

Erősen ajánlott az előadásdiákkal, órai ábrákkal együtt olvasni!

I. A 2D KÉPFELDOLGOZÁS ALAPJAI

1. Számítógépes látás alapok

1. Ismertesse a kép fogalmát: mire vagyunk kíváncsiak? Mit nevezünk képnek? Milyen képleírási lehetőségek állnak rendelkezésünkre? Mit jelent a videofolyam?

Kép: Vizuális információkat tartalmazó összefüggő adathalmaz.

Leírás általában **pixel/voxel** tömbként történik. 2 (pixel) vagy 3 (voxel, *volume pixel*) dimenziós képpontokat használunk, vonalkamera esetén 1 dimenziósakat.

Egyéb leírási formák:

- Láncok (pl. RLE)
- Topológiai leírások (gráfok)
- Relációs struktúrák
- Hierarchikus leírások

Videofolyam: Tipikusan képek sorozata, mozgásleírás (pl. MPEG4 tömörítés) címesemények sorozata. Általában ezt dolgozzuk fel, van (*a priori*) ismeretünk az előző képről, változást elemezzük.

2. Ismertesse a fényérzékelés fontosabb eszközeit. Részletezze a CCD és a CMOS érzékelők működését, típusait, előnyeit és hátrányait.

Fényérzékelés eszközei:

- Fotodióda
 - CCD
 - CMOS (Főként APS)
- Fotoellenállás
- Kvantumeszközök
- Kémiai érzékelők
- Hőmérsékletmérésen alapuló eszközök

CCD

1969-ben fejlesztették ki a Bell Labsnál, analóg léptetőregiszternek. Működésének lényege, hogy adott órajel hatására a bemeneti oldalon levő töltést mozgásra lehet bírni, azaz léptetni lehet a kimeneti oldal felé. Gyorsan kiderült, 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 fent leírt eszköz hivatalos neve a Full-Frame Transfer (FFT) CCD. Gyakran a sebesség növelése érdekében nem egy kiolvasási vonalat (kiolvasó regisztert) használnak, hanem többet, ennek neve a Frame Transfer CCD. Ezeknél az érzékelőknél nincs lehetőség elektronikus zár alkalmazására. A másik, elterjedtebb CCD típus az interline CCD. A fő különbség a kettő közt, hogy a Full-Frame Transfer CCD nem tud addig újabb képet készíteni, amíg a kiolvasás végre nem hajtodik, ami azért időbe telik. Ez egy tükörreflexes, élőkép nélküli gépnél még nem probléma, de mi van akkor, ha folyamatos képet szeretnénk látni, pl. kompakt fényképezőgépeknél vagy videokameráknál? Ilyenkor jön az interline CCD a képbe.

Összefoglalva:

- Fullframe-Transfer CCD: pixelenként olvasás (külső zár szükséges, amíg shiftelgetünk)
- Frame-transfer CCD: külön fényérzékelő és CCD réteg (drága), fotoérzékelők egyszerre dobják át a CCD-re
- Interline CCD: egy Si-réteg, CCD a fotoaktív oszlopok között (olcsóbb, de rosszabb felbontás)

Előnye

- Nagy érzékenység
- Kevés zaj
- Interline felépítés esetén elektronikus zár megvalósítható (ezt ki is használták pl. a Nikon D70-nél is, ahol 1/500s-tól már nem a mechanikus, hanem az elektronikus zár működik).

Hátránya

- Bonyolult előállítás, emiatt drága
- A kiolvasási elektronikának több kiolvasási csatorna esetén tökéletesen megegyezőnek kell lennie, egyébként sávósodás lép fel (banding)
- Könnyen létre jöhet az ún. Blooming effektus: ha egy elektródán túl sok töltés halmozódik fel, egyszerűen átfolyik a mellette levő elektróda területére (ez ellen számos anti-blooming eljárás létezik, általában a CCD-k adatlapján szerepel ennek a hatékonysága)
- Magas fogyasztás, emiatt nagyobb melegedés (és nagyobb termikus zaj)

CMOS

A CMOS-nál a megnevezés csak és kizárólag a gyártástechnológiát jelöli. Ahogy a leírásból is látszik, az első esetben aktív pixelekről beszélünk, azaz minden egyes képpont saját erősítővel rendelkezik (töltés-feszültség konverter, amely a CCD-nél a kiolvasó egységben volt megtalálható). A kiolvasás mátrix elven történik, minden képpontot külön-külön meg lehet címezni. Mivel minden egyes pixel külön címezhető, illetve a vezetékhálózaton keresztül bármilyen sorrendben összekapcsolható a chip további erősítőfokozataival, nagyon könnyű a kép egy részletét vagy éppen egy alacsonyabb felbontású képet is kiolvasni. A CMOS érzékelőre nagyon könnyű integrálni egyéb áramköri elemeket. Általában a CMOS érzékelők tartalmazzák az analóg-digitális átalakítókat is (míg ez a CCD-nél külön áramkör volt), sőt egyes esetben elő-feldolgozást is végeznek (szenzor szintű zajszűrés például). A nagyobb integrálásnak köszönhetően alacsonyabb az előállítási költségük, mint CCD társaiknak.

Hátrányok:

- Nagyobb zaj: a pixelek egyedi erősítőit nem lehet pontosan beállítani, ezért ezek extra zajt adnak a képhez (pix-pattern noise). Erre a problémára a Canon talált tökéletes megoldást hardver szinten (és lassan minden gyártó alkalmaz hasonló megoldásokat),
- Interferencia érzékenység: a nagyszámú aktív elem sokkal érzékenyebb a környezetből érkező elektromágneses zavarokra, mint a CCD,
- Az aktív elemek csökkentik az érzékelő hatásos méretét (mint az Interline CCD-knél), de itt is segítséget nyújtanak a mikrolencsék,
- Az elektronikus zár nem, vagy nehezen valósítható meg. Jól látható a CMOS érzékelőkkel készült videofelvételeken ennek a hatása: a kiolvasási sebesség miatt jól érzékelhetően elcsúszik a kép (nem azonos időpillanatban történik a teljes kép kiolvasása meg, mint a CCD-knél), és ha pl. egy mozgó autót fényképezünk, akkor az eredetileg kb. téglalatest forma szétescsúszik paralelogrammává.

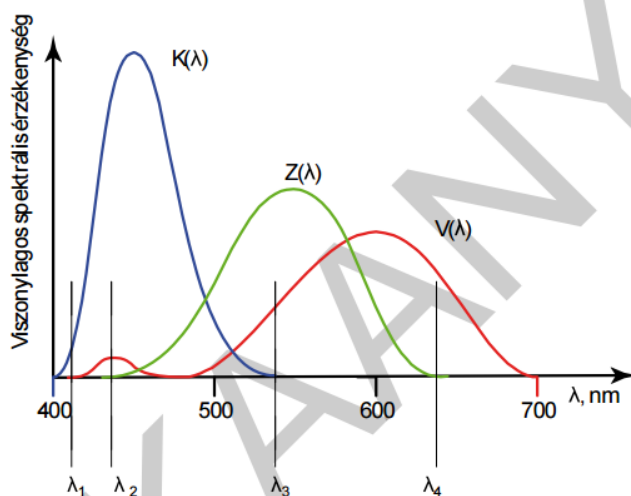
Előnyök:

- A nagy integrálhatóság miatt alacsony ár,
- Kis fogyasztás, kisebb hőtermelés (alacsonyabb termikus zaj).
- Kiseb késleltetés
- Nagyobb sebesség

3. Ismertesse a színlátás alapjait. Milyen módszereket használunk többmódusú (színes) érzékelésre? Mit jelent a Színhőmérséklet?

Az idegrendszer a retinán található különböző típusú csapok fényre adott válaszainak együttes eredményeként különíti el a színeket, így okozva a színérzetet az élőlények számára. A csapok a látható fénytartomány bizonyos szeleteire érzékenyek, viszont csak a beérkező fény mennyiségéről adnak információt az idegrendszernek, a beérkező fény hullámhosszáról nem. Az emberek számára a látható színtartományt hozzávetőlegesen a 380 - 740 nm hullámhosszú elektromágneses sugárzás jelenti. Ezt a színtartományt az emberi szem három különböző típusú csappal fedi le, más fajoknál mind a látható színtartomány, mind a csapok száma eltérő. Emberi szem: 3 szín szenzor + 1 fényerősség szenzor.

A színérzékelők sötétben rosszul működnek, ilyen fontosak a fényerősség érzékelők. Periférián alig látunk színt, az agy kompenzál, ahogy a vakfoltnál is. A számítógépes rendszerek erre még nem képesek.

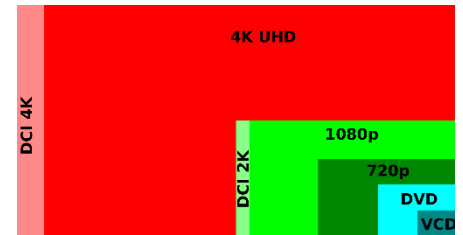


Színhőmérséklet: A látható fény egy jellegzetessége. Egy fényforrás színhőmérsékletét az általa okozott színérzet és egy hipotetikus feketetest-sugárzó által létrehozott színérzet alapján határozzák meg. Izzólámpák esetében, lévén, hogy a fény izzásból származik, a színhőmérséklet jól egybe esik az izzószáll hőmérsékletével. A nem hőmérsékleti sugárzás elvén működő fényforrások, mint például a fénycsövek esetében közvetlen fizikai jelentése nincsen. Ezért ilyenkor inkább korrelált színhőmérsékletről beszélünk. Elterjedt jelölése: CCT (Correlated Color Temperature)

4. Ismertesse a képi tartalom fontosabb jellemzőit, korlátait: felbontás, bitmélység, interlace, **mozgás**, digitális zaj, tömörítési problémák.

Felbontás:

A képfelbontás megmutatja, hogy a képünk hány pixelből áll (pl. 1920x1080). Nagyobb felbontás esetén jobb a kép minősége, azaz több részlet jelenik meg rajta. A megjelenített kép minősége viszont függ a megjelenítő felület nagyságától. Ugyanaz a kép sokkal élesebbnek tűnik, ha kisebb méretben jelenítjük meg. A képernyők pixelsűrűségének mértékegysége a képpont/hüvelyk.



Bitmélység: A bitmélység adja meg, hogy a kép egyes képpontjaihoz mennyi színinformáció áll rendelkezésre. Minél több a képpontonkénti információ, annál több szín jeleníthető meg a képen, és annál pontosabb a színmegjelenítés. Az 1-es bitmélységgel rendelkező képhez például két lehetséges érték tartozik: a fekete és a fehér. A 8-as bitmélységgel rendelkező kép lehetséges értékeinek száma 2^8 , vagyis 256. A 8-as bitmélységgel rendelkező, szürkeárnyalatos képen 256 szürkeárnyalat szerepelhet. A RGB rendszerű képek három színcsatornából állnak. Egy (csatornánként) 8 bites RGB-kép esetén 256 érték tartozik minden csatornához, ami azt jelenti, hogy a kép több mint 16 millió lehetséges színértékkel rendelkezik. A csatornánként 8 bites (8 bpc) RGB képeket szokták 24 bites képeknek is hívni.

Interlace: A televíziózásban a sávszélesség kihasználása, optimális kitöltése nagy feladat. A sávszélesség igény csökkentésének egyik megoldásként hozták létre az interlaced (sorváltott, összefűzött) megjelenítést. Az eljárás célja, hogy minőségvesztés nélkül (vagy minimális veszteség mellett) a lehető legkisebb sávszélesség felhasználásával lehessen az adásokat továbbítani. Az elgondolás azon az alapon nyugszik, hogy amennyiben egy teljes képkockát (keret, frame) felbontunk két fél keret, így egy időegység alatt csupán egy fél keret kell átvinni, amit aztán a vevő készülék többszöri villogtatással épít egybe a második fél kerettel, így szemünk nem veszi észre az eltérést és egészként érzékeli a képet. Első körben csupán minden páratlan pixelsort küldenek át, majd a párosakat. A hagyományos CRT TV-k, monitorok képesek megjeleníteni az ilyen jeleket, átlagosan 50-60 félkép/másodperc sebességgel, így a teljes kép 25-30 fps sebességű lesz.

Mozgás: Időben egymás után következő képeken bekövetkező elmozdulás, pixelváltozás?

Digitális zaj: A digitális fényképezőgépek érzékelőinek alapvető építőeleme a photosite. A szenzornak ezen része érzékeli valójában a fényt. Minden pixelre legalább egy photosite jut (van ahol kettő). Ezek az elemi egységek a fényt olyan módon érzékelik, hogy az expozíció ideje alatt rájuk eső fotonokat elektronokká (töltéssé) alakítják. Ez a töltés az expozíció ideje alatt halmozódik fel és megőrződik az expozíció után. Ha az expozíció befejeződött, akkor minden egyes photosite töltését megméri, majd a mért értéket digitális értékke alakítják (digitalizálják). Ezt a mérési folyamatot nevezzük kiolvasásnak. A digitális fényképezőgépekkel készített képek zaja több különböző forrásból ered:

- **Sötét zaj:** A Sötét Zaj (Dark Noise) a szenzorban (pontosabban a szilícium-szilíciumdioxid átmeneteknél) a hőtől keletkezett szabad elektronok felhalmozódása a szenzor photosite-jaiban. Ez jellegzetes "grízes" zajként jelentkezik a képen. Az ehhez kapcsolódó kifejezés még a "dark current", amely ezeknek az elektronoknak a keletkezési sebességét jelenti.
- **Kiolvasási zaj (Bias Noise):** Ahhoz hogy képet alkossunk a szenzor photosite-jaiban tárolt töltésekből, minden egyes photosite töltését meg kell mérni és a mért értéket digitalizálni kell. Ez a mérés a szenzor kiolvasási folyamatának része. Azonban az eljárás távolról sem tökéletes. Az egyes photosite-ok töltése túlságosan kicsi ahhoz, hogy erősítés nélkül meg lehessen mérni, és ez az erősítés a probléma (zaj) forrása. A kiolvasást végző erősítők valamennyi zajt elengedhetetlenül hozzáadnak a photosite által tárolt és általuk felerősített töltéshez.
- **Foton zaj:** A foton zajt a szenzorra érkező fotonok egyenetlen "érkezési ideje" okozza. Ha a fotonok konstans rátában érkeznek, (mintha csak egy szállítószalag szállítaná őket a photosite-okba) akkor nem lenne foton zaj. Azonban a valóságban a fotonok rendszertelenül érkeznek. Az egyik photosite elég szerencsés ahhoz, hogy az expozíció ideje alatt 100 foton találja el, addig előfordulhat, hogy a mellette lévő, ugyanazon idő alatt csak 80 foton éri el. Ha egy egyenletesen megvilágított felületet fényképezünk, akkor a foton zaj, a szomszédjához viszonyítva, rendellenesen sötét pixelek formájában jelentkezik.
- **Véletlen zaj:** Az előbbi három okon felüli zajt általában a fényképezőgép elektronikájában fellépő feszültség vagy áramingadozások, elektromágneses interferencia és még sok más tényező okozhatja. A véletlen zaj képről képre más és más. Azonban bármi is okozza, általában elenyészően kicsi.

A fenti zaj típusok Gauss zajok, mivel lényegében korrelálatlan valószínűségi változókból erednek. Érdemes még megemlíteni a **só-bors zajt**, amely véletlenszerű helyeken fekete vagy fehér pixeleket jelent. Nem Gauss jellegű hiba, nem is javítható azokhoz hasonlóan, hanem medián szűrővel érdemes. Jellemzően az ADC és az átvitel hibái okozzák.

