvő funkcionalitás,
- fokozott igény a biztonságra,
- élömunka költségének relatív növekedése,
- tömegtermelési technológiák elterjedése.

Könyvünk szempontjából legfontosabb annak megértése, hogy milyen módon képesek a modern megfigyelő rendszerek a különböző videóelemző feladatok elvégzésére, hogyan váltak a zártláncú televíziós rendszerekből (closed circuit television system - CCTV) automatikus, intelligens felügyeleti rendszerek.

10.2. A rendszerek összetevői és azok fejlődése

Az első videómegfigyelő rendszert Németországban telepítette a Siemens AG a V-2 rakéták indításának megfigyelésére 1942-ben [23]. Néhány évvel később az első kereskedelmi rendszer Vericon néven jelent meg az USA-ban [81]. A korabeli ismertetések szerint ipari folyamatok közeli megfigyelésére ill. orvosi műtétek diákok általi megfigyelésére ajánlották a készítői.

A CCTV rendszerek eleinte analóg kamerákat, analóg jelátvitelt, analóg képmegjelenítést és rögzítést alkalmaztak, napjainkban azonban minden komponens digitális változatban is megvásárolható és számítógépes hálózatok segítségével összetett, kiterjedt hálózatokban összekapcsolhatók. Az analóg technikának természetesen jelentős minőségi és mennyiségi korlátai vannak a digitálissal szemben, az elmúlt évtizedben megfigyelhető funkcionális fejlődés is egyértelműen a teret hódító digitális technika következménye.

A következőkben először a kamerákat majd a rendszer többi elemét (hálózat, képrögzítők, felügyeleti rendszerek, kiegészítő berendezések) mutatjuk be.

10.2.1. Biztonsági kamerák

Napjainkban, a 2010-es évek elején a meglévő biztonsági rendszereknél leggyakrabban analóg, színes vagy fekete-fehér kamerákat alkalmaznak, amelyeknek felbontása jó esetben eléri a PAL vagy NTSC szabvány felső határát, de a világszerte üzemben lévő kamerák nagyon nagy része csak 320×240 -es pixelfelbontásra képes. Természetesen ennek súlyos következményei vannak: amennyiben az alacsony felbontás nem párosul kicsiny látómezővel (azaz nagy nagyítással), a személyek azonosítására igen csekély esély marad. A kisfelbontású kép minőségét tovább fogja rontani a termikus zaj (elsősorban éjszakai felvételek esetén) ill. a

tömörítési eljárás kódolási hibája. (A korábbi rendszereknél az analóg rögzítésnél is minőségi romlás jelentkezett, ma pedig az analóg jeleket digitalizálják és veszteségesen tömörítik a tároláshoz, míg a digitális jelek eleve tömörítve kerülnek továbbításra, bár meg lehet, hogy a tárolás során újratömörítik azokat.) A tömörítési hibával bíró felvételek felbontásának javítására ismertek elméleti módszerek (pl. [73]), a gyakorlatban azonban ezek egyelőre nem terjedtek el széles körben.



10.1. ábra. PTZ (Pan-Tilt-Zoom) kamera, PTZ irányító egység, dome kamerához rádiós kapcsolattal, infra megvilágítóval felszerelt kamera

A rendszerek tervezése, telepítése során gyakorlott szakemberek választják meg az optika látószögét ill. azt, hogy szükség van-e a terület infra megvilágítására. Míg a hagyományos fényképezőgépek esetén a képérzékelő lapka elé infra szűrőt szerelnek (lásd 2.3. ábra), ezzel szemben a biztonságtechnikában ezek a szűrők ki-be kapcsolhatók. Ugyanis nappali fényviszonyok között színi hibát okoz a közeli infra tartomány (amire a fényérzékelő lapkák érzékenyek), míg gyenge megvilágítás mellett – kiegészítő infra lámpák használatakor – az információ legnagyobb része ebben a tartományban gyűjthető. Az első infra kamera megalkotása Tihanyi Kálmán¹ nevéhez fűződik.

Fontos megjegyeznünk, hogy az infravörös (*Infra Red - IR*) tartományt (0,7-1000 mikrométer) több keskenyebb sávra lehet osztani. A felosztás tartományainak határai és elnevezései attól függenek, hogy a technika milyen területéről van szó, így a CIE (International Commission on Illumination), az ISO 20473 szabvány, a csillagászat, a telekommunikáció ill. a szenzorgyártók szerint különböző kifejezéseket és tartományokat különböztetünk meg.

A digitális IR kamerákat durván három fő csoportba lehet besorolni az érzékenységi tartományuk szerint:

- A rövid-hullámú infra kamerák 0.9-1.7 mikron között érzékenyek, ez a látható tartományhoz igen közel van, gyakran használják a közeli infra kifejezést rájuk (Near Infra

¹Tihanyi Kálmán (Üzbég, 1897. április 28. - Budapest, 1947. február 26.) fizikus, villamosmérnök 1929-ben dolgozta ki és szabadalmazta különleges, infravörös sugarakra is érzékeny kameráját, melynek prototípusát repülőokban való felhasználásra az angol Légügyi Minisztérium számára készítette el.

Red - NIR) . Az aktív LED-es infra megvilágítást használó biztonsági kamerák leginkább ide sorolhatóak, bár érzékenyséjük inkább csak a NIR tartomány szélét érinti.

- A közép-hullámú kamerák tipikusan 2-5 mikron között érzékenyek, a légkör elnyelő hatása miatt ebben a tartományban nem adnak túlságosan részletgazdag képet. Tipikusan nagy hőmérsékletek esetében használják ezeket a kamerákat.
- Az egyre népszerűbb hosszú-hullámú kamerák érzékenysége valahol 7-12 mikron között van, ahol a légkör elnyelése minimális. A középhullámú kamerákhoz hasonlóan pontos hőmérsékleti adatokat lehet meghatározni velük akár kicsi, akár nagy hőmérsékleti tartományokról van szó.

A PAL és NTSC szabványok nyújtotta képfelbontás lehetőségeit jelentősen meghaladják napjaink korszerű, nagyfelbontású biztonsági kamerái: 1, 2 de akár 5 *megapixel*es kamerák is elérhetők a piacon. Természetesen a pixelszám növekedésével az adatátvitel korlátai miatt – a modern videótömörítő eljárások ellenére – az időbeli felbontás (frame per second - FPS) nem feltétlenül éri el a PAL (25FPS) vagy az NTSC (30FPS) szabvány által előírt szintet. Megapixel kamerák esetén már nem analóg szabvány szerint küldik át a jeleket a képrögzítő vagy feldolgozó egység felé, hanem USB, FireWire vagy Ethernet csatolót használnak. Praktikus okokból videómegfigyelő kamerarendszereknél szinte kizárólag az utóbbit alkalmazzák, ill. ennek PoE (Power over Ethernet) változatát, ahol a kamera tápellátása is a hálózati kábelon keresztül történik.

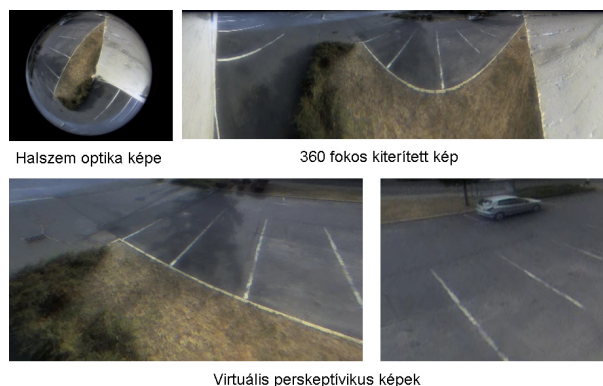
Bár viszonylag korán elérhetőek voltak *széles látószögű, halszem* vagy más típusú *panoráma optikák*, az elégtelen érzékelő felbontás miatt alkalmazásukra csak ritkán került sor. Az érzékelő lapkák felbontásának növekedésével ill. az árak csökkenésével várhatóan nő a szerepük a biztonságtechnikában [19], a közeljövőben felhasználásuknak két esetben lehet számottevő jelentőségük:

- Amikor egy nagyobb terület egyidejű megfigyelésére van szükség, akkor a pásztázó kamerák kiválthatók egy panoráma kamerával, amely az egész területet egyszerre megfigyeli.
- Pásztázó kamerákkal kombinálva elérhető, hogy az egész területet belátó kamerán detektált mozgásokra irányítsuk a pásztázó, nagyítható kamerát, részletgazdag képet kapva a mozgó objektumokról.

Egy épület sarkára felszerelt halszem optika képét láthatjuk a 10.2. ábrán a kép geometriai transzformációja előtt és után.

