

Curs 1

1. Introducere

Putem avea o imagine inițială despre tendințele în domeniu dacă accesăm informațiile de la adresele următoare:

- <https://heartbeat.fritz.ai/the-5-computer-vision-techniques-that-will-change-how-you-see-the-world-1ee19334354b>
- <https://blog.robotiq.com/top-10-challenges-for-robot-vision>
- <https://www.skillrary.com/blogs/read/future-of-computer-vision>
- <https://www.technologyreview.com/f/614551/ai-computer-vision-algorithms-on-your-phone-mit-ibm/>

De la telepatie la imunitatea totală față de boli, directorul executiv al Facebook **Mark Zuckerberg** nu s-a sfătuit să facă predicții în cadrul unei sesiuni de întrebări și răspunsuri de pe profilul său de Facebook din 30 iunie 2015. Dar unul dintre visele lui Zuckerberg relative la tehnologie este pe cale să devină realitate: un calculator să poată să descrie o imagine în limbaj natural, de exemplu în limba engleză. Zuckerberg crede că în acest fel se poate schimba modul în care oamenii, în special cei cu probleme de vedere, interacționează cu calculatoarele. Ar fi foarte interesant dacă am putea construi un calculator care să înțeleagă ceea ce se află într-o imagine și să poată să îi spună asta unui om care nu poate să vadă. El speră atunci, în 2015, că acest deziderat se va fi putut realiza în următorii 10 ani.

Imaginiile sunt și un concept cu caracter informațional. Oamenii primesc pe cale vizuală cea mai mare parte din informația pe care sistemul lor senzorial o achiziționează.

Oamenii preistorici au pictat, mult înainte de a scrie, transmițând în acest mod informații selectate. Probabil că am putea identifica aceste picturi din peșteri ca primele imagini artificiale cu aplicabilitate grafică. Multe mii de ani imaginile artificiale au fost create de oameni prin metode grafice și au fost interpretate în mod natural tot de către oameni. Apariția și ulterior dezvoltarea calculatoarelor a oferit posibilitatea ca imaginile artificiale să fie create și memorate sub formă numerică. Noul domeniu este **grafica pe calculator** și el este astăzi foarte cunoscut și utilizat. Informația, sub formă de imagine de sinteză, furnizată de o aplicație de grafică pe calculator este legată de lumea înconjurătoare numai în măsura în care se dorește ca această realitate să fie mai mult sau mai puțin copiată. Decodificarea informației transmise prin aceste imagini este făcută de oamenii care le privesc și care sunt utilizatorii finali ai aplicației. Din acest punct de vedere, imaginile au o rezoluție din ce în ce mai bună pentru a putea reprezenta cât mai bine detaliile. Numărul de culori este deasemenea din ce în ce mai mare pentru ca fidelitatea reprezentărilor să fie susținută și din acest punct de vedere. Se crează în acest mod o realitate virtuală aflată sub controlul total al programatorului care a proiectat-o. Dacă ar fi necesar atunci s-ar putea oferi un text explicativ care să descrie toate obiectele incluse în scena virtuală reprezentată printr-o astfel de imagine pentru că totul este cunoscut aprioric de către programator.

Dezvoltarea tehnicii a creat, chiar înaintea calculatoarelor, senzori care să ofere semnale electrice dependente de interacțiunea dintre lumină și o scenă cu obiecte. A apărut astfel posibilitatea de a produce imagini artificiale care ulterior să fie afișate pe dispozitive speciale și să fie privite de oameni. Conținutul acestor imagini achiziționate este puternic și direct legat de scena pe care o reprezintă. Calitatea imaginilor de acest tip nu este, în majoritatea cazurilor, satisfăcătoare și ca urmare semnalele corespunzătoare trebuie prelucrate pentru a elimina zgomotele sau pentru a

amplifica alte caracteristici utile, de exemplu contrastul. Cele două etape, **achiziția și prelucrarea imaginilor**, își păstrează scopurile generale, dar își adaptează tehniciile și metodele, și atunci când sistemul este completat cu un calculator și formatul imaginii este unul numeric. Imaginele achiziționate pot fi afișate, ca și imaginile de sinteză, pentru a fi privite și analizate de un observator uman. Dacă se dorește să se determine ce conține imaginea achiziționată fără a face apel la un observator uman atunci apare o problemă nouă, interpretarea sau înțelegerea imaginilor folosind tehnici de inteligență artificială. Rezolvarea acestei probleme presupune adăugarea unei noi etape după achiziția și prelucrarea imaginii. Noua etapă realizează o decodificare a informației conținute într-o imagine și poate fi asociată termenului generic de **recunoaștere a imaginilor**. Pe lângă fotografie și cinematografie, din punct de vedere istoric, primele aplicații care au necesitat achiziția și prelucrarea imaginilor au fost aplicațiile militare, spațiale și medicale. Cele trei etape amintite anterior, achiziția, prelucrarea și recunoașterea imaginilor, se constituie în problemele principale ale domeniului vederii artificiale. O tendință naturală ar fi aceea de a lucra cu imagini cu o cât mai bună rezoluție spațială și cu cât mai multe culori. În multe cazuri, echipamentele disponibile azi nu pot să-și îndeplinească sarcinile atât de repede pe cât este necesar și atunci vom constata că în aplicațiile de vedere artificială imaginile au o rezoluție redusă și sunt reprezentate cu nivele de gri. Privind numai imaginile folosite, s-ar putea trage concluzia că vederea artificială este mai puțin spectaculoasă decât grafica pe calculator. Totuși, identificarea și localizarea automate ale unui obiect, care pe baza acestor informații va fi manipulat corespunzător de un robot, este un exemplu de aplicație care ne poate trezi interesul.

Deși lucrează cu imagini și folosesc multe tehnici și metode comune, cele două domenii amintite mai sus, grafica pe calculator și vederea artificială, păstrează o complementaritate care le diferențiază principal.

Sistemele de vedere artificială au fost dezvoltate având ca model inevitabil sistemele biologice și în special sistemul uman. O astfel de abordare este limitată din start de două aspecte. În primul rând, **sistemul uman dispune de un avantaj net în ceea ce privește modul de prelucrare a informației** pentru că se bazează pe un număr impresionant de unități de prelucrare (nu de calcul), neuronii, care procesează în paralel informația codificată într-o manieră mult mai complexă decât aceea numerică și binară. Sistemele artificiale sunt obligate să folosesc formatul numeric pentru informație pentru că echipamentele care sunt disponibile sunt calculatoarele numerice. În al doilea rând, tentativa de a copia propriul nostru mod de gândire este limitată de capacitatea noastră discutabilă de **autocunoaștere**. Noi oamenii, recunoaștem cu ușurință un obiect, chiar dacă este parțial ascuns de alte obiecte, dar nu putem să explicăm cum am realizat această recunoaștere și ca urmare nu putem să copiem metoda într-un algoritm sau într-o tehnică de inteligență artificială. Mai mult, datorită capacitații de învățare, de generalizare și a intuiției oamenii pot recunoaște și obiecte pe care nu le-au mai văzut. De exemplu, oricine se poate pronunța în legătură cu un fruct chiar dacă el este exotic, spre exemplu avogado, și nu îl va confunda niciodată cu un bec chiar dacă nu l-a mai vazut niciodată! Pentru sistemele de vedere artificiale recunoașterea se reduce, în cele mai multe aplicații actuale, la o clasificare adică la identificarea unui obiect inițial necunoscut prin detectarea asemănării cu unul din obiectele învățate inițial.

2. Achiziția imaginii

2.1 Introducere

Indiferent de domeniul de aplicație, utilizarea calculatoarelor impune un format numeric pentru informația prelucrată. Vederea artificială respectă această regulă și lucrează cu imagini numerice.

Pentru a obține imaginea numerică (digital image) este nevoie de două elemente: **senzorul și convertorul (digitizer)**.

Senzorul este un element fizic care **percepe o radiație electromagnetică** într-o anumită bandă de frecvență (ultraviolete, infraroșu, vizibilă, raze X). **Ieșirea senzorului este o mărime electrică**, în general o tensiune, proporțională cu energia radiației electromagnetice primită de senzor. În funcție de aplicație, banda de frecvență și senzorul aferent sunt diferite.

Convertorul este un dispozitiv electronic care primește mărimea electrică furnizată de senzor și o transformă într-un **format numeric**.

Pentru aplicațiile de vedere artificială, cel mai frecvent, senzorii sunt identificați cu numele generic de camere de luat vederi. Realitatea observată se prezintă atunci sub forma unei **scene tridimensionale** care este iluminată natural sau artificial. Lumina reflectată este absorbită de senzor a cărui suprafață activă este **bidimensională**. Desigur, se pot imagina și alte situații particulare, dar cazul prezentat mai sus este cel mai frecvent și el are avantajul că poate fi pus ușor în corespondență, și deci înțeles, cu vederea umană. Cele mai cunoscute camere de luat vederi au fost, în ordinea apariției și utilizării istorice, camerele vidicon și camerele matriceale (solid state).

Lumina este focalizată pe suprafața senzorului prin intermediul unui sistem de **lentile** la nivelul căruia are loc o transformare optică care depinde de natura și caracteristicile lentilei, respectiv ale luminii și sursei de lumină. O condiție pentru obținerea unei imagini ideale este ca planul de focalizare să coincidă cu planul senzorului.

Camerele vidicon (numai sunt utilizate azi, dar au stat la baza dezvoltării unor standarde care încă mai produc efecte și astăzi), au fost proiectate pe baza principiului fotoconductivității. Pe suprafața tubului vidicon se crează o distribuție de zone cu conductivitate variabilă în funcție de intensitatea luminii cu care aceste zone vin în contact. Un fascicol controlat de electroni care baleiază suprafața tubului va genera un semnal electric variabil în timp în funcție de intensitatea luminoasă a zonelor din scena reală. Controlul modalității de baleiere permite păstrarea unei corespondențe cunoscute între **poziția geometrică** a zonei de pe suprafața tubului (x,y) și valoarea semnalului electric la un anumit **moment de timp t**. Pe lângă această corespondență **spațiu – timp**, apare și o a doua corespondență între nivelul **intensității luminoase** corespunzătoare zonei (x,y) și **mărimea semnalului electric u(t)**.

Camerele matriceale sunt formate din elemente fotosensibile, senzori individuali. Fiecare element fotosensibil poate fi privit ca o capacitate individuală care acumulează o sarcină mai mare sau mai mică în funcție de intensitatea luminoasă primită. Citirea sarcinilor nu se poate face în același timp pentru toate elementele (este evident că nu putem avea practic, din punct de vedere tehnologic, un circuit electronic cu sute de mii sau milioane de pini) și ca urmare un proces de **baleiere** rămâne necesar. Aici, baleiera este de fapt un transfer al sarcinilor către dispozitivul care face efectiv citirea. Rezultatul va fi tot o mărime electrică variabilă în timp $u(t)$. **Corespondența dintre poziția (x,y) a senzorului individual în matrice și momentul de timp t la care se consideră valoarea $u(t)$ este determinantă pentru geometria imaginii. Corespondența dintre intensitatea luminoasă corespunzătoare poziției (x,y) și valoarea mărimii electrice $u(t)$ este responsabilă pentru generarea unui nivel de gri sau a unei culori corecte.**

Fie un punct **P(X, Y, Z) din scena tridimensională**. Lumina provenită de la o sursă este parțial absorbită și parțial reflectată de acest punct. Lumina reflectată din punctul P este focalizată de sistemul de lentile al camerei pe suprafața senzorului, în punctul **P'(x,y)**. **Corespondența dublă P – P' constituie problema transformării geometrice**. Este intuitiv clar că pentru un sistem de lentile

cunoscut și un punct P dat se poate determina în mod unic punctul P' corespunzător de pe suprafața senzorului. **Invers problema este mai complicată**, dacă se dă un punct P' de pe suprafața senzorului atunci există o infinitate de puncte P din scena tridimensională al căror corespondent poate fi P' . Toate aceste puncte P sunt așezate pe dreapta care pleacă din P' și trece prin centrul lentilei.

În urma conversiei analog numerice se va obține o matrice cu elemente numerice **$I(i,j)$ numite pixeli (picture element)**. Această conversie analog numerică presupune două procese. **Eșantionarea, sau discretizarea spațială**, este procesul care pune în evidență corespondența dintre momentul t la care se face conversia și poziția (i,j) corespunzătoare din imagine. **Cuantificarea, sau discretizarea în amplitudine**, este procesul care asociază fiecărui nivel al mărimii continue $u(t)$ o valoare numerică $I(i,j)$ pe care o numim nivel de gri sau culoare.

2.2 Sistemul de vedere uman

În foarte multe domenii de aplicații este prezentă o tendință inevitabilă, aceea de **a folosi ca model sistemele biologice**, în special sistemele umane. Aplicațiile robotice și de vedere artificială, prin natura lor, sunt într-o măsură și mai mare afectate de această tendință. Despre avantajele unei astfel de abordări nu este necesar să insistăm. Putem să subliniem însă limitele proiectelor bazate pe astfel de modele. O primă limitare pare mai degrabă una de ordin filozofic și este legată limitele autocunoașterii în general. Putem să analizăm în detaliu o multitudine de procese fizice și chimice care au loc la nivelul diferențelor sisteme de locomoție și perceptie umane. Problemele se complică în momentul în care dorim să ne folosim propriul creier pentru a afla cum folosim noi însine informațiile disponibile pentru a lua deciziile necesare. Este o problemă extrem de complicată să ne explicăm cum reușim într-un timp foarte scurt să apreciem dacă ceea ce vedem este un scaun sau un fotoliu. Dacă printre-un miracol am intra în posesia "schemei" utilizate atunci ne-am lovi de o a doua limitare. Cu siguranță respectiva "schemă" nu se bazează pe folosirea informației numerice aşa cum suntem noi obligați să facem ca utilizatori de calculatoare. Este instructiv să conștientizăm faptul că încercăm să copiem abilitățile sistemelor biologice bazându-ne pe posibilitatea oferită de sistemele de calcul de a aduna și deplasa cifre binare, 1 și 0.

Sistemul vedere uman are trei componente principale:

- senzorul (**ochiul**),
- calea de transmisie (**nervul optic**) și
- unitatea de prelucrare (**creierul**).

Principial funcționarea poate fi descrisă simplu. Lumina este focalizată cu ajutorul unei lentile pe senzorul propriu-zis (retina). Printre-un proces bioelectrochimic se generează semnal electric ca răspuns la stimulul luminos. Semnalul este trimis pe nervul optic până la creier unde se crează forme neurologice pe care noi le percepem ca imagini și le interpretăm punându-le în legătură cu lumea exterioară. Lumina pe care o poate percepe ochiul uman ocupă **o mică parte a spectrului undelor electromagnetice** (figura 2.1) și corespunde doar frecvențelor cuprinse între aproximativ **400 nm și 800 nm**. Lentila este formată din apă (65 %), grăsimi (6%) și proteine (într-un procent mai mare decât oricare alt țesut). Ea este colorată ușor în galben și absoarbe sub **10 % din spectrul vizibil al luminii**. Razele infraroșii și cele ultraviolete sunt absorbite de proteinele din structura lentilei.