Tömörítési problémák

Tömörítési hiba: A veszteséges tömörítése során fellépő észrevehető minőségromlás. Általában a veszteséges tömörítés kvantálási lépése során fellépő probléma. Transzformációs kodekeknek általában a kódoló transzformációs terének bázisára jellemző a hiba. A blokk-alapú DCT transzformáció során, amit például az elterjedt JPEG tömörítés használ, többfajta tömörítési hiba szokott előfordulni. Ezek közé tartozik az elmosódott területen kontúrosodás megjelenése, a görbe vonalak mentén lépcsőzetes zaj megjelentése, „légypiszkok” a szélek mentén, sakktáblaszerű hiba (másképpen blokkosság) a kép „mozgalmas” részein.

Lehet, hogy egy tömörített kép az ember számára még teljesen szép, de az algoritmusok megbolondulnak tőle (pl. jpeg tömörített képen élkeresés). Lehetőleg ne tömörített képet dolgozzunk fel!

5. Mit jelent a színtér? Miért van szükség többféle színtérre? Mit jelent a szürkeárnyalatosítás? Röviden ismertesse a fontosabb színtereket, azok feladatát.

Színtér: A színterek a színek ábrázolására használható virtuális térbeli koordináta-rendszer, ahol az egyes színek tulajdonságait azok koordinátái fejezik ki. Színességi koordinátákból épül fel a CIE XYZ színtér. Valamennyi további színtér különbözik ettől; koordinátái általában:

- egy színezeti
- egy világossági
- egy színtelítettségi jellemző.

A színtérben az ábrázolható színek valamilyen rend szerint kerülnek elhelyezésre (például az alapján, hogy a színtér alapszíneinek milyen arányú keverésével állíthatók elő), és a pozíciójukat meghatározó koordinátákkal kerülnek azonosításra (például az RGB színtérben a (255,0,128) koordinátán a maximális vörös, nulla zöld, és a maximális felének megfelelő kék komponensek összeadásából keletkező szín található).

A kamerák által rögzített színes kép általában RGB színtérben készül, mivel az eltérő hullámhosszokhoz tartozó érzékelés gyártástechnológiailag megoldható. Az emberi szem az *R*, *G* és *B* komponenseket nem azonos súllyal/felbontással érzékeli. Ezt tömörítési és adatátviteli eljárások előszeretettel ki is használják.

A számítógépes képfeldolgozás tipikusan nem RGB-ben oldható meg. A (fényesség, színezet, telítettség) leképezések az alkalmazások jelentős részében jobban használhatók. A leképezést többféleképpen meg lehet valósítani, ezért létezik HSL, HSI, HSV, L^*u^*v stb.

A mai TV-k tartalmazzák az alkalmazott LED-ek/LCD stb. karakterisztikáinak megfelelő transzformációhoz szükséges adatokat (*Chromaticity coordinates*), hogy a bemenő jelet a forrás minden kijelzőn a lehetőségekhez mérten azonosan jeleníthesse meg. Ehhez is szükséges *többféle színtér*, minden kijelző típusához lényegében saját.

Szürkeárnyalatosítás: Pixelenként az RGB komponensekből intenzitást, világosságot képezünk. Az egyes színcsatornák különböző súlyozásúak lehetnek, így egy színes képből többféle szürkeárnyalatos is előállítható. Leggyakoribb a 0.30 0.59 0.11 súlyozás, ebben az esetben $I=0.30R+0.59G+0.11B$.

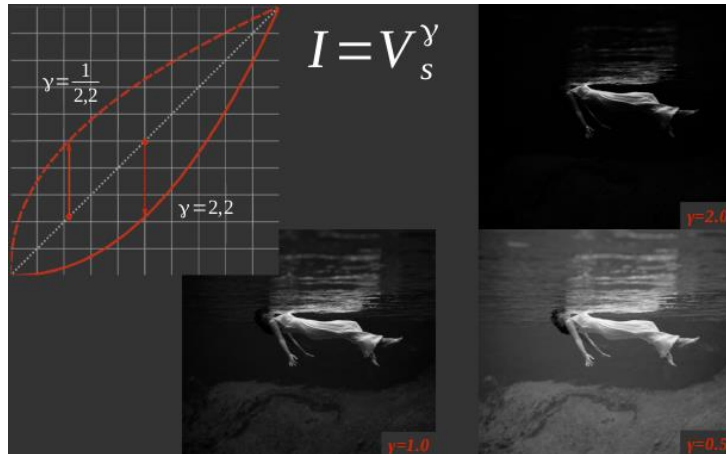
Színterek

- A **televíziós technikában** az RGB komponenseket veszi fel a kamera, ebből egy olyan világossági értéket (jele *Y*, neve *luma*) képeznek, amely kompatibilis a fekete–fehér televízió világossági értékével. Ezek után képezik az *R-Y* és a *B-Y* színtonkomponenseket. Így az analóg televízió egy világossági és két színességi (*chroma*) jelet továbbít (ez három merőleges térbeli vektor). *Megjegyzés:* a színtonkomponenseket magasabb frekvencián modulálják a világosság fölött, ami a fekete-fehér megjelenítőn nem látszik). Az alapötlet egyébként a mai digitális rendszerekben is használatos, pl. *YCbCr* színtérnél a színezeti adatokat 2-4 pixelre összevonhatják, hogy csökkenjen a szükséges sávszélesség (az emberi szem pedig alig veszi észre az eltérést).
- A **képfeldolgozás technikájában** a HSL, a **HSV** (fényerő, színezet, telítettség) és még további színterek 1970 után.
- **Számítástechnikában:** RGB és változatai (sRGB, AdobeRGB). A digitális fényképezőgépek legtöbbször az sRGB (standard RGB) rendszert használja. Az ADOBE virtuális alapszíningerek segítségével kiterjesztett színteret (*wide gamut*) használ. Az **RGB additív** színkeverési eljárás, amely azzal definiálja a színt, hogy a 3 alapszínű fényből mennyit kell összekeverni a kívánt szín eléréséhez. Alapból sötét van és fényeket adok hozzá.
- **Nyomdatechnikában:** CMYK. Négy alapszínnel dolgozó **szubtraktív** színkeverési technika.
- **Tömörítési színterek:** pl. *YCbCr*-t használ a JPEG

6. Mit jelent a színekorrektció? Ismertesse a fontosabb szín és fényességgorrektációs eljárásokat: Gamma korrekció, kontraszt, fényesség. Miért van ezekre szükség, mire jók?

Színekorrektció: A kép színeinek módosítása, valamilyen szempontból jobbá tétele, tipikusan az emberi szem számára és nem a gép számára. A számítógépen tárolt kép színeivel, telítettségével, kontrasztjával, az egyes színösszetevők arányaival történő műveletek sorozata, melynek során kiküszöböljük a digitalizáláskor keletkezett színeltéréseket és színegyensúlyi eltéréseket az eredeti képhez képest.

Gamma korrekció: Nemlineárisan látunk, de lineárisan tárolunk, kompenzálni kell. A kontraszt értékeket optimalizáló eljárás. A Gamma-korrekcióval a kép fényessége és színeinek élése szabályozható. Ezáltal a pontos részletek a sötétben játszódó jelenetknél jobban láthatók, és a kép mélysége is megmarad. Ha a Gamma korrekció értéke magas, a kép világosabb tónusú lesz kevésbé telt színekkel, míg ellenkező esetben sötétebb árnyalatok és valamivel teltebb színek jellemzik a képet.



Kontraszt: A kép világos és sötét részei közötti tónuskülönbség. Ha két kép azonos motívumot ábrázol és az egyik két adott részlet tónusának különbsége nagyobb mint a másikon, akkor ennek nagyobb a kontrasztja. A leggyakrabban használt algoritmus a hisztogram kiegyenlítése (histogram equalization- HE). A módszer az intenzitások előfordulási gyakoriságának függvényében úgy nyújtja a hisztogramot, hogy az kitöltse a rendelkezésre álló intenzitástartományt és a szomszédos intenzitások távolsága azok előfordulási gyakoriságának függvényében alakuljon. Bal oldali kép. Magyarul, nagyobb lesz a kontraszt, ha a világosabb színeket tovább világosítjuk, a sötét színeket tovább sötétítjük.

Fényesség: Relatív világosságérzet, a felület becsült reflektanciája vagy transzmittanciája. Ha növeljük egy kép fényességét, akkor minden színt egyformán világosítunk. Jobb oldali kép.



7. Ismertesse a képi tartalom tárolásának fontosabb problémáit, lehetőségeit, módszereit. Hogyan használjuk a képi adatbázisokat?

Képek tárolása

- Tömörítetlen: BMP, PNM
- Veszteség nélkül: GIF, JPEG-LS/2000, PNG, TIFF, H.264 lossless...
- Veszteséges: JPEG, PNG, H.264...
- Konténer formátumok: (PNG), TIFF, AVI, MOV, MP4, OGG, WMV...

A konkrét tömörítési eljárás a fejlécben tárolva, a codec az, ami dekódolni tudja.

Képek és főleg a videók tömörítetlen tárolása hatalmas tárterületet igényel, így rendkívül drága és néhány speciális esetet leszámítva fölösleges is. Egy 24 bites 1920x1080 felbontású tömörítetlen kép kb. 6 MB helyet foglal, másodpercenként 60 képkockával számolva egy 1 órás videó tárhely igénye 1 TB felett van, valamint már a háttértár írási sebessége is szűk keresztmetszet lehet.

A veszteségmentes tömörítés a képben levő sok pontosan azonos színű pixel jelenlétét használja ki, azok helyett egy "rövidítést" alkalmaz, így a képfájl mérete jelentősen csökkenthető lehet. A tömörítés lehetővé teszi a tömörített adatból az eredeti adatok pontos rekonstrukcióját. A veszteségmentes tömörítés nem tud valamilyen tömörítési arányt garantálni minden lehetséges bemeneti adatra. Más szavakkal kifejezve, bármely (veszteségmentes) adattömörítési algoritmus esetében lesz olyan bemeneti adathalmaz, aminek a méretét az algoritmus nem képes csökkenteni, sőt akár növelheti is.

A veszteséges tömörítés nem teszi lehetővé a tömörített adatból az eredeti adatok pontos rekonstrukcióját, ám (az emberi szem számára) egy „elég jó” rekonstrukciót igen. Videók esetén akár 1:100 tömörítési arány is elérhető látható minőségromlás nélkül. Az emberi szem nem képes több millió színárnyalat megkülönböztetésére, így általában csökkentik a bitmélységet. A veszteséges prediktív kodekekben az adott mintát megelőző és/vagy követő minták segítségével próbálják megjósolni az aktuális hangmintát vagy képkockát. A megjósolt adat és a valós adat közti eltérést (a jóslás reprodukálásához szükséges egyéb információkkal együtt) kvantálják és kódolják. Ha egy videóban csak a kép közepén van mozgás, felesleges az egészet eltárolni újra. A gépnek nem mindig jó, pl. hibásan éleket talál a tömörítési négyzetek szélén.

Képi adatbázisok

A képekhez rendelhetők különböző adatok, melyek megkönnyítik az utólagos rendezést, keresést. Ilyen adatok:

- Alapadatok: méret, felvétel/módosítás időpontja, felvétel helye, kamera beállítások (metadata)
- Méréseken alapuló adatok (pl. objektumok száma, színek, méretek)
- Intelligens felismerés (pl. ember, virág, gépjármű)
- Jogosultságok
- Felhasználói kulcsszavak (google)
- Miniaturók

2. Javítás képtartományban

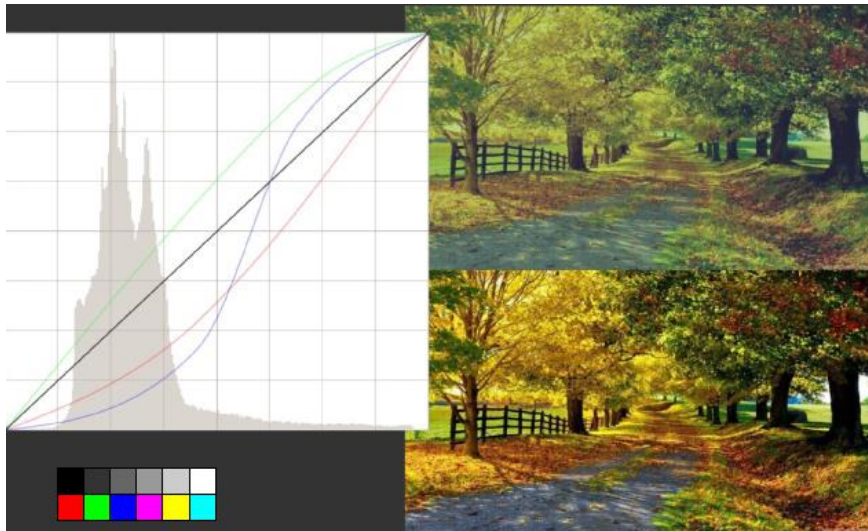
8. Mit jelent a hisztogram? Ismertesse a hisztogramtranszformáció működését és felhasználási lehetőségeit.

Hisztogram: A kép hisztogramja a fényességértékek vagy az egyes színcsatornákon felvett értékeinek eloszlását (sűrűségfüggvényét) mutatja. A lehetséges színértékek tengelyére az egyes színértékek gyakoriságát viszik fel. Minél magasabb ez, annál többször fordul elő az adott színérték a képen.

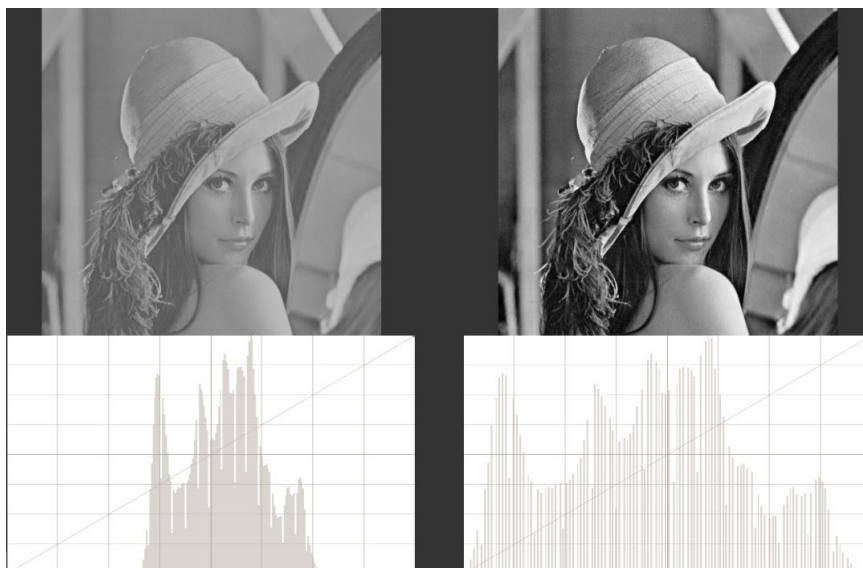
A digitális képfeldolgozásban a hisztogram a kép színértékeiről vagy szürkeségi fokozatáról készül. Ez alapján adatok nyerhetők az előforduló színekről, kontrasztokról és fényességekről. Egy színes képről több hisztogram is készíthető az egyes színcsatornáknak, vagy az összes szín szerint. Mivel a legtöbb eljárás a fekete-fehér képeket támogatja, ezért inkább az egyes színcsatornák hisztogramját használják, ami lehetővé teszi a kép feldolgozásának azonnali folytatását. A színcsatornák száma az alapszínek számától függ: RGB esetén három, CMYK esetén négy.

A digitális fényképészet gyakran használ hisztogramot. A jól felszerelt digitális fényképezőgépek valós időben mutatnak hisztogramokat, hogy így segítsék a képi egyensúly megtalálását. Ez pontosabb képszerkesztést tesz lehetővé, mint ami a kamerakép alapján tehetőes lehetne. Felismerhetők és javíthatók például a világítás hibái, ha a kép túl sötét vagy túl világos lenne. A kép későbbi feldolgozásában sokat számítanak a rajta levő kontrasztok és a fényességek, ezért érdemes a hisztogramokat figyelni.