A kamerák és a megfigyelő rendszer többi komponense közti digitális kommunikációra való áttérés nem jelenti automatikusan a képminőség javulását ill. a funkciók bővülését. *Digitális* biztonsági kamera² esetén csupán arról van szó, hogy nem analóg protokollal kerül a kamera képe a képrögzítőbe, képmegjelenítőbe, hanem digitális kódolás által, digitális tömörítést alkalmazva. Bár egyre jelentősebb a digitális kamerák aránya az értékesítésekben, egyes

²A digitális biztonsági kamerákat gyakran illetik *IP kamera*, *Internet kamera* vagy *hálózati kamera* néven gyártótól függően.



10.2. ábra. Halszem optikával rögzített panoráma kép és transzformált változatai

korai előrejelzések szerint [45] 2011-ben az új kamerák közel harmadát még mindig analóg kamerák fogják kitenni, míg más frissebb piaci felmérések még konzervatívabb képet festenek [39]. Természetesen a már üzemben lévő egységek döntő többsége még hosszú évekig analóg rendszerű lesz. Az analóg-digitális átmenetet könnyítik meg a *hibrid* rendszerek, amikor többféle komponens együtt kerül felhasználásra egy nagyobb rendszerben.

Intelligens kameráknak azokat a digitális kamerákat hívjuk, amelyek képesek a képek valamilyen előfeldolgozására vagy magasabb szintű értelmezésére. Ilyen funkció lehet pl. rendszámfelismerés, arcdetekció vagy például tiltott területen való mozgás jelzése. Magyarországon több cég is készít intelligens biztonsági kamerákat [38].

10.2.2. Egyéb komponensek

Bár a megfigyelő rendszerek minőségének és használhatóságának kulcsfontosságú elemei maguk a kamerák, a képi jel továbbításáért, megjelenítéséért és tárolásáért felelős további eszközök is jelentősen befolyásolják a képi minőséget és a rendszerek funkcionalitását. Napjainkban együttesen vannak jelen az analóg és digitális rendszerkomponensek, röviden ezeket tekintjük át a következőkben:

- Hálózat: Az analóg elektromos jelek átvitelét túlnyomórészt BNC kábelon oldják meg, digitális átvitel esetén Ethernet ill. Power over Ethernet hálózaton továbbítják a kamerák képeit és az egyéb kommunikációs adatokat.
- Digitalizálók: Az analóg jeleket ma már a legritkébb esetben rögzítik analóg tárolóra (azaz videómagnóra), tehát analóg hálózatok esetén is a tárolás előtti fázisban digitális átalakításra van szükség. Sok esetben a tárolást végző digitális tárolóba vagy a videófeldolgozást végző számítógépbe vannak a digitalizáló áramkörök beépítve (lásd 10.3. ábra). Ezek a berendezések általában nem csak az analóg videó jel, hanem hang digitalizálására, kamerák vezérlésére is alkalmasak, ill. egyéb digitális ki- és bemeneti csatornákkal is el vannak látva.
- Videó szerverek: A hibrid működést segítik elő az ún. *videó szerverek*, amelyek analóg fix vagy PTZ (Pan-Tilt-Zoom) kamerák analóg jeleit konvertálják IP hálózaton tovább-

bítható jelekké. Tehát itt már a hálózat egy része biztosan digitális. Nem csak a képek analóg–digitális átalakítása, kódolása a feladatuk, hanem a PTZ kamera vezérlését is lehetővé teszik az Internet segítségével.

- Kvadok, multiplexerek (képátalakítók): A kvadok az analóg képet lekicsinyítik és négy negyed méretű képből egy teljes méretűt raknak össze. Így egy monitor segítségével egyszerre négy kamera képét lehet megjeleníteni. A multiplexerek (időosztásos multiplexelés) több bemenő analóg képből egy analóg videófolyamot állítanak elő, amiben egy-egy időszegmens egy-egy kamera képeit tartalmazza. Mindkét eszköz több analóg kamera képének gazdaságos megjelenítését és tárolását teszi lehetővé az időbeli felbontás vagy a pixelfebontás rovására.
- Videó rögzítők: Az analóg képeket eleinte videómagnókkal (VCR – Video Cassette Recorder) rögzítették, később az analóg jeleket is fogadó, de digitális kódolást alkalmazó digitális videórögzítők (DVR – Digital Video Recorder) jelentek meg. A hálózati videórögzítők (NVR – Network Video Recorder) már IP-n képesek a jeleket fogadni és a DVR-ekhez hasonlóan távolról elérhetők és karbantarthatók.
- Egyéb komponensek: Az integrált biztonsági rendszerekben használhatók még: infra-, tűz-, füst-, gáz-, és nyitás érzékelők, ill. beavatkozást végző eszközök, mint például elektromos zárok, sorompók, vészleállítók, oltóberendezések. Az eseménydetekció megbízhatóságát jelentősen lehet növelni a többféle szenzor egy célból való felhasználásával ún. jelfúzió segítségével.



10.3. ábra. 8 csatornás DVR, 16 csatornás számítógépbe helyezhető PCI digitalizáló kártya és a hozzá való VGA-BNC átalakító

10.3. Biztonsági képek feldolgozása

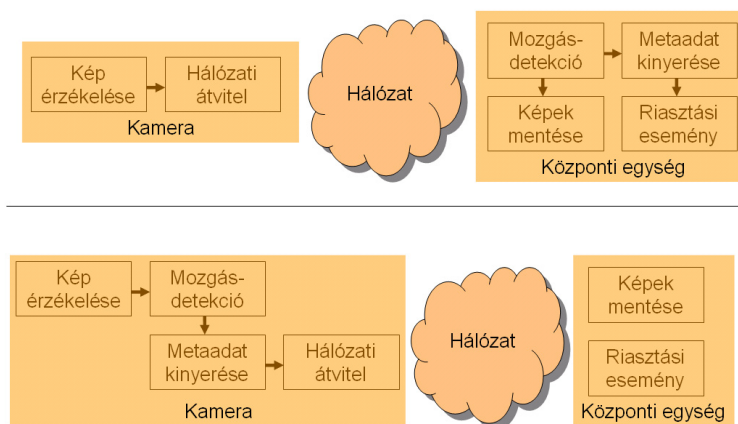
Az első analóg videómegfigyelő rendszerek nem voltak képesek a képek feldolgozására, a feladatuk egy terület távoli megfigyelése és a képek archiválása volt. Azonban a kamerák számának növekedése, a magas élömunka költségek nyilvánvalóvá tették, hogy a képek automatikus feldolgozása a hatékony működés érdekében elengedhetetlen.

Egyes becslések szerint 20 milliót is meghaladja a nyilvános helyeken elhelyezett kamerák száma a világban, csak Nagy-Britanniában több mint 4 millió kamera van üzemben [60]. A sűrűbben lakott területeken – pl. Londonban – könnyen előfordul, hogy naponta 300 esetben

kerül rögzítésre egy személy átlagos, hétköznapi tevékenysége [61].

Ezzel szemben a rengeteg kép megfigyelése lassan megoldhatatlan feladatot ró az operátorok számára. Megfigyelések szerint kb. 20 perc után a képzett megfigyelő személyzet figyelme annyit lankad, hogy igen nagy valószínűséggel nem vesz észre jelentős mozgásaktivitást sem [28].

Mindennek köszönhetően természetessé vált, hogy egyre több képelemző funkciót építenek be a megfigyelő rendszerekbe. Természeteszerűleg a központi egységek intelligenciáját célszerű megnövelni, hiszen így a rendszer többi része változatlan maradhat. Azonban egy bizonyos határon túl – amikor a kamerák száma eléri a több százat vagy ezret – ez a modell egyre nehezkesebbé válik, mivel túlzottan megnő a képátvitel sávszélesség igénye, túl nagy számítási teher hárul a központi számítógépekre. Ezeknek a problémáknak a kiküszöbölését elosztott hálózati architektúrával, elosztott számítási modellel lehet elérni, ahol bizonyos számításokat maguk a kamerák végeznek (lásd 10.4. ábra). Ilyen esetekben nő meg a beágyazott rendszerek, az intelligens kamerák szerepe.



10.4. ábra. Központosított és elosztott képfeldolgozási struktúrák a megfigyelő rendszerekben

A következőkben a megfigyelő rendszerekben alkalmazott képfeldolgozási lépések elméleti megközelítését vázoljuk fel. A gyakorlatban ettől jelentősen eltérhetnek a különböző kereskedelmi rendszerek megoldásai, mivel sokszor speciális, egyedi szempontokat (számítási architektúra, memória korlátok, energia fogyasztás, számítási komplexitás, stb.) kell figyelembe venni. A feldolgozás célja, hogy a képeken megfigyelhető jelentős mozgásokat vagy jelenségeket megtaláljuk, azok jellegét, tulajdonságait elemezzük és ezáltal meta adatokat generáljunk, ami alapján az operátor vagy biztonsági rendszer valamely komponense megfelelő beavatkozást tesz (pl. elindítja a videók rögzítését, riasztást tesz, stb.).