Din punct de vedere senzorial, la nivelul retinei, există o ierarhizare concretizată în specializarea elementelor individuale senzoriale. **"Conurile"** sunt specializate în **vederea diurnă**, sunt sensibile și diferențiate pe **culori**, oferă o foarte mare rezoluție (sensibilitate la detalii) și sunt concentrate în zona centrală a retinei. **"Bastonașele"** sunt specializate în vederea în penumbră, sunt sensibile la

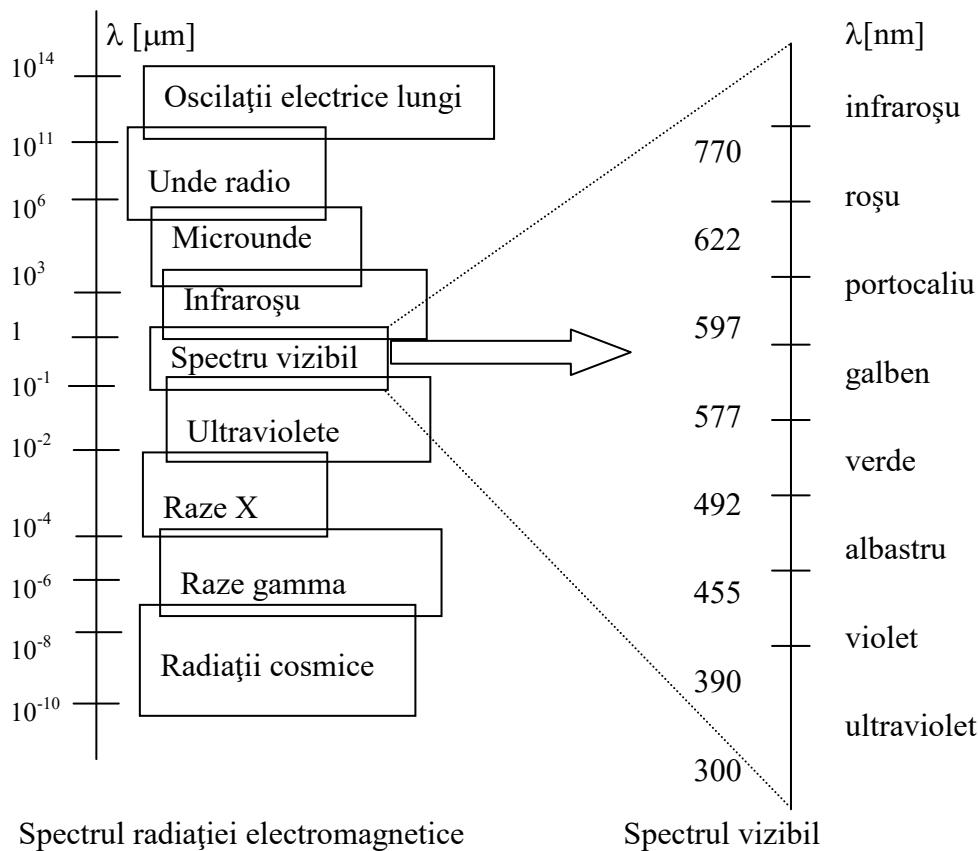


Figura 2.1 Poziția spectrului vizibil în cadrul spectrului radiației electromagnetice

strălucire nu la culoare, sunt distribuite pe toată retina, sunt utilizate în special pentru vederea periferică și oferă o rezoluție mai redusă. „Bastonașele” (75 - 150 milioane) sunt mult mai numeroase decât „conurile” (6 - 7 milioane), dar au conexiuni pe grupe de senzori și nu individuale ca acestea din urmă. Acest tip diferit de conectare este o justificare a rezoluțiilor diferite oferite de cele două grupe de senzori. În apropierea locului de conectare a nervului optic la retină există o zonă lipsită de senzori, zona oarbă. Creierul este capabil să completeze cu informație, prin extrapolare, porțiunea corespunzătoare din imagine astfel încât noi aflăm despre această problemă din cărți și nu o sesizăm direct. Modalitatea a fost împrumutată la fabricarea senzorilor electronici când pentru a nu arunca un senzor care are milioane de celule individuale se completează informația aferentă celor câtorva celule individuale defecte cu informația memorată în celulele adiționale de memorie.

Zona cu cea mai mare sensibilitate de pe retină este zona centrală unde sunt concentrate cele mai multe “conuri”. Această zonă (**fovea**) are o formă aproximativ circulară cu o rază mai mică de 1 mm (aproximativ 0,75 mm). Pentru a putea face o comparație cu senzorii electronici, o putem echivala cu o zonă dreptunghiulară de 1,53 mm x 1,15 mm care păstrează raportul de 4/3 între dimensiuni. Dacă acceptăm că densitatea “conurilor” în această zonă este de aproximativ 150.000 de elemente pe mm^2 , atunci obținem un senzor echivalent de 265.000 de elemente. Dacă luăm în considerare un senzor comun CCD care are o rezoluție de 756 x 581 vom găsi un număr mai mare de celule individuale, distribuite și pe o suprafață mai mare. Cum la nivelul tehnologic din mileniul trei există senzori cu mai multe milioane de celule individuale putem concluziona că **ochiul omenesc nu depășește acești senzori nici ca număr absolut de celule nici ca densitate pe unitatea de suprafață**.

Distanța dintre centrul focal al lentilei și retina variază între 14 mm și 17 mm, după cum se focalizează asupra unor obiecte mai apropiate sau mai depărtate ceea ce conduce la deformarea

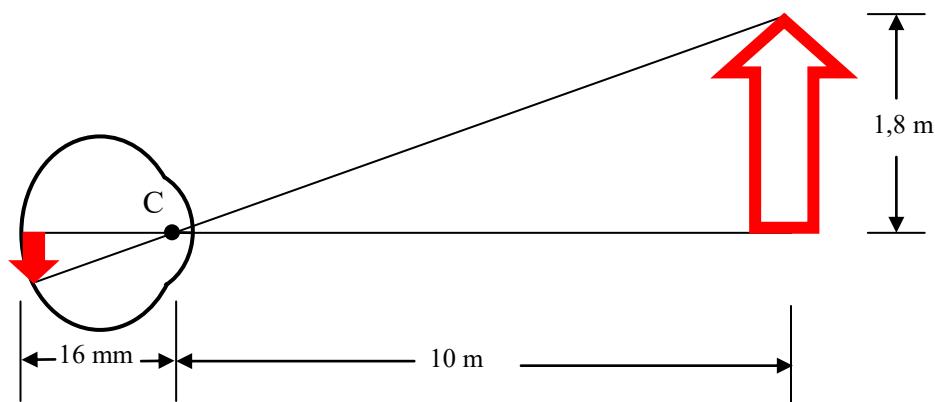
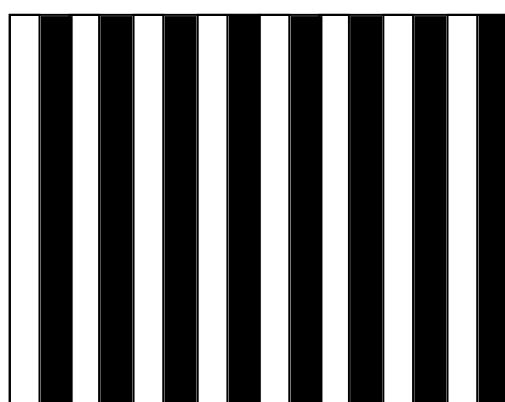


Figura 2.2 Influența distanței focale asupra formării imaginii

lentilei. Ca urmare un om de 1,80 m privit de la 10 m distanță va crea o imagine de aproximativ 2,5 mm (figura 2.2).

Sistemul vizual uman reușește să ofere o capacitate foarte mare de **adaptare** relativ la limitele inferioară și superioară acceptabile pentru intensitatea luminii incidente pe ochi (aceste limite se află într-un raport de 10^{10}). Acest lucru este justificat de caracteristica logaritmică a ochiului (strălucirea subiectivă este o funcție logaritică de intensitatea luminii incidente pe ochi), dar trebuie făcută precizarea că sistemul are nevoie de un timp de adaptare la trecerea de la un nivel de iluminare la un altul mult diferit. Sistemul vizual uman poate sesiza numai aproximativ **20 de schimbări** în strălucirea unei zone mici dintr-o imagine, iar pentru **întreaga imagine sunt necesare în jur de 100 de nivele de gri** pentru ca aceasta să fie apreciată ca realistă. Capacitatea sistemului vizual uman de a distinge detalii (rezoluția spațială) este limitată de mărimea celor individuale, a "conurilor" și "bastonașelor". Este evident că nu se pot sesiza detalii mai mici decât dimensiunea unei astfel de celule. Obiectele mai mici pot genera o imagine de aceeași mărime dacă sunt privite de la o distanță mai mică. **Rezoluția spațială depinde de condițiile de iluminare, crește cu nivelul strălucirii și este mai mare pentru imaginile monocrome.**

Rezoluția spațială (care poate fi definită și drept **capacitatea de a separa doi pixeli vecini dintr-o imagine**) este un concept legat **frecvența spațială**. Dacă privim o imagine formată din benzi verticale negre pe fond alb (figura 2.3) atunci se pune problema câte astfel de schimbări de strălucire putem sesiza. Este clar că este important să precizăm și distanța de la care privim. Pentru a elimina efectul distanței putem calcula această frecvență relativ la 1 grad spațial din câmpul de vedere (**mărimea pumnului privit de la o distanță egală cu lungimea brațului întins este de aproximativ 1 grad spațial**). În aceste condiții se poate aprecia că frecvența de tăiere pentru



$f = 8$

Figura 2.3 Imagine de test pentru rezoluția spațială

sistemul vizual uman este în jur de **50 de ciclii (un ciclu este o schimbare completă în semnal) pe grad.**

Pentru aprecierea percepției imaginilor în mișcare se definește **răspunsul temporal** al sistemului vizual uman care este capacitatea sistemului de a răspunde la informația vizuală primită ca funcție de timp. Ca o măsură a performanțelor, putem preciza că pentru o imagine afișată pe un monitor sistemul vizual uman nu poate percepe conștient pâlpâiri (flicker) dacă frecvența de afișare a cadrelor este mai mare de **60 Hz**.

Dacă, între anumite limite, fiziologia și aprecierea performanțelor sistemului senzorial de vedere uman sunt accesibile, mai dificilă se arată posibilitatea de înțelegere a modului cum această informație este prelucrată și utilizată de sistemul de prelucrare, care are ca element central creierul uman. Pe baza **informației vizuale, care reprezintă mai mult de 80% din totalul intrărilor senzoriale**, omul reușește, într-o primă etapă, să identifice și să localizeze obiectele din mediul înconjurător. Desigur, înțelegerea scenelor private este o operație mult mai complexă și o putem defini ca totalitatea operațiilor care mai pot fi făcute după simpla clasificare a obiectelor.

Putem remarcă faptul că sistemul de vedere este integrat în sistemul global care este omul. În sprijinul acestei afirmații putem face o observație interesantă. Pentru orice sistem de vedere este important să poată localiza obiectele din spațiul aplicației pe baza informației din imaginile achiziționate. Pentru vederea artificială, realizarea acestei **corespondențe între coordonatele globale și coordonatele imagine** se bazează pe o etapă inițială care se numește **calibrarea camerelor** de luat vederi. Calibrarea camerelor produce o matrice de transformare a coordonatelor. Această matrice este valabilă atât timp cât poziția și orientarea camerelor sunt fixe în raport cu anumite repere de referință. Este evident că în creierul omului astfel de corespondențe nu pot fi realizate matriceal! Totuși plasarea celor doi senzori, care sunt ochii, într-o poziție relativ fixă în raport cu scheletul nostru, și deci și cu sistemele noastre de locomoție și manipulare, este o condiție care susține abilitatea noastră de a localiza și manipula corect obiectele. Pentru configurația fixă care este proprie fiecărui individ, corespondența imagine – mediu înconjurător este învățată și permite o foarte bună viteză de reacție. Dacă încercați să manipulați obiecte pe baza unor imagini dintr-o oglindă sau achiziționate cu o cameră TV și afișate pe un monitor, veți avea dificultăți și performanțe modeste. O nouă corespondență imagine – mediu înconjurător trebuie construită și această sarcină necesită un timp de adaptare. Acesta este motivul pentru care nu ne aşteptăm la performanțe remarcabile de la un sistem de vedere plasat în vârful unor antene care se orientează permanent, ca în filmele SF! Nici melcul nu este un vitezist!

Pentru sistemul de vedere uman putem face o apreciere generală a performanțelor oferite:

- **Măsurarea distanțelor este făcută calitativ**, rezultând mai degrabă informații de tipul “mai aproape”, “mai departe”, “mai scurt” sau “mai lung” decât dimensiuni exprimate în unități de măsură. Va putea convinge încercând să apreciați dimensiunile obiectelor din jurul dumneavoastră, în centimetrii. În mod normal veți avea aproximări de circa 10%. Nu avem în vedere iluziile optice care pot fi induse prin diferite mici trucuri.
- **Sesiarea detaliilor tridimensionale (relief) este foarte bună.**
- Aprecierea direcției (orientarea muchiilor) este bună și se face tot **calitativ și nu cantitativ**.
- **Achiziția și prelucrarea** imaginilor în mișcare se face cu o viteză mai mare de **10 imagini într-o secundă**. Este o viteză apreciată ca timp real relativ la acțiunile (de exemplu deplasare, manipulare, supraveghere) pe care le poate întreprinde o persoană.
- **Sesizarea contururilor obiectelor se face cu foarte bună precizie.**
- **Rezoluția (sesizarea detaliilor)** este foarte bună și este superioară rezoluției așteptate în raport cu numărul celulelor vizuale individuale.
- **Culoarea** este percepță în banda de lungimi de undă **400 nm – 800 nm**.

Curs 2

2.3 Senzori de imagine

Prezentarea principală care urmează nu se adresează electroniștilor interesați de soluțiile tehnice utilizate în proiectarea și fabricarea senzorilor. Se dorește introducerea noțiunilor minime necesare pentru înțelegerea fenomenelor care au loc la nivelul diferitelor tipuri de senzori. Această înțelegere constituie baza pentru aprecierea parametrilor unui senzor de imagine și, ca urmare, pentru posibilitatea alegerii senzorului potrivit unei anumite aplicații de vedere artificială.

2.3.1 Vidicon

Acest dispozitiv aparține istoriei, **nu se mai folosește**, dar prezentarea lui foarte scurtă, poate facilita înțelegerea noțiunii de **baleaj**. Primele tipuri de senzori pentru camerele de vederi s-au bazat pe tuburile fotosensibile cu baleaj. Caracteristicile lor, ca și ale tuburilor cinescoape – dispozitivele complementare de afișare a luminii, au influențat însă decisiv standardele de televiziune care sunt valabile și astăzi.