A hisztogramok egy klasszikus felhasználása a színegyensúly beállítása (**equalizing**), lásd az ábrát. A hisztogramot és a megfelelő színeket egy alkalmas függvénnyel transzformálják. Jobban kiegyensúlyozza a színeket, mint a kontrasztok erősítése. Az ábrán bal alsó sarokban látható néhány szín a korrekció paramétereinek beállításához használható referencia, az erről készült fényképet úgy kell korrigálni, hogy a képen az eredeti értékeket kapjuk vissza (persze ilyen komolyságnál már arra is vigyázni kell hogy a minta megfakul idővel stb.).



Gyakori művelet a hisztrogram széthúzása (pl. röntgen felvétel).



9. Milyen fontosabb zajok jellemezhetik a képeket? Sorolja fel a lehetséges eljárásokat ezek eltüntetésére? Ismertesse a konvolúciós ablakkal történő képjavítás módszerét.

Gauss zaj: Az erősítő képzaja. Gyenge megvilágítás, magas hőmérséklet és a fényképező elektronikája által okozott zaj.
Só-bors zaj: A kép fekete és fehér pixelekkal van "megszórva". A/D konverzió, átviteli hibák okozzák.

Eljárások a zajok eltüntetésére

- Egyszerű átlagolás
- Konvolúció
- Gauss szűrés
- Simító szűrők
- Élesítő szűrők
- Rank szűrők
- Median

Konvolúciós ablak

A kép adott pixelét a környezetében lévő pixelek függvényében módosítjuk. Pixelről pixelre haladunk.

Az ablak (kernel, maszk) egy mátrix, elemei a súlyok. Lineáris esetben minden pixelt a környezetének lineáris kombinációjával helyettesítjük. A lineáris kombinációval történő leírást konvolúciós kernelnek nevezzük.

A képjavítás menete:

1. Az input minden pixelére egy ablakot/kernelt helyezünk úgy, hogy annak origója az adott pixelre essen.
2. Az input kép kernelében lévő pixeleit megszorozzuk a kernelben szereplő súlyokkal.
3. Az eredmény: az input helyzetének megfelelő pixel értéke a súlyozott értékek összege (esetleg skálázva).

10. Hogyan tudunk simító szűrőt készíteni? Hogyan hat a képre a konvolúciós kernel mérete és a kernelen belüli értékek elrendezése? Milyen előnyei vannak az alábbi kernelnek: $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$? Ismertesse a Gauss szűrő lényegét.

Simító szűrő: A simítás lényege, hogy a képpontok közelebb kerülnek környezetük átlagához, azaz a kép „simább” lesz. Vesszük a sima átlagoló szűrő kernelét, de különböző súlyozásokat használunk. A kernel elemei pozitívak.

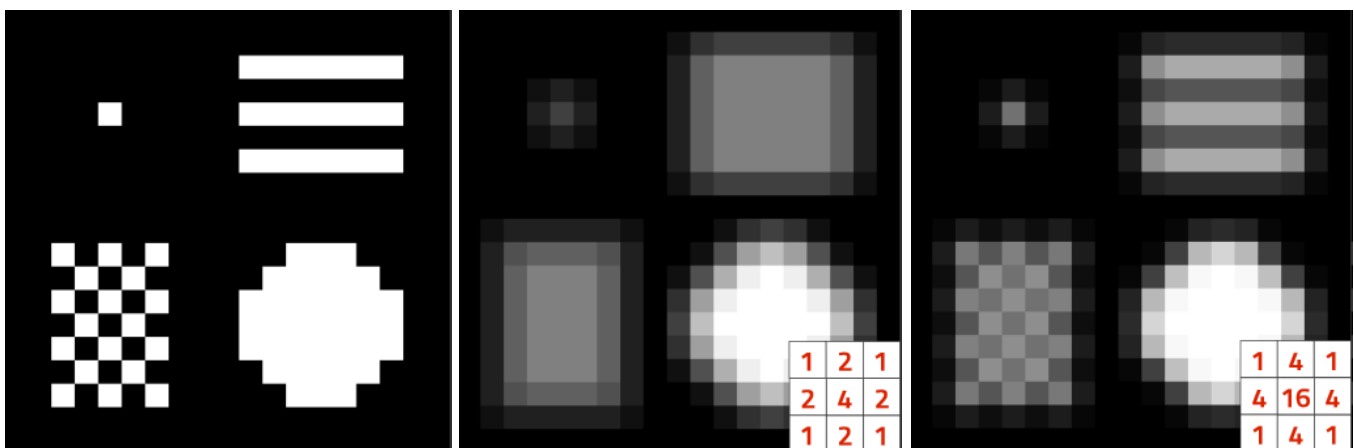
Konvolúciós kernel mérete: Minél nagyobb a kernel, az apróbb dolgok annál inkább eltűnnek, elmosódnak.

$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$ kernel előnyei

FPGA-ban nagyon könnyű megvalósítani, mivel 2-vel és 4-gyel kell szorozni, 16-tal kell osztani a végén. Ezek a műveletek a 2-hatványok miatt eltolások, FPGA-ban csupán huzalozással megoldhatók, gyors!

Gauss szűrő

Kernelméret 3σ . Kernel: $\begin{bmatrix} 1 & 4 & 1 \\ 4 & 16 & 4 \\ 1 & 4 & 1 \end{bmatrix}$. Egyenlő súly helyett érdemesebb a közelebbi pixeleket nagyobb súllyal számolni. A Gauss szűrő simító jellegű. Szépen elkeni a képet, viszont nem túl gyors (36-tal osztás). Algoritmusokban van szerepe, pl. a Canny-féle éldetektorban használjuk.



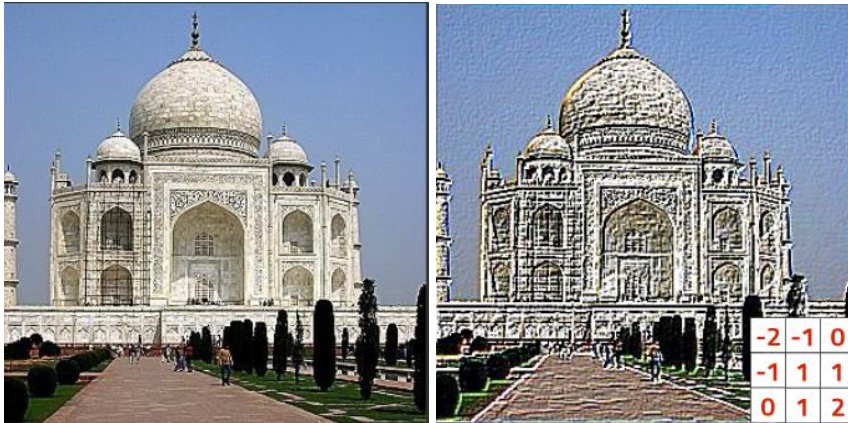
11. Hogyan működnek az élesítő szűrők? Miben különbözik egy élesítő és egy élkereső szűrő? Milyen változtatást hajt végre a képen az alábbi szűrő: $\begin{bmatrix} -2 & -1 & 0 \\ -1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$?

Élesítő szűrő: Kiemeli a környezeti átlagtól való eltérést (azaz az éleket). A kernelben vannak pozitív és **negatív súlyok is**. Minél nagyobb a középső érték annál inkább hasonlít az eredeti képre az új. Hátrány: a zajokat is kiemeli.

Élesítő és élkereső: Élkereső szűrő esetén a kernel súlyainak összege nulla, míg élesítőnél nem. A nullával nem osztunk, hanem eltoljuk egy félszürkeárnyalattal.

$\begin{bmatrix} -2 & -1 & 0 \\ -1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$ szűrő

Ez térbeli kiemelésre használatos. Gyakorlatilag azonban ez is élkiemelés. Csak önmagában, mint élesítés nem használják.



12. Milyen rank-szűrőket ismerünk, ismertesse ezek használatát. Milyen előnyei vannak a median szűrőknek. Hogyan hat a szűrésre a kernelméret és a kernel többszörös végrehajtása?

Rank-szűrők

A simító szűrők a zajhatásokat csak szétszítják (szétkenik, kiátlagolják), ezzel szemben a Rank-szűrők szinte teljesen képesek megszabadítani a képet ezektől a pontszerű hibáktól.

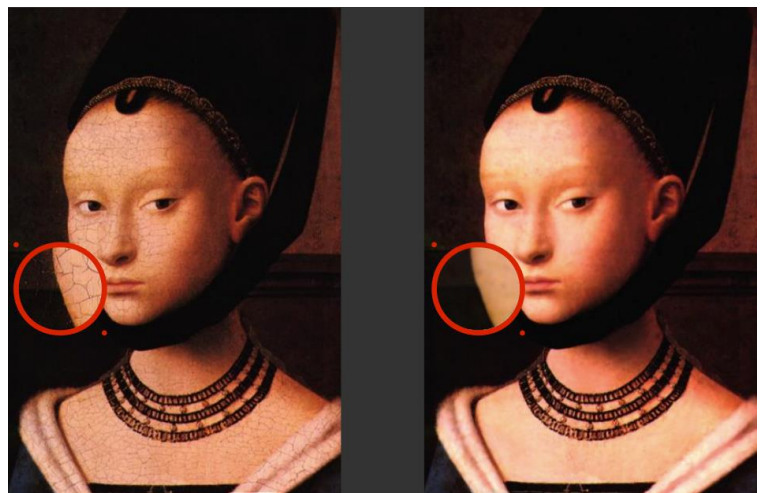
Működési elv: Vesszük minden pixel meghatározott környezetét. A pixel és a környezet intenzitásértékeit nagyság szerinti növekvő sorrendbe állítjuk. E sorból vegyük az n-edik elemet; ez lesz a pixel új világosságértéke.

Ha $n=1$, akkor mindig a minimális értéket választottuk, ha $n=k$, akkor mindig a k-adikat. Median szűrésről akkor beszélünk, ha $n=k/2$, vagyis a sorba rendezett világosságértékek közül éppen a középsőt választjuk. A vizsgált pixelkörnyezet mérete, alakja sokféleképpen választható.

- k-adik szomszéd
- Minimum szűrő
- Maximum szűrő
- Median szűrő

Median szűrő előnyei

Ez a szűrés a kiugró zajcsúcsokat teljesen eltünteti, és nem csak elkeni, mint a lineáris szűrés. (pl. só-bors) Az élek megmaradnak, a hibák elsimulnak.



Kernelméret, kernel többszörös végrehajtása

Általában 3x3, 5x5-ös kernelméreteket használnak. Van 7x7-es is, de ez valós időben nem mindig használható, lassú.

A nagyobb kernelmérettel végrehajtott szűrés hasonló eredményt okoz, mint egy kisebb méretű kernel egymás utáni többszöri végrehajtása, viszont utóbbi gyorsabb. A medián szűrőt például sokszor egymás után hajtják végre, így teljesen eltűnnek a zajcsúcsok.

Olimpiai szűrő

Az olimpiai szűrő, a medián szűrőhöz hasonlóan, a kiugró intenzitás értékeket zajforrásból származónak tekinti. Egyes sportok olimpiai pontozási módszerét követi. Sorba rendezi a kernel alatti elemeket, majd a középsőtől legjobban elütőket eldobja. Paraméterként megadható, hogy a legnagyobb és legkisebb elemekből mennyit hagy figyelmen kívül.

13. Ismertesse az élkeresés alapvető módszereit: DoG, első és második deriváltak, Canny, Frei & Chen (a konkrét kernelok nem szükségesek). Hogyan lehet az élkeresés eredményén javítani ha több-féle eljárásra is lehetőségünk nyílik (pl. Sobel)? Hogyan hat a kernelméret az élkeresésre?

DoG(Gaussok különbsége)

Az eredeti képen két különböző Gauss szűrést végzünk. (Az egyik nagyobb mértékben simít, mint a másik.) A végső képet úgy kapjuk meg, hogy az eredeti képen minden pixelt a két Gauss szűrt kép különbségével (pixelenként természetesen) helyettesítünk. A nullátmenetek (ahol a pozitív értékekből negatív lesz és fordítva) fontosak a számunkra, ezek gyakorlatilag az élek.



Deriváltak

Ha a derivált nagy $\rightarrow$ él

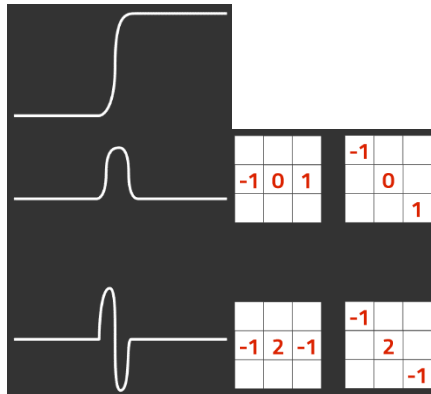
Első derivált: élnél csúcs

- [X X X ; -1 0 1 ; X X X]
- [-1 X X ; X 0 X ; X X 1]

Második derivált: élnél dupla csúcs

- [X X X ; -1 2 -1 ; X X X]
- [-1 X X ; X 2 X ; X X -1]

X: don't care



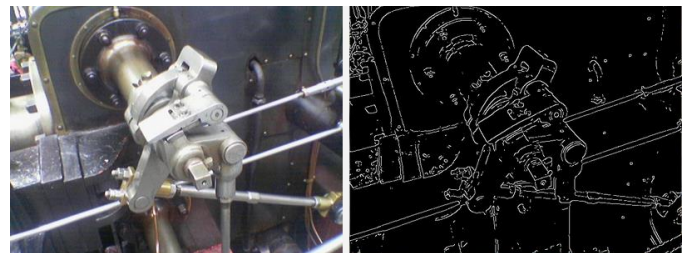
Órai diákban sok példa van. Ezek a példák vízszintes vagy átlós éleket keresnek.

Canny éldetektor

Elsősorban lépcsős élekre, a kép Gauss zajjal terhelt.

Lépések:

1. Gauss szűrés
2. Derivatív szűrés két/négy irányban
3. Nem maximumok törlése (gradiens irányban)
4. Küszöbözés: erős élek, gyenge élek



Pontatlanabb eredmény, nem szubpixeles. Könnyű megtalálni rajta az éleket, sokan használják.

Frei & Chen

Többféle élkeresési módszert átlagol. Kevésbé érzékeny a zajokra mint más élkereső eljárások.

Összesen 9 konvolúciós ablakot/kernelt használ(3x3 méret), ezek a legfontosabb irányú kernelok.

Többféle élkeresési eljárás

Sobel

Több módszer eredményéből:

- maximum kiválasztása
- eredmények összegzése
- eredmények normája (négyzetösszegekből vont négyzetgyökök)

Pl: adott képen horizontális és vertikális deriválás, majd ezek maximumának meghatározása, végül a választott összegzési módszer használata.

Kernelméret és az élkeresés

A kernelméret függ a keresendő objektum méretétől. A kernelméret csökkenése lehetőséget ad az apróbb tárgyak éleinek meghatározására is. Növekedésével az apró tárgyak eltűnnek az éldetektálás során.

14. Milyen jellegű geometriai torzítások terhelhetik a képet? Ismertesse a perspektív ill. a radiális és tangenciális torzítás hatását és kezelésének módját. *Hogyan tudjuk visszaállítani az eredeti arányokat egy ferde kamerával követett síkbeli képen?*

Perspektív

Gyakorlatban: Ha egy épületet az utcán állva fotózunk, gyakran azt tapasztaljuk, hogy a képen a függőleges vonalak összefutnak. Ez a perspektivikus torzítás, ami általában hibának számít. Még erőteljesebb a hatás, ha nagy látószögű objektívvel, viszonylag közletről fényképezünk. Bár az utómunka során ezt a hibát elvben korrigálni lehet, de az újabb torzulásokkal járhat. Ezért a legjobb a felvételkor elvégezni a korrekciót. A gép nem tudja jól elkülöníteni az elől és hátul lévő dolgokat.