10.5. ábra. A képelemzés tipikus szekvenciális lépései videómegfigyelő rendszerekben

10.3.1. Kamera kalibráció

A teret megfigyelő kamera leképezése egy mátrix-szal egyértelműen reprezentálható: Ha ismert az ún. kamera mátrix, akkor egy tetszőleges 3D-s pont képének koordinátáit elő lehet állítani, illetve egy képpontból visszavetítést is tudunk végezni (egy kamera esetén természetesen bizonyos korlátokkal).

$$\mathbf{x}_i = \mathbf{X}_i P, \quad (10.1)$$

ahol x_i egy pont képernyő koordinátája, X_i pedig a világi pont koordinátája. P kamera mátrix meghatározásához egy túlhatározott egyenletrendszert kell megoldani, amihez minimum 6 pontpár koordinátáinak ismerete szükséges. Ehhez kalibrációs mintát használunk, ahol könnyen detektálhatók biz. karakterisztikus pontok, amelyeknek ismerjük a geometriáját. Amennyiben pontos kalibrációt szeretnénk végezni, a kamera optikájának torzítását is figyelembe kell venni, amire az ún. Gold Standard algoritmus a leggyakrabban használt módszer [32].

Amennyiben kalibrált kamera képeit dolgozzuk fel, lehetőségünk van az objektumok tényleges méretének meghatározására, távolság és sebesség mérésre (lásd [16]), illetve többkamerás rendszerekben könnyen elvégezhető a különböző kamerák képein lévő objektumok egymásnak való megfeleltetése. A gyakorlatban azonban kamera kalibrációt csak igen ritka esetben végeznek biztonsági rendszereknél, mivel költséges műveletről van szó (különösen a kamerák nagy számát tekintve), ill. a kamerák a hosszú üzemidő során nagy eséllyel elállítódnak és így a kalibrációt időnként meg kellene ismételni.

A probléma megoldására különböző automatikus tér-geometria meghatározó módszereket dolgoztak ki [34], illetve olyan robosztus képelemző eljárások alkalmazására törekednek, amelyek nem igénylik kalibrált kamerák alkalmazását.

10.3.2. Változás- és mozgásdetekció

Egy biztonsági kamera képpontjainak értéke számos okból változhat meg anélkül, hogy ténylegesen objektumok (személyek, járművek, állatok) mozognának a képen. Például gyakran előfordul a fényviszonyok megváltozása, mint pl. lámpák ki-be kapcsolása, felhők, árnyékok megjelenése, zajt okozhat a csapadék (eső, hó), ködfoltok, ill. füst megjelenése. Ezek olyan jelenségek, amik zavarják a képen történő objektumok mozgásának detektálását, követését, ezért olyan eljárásokra van szükségünk, amelyek képesek a zavaró tényezőktől függetlenül a lényegi tartalom kiemelésére.

Alapvetően megkülönböztetünk változásdetekciós és mozgásdetekciós eljárásokat: az előbbi célja a lényegi változás megtalálása a képen, míg a mozgásdetekciós módszereknél azt szeretnénk, hogy csak a mozgásból adódó eltéréseket vegyük észre. E kettő jelenség megkülönböztetése nem triviális feladat, egyszerűbb számítási modellek esetén olyan háttérmodellt alkotnak, ami magába foglalja az apróbb változásokat, a nagyobb vagy szokatlanabb eltéréseket pedig a mozgás kategóriájába sorolják. Mozgáselemzésről akkor beszélünk, ha a mozgás tulajdonságairól (pl. irányáról, a mozgásvektorok nagyságáról) is gyűjtünk információt.

Tekintsük át a változásdetekció alapvető módszereit - egycsatornás képekre megfogalmazva. Ezeknél a módszereknél végeredményül egy olyan bináris képet szeretnénk meghatározni,

ahol a változó részek (előtér) és a statikus részek (háttér) vannak megjelölve:

- Két egymást követő kép abszolút különbsége egyszerűen megkapható:

$$d(t) = \|f(t) - f(t-1)\|. \quad (10.2)$$

$d(t)$ küszöbölésével tudjuk eldönteni, hogy adott pozícióban volt-e változás avagy sem:

$$v(t) = \begin{cases} 1 & \text{ha } d(t) > k \\ 0 & \text{ha } d(t) \leq k \end{cases} \quad (10.3)$$

ahol a bináris $v(t)$ maszk értékére k előre meghatározott küszöbölési konstans jelentős hatással van. A módszer előnye, hogy egyszerűen implementálható és gyors eljárás, legnagyobb hibája, hogy az apertúra problémára nagyon érzékeny, hiszen a homogén színű mozgó részekenél nincs jelentős különbség két egymást követő képkocka között és ennél fogva ott nem detektál mozgást.

- Detekció háttérkép tanulása által:

$$d(t) = \|f(t) - h(t)\|, \quad (10.4)$$

ahol

$$h(t) = \alpha f(t) + (1 - \alpha)h(t-1). \quad (10.5)$$

Itt $v(t)$ meghatározására $d(t)$ küszöbölését az előző módszer szerint végezzük, de alapvető különbség, hogy egy időben frissülő $h(t)$ háttérmodelt használunk. α -val tudjuk szabályozni, hogy milyen gyorsan tanulja meg a rendszer a háttérét. Probléma akkor jelentkezik, ha a háttér bizonyos része gyorsabban, mások lassabban változnak...

- A háttér Gauss függvényvel való modellezése azt a célt szolgálja, hogy a kép minden pixelének ingadozását is modellezni tudjuk:

$$h(t) = \mathcal{N}(\mu(t), \sigma(t)^2). \quad (10.6)$$

Természetesen μ várhatóérték és σ szórásnégyzet frissítését folyamatosan végezzük:

$$\begin{aligned} \mu(t) &= \alpha f(t) + (1 - \alpha)\mu(t-1) \\ \sigma(t)^2 &= \alpha(f(t) - \mu(t))^2 + (1 - \alpha)\sigma(t-1)^2. \end{aligned} \quad (10.7)$$

A különbséget a háttérmodell várható értéke és a pixel aktuális értéke alapján számítjuk ki:

$$d(t) = \|f(t) - \mu(t)\|, \quad (10.8)$$

míg a küszöbölésnél figyelembe vesszük a szórásnégyzet értékét is:

$$v(t) = \begin{cases} 1 & \text{ha } d(t) > k\sigma(t) \\ 0 & \text{ha } d(t) \leq k\sigma(t) \end{cases} \quad (10.9)$$

azaz ha az aktuális eltérés a megszokott értéktől nagyobb mint a szokásos szórás k -szorososa, akkor a pixel az előtérhez tartozik.

- Nagyon sok esetben az vehető észre, hogy egy pixel értéke nem egy Gauss függvény várható értéke körül ingadozik, hanem több, jelentősen eltérő értéket is felvehet, és leginkább ezen jellemző értékek körül tapasztalunk ingadozást. Pl. egy hullámzó vízfelület meg-megcsillan, a szél mozgatja egy fának a levelét. Ilyen esetekre az ad megoldást, ha több Gauss függvényből állítjuk elő a "potenciális" háttérmodellt:

$$h(t) = \sum_{k=1}^N \omega_{k,t} \mathcal{N}(\mu_k(t), \sigma_k(t)^2). \quad (10.10)$$

Ebben az esetben ω jelenti az egyes Gauss függvények súlytényezőjét, $\mu_k(t)$, $\sigma_k(t)^2$ frissítését is a korábbiaknak megfelelően végezzük, de csak azokra a függvényekre, amelyek egyeznek az aktuálisan megfigyelt értékkel. *Egyezésről* akkor beszélünk, ha $f(t)$ és $\mu_k(t)$ távolsága kisebb mint $2.5\sigma_k(t)$.

A "tényleges" háttérmodellbe azok a Gauss függvények tartoznak, amelyek adott T esetén az első b valamelyikének megfelelnek:

$$B_t = \arg \min_b \left(\sum_{k=1}^b \omega_{k,t} > T \right), \quad (10.11)$$

Végül pedig egy adott pont akkor tekinthető a háttér részének, ha van rá egyező tényleges háttérmodell. Ellenkező esetben a pontot - mint mozgó objektumot - jelezni kell. Ezt a megközelítést először [79]-ben olvashattuk, azóta számtalan variánsa jött létre.

Az utóbbi módszer már meglehetősen jól szűri a pixelek értékének véletlenszerű ingadozásait, változásait, ezért ezt a módszert már nevezhetjük mozgásdetekciónak. Mindegyik módszer esetében igaz, hogy az eredmény többé-kevésbé zajos lesz, ahhoz, hogy az apró pontokat vagy lyukakat ki tudjuk szűrni, kézenfekvő a bináris morfológia [65] alkalmazása:

- az *izolált pontok törlésével* az önálló, egy pixelből álló objektumok törölhetők;
- az *erózió* alkalmazásával az egy vagy néhány pixelből álló objektumokat tudjuk a bináris maszkról eltávolítani;
- az ún. *zárás* műveletével pedig az apróbb lyukakat tudjuk betömni.