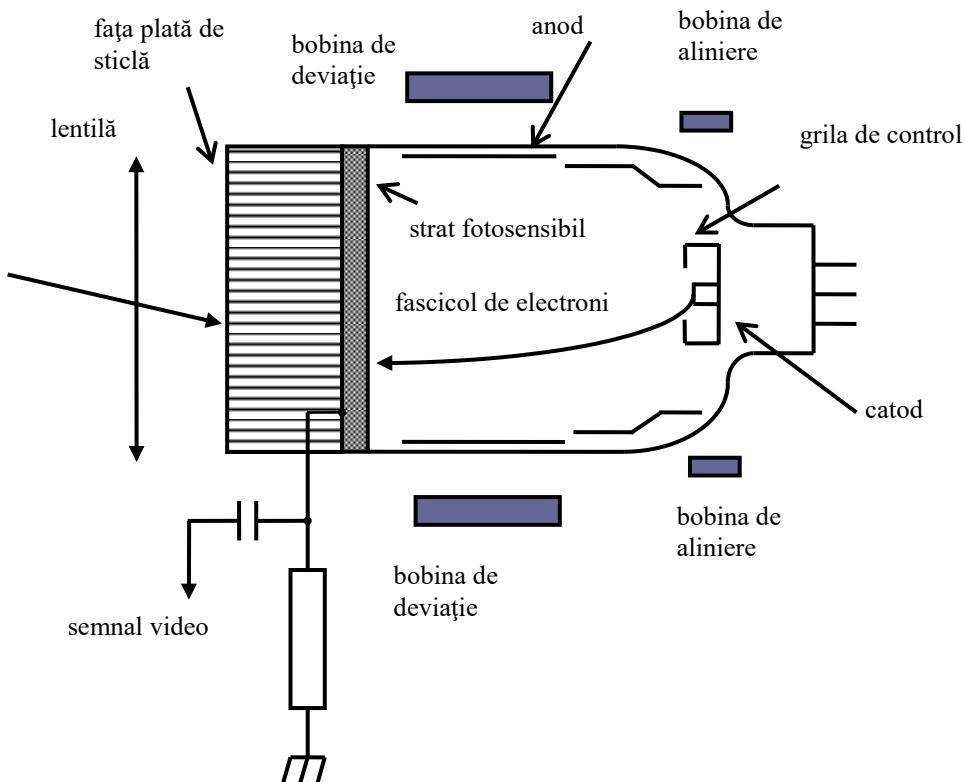


Figura 2.4 Tubul Vidicon

Spre exemplificare prezentăm modul principal de funcționare al tubului Vidicon (figura 2.4). La tubul Vidicon în secțiunea de formare a imaginii, lumina eliberează electroni la trecerea prin fotocatod, prin efect fotoelectric. Acești electroni, ajung la o țintă de sticlă încărcată pozitiv. Ca urmare a bombardamentului cu electroni apare o emisie de sarcini pozitive proporțională cu lumina incidentă la intrare. În secțiunea baleaj, spatele țintei este măturat cu un fascicul de electroni emis de un tun. Acești electroni sunt absorbiți proporțional cu sarcinile pozitive. Fasciculul reflectat este preluat de catod în amonte de secțiunea amplificare. Tubul Vidicon are o țintă de sticlă, fundul tubului, acoperită de un film transparent conductor (electrod) și de un **strat fotosensibil granulat** a

cărui rezistență variază local sub acțiunea luminii. Anodul accelerează electronii către țintă. Stratul fotosensibil are un potențial slab pozitiv, astfel încât să creeze un câmp electric local. Grila este încărcată pozitiv, ceea ce are ca efect frânarea electronilor care trec prin ea. Acești electroni ajung la țintă, stratul fotoconductor, cu viteză aproape nulă. Acolo unde ajung fotonii, pe țintă, rezistivitatea locală scade. Deci curentul rezultat ca urmare a prezenței electronilor pe țintă, va depinde de rezistivitatea stratului, deci de fluxul luminos, și de perioada de baleaj. Se poate lucra cu o perioadă de baleaj constantă (de exemplu 40 ms) pentru ca dependența să se referă numai la fluxul luminos.

Ca dezavantaje ale acestui senzor putem aminti:

- Fenomenul de **blooming** (influența luminii care lovește stratul fotosensibil într-un punct la un moment se răsfrâng și asupra zonelor vecine),
- **Reamanență** importantă (efectul influenței incidente nu dispără instantaneu ceea ce limitează numărul de imagini care pot fi recepționate într-o secundă),
- Tubul poate fi distrus printr-o iluminare deosebit de puternică și din această cauză se folosește greu împreună cu un laser,
- Dimensiune și greutate mari,
- Fiabilitate scăzută datorită încălzirii,
- Consum electric mare,
- **Probleme de reglare a geometriei imaginii, în special la colțuri,**
- Sensibilitate relativ scăzută,
- Dependență neliniară între lumina incidentă și semnalul de ieșire furnizat.
- Senzorul oferă o calitate foarte bună a imaginii din punct de vedere al rezoluției și a culorii.

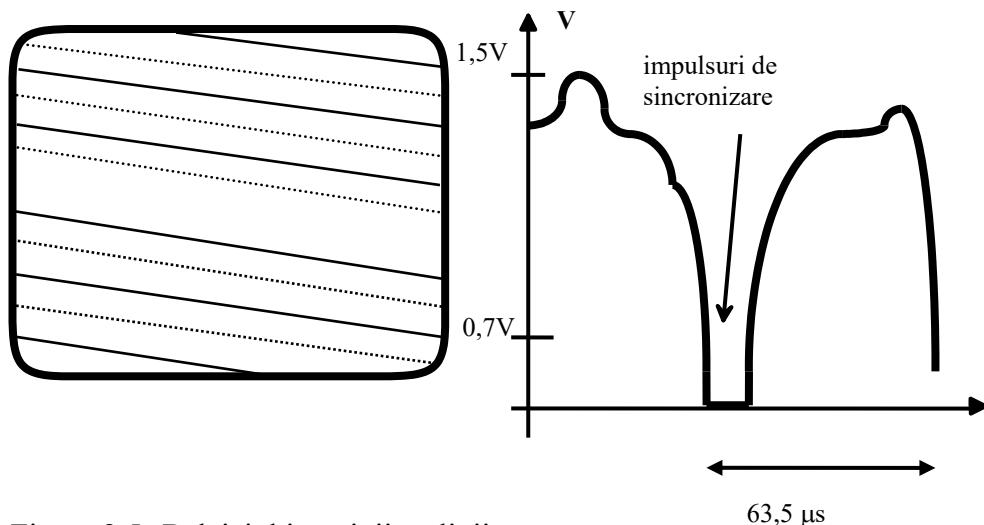


Figura 2.5 Baleajul imaginii pe linii

Pentru a nu avea probleme cu geometria imaginii, ca și la tubul cinescop, ar fi fost necesar ca toate punctele de pe suprafața fotosensibilă să fie egal depărtate de tunul de electroni care face baleajul. Din acest punct de vedere suprafața ideală ar fi fost un pătrat sferic (un petec de pe suprafața sferei mărginit de patru arce egale). Dar, aşa cum vom arăta mai târziu, cu o singură imagine, respectiv cu un senzor de imagine, nu se poate obține informația despre profunzime, adică distanța până la punctele din imagine. Aceasta este explicația naturală a faptului ca noi oameni avem doi ochi. Cum imaginile au fost în totalitate și sunt și azi în bună măsură destinate sistemului nostru vizual (să fie private), un astfel de format pătrat nu este convenabil. **Un prim compromis s-a făcut acceptând un raport între dimensiuni de 3/4.** Acest raport a fost folosit în epoca folosirii tubulilor cu baleaj,

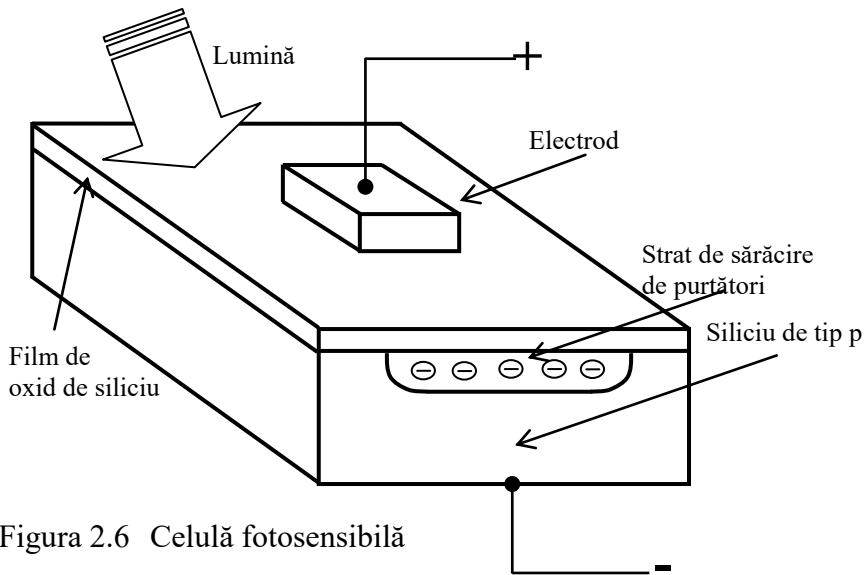
dar a fost preluat și de tehnologiile electronice ulterioare. Standardele de televiziune dezvoltate în această perioadă au ținut cont de performanțele sistemului vizual uman care “vede” imaginile, achiziționate și afișate, dar și de posibilitățile oferite de tuburile cu baleaj.

Standardul corespunzător pentru semnal video monocrom care a fost folosit în SUA și Japonia s-a numit RS-170, un subset al standardului NTSC (National Television Systems Committee). În Europa s-a utilizat standardul CCIR (International Radio Consultative Committee) care este similar dar incompatibil cu RS-170. Ambele standarde au unele inconveniente pentru aplicațiile de vedere artificială pentru că ele au fost dezvoltate pentru aplicații comerciale de televiziune. Spre exemplu în standardul RS - 170 baleierea se face cu o rețea de 525 linii pentru 60 de semicadre pe secundă (30 de imagini întregi pe secundă) și se crează semnalul video propriu-zis: semnale de sincronizare și semnal analogic, impulsurile de sincronizare pentru linii au atunci o frecvență de 15750 Hz. **Pentru a evita efectul de pâlpâire s-a propus o baleiere întrețesută (figura 2.5), împărțind cele 525 de linii în două cadre ("frame") de câte 262½ linii.** Din cele 525 de linii se pierd 22½ pentru întoarcerea baleajului din colțul din dreapta jos în colțul stânag sus. Similar, pentru standardul CCIR baleierea se face cu 625 de linii pe imagine și **50 de semicadre pe secundă** se ajunge la o frecvență de linii de 15 625 Hz. Deci **o imagine completă se obține în 1/25 secunde**, deci în 40 ms. Semnalul de la o linie a imaginii include informația de imagine, impulsuri de sincronizare pentru sfârșitul liniei (sincronizare orizontală) și perioada de "blanking".

2.3.2 Fotodiode și fototranzistori

Cele mai simple componente electronice sensibile la lumină sunt fotodioda și fototranzistorul. O fotodiodă este o joncțiune pn, polarizată în sensul de blocare, la care **unul dintre straturi, să presupunem că p, este subțire și poate fi expus la lumină**. Fotonii eliberează perechi electron - loc la zona de tranziție. Aici câmpul electric este foarte intens, locurile și electronii vor migra și se va stabili un curent proporțional cu fluxul luminos. Alte fotodiode își expun luminii stratul lor n. Un fototranzistor funcționează după un principiu asemănător: fotonii eliberează perechi electron - loc în joncțiunea bază - colector a unui tranzistor npn. Acest flux de perechi va crea un curent de bază proporțional cu fluxul luminos și deci unul de emitor mărit cu factorul de amplificare.

O capacitate MOS, privită ca o celulă fotosenzorială (figura 2.6), utilizează proprietățile fotoelectrice ale materialelor semiconductoare cu siliciu care, în anumite condiții, eliberează electroni (-) și goluri (+), proporțional cu fluxul de fotonii. Sarcinile negative sunt colectate de câmpul electric al unei joncțiuni de siliciu dopat (impurificat). Prin concatenarea unor celule



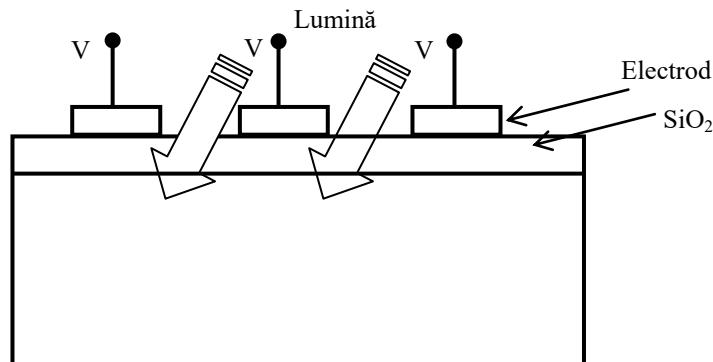


Figura 2.7 Dispunerea liniară a celulelor individuale

individuale se pot obține senzorii liniari și matriceali (figura 2.7). Senzorii individuali sunt izolați între ei și sunt dispuși echidistant sub forma unui vector sau a unei matrice. **Într-un astfel de senzor senzorul individual este chiar pixelul (picture element), dar acesta nu va corespunde practic niciodată cu pixelul din imagine.** Culegerea sarcinilor individuale, transferul lor de-a lungul senzorului și transformarea lor într-un semnal electric sunt operații complexe și delicate realizate prin tehnici diverse care diferențiază tipurile de senzori. În final semnalul analog este furnizat într-un format standardizat. **Conversia analog numerică poate avea loc în camera care conține senzorul sau în afara ei.**

Într-una dintre variantele tehnologice, un strat semiconductor de tip p este acoperit cu un strat subțire de oxid de siliciu. Acest ultim strat este acoperit la randul lui cu benzi conductoare pe post de electrozi. Senzorul este de fapt compus din capacități dispuse echidistant, fiecare capacitate fiind constituită din trei părți: electrodul, metalic; izolantul, din oxid ; un strat inferior de siliciu impurificat p. Dacă se aplică din exterior pe electrozi un potențial V pozitiv în raport cu bara impurificată de siliciu, atunci fiecare capacitate se încarcă și se crează un câmp electric local. Sub fiecare electrod apar acum zone de tranziție. Fotonii elibereză în aceste zone perechi loc - electron. Câmpul local respinge locurile, atrage electronii sub fiecare electrod. **Această acumulare depinde de fluxul luminos pe perioada cât s-a făcut integrarea.**

Aceeași celulă elementară poate fi folosită pentru obținerea senzorilor color care disting culorile componente ale culorii albe. Anumite sisteme disting culorile de bază RGB, roșu, verde și albastru, în timp ce altele lucrează cu culorile complementare YCM, galben, cian și magenta (figura 2.8). Pentru a diferenția culorile sau complementele lor se adaugă **filtre individuale pentru pixeli și culori.**