Radiális (hordó)

Ezt a fajta torzítást az okozza, hogy az eszközökben az olcsóbb kivitelezhetőség miatt ideális parabolikus lencsék helyett gömbszerű lencsákat alkalmaznak, valamint a képalkotó általában nem a lencse fókuszpontjában helyezkedik el. Ez a torzítás okozza az ún. hordó, vagy halszem effektust. A lencse közepétől távolabb eső fénysugarak jobban elhajlanak, mint a lencse közepén. Ez olcsó webkamerák esetén jól megfigyelhető, drága, csúcskategóriás eszközöknél igyekeznek a matematikailag ideális jobban megközelítő lencsét előállítani, illetve a képalkotót a fókuszpontba helyezni. A radiális torzítás a kép közepén 0, és a kép szélei felé haladva növekszik.

Tangenciális

Ez a fajta torzítás a gyártás hibáiból adódik, a lencse nem párhuzamos a képalkotóval. Ez a torzítás két paraméterrel adható meg.

15. Milyen interpolációs technikákat ismer? Ismertesse a legközelebbi szomszéd, a bilineáris és a biköbös interpolációs technikák alapelvét (képletek nem szükségesek).

Ebben az értelemben az adott mennyiségű képpontból nagyobb (vagy kisebb) felbontású, több (vagy kevesebb) képpontot tartalmazó képet állítunk elő. Motiválhatja például a kép nagyításának igénye. Nagyításnál egyértelműen hiányzó információt pótolunk, kicsinyítéskor viszont a pixelek elhelyezkedésének változása miatt szükséges interpolálni (azonos képtartalmat kevesebb, jellemzően más pozícióba kerülő adatponttal ábrázolunk).

- **Legközelebbi szomszéd:** A nagyított kép egyes pixeleinek értékeit az alapján határozzuk meg, hogy melyik eredeti pixelhez vannak legközelebb, azzal tesszük őket egyenlővé. *Pixeles* képet eredményez.
- **Bilineáris:** Az új pixelek értékét mindkét irányban lineárisan interpoláljuk (név innen), azaz egyenest illesztünk a két közeli eredeti pixel érték közé, és a keresett érték az egyenes pixelnek megfelelő helyén felvett értéke lesz. Az egyenesillesztést tehát a legközelebbi 4 pixel alapján hajtjuk végre. Furcsa *artifact*-okat tud okozni, viszont nagyon gyors. Grafikus kártyák ezt támogatják.
- **Biköbös (bicubic):** Lassabb a bilineárisnál, de szebb eredményt is ad. A legközelebbi 16 pixel alapján számol. Figyelembe veszi a szomszédos pixel értékeket, az x-y irányú deriváltakat és kereszt deriváltakat, az ezekből következő meredekségi kényszerekre felírt egyenleteket oldja meg.
- Láncczos (jól közelíti a vonalakat)
- Komplex megoldások
 - Görbeillesztés
 - Fraktál
 - Scale 2x, 3x, 4x

16. Mit jelent a képi matematika? Milyen műveleteket alkalmazhatunk a gyakorlatban? Adja meg, hogy melyik eljárás milyen feladatra használható.

Képi matematika esetén végezhetünk műveleteket **kép-érték** vagy **kép-kép** között.

Lehetséges műveletek

- összeadás/kivonás/átlagolás
- szorzás/osztás/normalizálás
- maximum/minimum
- logikai műveletek (főleg bináris képeknél)

Gyakorlati alkalmazás

- Kép-érték esetben az összeadás fényerő állítást jelent, a szorzás brightness/contrast állítást.
- A szorzás művelet használható textúrázásra (maszkolásra).
- Az összeadás/átlagolás használható zajcsökkentésre.
- A kivonás használható különbségképzésre, a háttér eltávolítására (mozgó/elmozdult objektumok keresésére).

3. Frekvenciatartomány

17. Mutassa be, hogy mit jelent egy kép frekvenciatere, hogyan kell értelmezni? (Képletekre nincs szükség). Hogyan hat egy objektum pozíciója és orientációja a frekvenciatartománybeli képre?

A kép frekvenciatere megadja azon komplex exponenciális komponensek együtthatóját, amelyekből a kép előállítható. Két dimenziós kép esetén a frekvenciatartomány is két dimenziós lesz. Mivel egy kép diszkrét pontokból áll, ezért a frekvenciatartomány periodikus lesz, azonban ennek csak az első periódusát ábrázoljuk, mivel a többi nem tartalmaz plusz információt. Továbbá mivel a pixelek a képen nyilvánvalóan valós értékűek, ezért a frekvenciatartományban komplex konjugált párokat kell kapnunk. Egy kép frekvenciatereben az egyes komponensekhez hasonlóan az egy dimenziós esethez két jellemző tartozik: a fázis és az amplitúdó. Lényeges különbség azonban, hogy egy komponenst már két koordináta azonosítja (k és l). Ezek a koordináták határozzák meg a képtartományban a szinuszos hullámok orientációját (ha az egyik nulla, akkor tengelyirányú), illetve periódusát. Eltolás a képtartományban nincs hatással az amplitúdókarakterisztikára a frekvenciatartományban (csak a fázisra). A forgatás már hatással van mind a két jellemzőre, de az amplitúdó karakterisztika esetén egy képtartománybeli forgatás az amplitúdó karakterisztikában is egy forgatásként jelenik meg.

18. Mit jelent az FFT? Hogyan működik az algoritmus? Mi a DCT, miben különbözik a DFT-től? Hogyan működik a FCT? Miért lehet jól alkalmazni ezt az algoritmust tömörítés során és hogyan?

FFT: Fast Fourier Transformation

Algoritmus működése: $2N$ méretű adathalmazon működik. Lényegében a problémát mindig megfelel a páros és páratlan komponensek mentén, egészen addig, amíg már csak egy pár marad, amire már triviálisan számolható a feladat. Az így kapott eredményeken azonban minden fokozat végén pillangó műveleteket kell végezni. Két dimenziós esetben először vízszintes, majd függőleges irányban.

Megjegyzés az értelmezéshez: Fourier transzformálni csak periodikus jeleket lehet, ezt úgy biztosítjuk, hogy a képet végtelen ismételtük az x és y irányokban (pusztán elméleti értelemben, nyilván).

DCT(Discrete Cosine Transform) vs DFT(Discrete Fourier Transform):

Koszinus transzformáció esetén olyan, mintha a kép tükörképe lenne a kép után x és y irányban, ennek megfelelően, nincsenek olyan hirtelen ugrások, mint DFT esetén (ez energiatömörítés, kisebb lesz a spektrum integrálja). További előnye, hogy valós értékeket kapunk, jobb tömörítést tesz lehetővé (kevesebb nagyfrekvenciás komponens a szélek miatt), illetve egyszerűbb is.

FCT: Fast Cosine Transform

1. Szimmetrikus függvény
2. DFT($2N$)
3. Pozitív tartomány

Adattömörítés

Jóval kisebb intenzitású nagyfrekvenciás komponensek, mivel nincs hirtelen átmenet a kép széleinél, mint DFT esetén. (kevesebb információt veszünk el a nagyfrekvenciás komponensek eldobásával)

19. Hogyan lehet frekvenciatartományban szűrést végezni? Mit jelent az alul- és felül-áteresztő szűrés? Milyen jellegű zajok kiszűrésére van lehetőség frekvenciatartományban, és hogyan?

Szűrés frekvenciatartományban: Az egyes frekvenciatartománybeli komponenseket meghatározott értékekkel szorozzuk meg, attól függően, hogy milyen ezeknek az értékeknek az eloszlása (milyen frekvenciákon csillapít, átereszt, erősít) az egyes szűrőket csoportosíthatjuk. A szűrés frekvenciatartományban egy egyszerű szorzással valósítható meg, ellenben a képtartományban, ahol a konvolúcióval számítható.

Aluláteresztő szűrő: Ebben az esetben alacsonyabb frekvenciás komponensek súlya nagyobb, mint a magasabb frekvenciásoké.

Felüláteresztő szűrő: Ebben az esetben magasabb frekvenciás komponensek súlya nagyobb, mint az alacsonyabb frekvenciásoké.

Zajszűrés frekvenciatartományban: Periodikus zajok nagyon jól kiszűrhetők, mivel frekvenciatartományban jól körülhatárolható komponensek okozzák őket. Elméletileg ez megoldható lenne képtartományban is, de túl költséges lenne, mert nagyon nagy kernel kellene hozzá a szűrő jellege miatt (lyukszűrő). Aluláteresztő szűrővel a nagyfrekvenciás zaj is jól kiszűrhető. Felüláteresztő szűrővel az élék, körvonalak jól meghatározhatók.

20. Ismertesse a konvolúció és a frekvenciatartománybeli műveletek kapcsolatát. Mit jelent a dekonvolúció? Mit jelent a Wiener dekonvolúció és mikor használjuk?

Konvolúció a frekvenciatartományban: A szűrés frekvenciatartományban egy egyszerű szorzással valósítható meg, ellenben a képtartománnyal, ahol az konvolúcióval számítható.

Dekonvolúció: A dekonvolúció során adott egy kép, amely valamely kernellel végzett konvolúció eredménye. A cél az eredeti kép meghatározása. A feladathoz elengedhetetlen a kernel nagyságrendi ismerete vagy sejtése. Tipikus felhasználási terület: defókuszált kép élesítése.

Wiener-dekonvolúció: Annyiban nehezebb problémát old meg, hogy feltételezi, hogy a konvolúció során előálló kép zajjal terhelt. Tehát nem a konvolúció eredményén dolgozik, hanem annak egy additív zajjal terhelt változatán. Egy frekvenciatartománybeli eljárásról van szó, amely azokon a frekvenciákon jobban csillapít, ahol a jel-zaj viszony rosszabb.

4. Bináris képek

21. Mit jelent a bináris kép? Mit jelent a küszöbözés? Milyen lehetőségeink vannak? Mit jelent a dupla-küszöbözés?

Bináris kép: Minden pixel csak két értéket vehet fel (megjelenítés tipikusan fekete/fehér színnel). A pixelekhez hozzárendelendő értéket valamilyen képi tulajdonság (fényerő, színintenzitás stb.) küszöbözésével határozzuk meg.

Küszöbözés (thresholding): A legegyszerűbb eljárás, aminek segítségével egy szegmentálást (objektum(ok) és a háttér szétválasztása) végezhetünk. Egy tulajdonsághoz hozzárendelünk küszöbértéket, majd a vizsgálandó kép minden pixelén végigmegyünk és összehasonlítjuk az értékeket. A végeredmény egy bináris kép. Például fényintenzitás alapján bizonyos érték fölött és alatt szétválasztom a képet.

Hisztogram egy kép egy adott tulajdonságáról (előfordulás: az adott értékkel ennyi számú pixel rendelkezett):



Szegmentálási lehetőségek

- Hisztogram alapú -> KÜSZÖBÖZÉS
- Osztályozás alapú
 - Színek
 - Formák: pl.: egyenes, kör
- Környezetfüggő
 - Lokális: inhomogén háttér esetén

Dupla-küszöbözés

Más néven hiszterézis küszöbözés. A probléma az, hogy elkülönítendő objektumok mellett más hasonló (kinézetű, színű stb) objektumok is vannak, melyek küszöbözés után megmaradnak. Ezért kettő küszöb használatával létrehozunk két képet, melyet aztán ÉS kapcsolatba helyezünk.

Példa

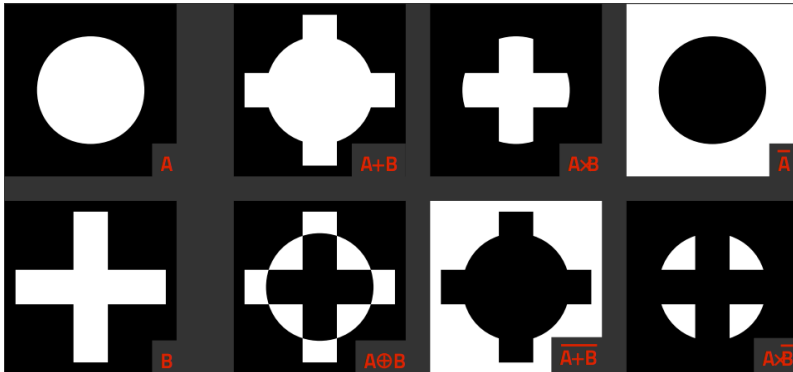
1. Eredeti kép
2. Eredeti kép küszöbözése
3. Eredeti képen élkeresés
4. Élkeresett kép küszöbözése
5. „4” maszkolása „2”-vel



22. Ismertesse a bináris képen végezhető logikai műveleteket és azok fontosabb feladatait. Ismertesse az objektum alapú logikát és felhasználási lehetőségeit.

Adott két kép: A, B

- Negálás: például kontúrkeresésnél
- AND: főleg képek maszkolásánál van jelentősége, két kép átfedésének megjelenítése (metszet)
- OR: unió (két kép által az adott képpont mátrixból együttesen elfoglalt képpontok)
- XOR: csupán egyikük által elfoglalt képpontok



Objektum alapú logika?

23. Mit jelent az erózió és a dilatació? Hogyan végzünk nyitást és zárást? Mire jók ezek a műveletek, mitől függ, hogy melyiket alkalmazzuk? Hogyan tudunk e műveletekkel kontúrt keresni? Mi az olvasztásmentes dilatació? Szürkeárnyaltos képen hogy lehet értelmezni ezeket a műveleteket?

Erózió: Csökkentés (fehér területek csökkentése fekete-fehér átmeneteknél). A pixelszerű zajok csökkentése.

Dilatáció: Szélesítés (fehér területek növelése fekete-fehér átmeneteknél) A változás mértéke függ a felhasznált kernel nagyságától. Nagy számításigény. A nulla zajokat tüntetem el, cserébe megnő az objektum mérete.

Zárás

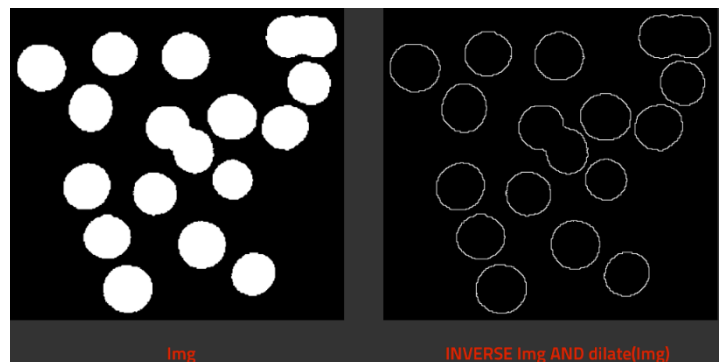
Dilatáció + erózió: nullzajok kiszedése (fekete pöttyök)

Nyitás

Erózió + dilatació: fehér pixelhibák kiszedése (Gyakran együtt használják a kettőt.)

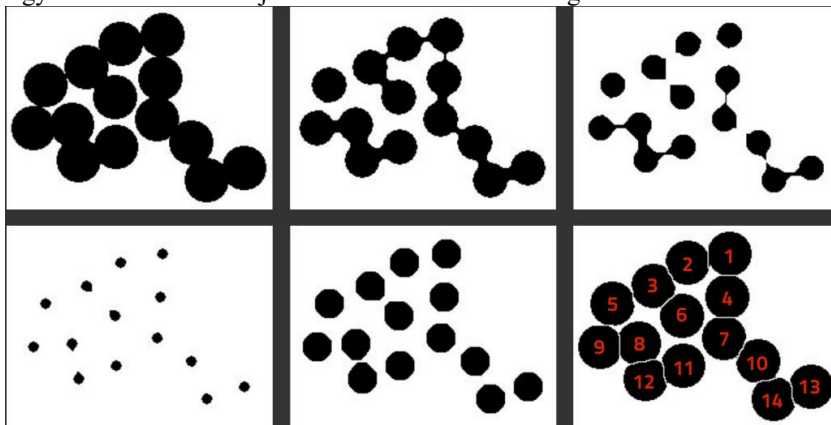
Felhasználás

- pixelhibák, pixelszerű zajok eltüntetése bináris képről
- kontúrkeresés: $\text{Img AND INVERSE}(\text{dilate}(\text{Img}))$
- élkeresés és lyukak kitöltése



Olvasztásmentes dilatació

Egymással érintkező objektumok éldetektálásánál segít ez a módszer.