A szakirodalom a fenti módszerek rengeteg variánsát tárgyalja, mivel számtalan olyan probléma van, amire ezek a módszerek sem adnak kielégítő választ. Ilyen például a kitakarások és árnyékok kezelése, a tanulási paraméter beállítása, vagy az előtér apertúra probléma. Ezek ismertetésétől terjedelmi okokból eltekintünk.

Hasonlóan nem tárgyaljuk azokat az eseteket, amikor nem álló kamerák képén, hanem mozgó kamerák (ill. mozgó objektumra, pl. gépjárműre szerelt kamerák) képén kell mozgásdetekciót végezni.

10.3.3. Optikai áramlás számítása

A kamera látómezőjében lévő pontok tényleges mozgását jellemző, a kép 2D-s síkjára vetített mozgásvektorok által alkotott vektormezőt *mozgásmezőnek* nevezzük. A levetítés által

meghatározott 2D-s $\mathbf{u} = (u_v, u_h)$ mozgásvektorok az adott pontok elmozdulását írják le t és $t+1$ időpontok között. A mozgásmező egy ideális vektormezőt jelöl, általában ennek csak egy közelítő becslését tudjuk mérni, mivel a 3D tér nem minden pontjának mozgását tudjuk meghatározni csupán a vetületi képek alapján. A gyakorlatban az ún. *optikai áramlást* tudjuk megmérni. A továbbiakban a mozgásbecslés alapvető feltételezéseit és módszereit tekintjük át. A számítások útján meghatározott optikai áramlás lehet:

- pixel alapú, ekkor a kép minden pixeléhez rendelünk vektort (sűrű mozgásmező),
- blokk alapú, amikor bizonyos méretű blokkokhoz tartozik egy-egy vektor,
- tulajdonság alapú, amikor csak kitüntetett pontokhoz rendelünk mozgásvektorokat
- objektum alapú, amikor a képen látható objektumok területe felett határozzuk meg a vektorokat.

Igen sokféle módszert dolgoztak ki az optikai áramlás becslésére. A módszerek alapvetően az intenzitás-megmaradás elvén alapulnak.

Az intenzitás-megmaradás elve

Ahhoz, hogy a képkockák között megfeleltetést tudjunk tenni, alapvetően azt kell hogy feltételezzük, hogy a 3D-s térben lévő objektumok kamerába vetülő színe két egymást követő felvétel során nem változik, a pontok csupán elmozdulnak kockáról kockára. Ez természetesen csak egy durva közelítése a valóságnak, hiszen akár a környezeti megvilágítást, akár az objektum saját fényét, akár a felületek irányát, akár a takarásokat vagy az objektumok alakjának változását tekintjük, mind sérti szigorú feltételezésünket. Azonban mégis ebből az egyszerű feltételből kiindulva olyan számítási modellek alkothatók, amelyek alkalmasak lesznek a mozgásmező közelítő becslésére. Az *intenzitás megmaradásának elve* tehát így írható fel:

$$f(x, y, t) = f(x+u_v, y+u_h, t+1), \quad (10.12)$$

ahol u_v illetve u_h (az optikai áramlás vízszintes és függőleges komponense) a pontok elmozdulását adja meg t és $t+1$ időpillanatok között.

Az intenzitás-megmaradás elvét felírhatjuk a következő egyenlettel is:

$$f(x(t), y(t), t+1) = c, \quad (10.13)$$

azaz egy c intenzitású pont pozíciója $x(t)$ és $y(t)$ függvények szerint változik az időben, miközben az intenzitás konstans marad. Vegyük az előbbi egyenletnek időbeli teljes deriváltját:

$$0 = \frac{d}{dt} f(x(t), y(t), t+1) = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial t} \frac{dt}{dt}. \quad (10.14)$$

Természetesen az x és y szerint deriváltak megfelelnek az adott pontban jellemző mozgásvektor vízszintes és függőleges komponensének (u_v, u_h). A fenti 10.14. egyenletet *mozgásgradiens megszorításnak* nevezik, és le lehet vezetni másfajta okfejtéssel is, amiről a [48]-ban

olvashatnak.

A mozgásdetekciós algoritmusok a következő kategóriák valamelyikébe sorolhatók:

- blokkegyezés alapú módszerek,
- Horn és Schunck algoritmus,
- Lucas és Kanade módszere,
- tér-időbeli szűrők,
- fázis korrelációs módszerek.

Ezek közül csak a blokkegyezés és a Lucas és Kanade módszeréről írunk röviden (bővebben lásd [48]).

Blokkegyezés alapú algoritmus

A 10.12. egyenletben definiált intenzitás megmaradás elvéből egyenesen következik, hogy két képkocka pontjai közti egyezőséget egyszerűen meghatározhatjuk keresztkorreláció vagy négyzetes különbség alapján:

$$\mathbf{u}(x, y, t) = \arg \max_{u_v, u_h} \sum_{i=-M/2}^{M/2} \sum_{j=-N/2}^{N/2} f(x+i, y+j, t) f(x+i+u_v, y+j+u_h, t+1), \quad (10.15)$$

illetve

$$\mathbf{u}(x, y, t) = \arg \max_{u_v, u_h} \sum_{i=-M/2}^{M/2} \sum_{j=-N/2}^{N/2} (f(x+i, y+j, t) - f(x+i+u_v, y+j+u_h, t+1))^2, \quad (10.16)$$

ahol M és N a keresési ablak méretét jelöli.

Ezeknek a módszereknek előnyük az egyszerű implementáció, hátrányuk pedig az ablakméret, azaz a keresési tér növelésével négyzetesen növekvő számítási komplexitásuk, ill. hogy nehezen viselik a kép különböző torzulásait. Ezt elkerülendő, különböző továbbfejlesztéseit alkalmazzák, lehetővé téve az eljárás invariáns viselkedését az átlagos intenzitás változására, vagy a kép affin transzformációjára. A keresés gyorsítására számos megoldás született a többfelbontású reprezentációtól kezdve különböző iteratív keresési megoldásokig [13].

Lucas és Kanade algoritmus

Lucas és Kanade népszerű módszere (LK) [55] szintén a gradiens megszorításból indul ki, és az előző módszerhez hasonlóan feltételezi, hogy lokálisan sima a keresett vektorfüggvény. A simaság azt jelenti, hogy egy tetszőlegesen választott pont környékén is nagyon hasonló u_v és u_h értékekre számítunk.

\mathbf{u} -t megkaphatjuk:

$$\mathbf{u} = -\mathbf{M}^{-1}\mathbf{b}, \quad (10.17)$$

ahol

$$\mathbf{M} = \begin{bmatrix} \sum G\left(\frac{df}{dx}\right)^2 & \sum G\frac{df}{dx}\frac{df}{dy} \\ \sum G\frac{df}{dx}\frac{df}{dy} & \sum G\left(\frac{df}{dy}\right)^2 \end{bmatrix} \quad (10.18)$$

$$\text{és } \mathbf{b} = \begin{bmatrix} \sum G\frac{df}{dx}\frac{df}{dt} \\ \sum G\frac{df}{dy}\frac{df}{dt} \end{bmatrix}.$$

$G(x, y)$ egy átlagoló függvény, ami tipikusan egy Gauss függvény alakját veszi fel, azaz a középponttól távolodva a távolság függvényében egyre kisebb súlyt alkalmazunk, miközben az összeadást egy adott pont környezetében elvégezzük. Mindehhez természetesen feltételezzük, hogy M invertálható. Belátható továbbá, hogy:

$$\det(M) = \lambda_1 \lambda_2 = 0, \quad (10.19)$$

azaz M nem invertálható, ha egyik vagy mindkét (λ_1 és λ_2) sajátértéke 0. Ennek elkerülése miatt van szükség erős simításra, azaz minél nagyobb területen való átlagolásra.

A Lucas és Kanade algoritmus fő előnye a gyorsasága – a blokkegyezés módszertől eltérően nem igényel kereső eljárást. A módszer fő korlátozását az jelenti, hogy a deriváltak segítségével csak kisebb elmozdulások becsülhetők. Ennek a legegyszerűbb feloldását az algoritmus többfelbontásos alkalmazása jelenti, amikor kép-piramist (lásd 3.4.1) generálunk a képekből, és a piramis különböző szintjeit mind felhasználjuk a becslésre. A piramis magasabb szintjein – ahol a kép kicsinyítve áll elő – egy-egy pixelnyi elmozdulás több pixelnyi távolságnak felel meg az eredeti képen.