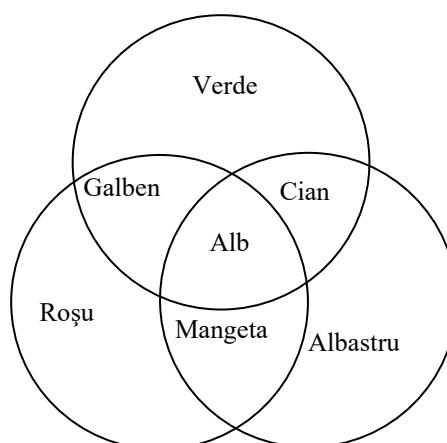


Figura 2.8 Relațiile dintre culorile de bază și culorile complementare

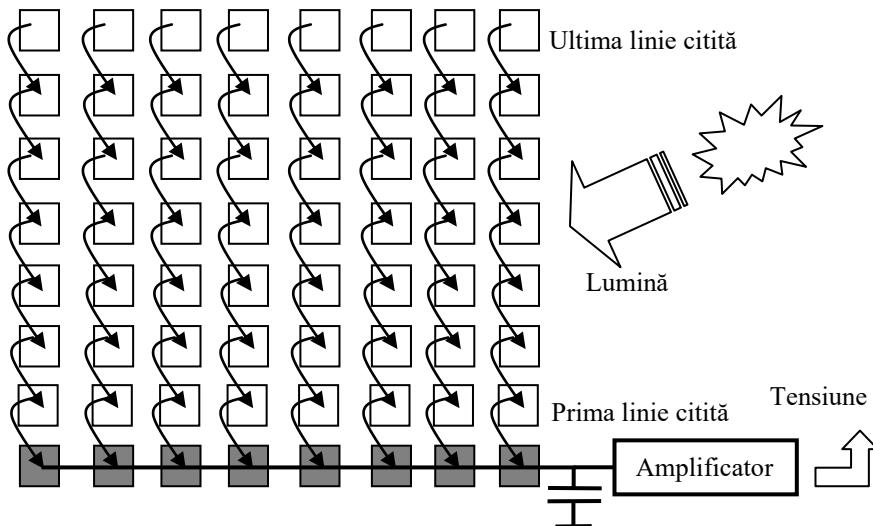


Figura 2.9 Funcționarea principală a senzorului CCD

2.3.3 CCD

Senzorii CCD (Charge coupled devices) captează **lumina în mici fotocelule** și și-au primit numele de la modul în care sarcinile sunt citite după expunere. Pentru aceasta (figura 2.9), mai întâi sarcinile din prima linie sunt transferate într-un registru de citire. De acolo, semnalele sunt preluate de un amplificator și ulterior de un convertor analog – numeric. După ce o linie a fost citită, sarcinile ei din registrul de citire sunt sterse. Următoarea linie va fi transferată în registrul de citire, iar toate liniile sunt transferate cu o linie mai jos. **Sarcinile din fiecare linie sunt cuplate astfel încât la fiecare transfer din linia curentă în linia următoare are loc și un transfer din linia precedentă în linia curentă.** În acest mod se poate citi o linie întreagă la un moment dat. Tehnologia CCD are acum zeci de ani și folosește un proces specializat VLSI bazat pe crearea unei rețele de electrozi de siliciu pe suprafața cipului. Nodurile rețelei sunt atât de mici și de apropiate încât permit păstrarea electronilor până când aceștia sunt mutați fizic din poziția în care lumina incidentă i-a generat, de-a lungul suprafeței cipului, până la un amplificator de ieșire. Pentru a realiza acest proces, rețeaua de electrozi este comandată de un ceas extern senzorului. Din punct de vedere tehnic este posibili, dar nu este rentabil din punct de vedere economic, să se integreze în tehnologia CCD alte funcții necesare funcționării camerei, cum ar fi circuite de ceas, logica de secvențiere, procesare de semnale etc. Aceste funcții sunt, în mod normal, implementate în alt cip. În acest fel se ajunge la soluții tehnice care presupun între 3 și 8 cipuri. Un alt punct nevralgic al tehnologiei CCD este necesitatea semnalelor de ceas cu o amplitudine și de o formă impuse care influențează decisiv performanțele finale ale sistemului. Un cip specializat care să genereze semnale corecte de ceas, ca formă și amplitudine, necesită tensiuni de alimentare nestandardizate și crește puterea consumată. Plecând de la o singură tensiune de alimentare, pentru a genera 5 sau 6 semnale diferite de alimentare, este nevoie de câteva regulatoare interne care evident cresc complexitatea soluției. Aceste inconveniente sunt prețul plătit pentru o imagine de foarte bună calitate.

Din punct de vedere istoric, tehnologia CCD a fost dezvoltată căutând soluții pentru alte probleme decât achiziția imaginilor. În anii 1960 calculatoarele nu dispuneau de sisteme de memorie ieftine și de mare dimensiune. Laboratoarele Bell au propus tehnologia CCD ca o modalitate de stocare de date. În 1974 Fairchaild Electronics au produs primul senzor de imagine CCD cu un format de 100 x 100 pixeli, iar în 1975 s-a produs prima cameră comercială bazată pe această tehnologie. Tot atunci Kurzweil Computer Products a realizat primul scanner bazat pe un senzor liniar CCD de 500 pixeli de la Fairchaild.

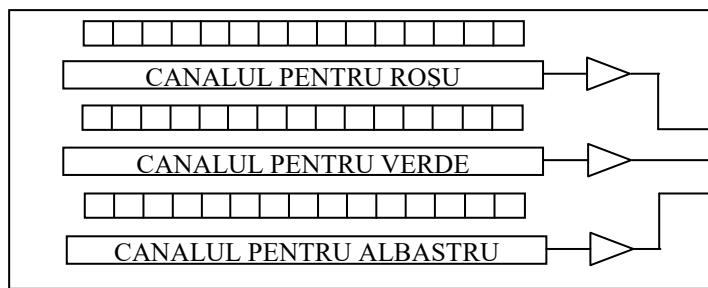


Figura 2.10 Structura unui senzor liniar

Principial, există patru tipuri de bază pentru senzorii CCD: liniari, interliniari, cadru întreg (full frame), transfer pe cadre (frame transfer).

Un senzor CCD **liniar** (figura 2.10) este alcătuit dintr-un sir de senzori dispuși pe o singură linie. Pentru a achiziționa o imagine folosind un senzor liniar este **necesar ca senzorul să se deplaseze cu viteza controlată de-a lungul imaginii**. Viteza de achiziție este redusă dacă se folosește această manieră. Structura electromecanică care asigură deplasarea se bazează pe motoare pas cu pas și crește, pe de o parte complexitatea sistemului, iar pe de altă parte riscul de alterare a geometriei imaginii. Utilizarea actuală a senzorilor liniari se concetează în fabricația scannerelor și a cititoarelor de coduri cu bare.

Celelalte trei variante de senzori sunt considerate generic ca senzori matriceali CCD pentru că formează zone senzoriale cu linii și coloane, de formă dreptunghiulară sau pătrată.

Un senzor CCD cu transfer interliniar (figura 2.11) are pentru fiecare pixel și un fotodetector și o zonă de stocare a sarcinii rezultate. Zona de stocare este formată prin ecranarea (opacizarea) unei părți din zona pixelului. Prin concatenarea zonelor opace se formează un canal vertical care permite transferul sarcinilor de-a lungul senzorului până la un registru orizontal de deplasare. Prin modul de funcționare varianta interliniară permite transferul rapid al sarcinilor din zona în care au fost acumulate sub influența luminii mai întâi în zonele opace de stocare, iar ulterior din linie în linie până la registrul orizontal de deplasare. Prin deplasarea vechilor sarcinilor se lasă

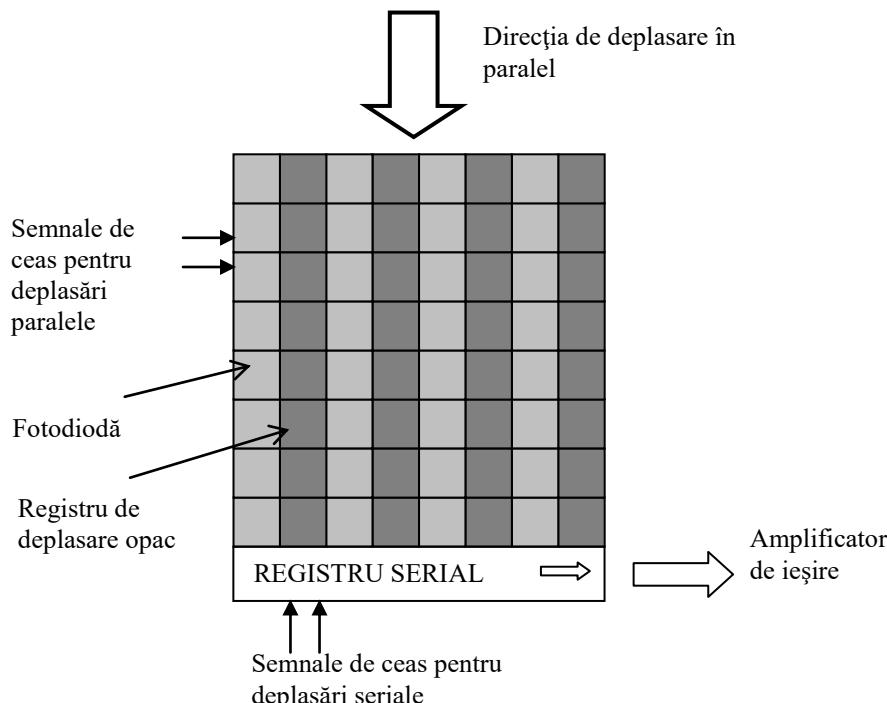


Figura 2.11 CCD cu transfer interliniar

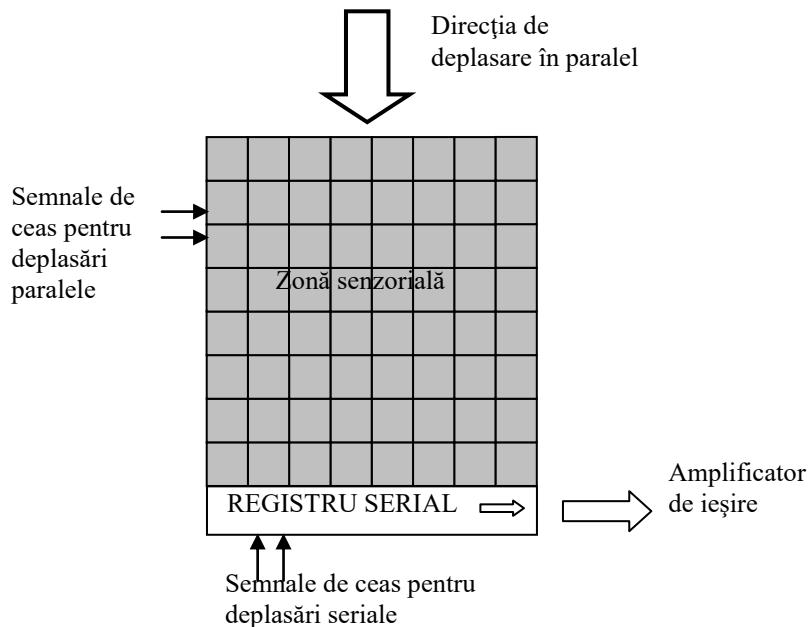


Figura 2.12 CCD cadru întreg (full frame)

loc pentru noilor sarcini acumulate ceea ce este un suport deosebit pentru achiziția de secvențe video. **Prin prezența zonelor de stocare și transfer se micșorează zona din senzor care este efectiv sensibilă la lumină.**

Senzorii **CCD de tip cadru întreg** (figura 2.12) folosesc toată zona pixelului pentru achiziția imaginii. În acest fel pe timpul transferului de sarcini nu se mai poate face integrare, deci nu se mai poate face acumulare de sarcini prin expunerea la lumină. Pentru a împiedica influența luminii pe timpul când are loc transferul de sarcini (ceea ce strică calitatea imaginii) se pot păstra **diafragme mecanice** în fața senzorilor. Există și situații în care aplicația, prin natura ei, elimină necesitatea diafragmelor mecanice, ca de exemplu atunci când durata și volumul luminii este controlat extern prin **lumină stroboscopică** de studio.

Varianta cu **transfer pe cadre** este similară cu varianta cadru întreg, dar **se ecranează (maschează) jumătate din matricea senzorială** astfel încât să fie aptă pentru stocarea sarcinilor. După terminarea perioadei de integrare, când elementele senzoriale elementare au înmagazinat sarcinile, are loc un transfer al sarcinilor către zona de stocare și ca urmare o nouă integrare se poate face fără o întârziere expresă. În acest mod, acest tip de senzori se pot folosi pentru achiziții rapide. Totuși, suprapunerea perioadelor de integrare cu acele de transfer de sarcini conduce la o scădere a calității imaginii (smear). Prețul de cost al acestor senzori este crescut datorită complexității induse de prezența celor două zone, senzoriale și de stocare.

Varianta interliniară reprezintă, din acest punct de vedere, o îmbunătățire pentru că se permite integrarea și transferul simultan, fără efectul de murdărire a imaginii.

Pentru citirea datelor din senzor se folosesc două metode: **citire progresivă și citire întrețesută**.

Într-un mod analog cu baleajul video de la străbunicul tub vidicon, și aici contează ordinea în care coloanele senzorului sunt transferate în registrul orizontal și ulterior la ieșirea din senzor. În varianta progresivă, liniile se citesc succesiv în ordinea în care apar în imagine. În varianta întrețesută, se citesc întâi liniile pare și ulterior liniile impare, după care are loc reintegrarea.

2.3.4 CMOS

CMOS este, ca și CCD, o tehnologie pe bază de siliciu și are proprietăți fundamentale relativ similare din punct de vedere al sensitivității în spectrul vizibil și aproape de infraroșu. Ambele tehnologii convertesc lumina incidentă sub formă de fotoni în sarcini electrice sub formă de electroni. Senzorii color pot fi fabricați în ambele tehnologii, în mod normal, prin adăugarea la fiecare pixel a unor filtre de culoare (de exemplu roșu, verde și albastru).

Tehnologia CMOS este o tehnologie de tip semiconductor metal – oxid și este arhitectura cea mai folosită pentru tehnica de calcul, unități centrale și module de memorie. Principal, senzorii de imagine CMOS (figura 2.13) performanți folosesc tehnica APS (active pixel) care a fost dezvoltată la NASA Jet Propulsion Laboratory la mijlocul anilor 1990.