Szürkeárnyaltos képen:

- erózió: minimum szűrő (sötétít)
- dilatació: maximum szűrő (világosít)

24. Mit jelent a 4- és 8-szomszédú távolság? Mik az előnyei és hátrányai az euklidészi távolsághoz képest?

Kernelek esetén a kernel középpontjától a pixelekben mért távolság a szélek felé haladva.

4 szomszéd					8 szomszéd					Euklidészi				
4	3	2	3	4	2	2	2	2	2	$\sqrt{8}$	$\sqrt{5}$	2	$\sqrt{5}$	$\sqrt{8}$
3	2	1	2	3	2	1	1	1	2	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{2}$	$\sqrt{5}$
2	1	0	1	2	2	1	0	1	2	2	1	0	1	2
3	2	1	2	1	2	1	1	1	2	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{2}$	$\sqrt{5}$
4	3	2	3	4	2	2	2	2	2	$\sqrt{8}$	$\sqrt{5}$	2	$\sqrt{5}$	$\sqrt{8}$

4 szomszéd: csak akkor tekintjük szomszédnak, ha van közös él, átlósan 2 a távolság.

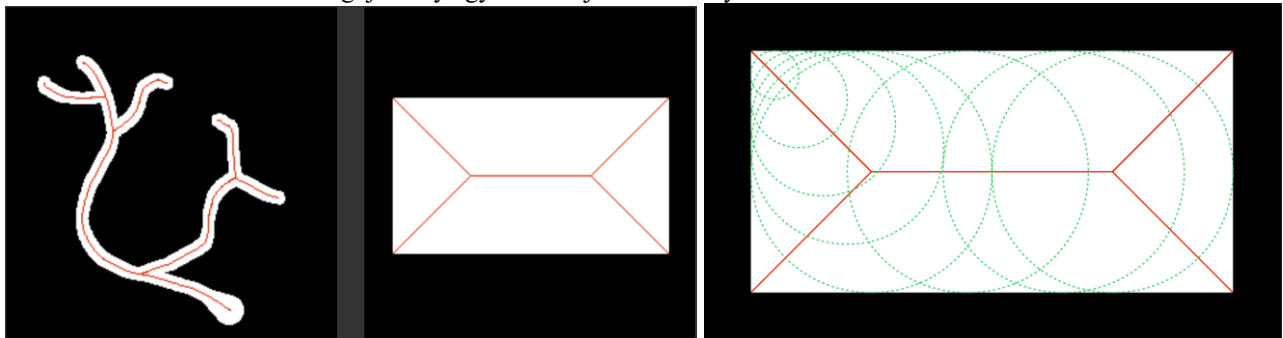
8 szomszéd: a csúcsnál érintkezőt is szomszédnak vesszük

euklidészi: pontos távolság (pitagorasz)

- 4/8 szomszéd: gyors, pontatlan
- EDM: pontos, számításigényesebb

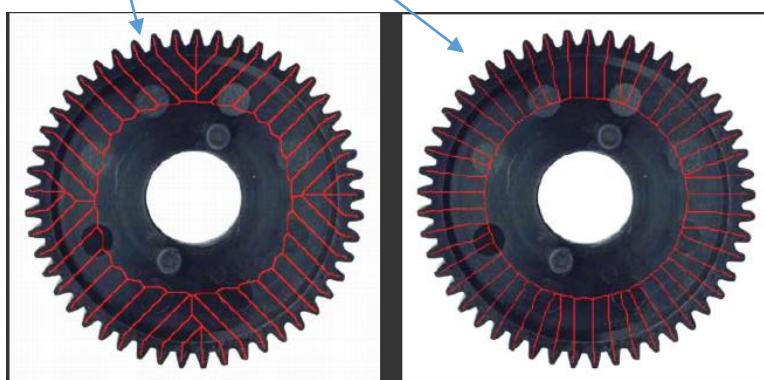
25. Ismertesse a csontváz definícióját és a csontvázásítás egy lehetséges megvalósítását.

Definíció: Maximális körök origója mely egy adott objektumba helyezhető.



Megvalósítás: iteratív erózió, de nincs erózió, ha kettészakadna az objektum vagy ha az objektum végén vagyunk.

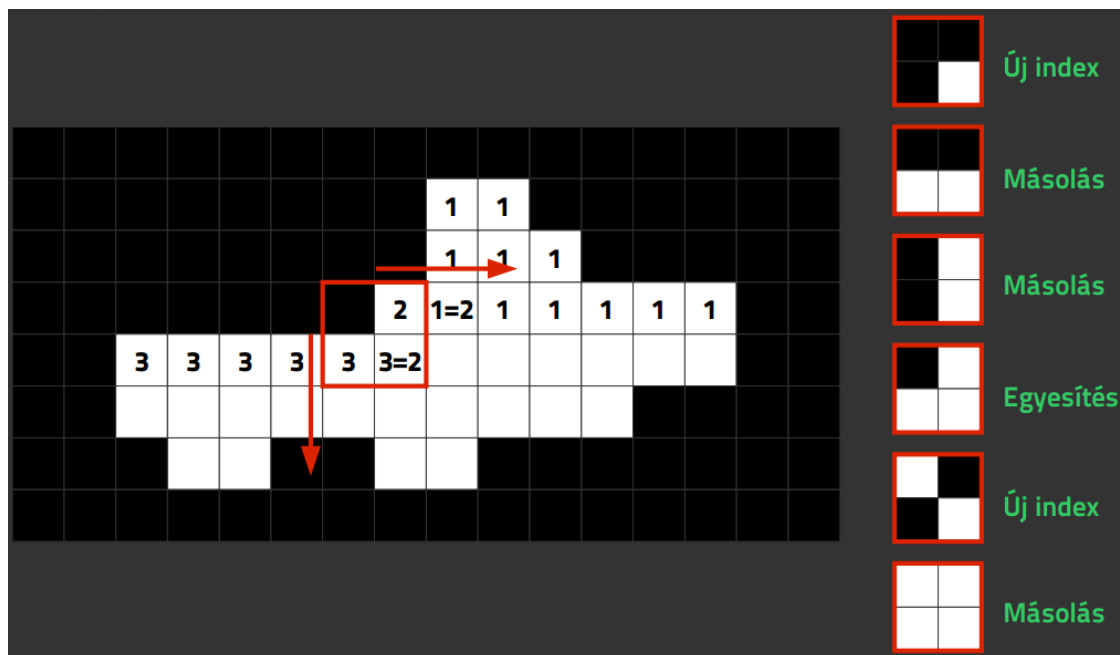
Lehet 4/8 szomszéd vagy EDM. Felhasználás: ujjlenyomat ellenőrzés



26. *Hogyan tudunk bináris képen objektumokat keresni és megszámolni? Ismertesse az eljárásokat.*

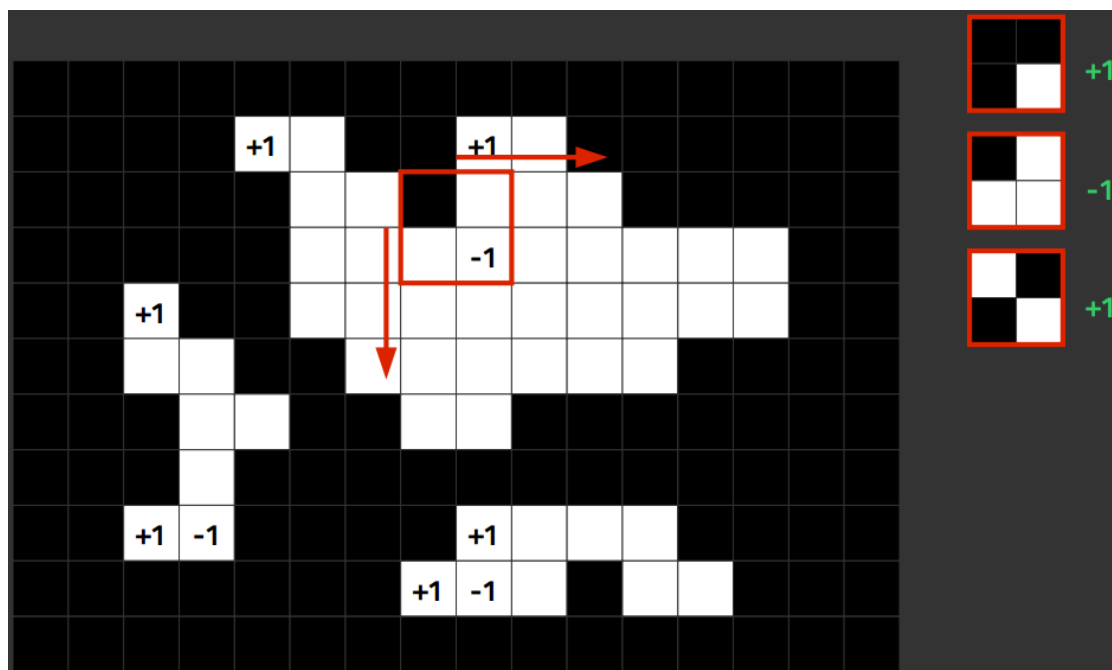
Objektumkeresés

Egy 2x2-es ablakkal végigpásztázunk minden pixelt és indexeljük az egyes pixeleket.



Objektumszámlálás

Egy 2x2-es ablakkal végigpásztázunk minden pixelt és növeljük, vagy csökkentjük (hogy az algoritmus ne számoljon egy objektumot többször) a számlálónk értékét.



5. Mérések

27. Hogyan definiálhatjuk egy objektum pozícióját a képen (pixeltartományban)? Ismertesse a pozíciómérés lehetőségeit. Mutassa meg, hogy lehet a pozíció értékét meghatározni bináris és maszkolt szürkeárnyalatos képeken.

Egy 2D-s képet ábrázolhatunk egy derékszögű koordináta-rendszerben, ahol az egyes pixelekhez hozzárendelhetünk egy (x,y) egész koordinátapárt. A koordináta-rendszer középpontja tetszőlegesen, feladattól függően megválasztható, de általában a bal felső sarokban lévő pixelhez rendeljük hozzá a $(0,0)$ -t. Egy objektum pozíciója az objektum egy jellegzetes koordinátapárjával jellemezhető. Ez a koordinátapár lehet:

- geometriai középpont (az objektumot befoglaló téglalap/kör középpontja)
- tömegközéppont

Tömegközéppont (x_c, y_c) meghatározása

A kép mérete: $M \times N$ pixel (M az oszlopok száma, N a soroké)

- $p_x(x)$ az x koordinátájú oszlopban a vizsgált objektum pixeleinek száma
- $p_y(y)$ az y koordinátájú sorban a vizsgált objektum pixeleinek száma

a) Bináris képekre

$$x_c = \frac{\sum_{x=0}^{M-1} p_x(x) \cdot x}{\sum_{x=0}^{M-1} p_x(x)}, \quad y_c = \frac{\sum_{y=0}^{N-1} p_y(y) \cdot y}{\sum_{y=0}^{N-1} p_y(y)}$$

b) Szürkeárnyalatos képekre

$I(x,y)$: intenzitásfüggvény

$$x_c = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y) \cdot x}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y)}, \quad y_c = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y) \cdot y}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y)}$$

Geometriai középpont (x_g, y_g) meghatározása

a) Meghatározás a befoglaló téglalap alapján

$x_{min}, x_{max}, y_{min}, y_{max}$: az objektum legszélső pixeleinek koordinátái

$$x_g = \frac{x_{min} + x_{max}}{2}, \quad y_g = \frac{y_{min} + y_{max}}{2}$$

b) Meghatározás a befoglaló kör alapján

- egyértelmű, ha 3 ponton érinti a kört
- kör átlójáig egyértelmű, ha 2 ponton érinti a kört.

28. Hogyan definiálhatjuk egy objektum orientációját a képen (pixeltartományban)? Sorolja fel az orientációmérés lehetőségeit, illetve röviden ismertesse ezek alapelvét.

Objektum orientációján egy objektum egy olyan 1D-s jellemzését értjük, mely irány-, szöginformációkat szolgáltat az adott objektumról. Objektum orientációja megadható a

- befoglaló téglalap arányaival és méreteivel
- legnagyobb távolsággal az objektumon belül
- középponttól vett legnagyobb távolsággal
- rá illeszthető legkisebb nyomatékú tengellyel

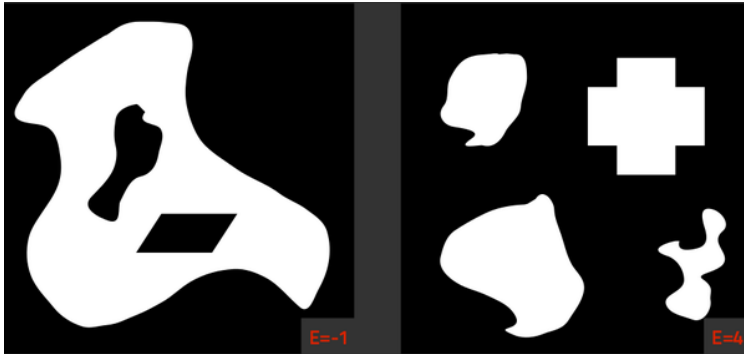
Orientációmérés lehetőségei

Például: Az autó elektronikában az önjáró autóknaál a sávkövetést és az autók közötti távolságtartást is orientációmérés alapján határozzák meg. Sávkövetés: egyenes festett vonalakat élkéréséssel keres az algoritmus és a vonalak a képen befelé haladva szűkülnek, ami irányt ad az orientációnak. Távolságtartás: Befogadó téglalap vagy trapéz arányaival meghatározható milyen távol van az elől haladó objektum.

29. Mit jelent az Euler-szám? Mire használható? Adja meg a mellékelt ábra Euler-számát.

Az Euler-szám egyike a topológiai tulajdonságoknak, melyek egy kép geometriai leírását segítik elő. Fontos része az ilyen tulajdonságoknak, hogy rubber-sheet jellegű transzformációkra invariáns. Az ilyen jellegű tulajdonságok jól használhatók formák keresésére, objektumok felismerésére, adatbázisbeli keresésre. Euler-szám fontos szerepet játszik például orvosi képfeldolgozásban, fertőzött sejtek felismerésében.

Euler-szám = (egybefüggő régiók száma) – (lyukak száma)



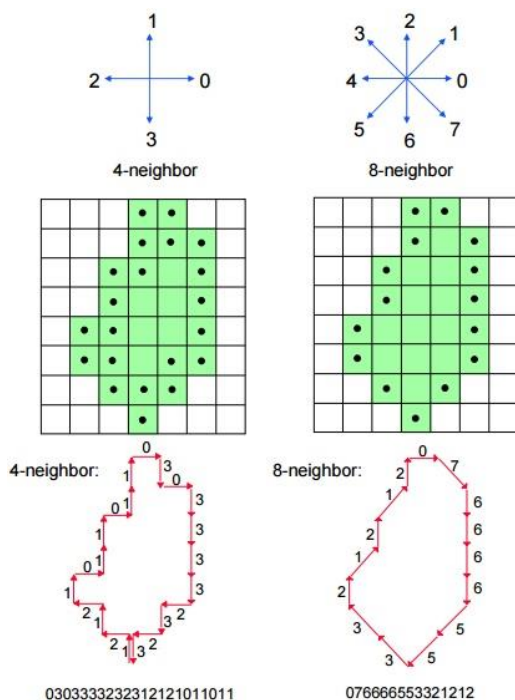
30. Mit jelent a lánc kód? Mire használható? Mi a különbség a 4-szomszédos és 8-szomszédos lánc kód között? Mik az előnyei és a hátrányai az így ábrázolt objektumoknak? Hogyan tudunk segítségével terület- és hossz-számítást végezni? Milyen problémák adódnak?