Mozgáskövetés

Az optikai áramlás számítása során csupán néhány képkockát használunk fel a mozgás megbecslésére, a legtöbb módszer tipikusan csak 2 képkocka elemzését végzi. Ugyanakkor az eddig ismertetett eljárások alapvetően nyers adatok (blokkok, pixelek) hasonlóságát vizsgálták a képkockák között, nem törekedtek objektum szintű elemzésre. A *mozgáskövetés* feladata objektumok közti kapcsolat megteremtése több képkockán keresztül, azaz segítségükkel lehetséges objektumok útvonalának végigkövetése egy videószekvencián. Azonban a gyakorlatban az objektumok elhatárolása a környezetüktől, körvonaluknak pontos meghatározása sokszor igen nehéz feladat, a különböző követő eljárások nem csak a képpontok színét, hanem sebességét, az ívek alakját és a képrészletek egyéb jellemzőinek paramétereit is beépítik modelljükbe. Alapvetően léteznek *top-down* módszerek, amelyek valamilyen – az objektumra jellemző – template-et, modellt fognak követni (pl. részecske követő) és vannak *bottom-up* megközelítések, amelyek a képet szegmentálják, és a különböző foltok követését végzik, így ezekből építkezve követik a tényleges objektumokat.

A legeredményesebb követő eljárások jelentős része vagy az ún. Kálmán³ szűrőt [47] [18], vagy pedig valamilyen részecske szűrő eljárást használ [62].

³Kálmán Rudolf Emil (Budapest, 1930. május 19.) amerikai-magyar villamosmérnök, matematikus. Munkásságának fő eredményeit a folyamatirányításban, szabályozáselméletben és operációkutatásban érte el az USA-ban. Szűrőssel kapcsolatos elképzeléseit eleinte erős kritika fogadta, ezért azokat először a mechanika területén publikálta, később azonban igen jelentős sikert ért el a villamosmérnöki és informatikai területeken.

10.3.4. Az elemzés magasabb szintű feladatai

Az eddigiekben azokat a legalapvetőbb képelemző feladatokat ismertettük, amelyek segítségével meg tudjuk állapítani, hogy a biztonsági felvétel mely részein volt mozgás és ezt a mozgást milyen mozgásvektorok jellemzik. Így lehetőségünk van pl. riasztást tenni, ha a kép egy meghatározott területén mozgás történik vagy egy adott zónában nem megengedett irányú mozgást tapasztalunk. Ezek után tekintsük azokat eljárásokat, amelyek egyrészt a mozgáselemzés lépéseire, másrészt egyéb képelemző, felismerő algoritmusokra építve növelik a biztonsági rendszerek funkcionalitását.

- Forgalomszámlálás: Amennyiben az utakat felügyelő kamerák képén nem túl jelentős a gépjárművek takarása, akkor az úttestbe épített indukciós forgalomszámláló berendezések könnyen kiválthatóak kamerás rendszerekkel. Ilyen módszert mutat be a [25] cikk.
- Elhagyott objektumok, elmozdult objektumok detekciója: Megfelelő háttérmodellel könnyen fel lehet ismerni, hogy egy adott pozícióban új objektum jelent meg, vagy valami elmozdult. Sok piaci alkalmazásba beépített funkció.
- Tömeg detekciója ill. mozgása: A kereskedelemben elérhető rendszerek egy része már képes tömeg detekciójára, a szakirodalom pedig bőségesen tárgyalja ezt a témát [88].
- Arcdetekció és arcfelismerés: Természetes igény, hogy a videófelveteleken lokalizáljuk az arcokat és lehetőség szerint felismerjük vagy azonosítsuk azokat. A felismerés csak akkor működik hatékonyan, ha nem túl sok arc közül kell kiválasztani a képen látott személyt. Az azonosítás feladata pedig csupán az, hogy leellenőrizzük, vajon a képen látott személy tényleg az-e, akinek mondja magát (pl. egy beléptető rendszerénél). Bővebben lásd [43].
- Rendszámfelismerés: A rendszámfelismerés közeli felvételeken rutinszerű művelet, azonban távoli, zajos, gyenge minőségű képeken a legtöbb kereskedelmi termék felhasználásra alkalmatlan. A napfény zavaró hatásának szűrésére ill. éjszakai használat-hoz sok esetben infra megvilágítást és infra szűrőket használnak.
- Szenzor-fúzió: Ahhoz, hogy egy nagyobb területet tudjunk megfigyelni, ill. hogy növeljük a megfigyelés hatékonyságát több, esetleg különböző típusú szenzor is használható. Kamerákon kívül előfordulhatnak mozgásérzékelők, mikrofonok, infra kamerák, ultrahang szenzorok, lidarok, radarok. A szenzor-fúzió feladata, hogy integrálja a szenzorokból szerezhető információt, és ezáltal több és megbízhatóbb adatunk legyen a megfigyelt objektumról.
- Biometrika integrálása: Feltehetően egyre nagyobb igény jelentkezik az iránt, hogy a videófelvetelek elemzéséből egyre magasabb szintű információt lehessen automatikusan kinyerni. Így az arcdetekción és felismerésen kívül a személyek magasságának, járásmintájának, hangjának elemzése is várhatóan egyre több intelligens rendszer részét fogja képezni.

A függelék

Egy CT képszelet DICOM fejléce

(csoport,elem) VR Leírás (elem hossz): érték

(0008,0000) UL Group 0008 Length (4): 414
(0008,0001) UL Group 0008 Length to End (RET) (4): 525646
(0008,0008) CS Image Type (22): ORIGINAL\PRIMARY\AXIAL
(0008,0016) UI SOP Class UID (26): 1.2.840.10008.5.1.4.1.1.2
(0008,0018) UI SOP Instance UID (50):
1.2.840.113619.2.1.1698.3142413933.2.45.993574425
(0008,0020) DA Study Date (8): 20010626
(0008,0021) DA Series Date (8): 20010626
(0008,0022) DA Acquisition Date (8): 20010626
(0008,0023) DA Image Date (8): 20010626
(0008,0030) TM Study Time (6): 163523
(0008,0031) TM Series Time (6): 163838
(0008,0032) TM Acquisition Time (6): 165333
(0008,0033) TM Image Time (6): 164208
(0008,0050) SH Accession Number (0):
(0008,0060) CS Modality (2): CT
(0008,0070) LO Manufacturer (18): GE MEDICAL SYSTEMS
(0008,0080) LO Institution Name (22): ??????????????????????
(0008,0090) PN Referring Physician's Name (0):
(0008,1010) SH Station Name (8): CT050C0
(0008,1030) LO Study Description (22): RESEARCH CISS PHANTOM
(0008,103e) LO Series Description (0):
(0008,1090) LO Manufacturer's Model Name (16): GENESIS_JUPITER

(0010,0000) UL Group 0010 Length (4): 94
(0010,0010) PN Patient's Name (16): ??????????????????
(0010,0020) LO Patient ID (8): ??????????
(0010,0030) DA Patient's Birth Date (0):
(0010,0040) CS Patient's Sex (2): 0
(0010,1010) AS Patient's Age (4): 000M

(0010,1030) DS Patient's Weight (8): 0.000000
(0010,21b0) LT Additional Patient History (0):
(0018,0000) UL Group 0018 Length (4): 262
(0018,0010) LO Contrast/Bolus Agent (4): NONE
(0018,0050) DS Slice Thickness (8): 3.000000
(0018,0060) DS KVP (4): 120
(0018,0090) DS Data Collection Diameter (10): 480.000000
(0018,1020) LO Software Versions(s) (2): 07
(0018,1100) DS Reconstruction Diameter (14): 480.0000000000
(0018,1110) DS Distance Source to Detector (16): 1099.3100585938
(0018,1111) DS Distance Source to Patient (10): 630.000000
(0018,1120) DS Gantry/Detector Tilt (8): 0.000000
(0018,1130) DS Table Height (10): 203.000000
(0018,1140) CS Rotation Direction (2): CC
(0018,1150) IS Exposure Time (4): 2000
(0018,1151) IS X-ray Tube Current (4): 140
(0018,1160) SH Filter Type (12): BODY FILTER
(0018,1190) DS Focal Spot(s) (10): 0.7 \0.9
(0018,1210) SH Convolution Kernel (4): SOFT
(0018,5100) CS Patient Position (4): HFS

(0020,0000) UL Group 0020 Length (4): 364
(0020,000d) UI Study Instance UID (48):
1.2.840.113619.2.1.1.322986825.674.993564792.356
(0020,000e) UI Series Instance UID (48):
1.2.840.113619.2.1.1.322986825.674.993564792.358
(0020,0010) SH Study ID (4): 1698 (0020,0011) IS Series Number (2): 2
(0020,0012) IS Acquisition Number (2): 45
(0020,0013) IS Image Number (2): 45
(0020,0020) CS Patient Orientation (4): L \P
(0020,0032) DS Image Position (Patient) (36):
-234.500000\ -240.000000\104.000000
(0020,0037) DS Image Orientation (Patient) (54):
1.000000\0.000000\0.000000\0.000000\1.000000\0.000000
(0020,0052) UI Frame of Reference UID (44):
1.2.840.113619.2.1.1698.993573323.1083601920
(0020,0060) CS Laterality (0):
(0020,1040) LO Position Reference Indicator (2): IC
(0020,1041) DS Slice Location (14): 104.0000000000

(0028,0000) UL Group 0028 Length (4): 168
(0028,0002) US Samples per Pixel (2): 1
(0028,0004) CS Photometric Interpretation (12): MONOCHROME2
(0028,0010) US Rows (2): 512