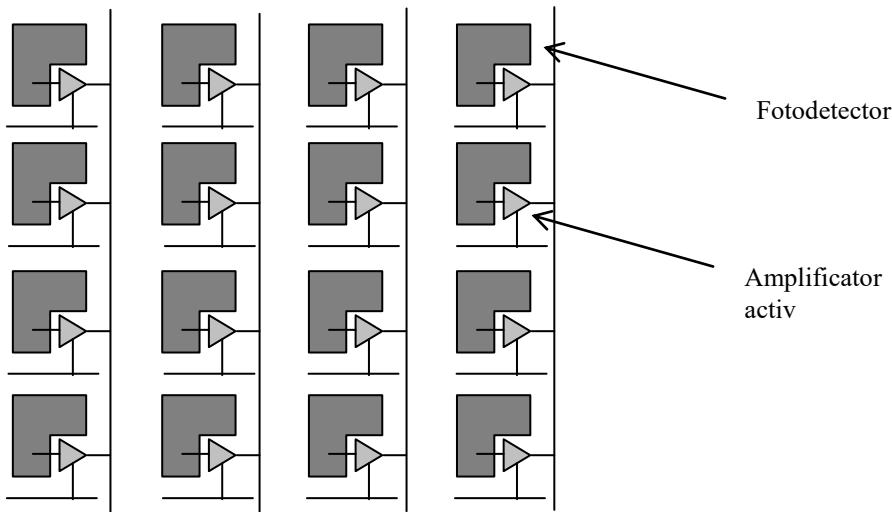


Figura 2.13 Senzor CMOS

Senzorii CCD sunt produși pe linii de fabricație specializate pe care nu se pot produce alte circuite integrate ceea ce crește prețul de cost. Alte linii de fabricație folosesc tehnologia CMOS Complementary Metal Oxide Semiconductor pentru circuite integrate pentru procesoare și memorii. De exemplu procesorul Pentium III conținea 10 milioane de elemente active de acest tip. Fabricând senzori de imagine pe astfel de linii de fabricație conduce la scăderea importantă a costului. Trebuie precizat că aici termenul CMOS se referă la modalitatea de fabricație și nu la o tehnologie specifică pentru senzori.

Principal, există două variante de bază pentru senzorii de imagine CMOS: pasivi și activi.

PPS (Passive pixel sensors) a fost prima variantă dezvoltată în anii 1960. La nivelul zonelor fotosensibile lumina sub formă de fotoni este convertită în sarcini, adică electroni. Sarcina acumulată pe timpul expunerii, integrării, este citită și amplificată. Senzorii sunt mici, atât cât să permită expunerea zonei fotosensibile și să includă conexiunile. Problema majoră a constituit-o, la acest tip de senzori, zgomotul materializat într-o rețea pe fundalul imaginii. Pentru a înlătura acest zgomot de fond sunt necesare etape de prelucrare suplimentare. APS (Active pixel sensors) a redus exact zgomotul amintit pentru varianta pasivă. Circuite specializate la nivelul fiecărui pixel determină și anulează zgomotul apărut. De la aceste circuite active vine și numele tehnologiei. Performanțele acestei variante de tehnologie CMOS se apropie de performanțele oferite de tehnologia CCD și permit realizarea de senzori de mare dimensiune și înaltă rezoluție. Tehnologia CMOS permite includerea în cipul senzorului a unor funcții suplimentare (inclusiv pentru micșorarea jitterului și stabilizarea imaginii sau compresia imaginii, pe lângă cele amintite anterior) care

necesită cipuri suplimentare la CCD. În această tehnologie se poate comuta rapid între achiziția de imagini (fotografii) și achiziția de secvențe video (filme). Trebuie subliniat că în acest ultim caz rămâne de rezolvat, la nivelul calculatorului cu care este cuplată camera, problema memorării în timp real a volumului mare de informație asociat secvențelor video. Prin prezența circuitelor suplimentare de eliminare a zgomotelor se micșorează procentul zonelor influențate de lumină din suprafața totală a circuitului (fill factor – procentul de acoperire). În acest fel sensitivitatea la lumină scade și apar probleme legate de calitatea imaginilor achiziționate în condiții de lumină puțină. Situația se poate corecta, extern, prin prezența surselor de lumina de tip flash și prin mărirea timpului de expunere. Din punct de vedere tehnologic, intern, se recurge la introducerea de microlentile pentru fiecare pixel, pentru a aduna mai multă lumină, și la reducerea circuitelor suplimentare.

Pentru că senzorii CMOS pot avea teoretic un nivel de zgomot mai mare decât senzorii CCD este nevoie de un timp de procesare mai mare între două imagini. Se pot folosi pentru aceasta procesoare de semnal (DSP) speciale. Prețul a fost un avantaj major pentru CMOS ceea ce a determinat tendința de a îndrepta cercetările în direcția producării unor astfel de senzori și de a le îmbunătății permanent performanțele.

2.3.5 Comparație CCD - CMOS

Tehnologia CCD are două elemente constitutive principale: **planul foto și memoria asociată**. Secvența de acțiuni parcursă în funcționarea senzorului este următoarea: expunerea planului video; transferul rapid în memoria asociată; citirea din memorie, câte o linie odată; în timp ce memoria este citită, planul foto este expus pentru următorul cadru. Expunerea simultană a tuturor pixelilor este un avantaj important; iar dezavantajul este structura mai complicată prin prezența memoriei. CMOS presupune o secvență mult mai simplă: se expune o linie cât timp este necesar pentru a fi oferită la ieșire; această linie este transferată în registrul de ieșire. Avantajul unei astfel de structuri este simplitatea, iar dezavantajul este că nu toți pixelii sunt expuși simultan, în același timp.

În tehnologia **CCD sarcinile se deplasează** de la un pixel la vecinul său pentru a fi citite. La **CMOS**, varianta APS, (APS - Active Pixel Sensor) **fiecare pixel are propriul amplificator** și circuit de ieșire deci poate fi citit independent ceea ce conduce la posibilitatea unor frecvențe de baleaj vertical, cadre pe secundă, variabile (se poate selecta o regiune de interes din imagine). Fiecare amplificator necesită 3 tranzistoare FET. Practic amplificatoarele au amplificări și offseturi ușor diferite ceea ce conduce la apariția unui zgomot fix pe imagine (FPN - Fixed Pattern Noise).

Problema amplificării se poate corecta cu o unitate specializată pentru fiecare pixel (UGA - Unity Gain Amplifier) care folosește o tabelă cu factori de corecție pentru amplificări. În acest mod apar 6 tranzistoare FET în plus ceea ce reduce zona activă a senzorului și crește prețul. Se folosește o tehnică specială de reducere a redundanței (ACS - Active Column Sensor) care micșorează numărul tranzistoarelor care au același rol și introduce un singur tranzistor FET la intrarea UGA pentru o coloană.

Problema offsetului (decalajului) se poate rezolva prin măsurarea valorii inițiale a fiecarui pixel și scăderea acesteia din valoarea de după expunere, în interiorul circuitului integrat (on-cip) sau în afara lui (off-cip). În majoritatea cazurilor se utilizează varianta off-cip ca și pentru problema amplificării, ceea ce crește complexitatea abordării. ACS are avantajul folosirii unui singur tact și a unei singure tensiuni de alimentare, în timp ce la tehnologia CCD sunt necesare mai multe tacte și mai multe tensiuni de alimentare. În ambele tehnologii, folosind timpi de integrare (expunere a senzorului la lumină) mici, se poate elimina "blur"-ul și se poate "îngheța" imaginea.

Diferența dintre cele două tehnologii apare în cazul imaginilor cu **linii verticale** care se mișcă rapid. Datorită expunerii linie cu linie, senzorul CMOS riscă să prezinte aceste linii fragmentate. Acest aspect are o deosebită importanță în aplicațiile industriale și științifice bazate pe vedere artificială (computer vision, machine vision). Uneori această deficiență se poate corecta prin program cu pierderea de timp de prelucrare inevitabilă.

În aplicațiile cu imagini în mișcare rapidă e nevoie de un **timp de integrare foarte mic** pentru a elimina blur-ul. Timpul mic de expunere cere un zgomot de fond foarte mic pentru a asigura o foarte bună gamă dinamică. CCD realizează acest deziderat nu numai la tempi de integrare mici, dar și la expuneri de ordinul minutelor. CCD folosește două tehnici: răcirea termoelectronica a senzorului la $-30F$ – $50F$ ($-35^{\circ}C$ - $45^{\circ}C$) - CCD se bazează pe captura electronilor liberi pentru că substratul rece reduce mobilitatea electronilor și deci zgomotul de fond; subțierea substratului și iluminarea lui din spate pentru a împiedica apariția electronilor rebeli. Această tehnică este utilizată în aplicații de astronomie, medicină și biologie.

Avantajele oferite de tehnologia CCD și prezentate până acum sunt legate de **operațiile la nivel de pixel**. CCD permite și construirea unui mozaic prin alăturarea zonelor din senzor în suprafețe mai mari, pe 3 sau 4 părți ale senzorului.

În concluzie, tehnologia **CCD oferă avantaje** care decurg din modul în care funcționează acest tip de senzor. Avantajele sunt importante în aplicații militare, științifice și industriale, dar mai puțin sesizabile pentru consumatorul obișnuit. **CMOS e mai ieftin** și se poate impune pe măsură ce va oferi performanțe asemănătoare. CMOS se poate produce pe liniile obișnuite de circuite integrate, ceea ce îi scade prețul în comparație cu CCD care necesită condiții speciale de fabricație și are o rată de defecte la fabricație mai mare. Din același motiv CMOS poate include mai ușor în cip circuitele adiționale: convertoare analog – numeric (ADC – Analog Digital Converter), procesoare numerice de semnal (DSP – Digital Signal Processor).

Curs 3

2.3.6 Terminologie

Trebuie făcută o diferență între **aparatele foto**, pentru fotografii, și **camerele video**, pentru filme. Aparatele fotografice numerice sunt capabile să achiziționeze o imagine de mare rezoluție pe care o depun în format numeric pe un suport de memorie. Ca o funcție suplimentară se oferă posibilitatea de a achiziționa o secvență scurtă formată din câteva cadre succesive. Camerele video achiziționează imagini cel puțin în cadență TV de 25 sau 30 de imagini într-o secundă. Imaginele, în format analogic sau numeric, sunt depuse pe suport de memorie. Întrebarea este de ce, atâția ani, nu s-au produs nici aparate foto care să facă filme și nici camere video care să facă fotografii? Explicația tehnică constă în imposibilitatea de realiză simultan cele două deziderate: rezoluția mare pentru detalii și numărul mare de imagini pe secundă pentru imaginile în mișcare. În momentul când aceste limite tehnice vor fi fost depășite vor interveni și criterii economice de satisfacere cu profit maxim a cerințelor de piață care nu vor permite trecerea bruscă la un număr prea mare de aparate care să facă simultan și fotografii și filme.

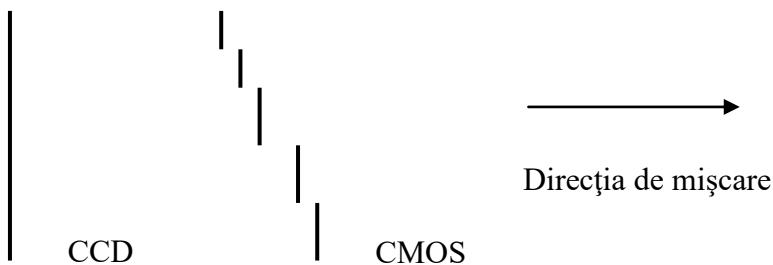
Blooming este fenomenul care apare atunci când capacitatea unui pixel este depășită și excesul de sarcini se răspândește către pixelii vecinii alterând calitatea imaginii finale. Există mai multe tehnici de reducere a acestui fenomen prin adăugarea în cip a unor circuite suplimentare care să asigure drenarea sarcinilor în exces.

Curentul de întuneric (Dark Current) este termenul care desemnează semnalul care este furnizat de senzor în lipsa luminii incidente. Acest curent rezidual se datorează acumulării de sarcini prin efect termic. Sarcinile acumulate prin acest efect variază de la pixel la pixel.

Pixeli de referință la întuneric (Dark Reference Pixels) sunt pixeli fotosensibili care sunt acoperți de un ecran metalic. Sarcinile acumulate în acești pixeli sunt datorate numai efectului termic și ele sunt folosite ca măsură a sarcinilor acumulate prin acest fenomen în pixelii propriu-zisi (pixeli activi) cărora le sunt asociati.

Gama dinamică (Dynamic Range) este raportul între valoarea semnalului de saturatie al senzorului (valoarea maximă oferită la ieșire) și semnalul de întuneric. Uzual valoarea se calculează în decibeli.

Factorul de umplere (Fill Factor) este raportul dintre aria zonei efectiv sensibile la lumină (zona activă) și aria totală a senzorului. Valoarea factorului de umplere poate fi crescută prin introducearea în cip a unor microlentile care să concentreze lumina pentru fiecare pixel în parte.



Zgomotul constant de fond (Fixed Pattern Noise) este zgomotul observat pe imagine în lipsa oricărei iluminări sau în prezența unei iluminări uniforme pentru întreaga imagine. Acest tip de zgomot poate fi înlăturat prin extragerea din semnalul util a valorii de întuneric corespunzătoare fiecărui pixel.

Sarcina întârziată la nivelul pixelului (Photodiode Lag) este definită ca procentul din sarcina acumulată pe perioada de integrare care rămâne la nivelul pixelului după terminarea perioadei de transfer. Sarcina rămasă netransferată va afecta liniile sau cadrele următoare.

Neuniformitatea răspunsului (Photoresponse Nonuniformity) este variația vârf la vârf care apare în semnalul de ieșire al unui senzor supus unei iluminări constante pentru toți pixelii săi. Pentru senzorii alb – negru această variație este foarte mică pentru că singurele valori variabile sunt rata de eficiență, curentul de întuneric și aria zonei active. Pentru senzorii color variația crește datorită imperfecțiunilor filtrelor de culoare care pot permite accesul unui volum mai mare de lumină pentru anumite zone în raport cu altele.

Rata de eficiență (Quantum Efficiency) a unui senzor este raportul dintre numărul electronilor generați și numărul fotonilor incidenti într-o perioadă de timp. Rata de eficiență depinde de lungimea de undă a luminii incidente.

Murdărirea imaginii (Smear) se datorează unui transfer (difuzie) nedorit de sarcini între un pixel și vecinii săi. Acest efect este mai important la iluminări cu lungime de undă mai mare.

Sensitivitatea senzorului este o măsură combinată a ratei de eficiență, la nivelul pixelilor, și a factorului de conversie a sarcinilor, acumulate și transferate, în tensiune, la nivelul amplificatorului de ieșie. Sensitivitatea este exprimată cel mai frecvent printr-o altă mărime: responsivitatea întregului senzor, care este raportul dintre lumina incidentă pe senzor și tensiunea rezultată.

2.3.7 Controlul timpului de integrare (shutter electronic)

Dacă aparatele fotografice clasice controlau **timpul de expunere** al filmului prin timpul de deschidere al diafragmei, această soluție nu mai poate fi aplicată cu succes în cadrul senzorilor pentru imagini. Controlul timpului de integrare, timpul în care se acumulează sarcinile electrice în fotocelulele individuale, se face electronic prin validarea senzorului la momentul ales și invalidarea lui după perioada dorită. Există câteva situații care justifică o funcție care să permită declansarea la comandă și pentru o durată prestabilită a achiziției.

Gama dinamică pentru un senzor de imagine poate fi definită și ca diferența între numărul electronilor pentru cazul lipsei de lumină (întuneric) și acela pentru cazul cu lumină maximă. Această gamă este finită și, de exemplu, pentru un raport de 2000 putem avea 50 pentru negru (zgomot de fond) și 100000 pentru alb. Această valoare nu se apropie de raportul natural dintre noapte absolută și lumina de la soare. Pentru a putea achiziționa imagini în condiții diferite de iluminare este nevoie de controlul intensității luminoase recepționate de senzor ceea ce se poate realiza prin ajustarea timpului de integrare. La scenele în mișcare apare "blur", imaginea are geometria alterată pentru că apar schimbări chiar în timpul achiziției unei imagini. Soluția este reducerea timpului de expunere. Pentru a putea primi evenimente care durează foarte puțin timp e nevoie de startarea achiziției unei imagini la comandă.