A lánc kód egy veszteségmentes tömörítési algoritmus bináris képekhez. Lánc kód segítségével alakfelismerést, sarokdetektálást végezhetünk. (A kódból egyértelműen látszik, hol vannak pl.: dudorok, bemélyedések.) Az objektum egy szélső pixelétől elindulva szomszédos, határ menti pixelekre lépünk. Attól függően, hogy milyen irányba lépünk tovább a pixelhez egy számot rendelünk hozzá. Ez a számsorozat alkotja a lánc kódot.

- 4-szomszédos: csak azok a pixelek számítanak szomszédosnak, amiknek van közös élük
- 8-szomszédos: közös él, vagy közös csúc
- 4-szomszédos lánc kód maximális hiba: 41% (45°-os átlós egyenes)
- 8-szomszédos lánc kód maximális hiba: 7.9% (~18-27°-os átlós egyenes)

Kerület = kódhossz

Hossz számításnál kerül elő az a probléma, hogy négyzetes pixelek esetén egy átlós lépés valóságos hossza $\sqrt{2}$ egység. 4-szomszédos lánc kód 2 egység hosszúnak, míg a 8-szomszédos esetben 1 egység hosszúnak veszi alapból. Ha szükséges akkor ezt kompenzálni kell.



31. Ismertessen szubpixeles eljárásokat. Hogyan tudunk pozíciót, területet, ill. területet mérni segítségükkel?

Interpoláció alapú eljárás, mely segítségével pixel alatti pontossággal illeszthetünk görbét egy objektumra. Megfelelő algoritmussal akár 0.1% pontosság is elérhető.

Eljárás szürkeárnyaltos képekhez

1. Szürkeárnyaltos képeket először is binarizáljuk.
2. Visszatérve az eredeti képhez (fekete-fehér kép alapján) az átmeneteknél lévő pixelekhez egy súlytényezőt (értéke lehet tört, megadja az interpoláció finomságát) rendelünk attól függően, hogy mennyire világos/sötét az adott pixel.
3. Megfelelő ablakozással (pl.: 2x2) végigpásztázzuk a határokat és súlyozásoknak megfelelően felbontjuk (kijelöljük a határpontot/határpontokat) a két szomszédos fekete-fehér pixel középpontját összekötő szakaszt.

Pozíciószámítás során a 3. lépésben meghatározott határpontok koordinátáit használjuk fel a képletekben. Kerület, területnél hasonlóan.

32. Ismertesse az egyenesekre vonatkozó Hough-transzformáció működését.

Hough-transzformáció

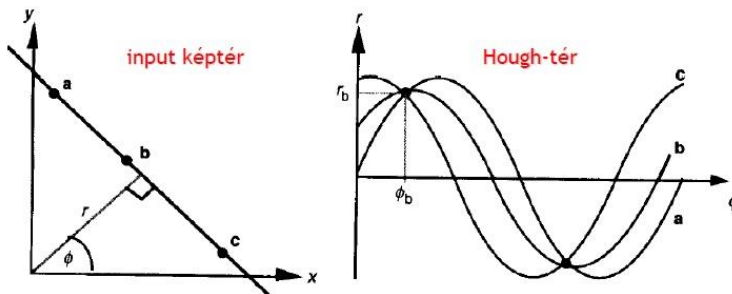
A Hough-transzformáció segítségével a képen általában az $f(x, y; a_1, a_2, \dots, a_n) = 0$, ahol $a_1, a_2, \dots, a_n$ paraméterekkel explicit alakban megadható görbéket keressük. A Hough-transzformáció alkalmazása célravezető, ha ismert alakú (és méretű) objektumokat keresünk a képen. Akkor is célszerű, ha az egyenesek részben takartak vagy zajosak.

Áttérés a Hough-térbe

Az input (kép)tér egy (x_i, y_i) pontjának az $r = x_i \cdot \cos\phi + y_i \cdot \sin\phi$ szinuszos görbe felel meg a Hough-térben. Az egy egyenesbe eső pontokhoz tartozó szinuszos görbék egy pontban metszik egymást.

Egyenesek meghatározása

- Egy (él)pont a képtérben megfelel egy szinuszos görbének a Hough-térben. Két pontnak két görbe felel meg.
- Két (vagy több) ilyen görbe metszéspontja által reprezentált egyenesre ekkor kettő (vagy több) szavazat esett.
- Az így kapott egyenes valamennyi rá szavazó ponton átmegy a képtérben.
- A Hough-tér küszöbölésével megkapjuk a képtér egyeneseit.



6. Szegmentálás

33. Mit jelent a szegmentálás? Mikor van rá szükség? Mik a főbb nehézségei? Mit jelent a szegmentálás során az idő-konzisztencia? A Szegmentálás paraméterterében milyen összetevőket alkalmazhatunk?

Szegmentálás: A képszegmentálás a képfeldolgozás egyik legfontosabb alaproblémája, mely a hasonló tulajdonságú pixelek homogén területekbe történő csoportosításával foglalkozik.

Szükséges/Felhasználás

- Felületleírás
- Sztereó látás (élek megfeleltetése => 3D rekonstrukció)
- Alakfelismerés
- Képi adatbázis indexelése, keresés

Nehézségek

- színhasonlóság
- mozgás (elmosódott élek)
- defókuszáltság
- hasonló textúrák

Egy videófolyam során egy adott szegmens/alak formája változhat, elfordulhat stb.

A Szegmentálás paraméterterében milyen összetevőket alkalmazhatunk? r , φ

34. Sorolja fel a fontosabb szegmentálási módszereket. Pár szóban mutassa be őket.

- Küszöbözés, hisztogram alapú
- Éldetektálás alapú
- **Klaszterezés:**
 - színcsatornák
 - térbeli eloszlás
 - időbeli eloszlás
- Régiónövelés
- Tömörítés-alapú módszerek
- Paraméteres megoldások, görbeillesztés
- Modell alapú
- Tanuló rendszerek

35. Ismertesse a küszöbözés alapú szegmentálási eljárásokat (beleértve a hisztogram-alapú szétválasztást). Milyen előnyök és hátrányai vannak? Milyen módon tudjuk meghatározni a küszöbszint(ek)et?

Intenzitás alapú, szín alapú?

$f(x,y)$ – a kép x,y koordinátáknál mért intenzitása

$p(x,y)$ – pont lokális tulajdonsága (pl. x,y középpontú szomszédság átlaga)

$g(x,y)$ legyen a küszöbözött kép:

$$g(x,y) = \begin{cases} 1, & \text{ha } f(x,y) > T \\ 0, & \text{ha } f(x,y) \leq T \end{cases}$$

ha T csak $f(x,y)$ -től függ $\rightarrow$ globális

ha T $f(x,y)$ -től és $p(x,y)$ -től is függ $\rightarrow$ lokális

ha T függ x,y -től is $\rightarrow$ dinamikus

Globális küszöbözés lehet az is, ha pl. egy globális hisztogram tulajdonságai (medián stb.) alapján határozzuk meg a később felhasználandó küszöbértéket. (Ez csak akkor működik jól, ha mi befolyásoljuk a megvilágítás körülményeit.)

36. Ismertesse a K-Means és a Mean Shift algoritmusokat. Milyen előnyeik és hátrányaik vannak? *Mikor használjuk őket?*

K-Means

Kiindulás:

- Klaszterek száma (K)
- Kiindulási klaszterközéppontok (magok)

Iteráció:

1. Pontokat a legközelebbi maghoz rendelni
2. Új klaszterközéppontok (átlag) számítása
3. Ismétlés, amíg „nagy” a változás

Előny:

- Egyszerű, gyors (helyes kiindulási értékek esetén)
- Könnyen implementálható

Hátrány:

- Szükség van a klaszterek számára
- Érzékeny a kiindulási középpontokra
- Érzékeny az outlier-ekre

Mean Shift

Kiindulás:

- Kernel típusa (Fiat, Gauss...)
- Kernel mérete (h)

Iteráció:

1. Pontot a kernelen belüli pontok átlagához tolni
2. Ismétlés, amíg „nagy” a változás

Előny:

- Általános, alkalmazásfüggetlen
- Csak egy paraméter szükséges (h)
- Robusztus

Hátrány:

- Lassú

37. Ismertesse a textúraalapú szegmentálás alapjait, módszereit (Gábor szűrő, frekvencia tartománybeli szűrések, egyéb megoldások).

Egy képen egy területnek állandó textúrája van, ha a lokális statisztikák vagy egyéb lokális tulajdonságok állandóak, lassan változnak, vagy megközelítőleg periodikusak. Ez lehet determinisztikus vagy véletlen is. Egy textúrált felületet homogénnek érzékelünk, de a gép számára ez a legkevésbé sem triviális.

A textúrák statisztikai leírása olyan textúra-jellemzőket állít elő, melyek egy régió belül homogének, így a szegmentálás viszonylag könnyen elvégezhető. Három fő típus:

- hisztogram (elsőrendű statisztika)
- együttes előfordulási (co-occurence) mátrix (másodrendű statisztika)
- textúra elemek eloszlása egy ablakon belül

38. Milyen régiónövelő eljárásokat ismerünk? Pár szóban mutassa be őket.

Kiinduláskor a pixelek egyenként külön halmazokat alkotnak. Összehasonlítjuk őket a szomszédaikkal, hasonlóság esetén összevonjuk a halmazokat. Hasonlóság lehet:

- Kiválasztott (szín)csatornák
- Defókuszáltság
- Mélység (Mélységi képen)

SRM (Statisztikus régiónövelés)

Gráf súlyozás a kül. alapján.

Split & Merge

- Szétválasztás csak hasonló (belül) elemekre
- Hasonlóság alapján egyesítés

II. A 3D KÉPFELDOLGOZÁS ALAPJAI

7. Matematikai Alapok, Projektív Geometria

39. Adja meg egy inhomogén lineáris egyenletrendszer általános alakját! Mutassa be a legkisebb négyzetek (LS) módszerét!

40. Hogyan oldható meg egy homogén lineáris egyenletrendszer LS módszerrel? Mit értünk szinguláris érték és vektor alatt?

Lineáris egyenletrendszer: $\mathbf{A} \cdot \underline{x} = \underline{b}$, ahol $\mathbf{A} \in \mathbb{R}^{m \times n}$; $\underline{x} \in \mathbb{R}^n$; $\underline{b} \in \mathbb{R}^m$

$\mathbf{A}$: az együtthatómátrix, ezt vizsgálhatjuk.

Az egyenletrendszer megoldása az oszlopvektorok lineáris kombinációja: $\underline{a}_1 x_1 + \underline{a}_2 x_2 + \dots + \underline{a}_n x_n = \underline{b}$

A lineáris egyenletrendszer megoldható, ha $\underline{b}$ előáll az $\mathbf{A}$ mátrix oszlopvektorainak lineáris kombinációjaként, azaz benne van $\mathbf{A}$ oszlopterében. A lineáris egyenletrendszer minden $\underline{b} \in \mathbb{R}^m$ vektorra megoldható, ha $\mathcal{O}(\mathbf{A}) = \mathbb{R}^m$.

LS módszer

Általában a paraméterek számánál több mérési eredmény áll rendelkezésünkre, de a mérési pontatlanságok és zajok miatt az egyenletrendszer nagyon kis valószínűséggel oldható meg. A megoldás legjobb közelítése az LS (Least Squares, legkisebb négyzetek) módszerrel kapható meg. Nevét arról kapta, hogy az eltérések négyzetösszegét igyekszik minimalizálni. Mivel az oszlopter és transzponált nulltér egymás merőleges kiegészítő alterei, bármely $\underline{b} \in \mathbb{R}^m$ vektor előáll

$\underline{b} = \underline{\sigma} + \underline{n}$ formában, ahol $\underline{\sigma} \in \mathcal{O}(\mathbf{A})$, $\underline{n} \in \mathcal{N}(\mathbf{A}^T)$. Ekkor $\mathbf{A} \cdot \underline{x} = \underline{\sigma}$

$\mathbf{A}^T \cdot \underline{b} = \mathbf{A}^T \cdot \underline{\sigma} + \mathbf{A}^T \cdot \underline{n} = \mathbf{A}^T \cdot \underline{\sigma} = \mathbf{A}^T \cdot (\mathbf{A} \cdot \underline{x})$

$\hat{\underline{x}} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \underline{b}$ a megoldás legjobb közelítése (optimális megoldás, LS becslő)

Az SVD (szinguláris érték) felbontás bázistranszformáció, arra használjuk, hogy a mátrixainkat olyan ortonormált bázisban írassuk fel, ahol diagonálisak. A diagonális mátrixokat azért szeretjük, mert *irányfüggő erősítéseket* jellemeznek (legnagyobb, legkisebb $\rightarrow$ rendszerek stabilitása).

$\mathbf{A} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T$, ahol $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$, $\mathbf{V} \in \mathbb{R}^{n \times n}$

$\mathbf{U}$ és $\mathbf{V}$ bázistranszformációs mátrixok, ortogonálisak (ezért ortonormált bázisokra visznek).

$\mathbf{\Sigma}$: a szinguláris értékekből ($\sigma_1, \sigma_2 \dots \sigma_m$) képzett diagonális mátrix. Konvencionálisan a szinguláris értékek csökkenő sorrendben szerepelnek.

A pozitív $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ számok az r -rangú valós (komplex) $\mathbf{A}$ mátrix **szinguláris értékei**, ha van olyan $\{v_1, \dots, v_r\}$ ortonormált bázisa a sortérnek (konjugáltjának) és $\{u_1, \dots, u_r\}$ ortonormált bázisa az oszlopternek (konjugáltjának), hogy

$\mathbf{A} \cdot v_i = \sigma_i \cdot u_i, \quad i = 1, \dots, r$

A v_i vektorokat bal, az u_i vektorokat jobb **szinguláris vektoroknak** nevezzük.

41. Adja meg a lehetséges geometriai transzformációk típusait, és a mátrixaik általános alakját! Milyen transzformációkat engednek meg és milyen tulajdonságokat őriznek meg az egyes típusok?

Euklideszi transzformáció

- Nincs skálázás

$$T_{eucl} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & t_{13} \\ \sin(\alpha) & \cos(\alpha) & t_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

Hasonlósági transzformáció

- Nincs irányfüggő skálázás
- Nincs nyírás

$$T_{simi} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & t_{13} \\ \sin(\alpha) & \cos(\alpha) & t_{23} \\ 0 & 0 & t_{33} \end{bmatrix}$$

Affin transzformáció

- Megőrzi az ideális pontokat.

$$T_{aff} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ 0 & 0 & t_{33} \end{bmatrix}$$

Projektív transzformáció

$$T_{proj} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix}$$

Geometriák:	Euklideszi	Hasonlósági	Affin	Projektív
Transzformációk				
Eltolás	I	I	I	I
Forgatás	I	I	I	I
Uniform skálázás	X	I	I	I
Nem uniform skálázás	X	X	I	I
Nyírás	X	X	I	I
Perspektív vetítés	X	X	X	I
Invariáns jellemzők				
Hossz	I	X	X	X
Szög	I	I	X	X
Hosszak aránya	I	I	X	X
Párhuzamosság	I	I	I	X
Egybeesés	I	I	I	I
Keresztarány	I	I	I	I

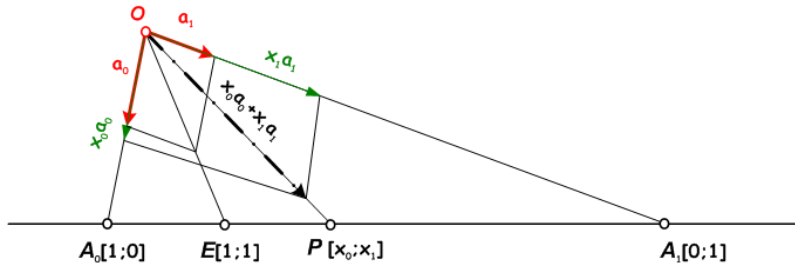
42. Mik a homogén koordináták és mik a használatuk előnyei? Mi az az ideális pont?