(0028,0011) US Columns (2): 512
(0028,0030) DS Pixel Spacing (26): 0.9375000000\0.9375000000
(0028,0100) US Bits Allocated (2): 16
(0028,0101) US Bits Stored (2): 16
(0028,0102) US High Bit (2): 15
(0028,0103) US Pixel Representation (2): 1
(0028,0106) US Smallest Image Pixel Value (2): 0
(0028,0120) US Pixel Padding Value (2): 32768
(0028,1052) DS Rescale Intercept (6): -1024
(0028,1053) DS Rescale Slope (2): 1

(7fe0,0000) UL Group 7FE0 Length (4): 524296
(7fe0,0010) OX Pixel Data (524288): Data starts at position 1382

Tárgymutató

- L_1 norma, 52
- átlagos négyzetes hiba, 35

- Abbé-féle képlet, 103
- ablakolás, 68, 86
- adatbázis, képi, 47
- aktív kontúr szegmentálás, 91
- aktív távérzékelés, 98
- aliasing hiba, 29
- aluláteresztő szűrő, 42
- alulexponált kép, 26
- Anger-kamera, *lásd* Gamma-kamera
- arcfelismerés, 131
- atlasz regisztráció, 81
- atmoszférikus ablak, 98
- atmoszférikus hatások, 98
- automatikus képregisztráció, 81

- Bag of words, 56
- Bayer szűrő, 20
- belső képjellemző, 79
- biztonsági kamera, 119
- BOLD effektus, 64

- Canny éldetektor, 91
- CCD, 20
- Chamfer-távolság, 84
- ciklotron, *lásd* elektrongyorsító
- CMOS, 20
- Cornelius Lanczos, 30
- csúcs jel-zaj viszony, 35
- CT, *lásd* számítógépes tomográfia

- DICOM formátum, 67, 73
- digitális kép, 66
- disszekció
 - virtuális, 70
- doboz szűrő, 39

- Dublin Core, 48

- egymodalitásos regisztráció, 81
- él hiszterézis, 90
- él-alapú szegmentálás, 88, 90
- előtér, 126
- elektromágneses tartomány, 97
- elektrongyorsító, 64
- elektronmikroszkóp, 104
- elmozdulási mező, 79
- endoszkópia, 58
 - virtuális, 70
- energiaminimalizáló szegmentálás, 91
- erőforráskutató műholdak, 99
- ETM+ kamera, 100
- Euler-Lagrange egyenlet, 92
- EUMETSAT űrprogram, 100

- fáziskontraszt mikroszkópia, 104
- fénymikroszkóp, 103
- feluláteresztő szűrő, 42
- felidézés, 53
- felszín megjelenítés, 69
- felszínillesztés, 82
- ferde orientáció, *lásd* szeletorientáció
- fluoreszcens mikroszkópia, 104
- fMRI, *lásd* funkcionális MRI
- foton emissziós tomográfia, 58, 61, 63
- Fourier transzformáció, 27
- Foveon érzékelő, 20
- frekvencia, 27
- funkcionális MRI, 58, 64
- fuzzy szegmentálás, 89

- Gamma sugárzás, 97
- Gamma-kamera, 58, 63
- Gauss függvény, 39, 126, 130

- Gauss piramis reprezentáció, 86
 GeoEye műholdak, 100
 geometriai transzformáció, 78
 GOES űrprogram, 100
 gradiens (kép), 90
 gradiens vektor folyam, 92
 GVF, *lásd* gradiens vektor folyam

 hálózati kamera, 121
 hálózati videórögzítő, 123
 háttér, 126
 hődiffúzió, 41
 hőkamera, 58
 hőtartomány, 98
 hallucináció, 30
 hasonlósági mérték, 78, 80, 82, 83, 86
 HIS, *lásd* kórházi információs rendszer
 hisztogramkiegyenlítés, 26
 Hounsfield érték, 61
 Hu momentumok, 117
 HVR kamera, 100

 ICP, *lásd* iteratív legközelebbi pont módszer
 idősorozat, 66
 IJK koordináta-rendszer, 67
 IKONOS műhold, 100
 implantátum, 12
 indexstruktúra, 47
 Infra Red - IR tartomány, 120
 intelligens kamera, 113, 122
 intenzitás-megmaradás elve, 128
 intenzitások abszolút különbsége, 86
 intenzitások hasonlóságán alapuló képregisztráció, 85
 intenzitások négyzetes különbsége, 86
 interaktív képregisztráció, 81
 interaktív szegmentálás, 88, 91
 Internet kamera, 121
 IP kamera, 121
 ipari PC, 113
 iteratív legközelebbi pont módszer, 85

 jel-zaj viszony, 35
 JPEG2000, 49

 kölcsönös információtartalom, 87

 köríves pásztázó rendszer, 99
 követés, 130
 külső képjellemző, 79
 küszöbölés, 88
 Kálmán Rudolf Emil, 130
 Kálmán szűrő, 130
 kép-piramis, 30, 130
 képfúzió, 64, 76, 77
 képi hibák, 33
 képillesztés, 76
 képjellemzők tere, 78, 79
 képpont, 66
 képregisztráció, 76
 atlasz, 81
 automatikus, 81
 felszín-alapú, 82
 intenzitáshasonlóság, 85
 interaktív, 81
 kalapot a fejre, 84
 kontúr-alapú, 82
 manuális, 81
 pont-alapú, 82, 85
 távolságtérkép-alapú, 84
 képszűrés, 32
 kórházi információs rendszer, 72
 kalapot a fejre algoritmus, 84
 kalibráció, 115, 125
 kamerarendszer, 118
 katadioptrikus, 17
 keresési stratégia, 78, 81, 86
 keresési tér, 78
 kereszt korreláció, 86
 keresztsávós pásztázás, 99
 kollimátor, 112
 kolonoszkópia
 virtuális, 70
 kontúrillesztés, 82
 kontrasztérzékenységi függvény, 36
 konvolúció, 28
 konvolúciós maszk, 90
 konvolúciós szűrő, 39
 koordináta-rendszer
 IJK, 67
 RAS, 67, 68

- WLD, 67
- korai látás, 9
- koronális orientáció, *lásd* szeletorientáció
- korrelációs együttható, 86
- kvad, 123
- kvantálás, 24
- kvantum-hatékonyság, 11
- látható fény tartomány, 97
- LabVIEW, 114
- Lanczos szűrő, 30
- LANDSAT űrprogram, 99
- Laplace operátor, 90
- level-set szegmentálás, 92
- livewire szegmentálás, 91
- LoG operátor, 91
- Lucy-Richardson algoritmus, 39
- lyukkamera, 13
- mágneses rezonancia képalkotás, 58, 63–65, 84, 86
- médián szűrő, 41
- manuális képregisztráció, 81
- manuális szegmentálás, 88
- metaadat, 47
- meteorológiai műholdak, 100
- MI, *lásd* kölcsönös információtartalom
- mikrohullám, 98
- mikroszkóp, 103
- mintavételezés, 24, 65
- MIP vizualizáció, 70
- modell-alapú szegmentálás, 92
- Moiré hatás, 29
- morfológia, 127
- mozgás-gradiens megszorítás, 128
- mozgásdetekció, 125, 129
- mozgásmező, 127
- mozgásvektor, 128
- MPEG-7, 50
- MRI, *lásd* mágneses rezonancia képalkotás
- MSS kamera, 100
- műholdas programok, 99
- multidetektoros CT, 61
- multimodális regisztráció, 79, 86
- multiplexer, 123
- multispektrális detektor, 99
- Near Infra Red - NIR tartomány, 98, 121
- Nobel-díj, 57, 58, 104
- nukleáris medicina, 58, 63, 65
- Nyquist kritérium, 28
- Nyquist, Harry, 28
- optikai áramlás, 127
- optikai átviteli függvény, 39
- optikai alapegyenlet, 15
- optikai hibák, 18
- optikai hullámhossz-tartomány, 98
- orvosi anatómiai képalkotás, 59, 61, 63, 77
- orvosi emissziós képalkotás, 59
- orvosi funkcionális képalkotás, 58, 59, 63, 77
- orvosi képarchiváló és -továbbító rendszer, 59, 72
- orvosi transzmissziós képalkotás, 59
- outlier szűrő, 41
- pánkromatikus szenzor, 99
- pásztázó lézer konfokális mikroszkópia, 104
- pásztázó rendszer, 99
- pásztázószondás mikroszkóp, 104
- PACS, *lásd* orvosi képarchiváló és -továbbító rendszer
- particionált intenzitás uniformitás, 86
- passzív távérzékelés, 98
- PDM, *lásd* pont-eloszlási modell
- PenTile, 24
- perfúziós vizsgálat, 64
- PET, *lásd* pozitron emissziós tomográfia
- PET/CT, 58, 64
- PIU, *lásd* particionált intenzitás uniformitás
- pixel, 66
- pont-alapú regisztráció, 82, 85, 92
- pont-eloszlási modell, 92
- pontosság, 53
- pontszóródási függvény, 37
- pozitron emissziós tomográfia, 58, 61, 64, 84, 86
- R,G,B értékek, 11
- Röntgen, 57, 59, 88, 97
- rádióhullám, 98