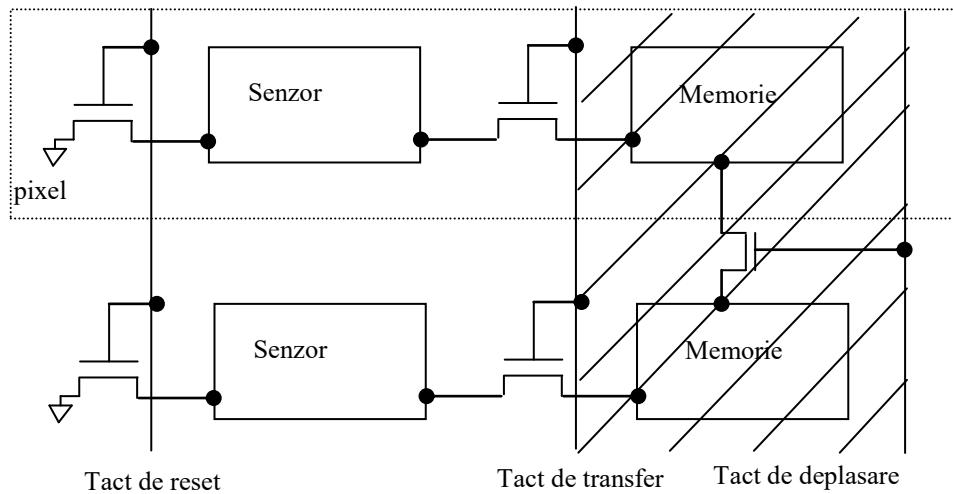


Figura 3.3 Transfer interliniar

Transferul interliniar (figura 3.3) presupune doi pixeli, ambii în planul imagine, unul propriu-zis și al doilea pentru memorare. Operația decurge astfel: pixelul e ținut în reset ignorând lumina, pixelul e activat, la sfârșitul perioadei dorite este mutat în zona de memorie, pixelul revine în starea reset. Mutarea se face aproape, în locații adiacente, și necesită doar un tact. Stocare se face în formă analogică. Zona de stocare, opacă, trebuie golită până la o nouă utilizare. **Transferul pe cadre** lucrează similar, dar cei doi pixeli nu sunt situați în acest plan. Operația decurge similar cu precizarea că deplasarea nu se face în zone adiacente, ci din pixel în pixel de-alungul planului imagine, până în zona de stocare. Va fi nevoie de mai multe tacte, proporțional cu numărul de elemente pe coloane (sute până la mii). Tehnica inciziei (slit shutter) nu folosește zone de memorie asociate. La sfârșitul achiziției sarcinile sunt transferate printr-un amplificator de coloană direct la ieșirea senzorului. Timpul de expunere poate fi foarte scurt sau cât timpul unui cadru dar trebuie facută sincronizarea cu ieșirea senzorului.

Transferul interliniar este cel mai rapid pentru că necesită o perioadă de tact pentru startarea expunerii și o perioadă de tact pentru derularea transferului. Aceeași metodă este cea mai imună la zgomot pentru că transferul se face în zone situate fizic în apropiere. În schimb sensibilitatea e redusă pentru că numai jumătate din lumină cade pe zona activă, restul cade pe zona opacă. Transferul pe cadre preia un procent mult mai mare din lumină incidentă pentru că memoria opacă este situată într-o zonă separată. și această metodă este rapidă ca și varianta interliniară. Problema este că lumina cade și pe zona care corespunde pixelilor care sunt în curs de a se transfera, ca urmare este alterată informația care este în tranzit. În acest fel scade imunitatea la zgomot. și pentru această tehnică este nevoie de două ori mai mulți pixeli decât rezoluția imaginii finale obținute. **Tehnica inciziei** oferă o mult mai mică complexitate, de două ori mai mică, și deci un preț mult mai mic. Viteza e bună, dar mai mică ca la transferul interliniar pentru că distanța de transfer e mai mare, cu cât traseul care străbate planul imagine până la amplificatorului comun este mai mare. Marea problemă este că pixelii nu sunt expuși simultan și deci pot apărea deformări la imaginile în mișcare, blur. Această distorsionare poate fi corectată printr-o prelucrare ulterioară.

În concluzie, pe principiul de cost minim, ca și pentru cazul imaginilor statice, se impune tehnica inciziei. Pentru scene în mișcare cu lumină suficientă se recomandă transferul interliniar. Pentru scene în mișcare cu lumină puțină, utilizatorul trebuie să aleagă între zgomot și distorsiuni. Pentru achiziția evenimentelor aleatoare se pot folosi toate cele trei tehnici, dar iarăși depinde cât de repede se mișcă scena.

2.4 Caracteristicile camerelor de luat vederi

2.4.1 Interpretarea datelor din foaia de catalog

Pentru a exemplifica parametru care caracterizează funcționarea unei camere de luat vederi am ales camera Pulnix CCD TMC-6 RGB. Camera este prezentată de producător ca o cameră de înaltă rezoluție pentru aplicații cu imagini color. Este recomandată pentru aplicații de teleconferință, aplicații industriale bazate pe vedere artificială (machine vision), recunoaștere de obiecte, cercetare medicală și supraveghere video. Putem comenta totuși, analizând caracteristicile specificate de producător, că pentru aplicații de videoconferință sau supraveghere video uzuale această cameră este o soluție prea performantă și deci prea scumpă. Se permite expunerea senzorului pentru un timp mai mare decât perioada TV normală (16,7 ms corespunzător la 60 Hz sau 20 ms corespunzător la 50 Hz) ceea ce este util pentru aplicații cu lumină puțină, cum sunt cele de microscopie. Se oferă o tehnologie proprie de control al timpului de expunere care să garanteze imagini de bună calitate și pentru scenele în mișcare rapidă. Discutăm în continuare informațiile furnizate în **foaia de catalog** a acestei camere.

➤ “Imager 1/2” interline transfer CCD (6.4 mm x 4.8 mm)”

6.4 mm x 4.8 mm sunt dimensiunile senzorului. Observăm că aceste dimensiuni sunt în raportul 3/4. Echipamentele de afișare video și cele senzoriale au fost puse în legătură cu formatul câmpului vizual uman. Numai dacă ne gândim că posedăm doi ochi dispuși pe orizontală, putem înțelege că avem un câmp vizual mai mult lat decât înalt. Pe de altă parte, tehnologiile bazate pe baleaj ar fi fost favorizate de formate de imagine, dacă nu circulare, măcar pătrate. În acest fel s-ar fi asigurat distanțe egale de la tunul de electroni până la punctele de pe suprafațele de afișare, respectiv senzoriale. Ca un compromis, s-a ales formatul de imagine cu un raport de 3/4 între dimensiunile verticală și respectiv orizontală. Au apărut astfel, dar au fost rezolvate, probleme de corectare a geometriei imaginii, mai ales în colțurile ei. Totuși, este cunoscut că atunci când testăm, de exemplu, calitatea imaginii afișate pe un display de calculator, zonele în care ne putem aștepta la deteriorări ale calității imaginii sunt tocmai cele menționate mai sus. Remarcăm că acest raport s-a păstrat și pentru formatele numerice standardizate 640 x 480, 1024 x 768, 1600 x 1200 etc. Ulterior, tehnologiile electronice au făcut posibilă lățirea imaginii către rapoarte mai mici, deci mai adaptate câmpului vizual uman. Este specificată aici și tehnologia electronică de fabricare a senzorului, CCD (Coupled Charge Device) și tehnica de citire a sarcinilor: transfer interliniar. Tehnologia CCD este cea mai răspândită tehnologie de fabricație a senzorilor de imagine. Senzorul este organizat pe perechi de coloane, una expusă luminii și cealaltă opacă. După prima fază, timpul de expunere, într-o a doua fază, se face transferul sarcinilor acumulate dintr-o coloană în cealaltă. Ulterior coloana opacă, care este defapt un registru de deplasare, este secvențial transferată (de aici termenul interliniar) într-un registru de deplasare orizontal. Acest registru va furniza imaginea linie cu linie.

➤ “Pixels 752(H) x 582(V)”

Se precizează aici numărul de linii, 582, și numărul de coloane, 752, al matricei senzoriale. Se observă că raportul dintre aceste numere este doar aproximativ 3/4. Acest lucru se întâmplă pentru că nu toate celulele disponibile sunt folosite efectiv pentru imagine, pentru că celula individuală nu este întotdeauna pătrată și pentru că celulele sunt apropiate dar nu sunt adiacente. Ca urmare, camera are un senzor de 437664 pixeli. Aici elementele imaginii, pixelii, sunt celulele senzoriale individuale. Se poate remarcă că nu există un raport de corespondență biunivocă între pixelii senzorului și pixelii imaginii afișate. Între aceste două reprezentări se interpun procesele de citire a senzorului, generare a semnalului analogic și conversia analog-numerică. Semnalele de ceas care guvernează aceste procese sunt responsabile de păstrarea geometriei imaginii. Numărul de pixeli

este parametrul cel mai important pentru definirea rezoluției oferite de senzorul camerei, indiferent de aplicația în care este utilizată.

➤ “**Cell size 8.6 µm (H) x 8.3 µm (V)**”

Se precizează dimensiunea celulei senzoriale individuale. Acest parametru este necesar în tehniciile de calibrare a camerelor pentru aplicații în care imaginile sunt folosite pentru măsurări dimensionale. Pentru o cameră de video conferință sau pentru un aparat de fotografiat nu interesează un astfel de parametru și nici nu este indicat, în majoritatea cazurilor, de către producător. Dimensiunea totală a senzorului nu coincide cu zona activă, aceea care percepă lumina. Dacă înălțimea și lățimea senzorului, pe de o parte și numărul de linii și numărul de coloane, pe de altă parte s-ar afla în raportul de 3/4 atunci ar fi convenabil să avem o celula senzorială individuală pătrată. Această condiție este parțial îndeplinită în acest caz.

➤ “**Colour filter Cy/Ye/Mg/G Complementary Colour Filter**”

Anumite camere compun culoarea din cele trei nuanțe de bază RGB, Red, Green, Blue respectiv roșu, verde și albastru. Alte camere diferențiază culorile complementare Cyan, Yellow, Magenta. Pentru a diferenția culorile de bază sau pe cele complementare se echipează celulele individuale cu filtre de culoare. Apariția a trei registri orizontali de deplasare, de exemplu, este o soluție care permite formarea unui semnal video care să respecte unul din standardele consacrate și în care semnalele de culoare apar într-o ordine prestabilită. Un grafic suplimentar oferit de producător (figura 3.4) dă indicații asupra sensibilității spectrale, în funcție de lungimea de undă, a senzorului. După acest grafic această cameră este mai sensibilă în infraroșu (partea din dreapta a graficului) decât ochiul omului. Ca urmare este de așteptat ca această cameră să “vadă” emisia unei telecomenzi TV, ceea ce noi nu reușim. Există și situații în care o sensibilitate crescută în zona de infraroșu să nu fie dorită pentru că se dorește obținerea de imagini care corespund numai spectrului vizibil din punct de vedere uman.

➤ “**Scanning 2:1 Interlace, field mode scanning 625 line 50 Hz**”

Modalitatea de afișare a imaginilor pe tuburile catodice se bazează pe viteza limitată de reacție a sistemului vizual uman. Imaginea este baleiată de la stânga la dreapta și de sus în jos cu o viteză atât de mare astfel încât să se formeze suficient de multe cadre într-o secundă pentru ca ochiul uman să percepă o imagine stabilă fără pălpăiri. Dispozitivele electronice utilizate atunci când s-au impus standardele de televiziune (primul utilizator major al imaginilor) nu permitau realizarea unei astfel de viteze de baleiere. S-a recurs la un truc, se baleiază imaginea pe rând, pe liniile impare (care formează semicadrul impar) și pe liniile pare (care formează semicadrul par). Această manieră se numește baleiere întrețesută. Evoluția ulterioară a performanțelor circuitelor electronice a înălțurat necesitatea baleajului întrețesut, dar datorită răspândirii foarte mari acesta este încă folosit. Specificația din foaia de catalog se referă, în mod analog, la maniera de baleiere a senzorului (imaginile se preia nu se afișează). Se indică faptul că semnalul furnizat de cameră respectă standardul (CCIR pentru Europa) cu imaginea împărțită pe două semicadre, cu 625 linii pentru un cadru și 50 de semicadre (în fapt 25 de imagini întregi) pe secundă. Nu toate cele 625 linii fac parte din imaginea efectivă, un număr de linii sunt necesare pentru revenirea baleierii din dreapta jos până în stânga sus. Alt standard (RS-170 pentru SUA și Japonia) micșorează numărul de linii la 525 linii pentru o imagine și mărește numărul de semicadre pe secundă la 60. Se pierde din rezoluția spațială, dar crește calitatea redării imaginilor în mișcare.

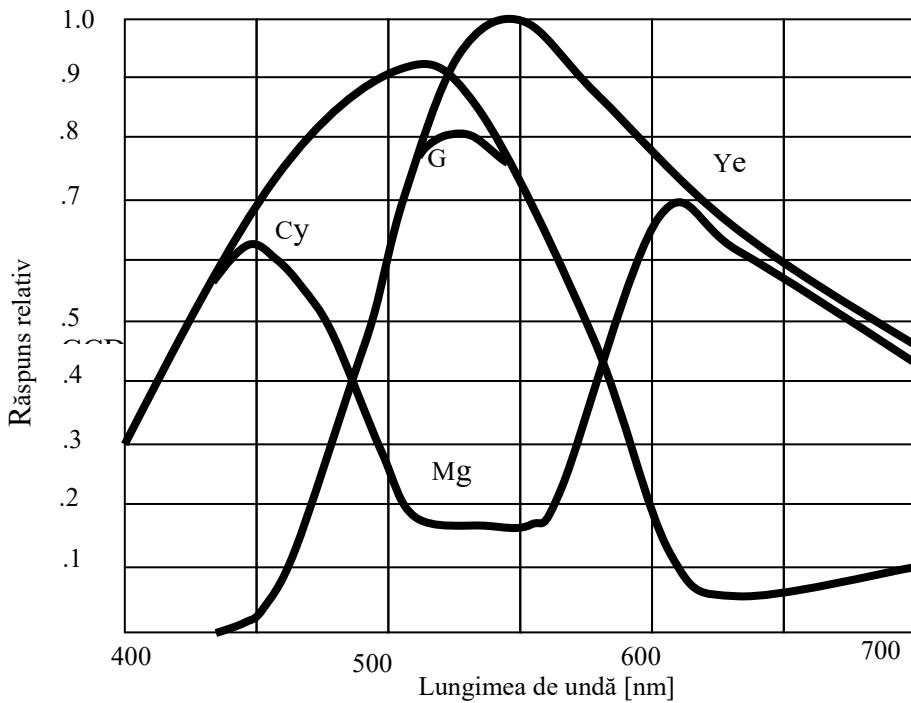


Figura 2.14 Răspunsul spectral al senzorului camerei

➤ “Sync fH = 15.625 KHz, fV= 50.000 Hz internal sync only (Ex version for external sync)”

Se specifică frecvențele de sincronizare cu care lucrează camera. Numărul de linii pe secundă este frecvența baleajului orizontal (H), sau frecvența de linii, și se obține multiplicând numărul de cadre cu numărul de linii dintr-un cadru: $25 \times 625 = 15625$. În varianta alternativă menționată mai sus s-ar fi obținut o valoare apropiată: $30 \times 525 = 15750$. Numărul de semicadre pe secundă se numește frecvență de baleaj vertical (V) sau frecvență de cadre. Această camera nu poate fi sincronizată cu un semnal de ceas furnizat din exterior, dar există altă variantă constructivă, marcată cu indicativul EX, care are această facilitate.