Az olyan koordináta-rendszereket nevezik homogénnek, amelyekben a pontot azonosító rendezett pár (hármast, négyest stb.) elemeit egy nullától különböző számmal megszorozva ugyanazt a pontot azonosító párt (hármast, négyest...) kapjuk.

$$(x_0, x_1) = (\mu \cdot x_0, \mu \cdot x_1)$$

A homogén koordináták igazi jelentőségét az adja, hogy használatukkal az ideális térelemek (pont, egyenes, sík) is megadhatók.

Az egyenes (1D-tér) projektív koordinátái



A projektív rendszert az egyenesen kívüli pontból induló két bázis-vektorral (a_0, a_1) adjuk meg. A két bázis egy tetszőleges lineáris kombinációja egy olyan vektort ad meg, aminek egyenese az adott egyenest egy P pontban metszi.

Definíció: Az $[x_0; x_1]$ rendezett pár elemei a P pont koordinátái. A definíció következménye:

$(x_0, x_1) = (\mu \cdot x_0, \mu \cdot x_1)$, mert a μ szorzó az $x_0 a_0$ és az $x_1 a_1$ vektor-komponenseket és ezzel az eredőt arányosan nyújtja meg. A vektor egyenese nem változik, ugyanazt a P pontot jelöli ki.

Ideális pont: A sík két egyenese vagy párhuzamos, vagy pedig van egy közös pontjuk. Ezt a kétféleséget megszüntethetjük azáltal, hogy az egyenes pontjainak az összességét egy *ideális pontnak* (végtelen távoli pont) mondott elemmel bővítjük, mégpedig két egyenes pontjainak az összességét akkor és csak akkor bővítjük ugyanazzal az új elemmel, ha a két egyenes párhuzamos.

Az egy síkban fekvő egyenesek ideális pontjai alkotják a sík **ideális egyenesét**. A párhuzamos síkok ideális egyenesei közősek. A tájképen a vízszintes síkok ideális egyenesének képe a horizont. A tér különböző állású síkjainak ideális egyenesei alkotják a tér (egyetlen) **ideális síkját**.

43. Adja meg a projektív térből a projektív síkra történő vetítés egyenletét! Mi az eltűnő pont?

Vetítés a projektív térből a projektív síkra: $P_3 \rightarrow P_2$

Egyenlet

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix}$$

Eltűnő pont

Párhuzamos P_3 -beli egyenesek metszéspontja, az adott irányban lévő ideális pont.

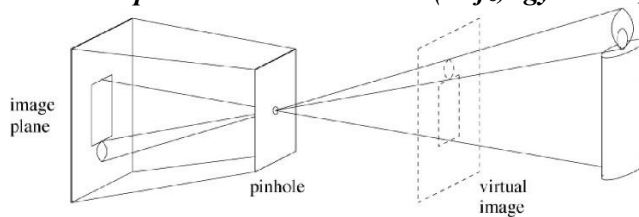
$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 0 \end{bmatrix}$$

A P_3 -beli ideális pont képe P_2 -ben nem biztos, hogy ideális pont lesz!

- Csak, ha (X, Y, Z) merőleges (t_{31}, t_{32}, t_{33}) -ra
- Ha az egyenesek párhuzamosak a képsíkkal!

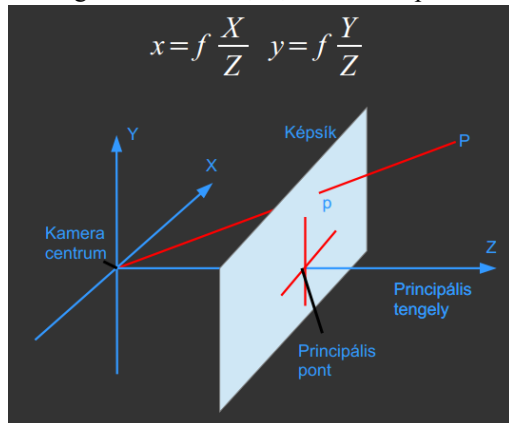
8. Kameramodellek, kalibráció

44. Ismertesse a pinhole kameramodellt! (Rajz, egyenletek, paraméterek, mátrixok)



Távolabbi objektumok kisebbek, párhuzamos egyenesek találkoznak.

A világ koordinátái: X, Y, Z A kép koordinátái: x, y



Vetítés egyenlete: $x = A \cdot [R \ t] \cdot X = P \cdot X$

A vetítés mátrixa: $P = A \cdot [R \ t]$

A vetítés mátrixa felbontható belső és külső paraméterekre.

- **Külső:** $T_{ext} = [R \ t]$ - A kamera helyzetétől függenek a világ koordinátarendszerben
- **Belső:** A - A kamera paraméterei (kameramátrix)
 - Pixelsűrűség (pix/mm) (k, l)
 - Fókusz táv (f)
 - Skew (Θ)
 - Principális pont (p_x, p_y)

$$A = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

45. Mi a kamerakalibráció célja? Ismertesse a 3D markeres kalibráció elvét és mutassa be a lépéseit!

Ha a képeket a későbbiekben analizálni akarjuk, akkor a beállítás pontosságát, a vizsgált terület méreteit és azonosítási pontjait fel kell térképeznünk. Szeretnénk a kamera belső és külső paramétereit kiszámolni!

Elve: Van egy 3D objektumunk, azon előre ismert markerek. A markerek képét akarjuk meghatározni.

Lépései:

1. Markerek megkeresése
 - Sarokdetektálással
 - Egyenesek metszéspontja
2. P meghatározása (projekciós mátrix)
3. A, R, t meghatározása (külső-belső paraméterek)
4. Becslések finomítása

46. Mi a kamerakalibráció célja? Ismertesse a sakktáblás kalibráció elvét és mutassa be a lépéseit! (azokat is, amelyek a 3D markeres kalibrációban ugyanazok)

Elve: gyakran nincsen 3D kalibrációs objektum, ekkor 2D objektumot használunk.

Ismert számú sakktáblászerűen elhelyezkedő markerünk van. Az egymástól mért távolságuk is ismert.

Lépései:

- Sarokdetektálás
 - Mivel nincs 3D információnk a markerekről, P mátrix nem határozható meg egyértelműen.
- Homográfia (H mátrix meghatározása DLT-vel)
- Kameramátrix meghatározása
- Torzítások figyelembevétele

47. Hogyan lehet a kamerák torzítását figyelembe venni a kalibráció során?

Csak a radiális torzítást vesszük figyelembe, mert általában ez a domináns.

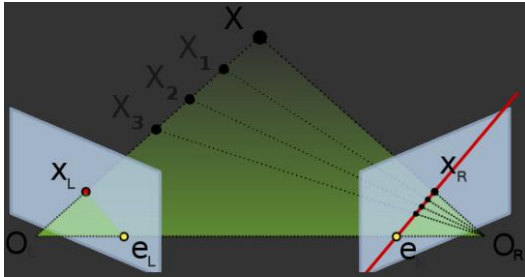
$$\begin{aligned}\delta_x &= x(k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots) \\ \delta_y &= y(k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots) \\ r^2 &= x^2 + y^2\end{aligned}$$

$$\begin{aligned}\tilde{u} &= u + (u - p_x)(k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots) \\ \tilde{v} &= v + (v - p_y)(k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots)\end{aligned}$$

Az ideális és a tényleges pont közti kapcsolat:

48. Hogyan néz ki a sztereó elrendezés? Mi az epipoláris egyenes? Mi az esszenciális és a fundamentális mátrix?

Sztereó elrendezés



Epipoláris egyenes: Azok az egyenesek, amelyeken mentén az elmozdulás történik! (piros)

- Nem biztos, hogy a képen is egyenesek (torzítás, stb)
- Rektifikáció

Az esszenciális mátrix (E) csak a két képsík koordináta rendszer közötti kapcsolatot írja le, a fundamentális mátrix (F) a pixelkoordináták közötti összefüggést is tartalmazza. Ez tehát azt jelenti, hogy az esszenciális mátrixhoz a külső, míg a fundamentális mátrixhoz a külső és belső paraméterek is szükségesek. F mátrix 3x3-as, de rangja 2 kell legyen, nem invertálható. A fundamentális mátrix segítségével felírható az úgynevezett epipoláris összefüggés egyenlete:

$P_L^T \cdot F \cdot P_R = 0$, ahol P_L^T a bal oldali képpont transzponáltja, P_R pedig a jobb oldali képpont. Abban az esetben, ha már vannak képpontjaink, ezen egyenletet felhasználva írható képpontpárosító algoritmus. A fundamentális és esszenciális mátrix egymásból kiszámítható, ha ismerjük a sztereó rendszer külső és belső paramétereit. Amennyiben valamelyik paraméterkészlet hiányzik, a mátrixokat csak megközelítőleg tudjuk kiszámolni.

49. Mi a sztereó kalibráció célja? Ismertesse a 7 és 8 pontos kalibrációs algoritmusok elvét és mutassa be a lépéseiket!

Képalkotás: $P_3 \rightarrow P_2$

Ha vissza akarjuk állítani a térbeli (3D) információt, legalább két kép kell. A két kép közti kapcsolatot fel kell tární!

7 pontos elve:

Fundamentális mátrix szinguláris mátrix kell, hogy legyen. Plusz egy megkötés $\rightarrow$ 7 pont elég

Lépések

1. DLT módszerrel általános megoldás, 2 dimenziós megoldáshalmaz
2. Megoldás alakja: $f = \lambda \cdot f_0 + \mu \cdot f_1$
3. $\det(F)=0$ feltételből a paraméterek számolása

8 pontos elve:

F szinguláris kell, hogy legyen: 8 pontos módszerrel nem lesz az!

Megoldás

1. Kiszámoljuk F SVD felbontását
2. A legkisebb szinguláris értéket nullára állítjuk
3. A kapott F' lesz az F-hez legközelebb lévő szinguláris mátrix

A 8 pontos kalibráció rossz eredményt ad zaj esetén! Érzékeny a skálára és az origó megválasztására!

Pontok normalizálása:

- A pontokat úgy toljuk el, hogy az átlaguk pont az origóban legyen!
- A skálafaktort úgy válasszuk meg, hogy a pontok RMS távolsága az origótól pont $\sqrt{2}$ legyen.

SOHA NE HASZNÁLJUK A 8 PONTOS ALGORITMUST NORMALIZÁLATLAN PONTOKKAL!!!

50. Ismertesse a RANSAC algoritmust (elv, működés, előnyök, hátrányok) és a felhasználási lehetőségeit!

RANdom SAMple Consensus

Felhasználási lehetőségek:

Univerzális algoritmus, iteratív robusztus becslési eljárás, nem csak kalibrációra, hanem nagyon sok egyéb paraméterillesztési feladatra használható. Kalibráció, 2D/3D alakfelismerés. Nem determinisztikus, csak egy adott valószínűséggel ad helyes eredményt, ez a valószínűség az iterációk számának növelésével növelhető.

Lépései:

Kiindulás: sok pontpárunk van

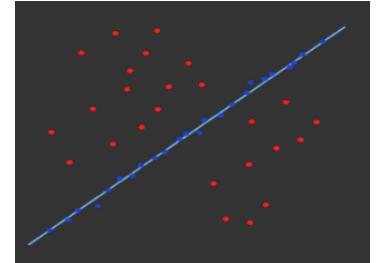
1. Véletlen 7/8 pontpár halmazok kiválasztása
2. 7/8 pontos kalibráció elvégzése minden halmazra
3. Megnézni, hogy hány pontpár illeszkedik kis hibával az adott fundamentális mátrixra
4. A legjobb kiválasztása

Előnyök:

- Egyszerű
- 50%-nál nagyobb arányú outlier (piros pontok) esetén is működik
- Zajos esetben is jól használható, ekkora mértékű zaj esetén az LS becslés teljesen csődöt mond

Hátrányok:

- Véletlenszerű, nem garantált, hogy megtalálja a jó megoldást!
- Ha szeretnénk biztosra menni, akkor nagyon sok véletlen jelöltet kell állítani: nagyon lassú tud lenni!



9. 3D rekonstrukció

51. Adja meg és röviden ismertesse a 3D rekonstrukció lépéseit!

2D képek → Pontfelhő (3D pontok halmaza, lehet színük is)

Pontfelhő → Magasabb szintű leírás (Primitívek; Görbék, felületek; Mesh)

1. Kalibráció
2. Rektifikáció (bővebben a következő kérdésben)
3. Elmozdulás (diszparitás) meghatározása
4. Visszavetítés 3D térbe (projekció/háromszögelés)

52. Mi a rektifikáció? Mi a jelentősége?

Megfeleltetések keresése a két képen nehéz.

Egyszerűsítés: Torzítuk el úgy a képeket, hogy köztük csak az egyik (általában az x) irányban legyen elmozdulás.

Vetítsük a képeket egy közös képsíkra (projektív transzformáció).

Az epipoláris egyenesek mind vízszintesek lesznek, így csak az egyik irányban kell megfeleltetéseket keresni!

Torzítatlan eset: Lineáris transzformáció.

Torzítás esetén először azt kell eltüntetni.

Torzítás térkép: Torzítás paramétereit alapján pixel szinten megadott transzformáció.



Sour

53. Mi a diszparitás? Milyen módszereket ismer a képzésére?

Az adott pixel és a másik képen megtalált párja közötti x irányú távolság.
Nem rektifikált képen hibás lesz! A vízszintes irány mentén minden pixelhez/kulcsponthoz megpróbálunk párt keresni a másik képen → Diszparitás kép

Módszerek:

- Pixel szintű
 - Block Matching
 - Optical flow
 - Belief Propagation
- Képjellemező alapú
 - SIFT, egyéb algoritmusok

54. Hogyan működik a Block Matching (BM) algoritmus? Milyen probléma megoldására használható még fel? Milyen módszereket ismer a diszparitás simaságának biztosítására?

Brute Force módszer:

Vesszük az adott pixel egy kis környezetét, és megkeressük a másik képen azt a környezetet, ami a legjobban illeszkedik. Több hibafüggvény közül választhatunk (pl. négyzetes hiba, abszolút hiba összege).

Probléma:

A BM algoritmus nem veszi figyelembe a közeli pixelek diszparitását. A diszparitásban gyakori, nagy szakadások lehetnek. A valóságban a diszparitás általában sima, mivel a kamerától való távolság is csak ritkán ugrik (objektumhatárok).

Megoldás:

A BM eredménye mellé vegyünk fel valamilyen globális, simasági kritériumot!

55. Hogyan lehet egy pont 3D koordinátáit megkapni, amennyiben a kamerák paraméterei, valamint a pont képei ismertek?

Reprojekció: $(x,y,d) \rightarrow (X,Y,Z)$

Egy kalibrált kamera esetén ismerjük azt a sugarat, amelyre az adott pixelhez tartozó 3D pont esik.

Két kép → pontpárok → két sugár van minden pixelhez (Háromszögelés)

56. Mi(k) a metrikus rekonstrukció lehetőségességének feltétele(i)? Mi a rekonstrukció bizonytalansága, ha ezek csak részben vagy egyáltalán nem teljesülnek?

A 3D pontok helyzetét egy előre rögzített koordináta-rendszerben, ismert mértékegység szerint (m, cm, inch) megadhatjuk. Feltétele, hogy mind a belső (A), mind a külső ($[R \ t]$) paraméterek ismertek legyenek.

Ha a külső paraméterek nem ismertek: A rekonstrukció csak egy skálafaktor erejéig lesz egyértelmű!

Kalibrálatlan kamera: Se a külső, se a belső paraméterek nem ismertek. A rekonstrukció egy ismeretlen projektív transzformáció erejéig lesz egyértelmű.

10. Képilllesztés

57. Ismertesse az optikai áramlás (optical flow) algoritmus célját, alapelvét és feltételezéseit! Írja fel az intenzitás áramlás egyenletét! Milyen problémái vannak az algoritmusnak?