- régió-alapú szegmentálás, 88
 régiönövelés, 89
 részecske szűrő, részecske követő, 130
 radar, 98, 101
 radiofarmakon, 58, 63, 64
 radioizotóp, *lásd* radiofarmakon
 Radon, 57, 60
 RAS koordináta-rendszer, 67, 68
 raszeteres kép, 46
 RBV kamera, 100
 rekonstrukció, 57, 58, 60, 61, 64
 rendszámfelismerés, 131
- sávkorlátolt jel, 28
 sávmenti pásztázó technika, 99
 sávszűrő, 42
 SAD, *lásd* intenzitások abszolút különbsége
 sagittális orientáció, *lásd* szeletorientáció
 sarokpont detektálás, 82, 92
 Shannon, Claude, 23
 Shepp-Logan szűrő, 61
 sinc függvény, 29
 snake, *lásd* aktív kontúr szegmentálás
 SPECT, *lásd* foton emissziós tomográfia
 spektrális reflektancia, 98
 spektrális reflektancia görbe, 98, 101
 spektrométer, 99
 spektrumátfedési hiba, 29
 spirál CT, 61
 SPOT űrprogram, 100
 SRTM modell, 101
 SSD, *lásd* intenzitások négyzetes különbsége
 színmélység, 24
 színskálás megjelenítés, 68, 105
 számítógépes tomográfia, 57, 58, 61, 65, 66, 68, 73, 88
 számítógéppel segített műtéttervezés, 62, 76, 78, 79
 szűrt visszavetítés, 61
 szeletorientáció, 68
 szelettávolság, 66
 szeletvastagság, 66
 szemantikai rés, 46
 szenzor-fúzió, 131
- színogram, 61
 sztereo képpár, 100
 sztereo képpárok, 101
 szubpixel, 24
 szuperfelbontás, 30
- többmodalitásos regisztráció, 81
 többszeletes rekonstrukció (vizualizáció), 69
 távolságtérkép-alapú képregisztráció, 84
 térfogat megjelenítés, 69
 túlexponált kép, 26
 Tihanyi Kálmán, 120
 TIROS űrprogram, 100
 TM kamera, 100
 transzaxiális orientáció, *lásd* szeletorientáció
 transzfer függvény, 70
- ultrahang képalkotás, 58, 66
 unitér transzformáció, 26
 UV sugárzás, 97
- vízválasztó szegmentálás, 89
 változásdetekció, 125
 valós-idejű feldolgozás, 115
 vektoros kép, 46
 voxel, 66
- WLD koordináta-rendszer, 67
 WordNet, 55

Irodalomjegyzék

- [1] Sz. V. Fomin A. N. Kolmogorov: *A függvényelmélet és a funkcionálanalízis elemei*. 1981, Műszaki Könyvkiadó.
- [2] National Electrical Manufacturers Association: Digital imaging and communications in medicine (DICOM), 2004. <http://dicom.nema.org/>.
- [3] I. Bankman: *Handbook of Medical Image Processing and Analysis*. 2008, Academic Press.
- [4] P. J. Besl–N. D. McKay: A method for registration of 3D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14. évf. (1992), 239–256. p.
- [5] G. Borgefors: An improved version of the chamfer matching algorithm. In *Proceedings of the 7th International Conference on Pattern Recognition* (konferenciaanyag), 2. köt. 1984, 1175–1177. p.
- [6] György Ábrahám: *Optika*. 1998, Pánem.
- [7] M. Bro-Nielsen: *Medical Image Registration and Surgery Simulation*. PhD értekezés (Informatics and Mathematical Modelling, Technical University of Denmark, DTU). Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 1996.
URL <http://www.imm.dtu.dk/~bro/phd.html>. IMM-PHD-1996-25.
- [8] L. G. Brown: A survey of image registration techniques. *ACM Computing Surveys*, 24. évf. (1992) 4. sz., 325–376. p.
- [9] BT.500: Methodology for the subjective assessment of the quality of television pictures. <http://www.itu.int/rec/R-REC-BT.500/en> megtekintve: 2011. május 1.
- [10] P. J. Burt–E. H. Adelson: The laplacian pyramid as a compact code. *IEEE Transactions on Communications*, 31. évf. (1983), 532–540. p.
- [11] S. Butterworth: On the theory of filter amplifiers. *Wireless Engineer*, 7. évf. (1930), 536–541. p.
- [12] T.F. Chan–L.A. Vese: Active contours without edges. *IEEE Transactions on Image Processing*, 10. évf. (2001) 2. sz., 266–277. p.

- [13] K.W. Cheng – S.C. Chan: Fast block matching algorithms for motion estimation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (konferenciaanyag), 4. köt. 1996, 2311–2314. p.
- [14] A. Collignon – F. Maes – D. Delaere – D. Vandermeulen – P. Suetens – G. Marchal: Automated multi-modality image registration based on information theory. In *Proceedings of Information Processing in Medical Imaging* (konferenciaanyag). 1995, 263–274. p.
- [15] T.F. Cootes – D. Cooper – C.J. Taylor – J. Graham: A trainable method of parametric shape description. *Image and Vision Computing*, 10. évf. (1992), 289–294. p.
- [16] Antonio Criminisi: Single-view metrology: Algorithms and applications. In *DAGM 2002 Symposium* (konferenciaanyag). 2002.
- [17] Gergely Császár – László Czúni: Visszacatolás alapú képi indexelés. In *Képfeldolgozók és Alakfelismerők IV Konferenciája* (konferenciaanyag). 2004, 29–34. p.
- [18] E. Cuevas – D. Zaldivar – R. Rojas: Kalman filter for vision tracking. *Measurement*, 2005. Auguszt., 1–18. p.
- [19] L. Czúni – B. Vágvolgyi – T. Szirányi – T. Greguss: A compact panoramic visual sensor for intelligent applications. In *Proceedings of the 4th Asian Conference on Computer Vision (ACCV2000)* (konferenciaanyag). Taiwan, 2000, 258–263. p.
- [20] László Czúni – Gergely Császár – Dae-Sung Cho – Hyun Mun Kim: New algorithms for example-based super-resolution. In *Proc. of the 11th International Conference on Computer Analysis of Images and Patterns, Lecture Notes in Computer Science, Springer Verlag* (konferenciaanyag). Paris, 2005.
- [21] Chris Dance – Jutta Willamowski – Lixin Fan – Cedric Bray – Gabriela Csurka: Visual categorization with bags of keypoints. In *Proceedings of ECCV* (konferenciaanyag). 2004.
- [22] DCTune 2.0: Perceptual optimization of JPEG images (and perceptual error metric). <http://vision.arc.nasa.gov/dctune/> meglekintve: 2011. május 1.
- [23] Walter Dornberger: *V-2*. 1954, Ballantine Books, 14. p.
- [24] P. Duhamel – M. Vetterli: Fast Fourier transforms: a tutorial review and a state of the art. *Signal Processing*, 19. évf. (1990), 259–299. p.
- [25] L. Eikvil – R.B. Huseby: Traffic surveillance in real-time using Hidden Markov Models. 2001, O–Tu3B. p.
- [26] GIMP - GNU Image Manipulation Program. <http://www.gimp.org> meglekintve: 2011. május 1.
- [27] A.A. Goshtasby: *2-D and 3-D Image Registration*. New York, NY, 2005, John Wiley & Sons, Inc.

- [28] M. W. Green: The appropriate and effective use of security technologies in U.S. schools. a guide for schools and law enforcement agencies. *Technical report, Sandia National Labs*, 1999.
- [29] Rudolph J. Guttosch: Investigation of color aliasing of high spatial frequencies and edges for bayer-pattern sensors and Foveon X3 direct image sensors. http://www.foveon.com/files/Color_Alias_White_Paper_FinalHiRes.pdf megtekintve: 2011. május 1.
- [30] J.V. Hajnal – D.L.G. Hill – D.J. Hawkes: *Medical Image Registration*. Reading, Massachusetts, 2001, CRC Press.
- [31] T. Hartkens – K. Rohr – H.S. Stiehl: Evaluation of 3d operators for the detection of anatomical point landmarks in mr and ct images. *Computer Vision and Image Understanding*, 86. évf. (2002), 118–136. p.
- [32] R. Hartley – A. Zisserman: *Multiple View Geometry in Computer Vision*. 2003, Cambridge University Press.
- [33] H. K. Hartline – H. G. Wagner – F. Ratliff: Inhibition in the eye of Limulus. *Journal of General Physiology*, 39. évf. (1956) 5. sz., 651–673. p.
- [34] László Havasi – Zoltán Szilávik: Using location and motion statistics for the localization of moving objects in multiple camera surveillance videos. In *Proceedings of the IEEE International Workshop on Visual Surveillance (ICCV)* (konferenciaanyag). 2009.
- [35] D.L.G. Hill – P.G. Batchelor – M. Holden – D.J. Hawkes: Medical image registration. *Physics in Medicine and Biology*, 46. évf. (2001. March) 3. sz., R1–R45. p.
- [36] How much information? 2003. A University of California, Berkeley felmérése. <http://www.sims.berkeley.edu/research/projects/how-much-info-2003/> megtekintve: 2011. május 1.
- [37] H.K. Huang – O. Ratib – A.R. Bakker – G. Witte: *Picture Archiving and Communication Systems (PACS) in Medicine*. Berlin Heidelberg, 1990, Springer-Verlag.
- [38] Intellio Kft. <http://www.intellio.eu> megtekintve: 2011. május 1.
- [39] IP/Networked video surveillance market: Equipment, technology and semiconductors, 2008. April. Report.
- [40] ISO/IEC JTC1/SC29/WG11 (2004-10). MPEG-7 Overview (version 10). <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm> megtekintve: 2011. május 1.
- [41] Kótai István: *A mikroszkóp használata*. Budapest, 1979, Natura.