➤ “TV resolution 450(H) x 450(V) TV lines”

Așa cum am menționat deja, numărul de pixeli este un parametru determinant pentru rezoluția imaginii oferite de senzor. Totuși și alți parametrii influențează calitatea semnalului camerei (de exemplu tehnologia senzorului, tehnologia și schemele de interconectare pentru celelalte componente electronice ale camerei, standardul TV respectat). Este nevoie atunci de un indicator sintetic care să cuantifice calitatea imaginii finale oferite de cameră. În mod frecvent se utilizează, în acest sens, rezoluția exprimată în linii TV. Aici ni se garantează că imaginea are o rezoluție echivalentă cu 450 linii TV atât pe orizontală (H) cât și pe verticală (V). Este recomandat ca acest parametru să fie folosit în special comparativ pentru diferite produse și mai puțin ca valoare absolută. Pentru aplicațiile științifice și industriale care presupun prelucrare și analiză automată de imagini, nici nu interesează prea mult calitatea unei imagini afișate pe un ecran ci interesează rezoluția imaginii numerice disponibile în memoria sistemului. Remarcăm că există o diferență între rezoluția senzorului și rezoluția TV. Fără a intra în detaliile teoriei eșantionării semnalelor putem afirma că rezoluția efectivă a unui senzor matriceal nu poate fi decât aproximativ jumătate din numărul de celule elementare pe fiecare direcție. Pentru camerele cu tub această rezoluție este dată de diametrul spotului de electroni relativ la mărimea zonei fotosensibile. Pentru evaluarea rezoluției TV se folosește un test standardizat (o miră de control), de exemplu cel numit EIA Test Pattern. Din nefericire acest parametru nu presupune utilizarea unor unități de măsură (mm sau inch) și deci

poate deveni ambigu. Pentru a face o astfel de conversie în linii pe milimetru, de exemplu, se pot folosi formulele:

$$\text{Rezoluția orizontală} = \frac{\text{Linii TV[H]} \times 1,33}{2 \times \text{Dimensiunea totală pe orizontală [mm]}}$$

$$\text{Rezoluția verticală} = \frac{\text{Linii TV[V]}}{2 \times \text{Dimensiunea totală pe verticală [mm]}}$$

Este evident că “1.33” provine de la raportul 3/4 care trebuie luat în calcul pentru că spațierea diferă astfel pe cele două direcții.

➤ “S/N ratio 50 dB (AGC off)”

Un alt parametru care caracterizează calitatea semnalului oferit de cameră este raportul semnal zgomot. Calitatea superioară este evidențiată de valorile mari ale acestui parametru pentru că logaritmul este funcție crescătoare, iar raportul crește atunci când crește numărătorul (semnalul) sau atunci când scade numitorul (zgomotul). Un prag acceptabil este valoarea de 40 dB, sub această valoare calitatea camerei devine absolut discutabilă. AGC semnifică controlul automat al amplificării. Dacă această funcție este activă (on) ea permite mărirea amplificării atunci când semnalul este slab (lumină puțină) și respectiv, scăderea amplificării când semnalul este prea puternic (lumină prea multă). Activarea AGC implică activarea unor circuite electronice care să controleze amplificarea ceea ce nu poate avea ca efect decât scăderea raportului semnal-zgomot (orică circuit suplimentar înseamnă o sursă suplimentară de zgomot). Producătorul își ia precauția de a garanta raportul semnal/zgomot numai atunci când AGC este inactivat.

➤ “Min. ilumination 5 lux at F 1.4 (AGC on)”

Acest parametru dă indicații despre iluminarea minimă care îi este necesară camerei pentru a furniza o imagine acceptabilă. Candela este unitatea de măsură pentru intensitatea luminoasă. O candelă corespunde la intensitatea luminoasă a unei suprafețe de $1/60 \text{ cm}^2$ a unui radiator negru aflat la temperatura de solidificare a platinei. Unitatea pentru fluxul luminos este lumenul. Un lumen reprezintă fluxul luminos emis de o sursă punctuală cu intensitatea de o candelă de-a lungul unui unghi spațial de un steradian. Unitatea pentru iluminare este luxul. Un lux reprezintă iluminarea corespunzătoare unui lumen pe metru pătrat. Valoare de 5 lux indică o cameră care are nevoie de relativ puțină lumină. Camerele mai sensibile sunt de nivelul 1 sau 2 lux. Camerele pentru aplicații speciale (de exemplu imagini medicale endoscopice) sunt camere de fracțiuni de lux. Există aplicații în care este utilă o oarecare insensibilitate, dar care permite și o imunitate la zgomote. De exemplu într-o aplicație industrială cu vedere pentru roboți poate fi preferabilă o cameră de 20 lux și iluminarea mai puternică a scenei de operare. Numărul F este o măsură a volumului de lumină care poate trece printr-o lentilă. Valorile sunt o normalizare a deschiderii diafragmei lentilei. Numărul F este definit ca raportul dintre distanța focală și diametrul deschiderii circulare a diafragmei. Valorile standardizate sunt 1,4 ; 2; 2,8; 5,6; 8; 11; 16. Valorile sunt în progresie geometrică cu rația $\sqrt{2}$. Fiecare avansare pe această scară corespunde reducerii la jumătate a suprafeței (care depinde pătratic de rază) deschiderii diafragmei și deci în aceeași măsură a volumului luminii care trece prin lentilă. Este natural ca iluminarea minimă necesară pentru cameră să depindă de volumul de lumină care este lăsat să treacă prin lentilă și aceasta este explicația pentru care se precizează și această valoare. Valoarea de 5 lux va garanta o imagine acceptabilă numai dacă se activează AGC, ceea ce în acest caz va determina o creștere a amplificării, cu o aşteptată scădere a raportului semnal-zgomot.

➤ “VBS & RGB (700 mV p-p)”

Camerele de luat vederi se cupleză la sistemele de achiziție și prelucrare de imagini sau la monitoarele TV. Cuplarea presupune ca semnalul de ieșire al camerei să fie acceptat ca semnal de intrare de către aceste sisteme sau monitoare. Sunt două probleme: standardul semnalului care determină în special frecvențele și modul de interpretare a semnalului, pe de o parte, și pe de altă parte nivelul de tensiune al semnalului. Aici ni se indică că este disponibil un semnal video complex standard (VBS) și trei semnale separate pe culori RGB. Ambele au valoarea vârf la vârf de 0.7 V. Aceste semnale împreună cu cele de masă (GND) și o sincronizare (SYNOUT) sunt disponibile în conectorul de ieșire al camerei. Dacă identificăm pinul “video out” din acest conector (foaia de catalog include întotdeauna o hartă a pinilor, iar aici această hartă îl indică ca fiind pinul 4) atunci semnalul respectiv poate fi furnizat (evident al doilea fir este masa GND) intrării video a unui monitor sau televizor standard CCIR și imaginea se va “vedea”, adică se va sincroniza. Similar pentru un monitor TV cu intrări RGB analogice.

➤ “AGC Max. 32 db, AGC, on-off switchable, manual gain control”

Producătorul specifică faptul ca AGC poate modifica amplificarea cu maxim 32 db și că activarea/dezactivarea lui se poate face direct cu un comutator. Se poate modifica și manual factorul de amplificare cu ajutorul unui semireglabil. Nu reiese dacă cele două posibilități de reglaj, comutatorul și semireglabilul, sunt disponibile direct pe panoul camerei sau numai prin demontarea capacului camerei. Acest lucru trebuie verificat direct.

➤ “Gamma 0.45”

Intrarea senzorului este o distribuție spațială de lumină, iar ieșirea lui este un semnal electric (o tensiune) variabil în timp. Funcția de bază a senzorului este tocmai stabilirea unei anumite funcții între ieșire și intrare. Această dependență este descrisă în general ca o funcție putere.

$$\text{Amplitudinea ieșirii} = (\text{Amplitudinea intrării})^{\gamma}$$

Sau sub formă logaritmică: $\gamma = \frac{\log (\text{Amplitudinea intrării})}{\log (\text{Amplitudinea ieșirii})}$

Un senzor liniar va avea valoarea 1. În majoritatea cazurilor senzorii CCD au $\gamma=1$ ca valoare standard și ca valoare optională $\gamma=0.45$. Alegerea se face tot cu un comutator de pe panoul camerei. Totuși, în cazul analizat, singura valoare specificată de foaia de catalog este $\gamma=0.45$.

➤ “Lens mount C/CS mount”

Camerele de luat vederi sunt furnizate, în majoritatea cazurilor, fără lentile. Se presupune că lentilele trebuie alese de către utilizator în funcție de aplicația în care dorește să folosească respectiva cameră. În funcție de dimensiunea obiectelor din scenă și de distanța la care acestea sunt plasate față de cameră se va calcula distanța focală a obiectivului. Monturile (dispozitivele mecanice de cuplare prin filetare) camerelor sunt standardizate. Utilizatorul trebuie să se asigure că obiectivul pe care îl achiziționează are același tip de montură cu camera. Aici se precizează două standarde: montura de tip C, la care distanța de la flanșă la senzor este de 17,5 mm, și montura de tip CS, la care distanța de la flanșă la senzor este de 12,5 mm. Există dispozitive mecanice suplimentare care pot face adaptarea chiar între monturi de tipuri diferite. Deși aparent montura pare un element auxiliar neimportant, fie și prin natura lui mecanică, totuși, mai ales pentru aplicațiile de măsurare, aceasta are o influență foarte mare asupra rezultatelor finale. În aceste aplicații intră în calcule distanța de la flanșă la senzor. Teoretic, axa care trece prin centrul lentilei și

este perpendiculară pe planul ei ar trebui să intersecteze senzorul în centrul său. Poziția astfel obținută poate fi definită ca originea imaginii. (Cum să definim o astfel de origine într-o imagine cu un număr par de linii și coloane?!). O altă problemă apare atunci când reglăm distanța focală. Axa amintită mai sus poate descrie o suprafață conică în spațiu și ca urmare centrul imaginii se poate deplasa cu câțiva pixeli. Este situația în care lentila nu este tot timpul paralelă cu planul senzorului ca în cazul ideal.

➤ “Power 12 Volt DC 450 mA”

Se specifică tensiunea de alimentare necesară, 12 volți curent continuu, și curentul consumat, 320 mA. Aceste valori sunt importante pentru a decide dacă putem alimenta camera folosind tensiunile disponibile în extensia de magistrală a unui calculator care găzduiește placa de achiziție și prelucrare de imagini la care se cuplează camera, sau dacă trebuie să folosim un alimentator extern suplimentar. Soluția folosirii unui semnal de alimentare disponibil în sistemul de calcul este mai comodă, dar poate limita numărul de camere utilizabile pe care altfel placa de achiziție și prelucrare de imagini le-ar putea suporta. Există deosemenea și riscul, pe care mulți producători par să-l ignore, de a avea în același conector semnale de alimentare și semnale video. Dacă unul din firele din conector se rupe, conform regulilor lui Murphy, probabilitatea ca un eventual scurtcircuit să se producă între alimentare și un semnal video se apropie de certitudine. Cum protejarea intrărilor și ieșirilor video este dificilă datorită benzii de frecvență a semnalelor, urmarea este mai întotdeauna distrugerea unor circuite. Folosirea unor conectori de calitate este, și din acest punct de vedere, o cerință importantă. În plus, este prudent să nu deconectați sau să conectați camerele în timpul funcționării. Pentru camerele multimedia tensiunea de alimentare este de cele mai multe ori 5 volți și alimentarea se face din calculator.

➤ “Temp. range – 10° C to +50° C”

Se precizează aici limitele de temperatură, dar nu se specifică, aşa cum se fac alți producători, limitele pentru funcționare, funcționare cu garantarea performanțelor și depozitare. Pentru aplicații speciale, aceste informații pot deveni utile.

➤ “Vibration and shock Vibration: 7G (11 Hz to 2000 Hz) Shock: 70 G”

Există aplicații în care rezistența la vibrații și şocuri este esențială. Putem aminti ca exemplu camerele montate pe brațe de robot, pe roboți mobili sau pe avioane și rachete. “G” este accelerarea gravitațională $9,8 \text{ m/s}^2$. Pentru vibrații se specifică accelerarea maximă și plaja de frecvențe a acestor vibrații. Pentru şoc se precizează numai valoarea maximă a accelerării. Valoarile oferite de această cameră sunt foarte bune și susțin afirmația făcută inițial că această cameră este una prea performantă și deci prea scumpă pentru aplicații uzuale de videoconferință și supraveghere video.

➤ “Size 42 x 32 x 132 mm”

Sunt evidențiat dimensiunile geometrice ale camerei într-o aproximare paralelipipedică.

➤ “Weight 210 grams”

Se precizează greutatea camerei în grame. Această informație este utilă în special pentru includerea camerei într-un sistem mobil, acolo unde o sarcină suplimentară este limitată (de exemplu roboți mobili, minidirijabile telecomandate). Dacă aplicația cere montarea camerei pe un suport reglabil, de exemplu un trepied, atunci trebuie considerată și aici greutatea camerei. Din motive comerciale producătorii de astfel de suporturi oferă produse diferențiate în funcție de greutatea suportată. Nu

încercați să plasați o cameră puțin mai grea decât valoarea garantată de producătorii unor astfel de suporturi, chiar dacă astfel ați face o bună economie la preț, pentru că s-ar putea să aveți probleme de poziționare corectă!

2.4.2 Interpretarea intuitivă a performanțelor camerelor de video conferință

Dacă pentru camerele dedicate aplicațiilor industriale și științifice este absolut necesar ca achiziția să fie precedată de o atență analiză a caracteristicilor oferite de producător în foaia de catalog, pentru camerele de video conferință este preferabil să se testeze direct calitatea produsului. Această manieră de testare este susținută de faptul că imaginile din astfel de aplicații sunt destinate în special vederii umane. În plus veți constata că foile de catalog ale acestor camere oferă mult mai puține informații decât acelea prezentate mai sus pentru camerele dedicate aplicațiilor industriale și științifice. Revistele de specialitate, de exemplu Advanced Imaging, ne oferă sfaturi utile pentru a aprecia calitatea unei camere multimedia.