Az intenzitásalapú képilllesztés alapelve, hogyha találunk a két képen két ugyanolyan színű területet, akkor ezek egymással megfeleltethetők (ez azonban csak erős feltételezés). Lehetséges módszer az optikai áramlás algoritmus, amikor is pontpárokat két, különböző időben készült felvételen keressük és a mozgást próbáljuk leírni. Az adott pixelt a következő képen az előző pozíciójának közelében keressük. Gradiensképzést alkalmazunk.

Feltételezések:

- Az egyes objektumok intenzitása időben állandó.
- A két képkocka közti elmozdulás kicsi.
- Az egymáshoz közel eső pixelek hasonlóan mozognak.

Az optikai áramlás nem működik:

- Homogén intenzitású területen.
- Túl nagy mozgás esetén.
- Apertúra probléma.

Az intenzitás áramlás egyenlete (az u és v sebességek nem ismertek, 1 egyenlet, 2 ismeretlen):

$$I_x \cdot u + I_y \cdot v + I_t = 0$$

58. Mutassa be a Lucas-Kanade algoritmus elvét és működését! Milyen módszereket ismer a gyors mozgások követésére optikai áramlás segítségével?

Az optikai áramlás algoritmusánál feltételeztük, hogy az egymáshoz közeli pixelek azonos objektumhoz tartoznak, ezért együtt mozognak. Ne a pixelt nézzük, hanem a környezetét! Több pont együttes kezelése esetén az alulhatározott egyenletből túlhatározott lesz (Lucas-Kanade módszer).

A környezet elmozdulásának négyzetes hibáját minimalizáljuk. Parciális derivált = 0 helyen optimum.

$$E = \sum (uI_x + vI_y + I_t)^2 \rightarrow \text{Min} \quad \frac{\partial E}{\partial u} = 0 \quad \frac{\partial E}{\partial v} = 0$$

$$\begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} \cdot \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -\sum I_x I_t \\ -\sum I_y I_t \end{bmatrix}$$

$$H\vec{u} = b \rightarrow \vec{u} = H^{-1}b$$

Az Lucas-Kanade egyenlet megoldható, ha a H sajátértékei nem túl kicsik (vagy nullák) és a H sajátértékeinek aránya nem túl nagy (H jól kondicionált).

Egyéb módszerek a gyors mozgások követésére:

- Iteratív Lucas-Kanade algoritmus.
- Optical Flow piramis.

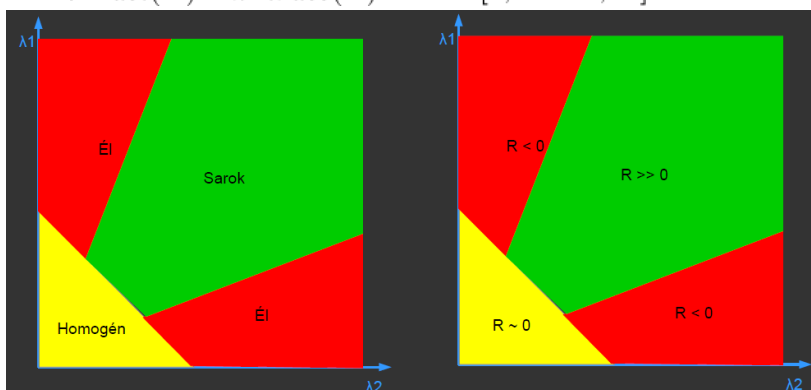
59. Mit nevezünk sarokszerű képrészletnek? Ismertesse a Harris operátor elvét és mutassa be a működését!

- Él: Ahol valamilyen irányban az intenzitásban egy ugrás van.
- Sarok: Ahol az intenzitás minden irányban ugrik!

Harris detektor alapelve:

- Vegyünk egy ablakot a pont körül.
- Mozgassuk el minden irányba, és számoljuk ki a hibát a kép és az ablak közt.
- Ha minden irányban nagy a hiba, akkor sarokpontunk van.

$$R = \det(H) - k \cdot \text{trace}(H)^2 \quad k \in [0,04 - 0,06]$$



60. Milyen képtranszformációkat ismer? Ezek közül melyekre invariáns a Harris operátor?

- Intenzitásváltozás:
 - Additív intenzitásváltozás: $I \rightarrow I + b$
 - Multiplikatív intenzitásváltozás: $I \rightarrow a \cdot I$
 - Csak részlegesen invariáns rá a Harris detektor.
- Forgatás:
 - A sajátvektorok változnak, de a sajátértékek nem!
 - Invariáns rá a Harris detektor.
- Skálázás:
 - Nincs invariancia!

61. Milyen elv alapján lehet skálainvariáns módon képjellemzőket detektálni? Mi az a DoG szűrő, és hogyan számoljuk ki? Mit ad meg a DoG szűrő válasza?

Sarokdetektálást végezzünk el több skálafaktor mellett. Ezt a skálafaktort tároljuk el az egyes jellemzőkhöz.

A DoG szűrő maximuma ott lesz a skálatérben, ahol az objektum mérete egybeesik a szűrő inflexiós pontjával. Ez lesz a jó skála.

62. Mi az a gradienshistogram és hogyan számolható? Hogyan használhatjuk fel SIFT leíró vektor generálására? Milyen módon biztosíthatjuk a rotáció invarianciát?
?

11. Tanuló Látás

63. Mit jelent az objektumfelismerés, illetve az osztályozás? Milyen nehézségekkel kerülünk szembe az egyes feladatok esetén?

Objektumfelismerés

Cél: Valamilyen szempontból fontos részletek megragadása. Általában bonyolultabb, mint egy képjellemző.

Tipikus példák:

- OCR (Optical Character Recognition)
- Arcfelismerés
- Markerek felismerése

Osztályozás

Cél: A képen különböző objektum kategóriák detektálása. Általános kategóriák (ember, szék, autó, stb.), nem konkrét objektumok.

Nehézség

- Az egyes kategóriákon belül nagy variáció (sokféle szék lehet) → Nehéz a releváns jellemzők megragadása.
- Különböző nézőpont
- Transzformáció

64. Mit értünk tanuló algoritmus alatt? Milyen elemekből épül fel? Milyen típusait ismeri a gépi tanulásnak?

A mesterséges intelligencia egy fajtája. Explicit programozás nélkül old meg bizonyos feladatokat. A probléma modellezése a fejlesztő feladata. A konkrét megoldást a tanító adatok alapján határozzuk meg. A paramétereket a tanítás során hangoljuk (kivéve hiperparaméterek). Általában költségfüggvényt minimalizál (pl négyzetes hiba).

Típusai

1. **Felügyelt tanulás** (Supervised Learning): A tanítást példák alapján csináljuk → Az algoritmus ismeri a tanító példákra adandó helyes választ.
2. **Felügyelet nélküli tanítás** (Unsupervised Learning): Itt is vannak példák, de a helyes válasz nem ismert.
3. **Megerősítési tanulás** (Reinforcement Learning): A működés során döntések sorozatát kell meghozni, de visszajelzés csak a sorozat végén lehetséges. Pl.: Járművezetés, Sakk

Feladatok

- **Felügyelt tanulás**
 - Osztályozás: Bináris/több állapotú címke eltalálása
 - Regresszió: Folytonos érték megbecslése
- **Felügyelet nélküli tanulás**
 - Reprezentációk tanulása
 - Klaszterezés (Clustering)
 - Dimenzió redukció
 - PCA – Főkomponens Analízis (Principal Component Analysis)
 - Mély Tanulás (Deep Learning)
 - Autoencoder, RBM (Restricted Boltzmann Machine) DBN (Deep Belief Networks)

Bináris osztályozás

Cél: Egy olyan becslő elkészítése, amely az adatvektorok alapján képes azok címkéit meghatározni.

- A becslő struktúrájáról mi döntünk
- A paramétereit tanulás segítségével határozzuk meg

Nem bináris eset

Építkezzünk bináris osztályozókból! Stratégiák:

- 1 vs. mindenki más típusú osztályozók
- 1 vs. 1 típusú osztályozók: szavazásos módszer

65. Milyen gyakori problémák adódhatnak gépi tanulás használata esetén. Milyen technikák/mérőszámok segítségével háríthatóak ezek el?

Underfitting: Az algoritmus nem elég komplex
Tanítási és tesztelési hiba is nagy

Overfitting: Az algoritmus túl komplex („magol”)
Kicsi tanítási hiba, rossz működés a gyakorlatban

Megoldás: Regularizáció, kereszt-validáció

66. Ismertesse a Bayes hálók felépítését! Mit jelképeznek a csomópontok, illetve az élek? Mi az inferencia, a diagnózis, illetve a predikció? Milyen problémák esetén használható?

Bayes háló

- Irányított körmentes gráf
- Csomópontok: Változók (általában diszkrét)
- Élek: Változók közti ok-okozati kapcsolatok (feltételes eloszlásfüggvény)

Inferencia

Új adat (bizonyíték) érkezik → A háló sűrűségfüggvényeit frissíteni kell! A csomópontok üzennek egymásnak (mindkét irányban): Belief Propagation

Diagnózis

A hálón visszafelé haladunk. A „tünetekből” következtetünk az okokra.

Predikció

A hálón előre haladunk. Az okok ismeretében következtetünk az okozatokra.

Hogyan lehet a kívánt változó értékét megjósolni/előállítani a többi változó segítségével?

Pl. Előző hét időjárása + mérések → holnapi időjárás

67. Ismertesse a KNN algoritmus működését. Milyen problémák esetén használható?

K Nearest Neighbor – k legközelebbi szomszéd. Az osztályozandó példához legközelebb eső k tanító adatot vesszük figyelembe. Az ő címkéjük alapján dönt az algoritmus az adott problémáról.

68. Adja meg a perceptron/neuron felépítését! Milyen problémák esetén használható? Hogyan képes dönteni, illetve tanulni? Hogyan terjeszthető ki nemlineáris esetre?

69. Ismertesse a Support Vector Machine (SVM) algoritmust! Mikor használható? Mi az a kernel függvény, és mire használható? Hogyan válasszuk ki?

Rés (Margin):

Az elválasztó hipersíkhöz legközelebb eső vektor távolsága a síktól. Minél nagyobb, annál „biztosabb” az elválasztás.

Support vector

Azok a vektorok, amelyek távolsága az elválasztó hipersíktól pont a rés mértéke.

- Mindig legalább kettő SV van – a sík két oldalán
- Miért? Mert különben a síkot arrébb lehet mozgatni, hogy mindkét vektortól ugyanolyan távolságra legyen → a rés növekedni fog.

Tanuló algoritmus

Az elválasztó hipersíkot nem a normálvektorával és az origótól vett távolsággal, hanem a support vektorok segítségével írja le. Maximális résű elválasztó síkot keres. A Perceptronnal ellentétben egyszerre veszi figyelembe az összes tanító adatot.

70. Mi a klaszterezés, mire jó? Ismertessen klaszterező algoritmusokat! Mi az erős és a gyenge hozzárendelés, és hogyan kapcsolódik az egyes algoritmusokhoz?

Klaszterezés

- Felügyelet nélküli tanulási módszer
- Feladat: Adott egy halmaz. Szedjük őket részhalmazokra (klaszterekre), úgy, hogy az valamilyen szempontból optimális legyen. Minden klaszternek sorszáma van → Ez lesz az adat címkéje.

Algoritmusok

1. **k-Means:**
 1. Vegyünk fel k db pontot (ezek lesznek a középpontok) véletlenszerűen
 2. A konvergenciáig:
 1. Minden pontot rendeljünk hozzá ahhoz a klaszterhez, amelyik középpontjához a legközelebb van
 2. A klaszterek új középpontja legyen a hozzá tartozó pontok átlaga
2. **Mixture of Gaussians (MoG)**
 1. A pontok eloszlását képzeljük el k db normális eloszlás összegeként
 2. Határozzuk meg ezeket az eloszlásokat a maximum likelihood elv segítségével

Erős hozzárendelés: Minden adat csak egy klaszterbe tartozik (k-Means)

Gyenge hozzárendelés: Az egyes adatok több klaszterbe is tartozhatnak, eltérő valószínűséggel (MoG)

71. Mi a dimenzió redukció? Mire használható a főkomponens analízis (PCA) algoritmus? Mutassa be a működését! Miben különbözik az LDA algoritmus?

Dimenzió redukció:

Sok pixel nem hordoz információt (képek szélei, háttér, stb). Szomszédos pixelek nem függetlenek (homogén régiók, stb).

Sok dimenzió → Sok számolás, felesleges bonyolultság, jó lenne csökkenteni.

Felesleges dimenziók eldobása, összefüggő dimenziók „összevonása”

PCA (Principal Component Analysis)

Tegyük fel, hogy a változóink nagyjából normális eloszlásúak, 0 középpértékkel .

- Keressük azokat az irányokat, amerre a legnagyobb a szórás.
- Az így definiált új változók függetlenek lesznek.

Cél: Keressük meg azt a bázistranszformációt, ami a standard bázisban felírt adatokat a főkomponensek által meghatározott bázisba viszi.

LDA: Lineáris diszkriminancia analízis. PCA felügyelt verziója.

Nem a teljes varianciát, hanem a különböző osztályok közötti szóródást maximalizáljuk a teljes szóródáshoz képest.

72. Röviden ismertesse a Bag of Words, illetve a konstellációs/rész alapú osztályozók elvét.

Bag of Words

Minden dokumentum egy zsák, amiben szavak találhatóak. A szavak előfordulása, vagy sűrűsége alapján következtetünk a dokumentum témájára.

Lépései

1. Vizuális szavak keresése
2. Vizuális szótár építése
3. Az egyes képek osztályozása

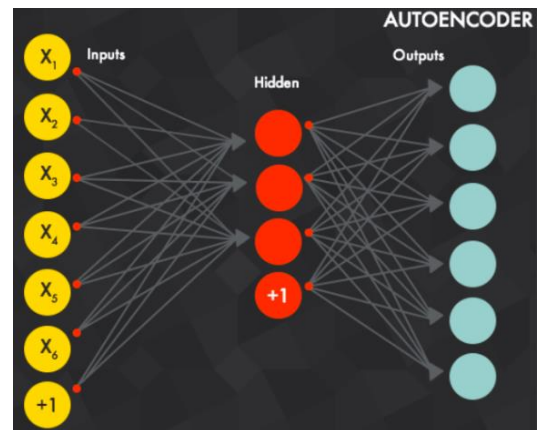
Hátránya

- Képeket osztályozunk: Jelen van / nincs jelen típusú válasz
- Vizuális szavak leírása
 - Csak a jelenlétről van információnk
 - A pozíció nem szerepel sehol
- A BoW nem túl jó lokalizációs (detektálás) feladatra

73. Mi az a mély tanulás, milyen előnyei és hátrányai vannak? Mutassa be az autoencoderek működését.

Cél: Emberihez hasonló tanulás.

Autoencoder: Neurális háló 1 rejtett réteggel.
Be- és kimeneti rétegek elemszáma megegyezik.
Rejtett réteg elemszáma mindig kisebb (tömörít).



74. Mutassa be a konvolúciós neurális hálók felépítését, működését, és előnyeit.

Konvolúciós neurális háló: Kifejezetten a gépi látásra fejlesztett neurális háló.

Gond: Minden pixel egy neuron $\rightarrow$ Rengeteg neuron, hatalmas paraméterszám.

A kép jellemzői alapvetően lokálisak. A neuronok között csak lokális kapcsolat legyen $\rightarrow$ Sokkal kevesebb paraméter.

Ha egy képjellemző detektálása a kép egyik részén fontos, akkor minden másik részen is fontos (elmozgatás invariancia).

Egy rétegen belül minden neuron paramétere azonos. Több millió $\rightarrow$ pár tucat paraméter/réteg

III. A VALÓSIDEJŰ LÁTÁS ESZKÖZEI ÉS PARADIGMÁI

Később...