- [42] Huang J. – Kumar S.R. – Mitra M. – Zhu W.J.: Image indexing using color correlograms. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (konferenciaanyag). San Juan, Puerto Rico, 1997, 762–768. p.
- [43] Anil K. Jain – Ruud Bolle – Sharath Pankanti: *Biometrics: Personal Identification in Networked Society*. 2005, Springer.
- [44] Lóki József: *Távérzékelés*. 2002, Debreceni Egyetem.
- [45] JP Freeman: Surveillance & security market trend, 2003. Report.
- [46] JPEG2000. <http://www.jpeg.org/jpeg2000/> meglekintve: 2011. május 1.
- [47] Kalman – Rudolph Emil: A new approach to linear filtering and prediction problems. *Transactions of the ASME – Journal of Basic Engineering*, 82. évf. (1960) Series D. sz., 34–45. p.
- [48] Zoltán Kató – László Czúni: *Számítógépes látás*. 2011, Typotex.
- [49] Bruno P. Kremmer: *A mikroszkóp használata*. 2008, Geobook.
- [50] Andrea Kutics – Akihiko Nakagawa: Semantic browsing and retrieval in image libraries. In *Proceedings of ICIAR, LNCS* (konferenciaanyag). 2004, 737–744. p.
- [51] K. Kyuma – E. Lange – J. Ohta – A. Hermanns – B. Banish – M. Oita: Artificial retinas – fast, versatile image processors. *Nature*, 372. évf. (1994) 197. sz., 259–299. p.
- [52] LabVIEW. <http://www.ni.com/labview/> meglekintve: 2011. május 1.
- [53] Mucsi László: *Műholdas távérzékelés*. Szeged, 2004, Libellus.
- [54] S. Lavallée: Registration for computer-integrated surgery: methodology, state of the art. In R. H. Taylor – S. Lavallée – G. C. Burdea – R. Mösges (szerk.): *Computer-integrated surgery, Technology and clinical applications*. 5 fejezet. 1996, MIT Press, Cambridge, MA, 77–97. p.
- [55] B. D. Lucas – T. Kanade: An iterative image registration technique with an application to stereo vision. In *Proceedings of Imaging Understanding Workshop* (konferenciaanyag). 1981, 121–130. p.
- [56] Magyar Bionikus Látásközpont. <http://lataskozpont.itk.ppke.hu/> meglekintve: 2011. május 1.
- [57] J. B. A. Maintz – M. A. Viergever: A survey of medical image registration. *Medical Image Analysis*, 2. évf. (1998) 1. sz., 1–36. p.
- [58] Matlab. <http://www.mathworks.com/> meglekintve: 2011. május 1.

- [59] C. R. Maurer–J. M. Fitzpatrick: A review of medical image registration. In R. J. Maciunas (szerk.): *Interactive image-guided neurosurgery*. Park Ridge, IL, 1993, American Association of Neurological Surgeons.
- [60] M. McCahill–C. Norris: *Estimating the extent, sophistication and legality of CCTV in London*. 2003, Perpetuity Press.
- [61] C. Norris–G. Armstrong: *The Maximum Surveillance Society: The Rise of Closed Circuit Television*. 1999, Oxford: Berg Publishers.
- [62] K. Nummiaro–E. Koller-Meier–L.V. Gool: An adaptive color-based particle filter. *Image and Vision Computing*, 21. évf. (2003), 99–110. p.
- [63] Timo Ojala–Markus Aittola–Esa Matinmikko: Empirical evaluation of MPEG-7 XM color descriptors in content-based retrieval of semantic image categories. In *Proceedings of the 16th International Conference on Pattern Recognition* (konferenciaanyag), 2. köt. 2002, 1021–1024. p.
- [64] OpenCV. <http://opencv.willowgarage.com> megtekintve: 2011. május 1.
- [65] Kálmán Palágyi: *Képfeldolgozás haladóknak*. 2011, Typotex.
- [66] C. A. Pelizzari–G. T. Chen–D. R. Spelbring–R. R. Weichselbaum–C. T. Chen: Accurate three-dimensional registration of CT, PET, and/or MR images of the brain. *Journal of Computer Assisted Tomography*, 13. évf. (1989) 1. sz., 20–26. p.
- [67] Pietro Perona–Jitendra Malik: Scale-space and edge detection using anisotropic diffusion. In *Proceedings of IEEE Computer Society Workshop on Computer Vision* (konferenciaanyag). 1987. November, 16–22. p.
- [68] William K. Pratt: *Digital Image Processing*. 2007, Wiley-Interscience.
- [69] W.H. Press–S.A. Teukolsky–W.T. Vetterling–B.P. Flannery: *Numerical Recipes in C: The Art of Scientific Computing*. 2nd. kiad. New York, NY, 1992, Cambridge University Press.
- [70] *Resolution from Contrast Modulation*. 2001, Video Electronics Standards Association (VESA), Display Metrology Committee, 76–77. p.
- [71] T. Roska–L. Chua: The CNN Universal Machine: An analogic array computer. *IEEE Trans. on Circuits and Systems-II*, 40. évf. (1993) 3. sz., 163–172. p.
- [72] János Schanda: *Szintervezés számítógépes felhasználás számára*. 2011, Typotex.
- [73] C. Andrew Segall–Rafael Molina–Aggelos K. Katsaggelos: High-resolution images from low-resolution compressed video. *Signal Processing Magazine*, 20. évf. (2003), 37–48. p.
- [74] C. E. Shannon: Communication in the presence of noise. In *Proc. Institute of Radio Engineers* (konferenciaanyag), 1. köt. 1949. January, 10–21. p.

- [75] L. A. Shepp–Y. Vardi: Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1. évf. (1982), 113–122. p.
- [76] Arnold W. M. Smeulders–Marcel Worring–Simone Santini–Amarnath Gupta–Ramesh Jain: Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22. évf. (2000. December), 1349–1380. p. ISSN 0162-8828. URL <http://portal.acm.org/citation.cfm?id=357871.357873>. 32 p.
- [77] M. Sonka–V. Hlavac–R. Boyle: *Image Processing: Analysis and Machine Vision*. 2007, CL-Engineering.
- [78] STATS. <http://www.stats.com/> megletekintve: 2011. május 1.
- [79] C. Stauffer–W. E. L. Grimson: Adaptive background mixture models for real-time tracking. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (konferenciaanyag), 2. köt. 1999, 246–252. p.
- [80] C. Studholme–D. L. G. Hill–D. J. Hawkes: An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32. évf. (1999. Jan.) 1. sz., 71–86. p.
- [81] Television rides wires. *Popular Science*, 1949. February., 179. p.
- [82] Video Quality Experts Group. <http://www.vqeg.org> megletekintve: 2011. május 1.
- [83] P. Viola–W. M. Wells III: Alignment by maximization of mutual information. In *Proceedings of IEEE International Conference on Computer Vision* (konferenciaanyag). Los Alamitos, CA, 1995. Jun., 16–23. p.
- [84] W.M. Wells–P. Viola–H. Atsumi–S. Nakajima–R. Kikinis: Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1. évf. (1996) 1. sz., 35–51. p.
- [85] Norbert Wiener: *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. 1949, Wiley.
- [86] R.P. Woods–J.C. Mazziotta–S.R. Cherry: MRI–PET registration with automated algorithm. *Journal of Computer Assisted Tomography*, 17. évf. (2003) 4. sz., 536–546. p.
- [87] Wordnet. A lexical database for English. <http://wordnet.princeton.edu/> megletekintve: 2011. május 1.
- [88] Beibei Zhan–Dorothy N. Monekosso–Paolo Remagnino–Sergio A. Velastin–Li-Qun Xu: Crowd analysis: a survey. *Machine Vision and Applications*, 19. évf. (2008) 5-6. sz., 345–357. p.