Dacă vizitați o expoziție sau un magazin de prezentare pentru a alege o cameră de luat vederi pentru internet (webcam) sau video conferințe trebuie să aveți în vedere următoarele eventuale trucuri la care apeleză prezentatorii de astfel de echipamente:

- Comercianții folosesc simulatoare de linie telefonică sau conexiuni directe. Cereți să vă conectați la un număr real și întrebați cum se vede acolo imaginea.
- Dacă se face transmisie în dublu sens, verificați că ceea ce vedeați este preluat la celălalt capăt cu exact aceeași cameră și nu cu una mult mai bună. E preferabil să vedeați ambele imagini.
- Vi se prezintă condiții de studio: lumini perfect plasate, scenă și persoane îmbrăcate în condiții ideale, persoanele sunt instruite să nu se miște rapid. Rugați persoana de la celălalt capăt să modifice aleator și repetat orientarea camerei și să bată din palme.
- Vi se prezintă condiții perfecte și fixe: lumina, focalizarea, fondul, poziția camerei. Dacă aveți doi-trei metri de cablu, deplasați camera, schimbați poziția, astfel încât scena să numai fie perfectă ca iluminare și fond, refocalizați singur.
- Activăți microfonul care a fost dezactivat sub motivul că este prea mult zgomot. S-ar putea să apară zgomot pe imagine.
- Camera poate fi conectată la un calculator mult mai performant decât acela pe care urmează să-l folosiți dumneavoastră. Dacă e posibil solicitați conectarea la alt sistem de calcul și verificați influența calculatorului propriu-zis și a monitorului.

Sunt utile și ușor de realizat teste cum sunt cele de mai jos:

- Focalizați pe față unei persoane. Verificați culoarea și detaliile feței, buzelor și ale ochilor. Verificați dacă puteți citi pe buze ceea ce spune persoana.
- Mișcați capul, întâi încet și ulterior din ce în ce mai rapid. Verificați dacă pragul de la care se degradează imaginea este acceptabil pentru aplicația avută în vedere.
- Același lucru pentru mișcarea palmelor și a degetelor. Dacă ați exersat anterior puteți avea o măsură reală a numărului de cadre pe secundă și a performanțelor compresiei.
- Priviți culorile în zona muchiilor din imaginea feței; trebuie să fie nuanțe clare, fără curcubeuri sau denaturări. Aduceți o miră de test pentru culori și puneți-o în fața camerei. O culoare dificilă este purpuriul. Un test dur este un curcubeu de culori. Într-o iluminare roșiatică sau galbenă, camerele ar trebui să aibă probleme. Dacă nu, atunci cineva a trucat cumva achiziția sau afișarea.
- Puneți o foaie albă de hârtie în fața camerei, în lumină puțină. Ar trebui să nu fie zgomot aleator sau fix. Luați rapid foaia. Cum reacționează autoreglajele, cât de repede, ce fenomene tranzitorii apar? Puneți foaia în fața camerei astfel încât să se vadă muchia. Deplasați și rotați foaia. Ce se întâmplă în zona muchiei? Puneți foaia în fața caemeriei. Depărtați și apropiati foaia. Rămâne culoarea uniformă?
- Puneți camera în contralumină puternică. Funcționează autoreglajul expunerii? Chiar dacă apar efecte nedorite puteți folosi camera în condiții controlate.

Curs 4

2.4.3 Lentile și camere

În general, camerele de luat vederi se livrează fără lentile (obiective). Se pleacă de la ideea că alegerea lentilelor se face de către utilizator în funcție de aplicația avută în vedere și considerând în principal mărimea obiectelor din scenă, distanța la care se află acestea în raport cu camera și rezoluția cerută de aplicație. Totuși, de multe ori, proiectantul unei aplicații va alege în același timp și lentilele și camerele de care are nevoie într-o aplicație. Pentru a face o astfel de alegere (figura 2.15) se folosesc următoarele noțiuni:

- Câmpul de vedere (FOV - Field of View) este zona din scenă cu obiecte care este “văzută” de cameră și eventual va fi vizualizată, de exemplu pe monitor.
- Rezoluția este o măsură a capacitatii de reprezentarea a detaliilor obiectelor din imagine.
- Contrastul este dat de diferența dintre nivelele de gri care corespund obiectului și respectiv fondului imaginii.
- Distanțele de lucru minimă și maximă sunt măsurate de la lentila obiectivului la obiect.
- Adâncimea câmpului (DOF – Depth of Field) este diferența dintre cea mai apropiată și cea mai depărtată poziții între care poate fi deplasat un obiect fără ca imaginea să se deterioreze inacceptabil (se observă blur). Adâncimea câmpului este distanța maximă pentru care se poate asigura focalizarea perfectă cu ajutorul lentilelor. Acest parametru capătă sens în asociere cu specificarea rezoluției și contrastului. DOF poate fi crescut prin închiderea diafragmei lentilei însotită de o creștere corespunzătoare a iluminării.
- Mărimea senzorului este dimensiunea zonei active a senzorului, care este de cele mai multe ori

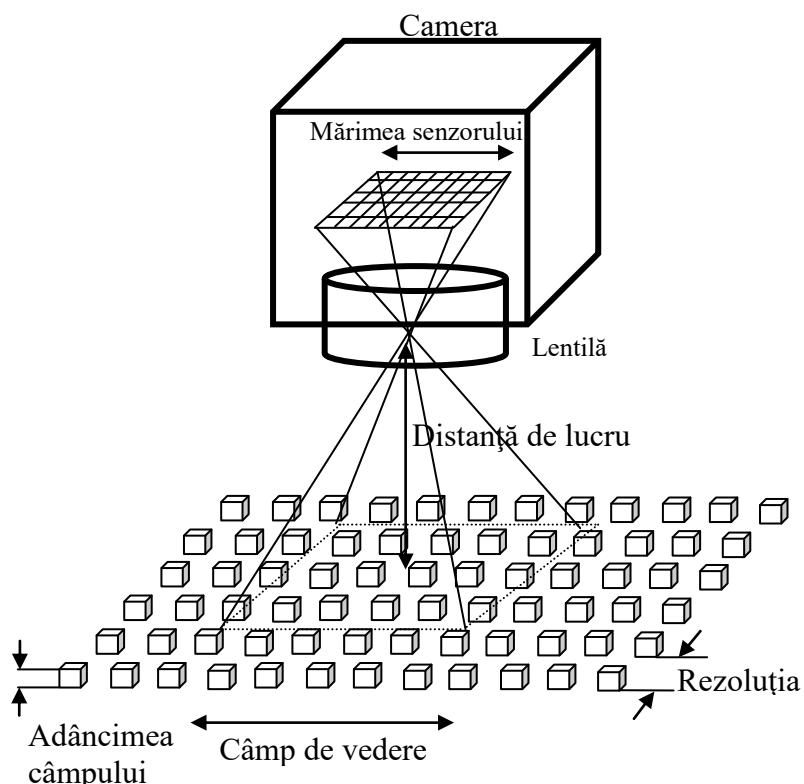


Figura 2.15 Ansamblul cameră - lentilă

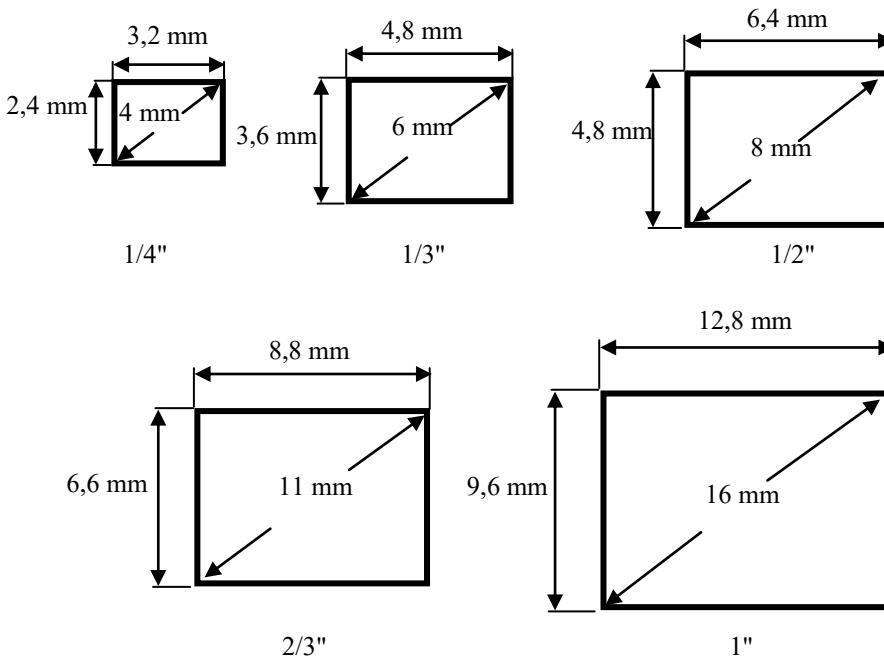


Figura 2.16 Diferite mărimi de senzori

specificată prin dimensiunea orizontală. Acest parametru este folosit în calcularea magnificării primare cerute pentru obținerea unui anumit câmp de vedere. Cele mai multe variante comercializate (figura 2.16) au un senzor cu raportul dintre dimensiunile verticală și orizontală în raportul 3/4.

- Magnificarea primară este definită ca raportul dintre mărimea senzorului și mărimea câmpului de vedere. Acest parametru este reglat prin intermediul sistemului de lentile.
- Magnificarea secundară este raportul dintre mărimea monitorului și mărimea senzorului.
- Magnificarea sistemului este magnificarea totală realizată de sistem în totalitatea sa și este produsul dintre magnificarea primară și magnificarea secundară.
- Numărul de linii pe milimetru este o măsură a rezoluției în sensul că măsoară perechile de linii pe milimetru care încă apar distinct (separate) în imagine. În acest fel rezoluția poate fi exprimată ca o frecvență. Legătura cu rezoluția devine mai evidentă dacă remarcăm că inversul acestui parametru este distanța minimă, în milimetrii, care trebuie să separe două regiuni din imagine pentru ca acestea să apară distincte. Specificarea poate fi făcută atât pentru lentile cât și pentru camere.
- Numărul de pixeli este un parametru care influențează decisiv rezoluția oferită de camera care include senzorul. Este evident că distanța minimă amintită mai sus trebuie să corespundă cel puțin la un pixel.
- Numărul de linii TV măsoară rezoluția folosind o miră formată din linii echidistante. Dacă mira acoperă câmpul de vedere, atunci numărul de linii TV este calculat numărând liniile și spațiile distinct vizibile în imagine. Parametrul nu are unități de măsură și nu trebuie comparat cu numărul de linii pe milimetru.
- Numărul F este o măsură a volumului de lumină care trece prin lentilă. Numărul F crește pe măsură ce deschiderea diafragmei scade (diafragma se închide). Închizând diafragma se produce o mărire a adâncimii câmpului de vedere, dar acest lucru se poate compensa crescând iluminarea câmpului de vedere. Numărul F este definit ca raportul dintre distanța focală și diametrul deschiderii circulare a diafragmei. Valorile standardizate sunt 1,4 ; 2; 2,8; 5,6; 8; 11; 16. Valorile sunt în progresie geometrică cu rația $\sqrt{2}$. Fiecare avansare pe această scară corespunde reducerii la jumătate a suprafeței (care depinde pătratic de rază) deschiderii diafragmei și deci în aceeași măsură a volumului luminii care trece prin lentilă.

Pentru alegerea unei lentile pentru o cameră, soluția cea mai simplă este să se folosească numai ecuația clasică a lentilelor:

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f},$$

unde v este distanța de la lentilă la imagine, u este distanța de la lentilă la obiect, f este distanța focală.

Notând factorul de mărire sau de magnificare cu M , avem:

$$M = \frac{\text{Marimea_obiectului_in_imagine}}{\text{Marimea_obiectului_real}}$$

și echivalent:

$$M = \frac{v}{u}$$

atunci:

$$f = \frac{u \cdot M}{M + 1}$$

Dacă se cunoaște factorul de magnificare și distanța până la obiect se poate calcula distanța focală necesară. De exemplu, pentru un obiect de 10 centimetri și un senzor matriceal de 8,8 x 6,6 mm, pentru o distanță de 0,5 metri factorul de magnificare va fi:

$$M = 8,8 / 100 = 0,088$$

iar lungimea focală cerută este: $f = 500 \times 0,088 / 1,088 = 40,44$ mm

Se va prefera o ușoară pierdere de rezoluție pentru acoperirea unei zone mai întinse și se va alege o lentilă cu lungimea focală mai mică, de 35 mm de exemplu.

Pentru a lua în calcul și rezoluția și acuratețea de măsurare așteptate, se pot calcula următoarele mărimi:

- Rezoluția camerei [μm] = $100 / \text{Numărul de linii [perechi de linii/mm]}$

- Rezoluția analogică:

- Rezoluția orizontală [perechi de linii/mm] =
 $(\text{Numărul de linii TV pe orizontală} \times 1,33) / (2 \times \text{Dimensiunea orizontală a senzorului [mm]})$

Este evident că “1.33” provine de la raportul 3/4 care trebuie luat în calcul pentru că spațierea diferă astfel pe cele două direcții.

- Rezoluția verticală [perechi de linii/mm] =
 $(\text{Numărul de linii TV pe verticală}) / (2 \times \text{Dimensiunea verticală a senzorului [mm]})$

- Rezoluția numerică:

- Rezoluția camerei [perechi de linii/mm] = Numărul de pixeli / (2 x Mărimea senzorului)

- Rezoluția camerei [μm] = 2 x Mărimea pixelului [μm]

- Magnificarea:

- Magnificarea primară PMAG =

Mărimea senzorului pe orizontală [mm] / Mărimea câmpului de vedere pe orizontală [mm]

- Magnificarea sistemului =

PMAG x (Mărimea diagonalei monitorului [mm] / Mărimea diagonalei senzorului [mm])

- Rezoluția:

- Rezoluția sistemului [μm] = Rezoluția camerei [μm] / PMAG

- Rezoluția sistemului [perechi de linii/mm] = Rezoluția camerei [perechi de linii/mm] x PMAG

Să presupunem că se dorește măsurarea un detaliu de aproximativ 3 mm dintr-o componentă mecanică cu o precizie de 25 μm . Considerăm că imaginea preluată de camera de luat vederi, cuplată la un sistem de achiziție și prelucrare de imagini, este în final afișată pe un monitor. Imaginea finală afișată nu poate depăși în calitate parametrii celei mai puțin performante componente din sistem. Dacă aplicația este o aplicație de vedere artificială cu prelucrarea automată a imaginii, atunci imaginea afișată nu va influența rezultatul final. În acest caz calitatea monitorului nu mai este critică. Dacă se face doar preluarea imaginii de la cameră și afișarea ei pe monitor atunci calitatea monitorului va influența evident rezoluția sistemului. Este necesar să fie cunoscute valorile parametrilor care definesc aplicația și calculate mărimile prezentate mai sus. Pe baza acestor informații se poate face o alegere a componentelor sistemului. Trebuie luate în calcul și condițiile de iluminare care influențează decisiv performanțele globale ale sistemului. Acolo unde este posibil, stabilirea unor condiții de iluminare controlate și constante (de exemplu în aplicații industriale de fabricație integrată) pot asigura o bază solidă pentru dezvoltarea unei aplicații performante. În aplicațiile în care condițiile de iluminare nu sunt controlabile și/sau sunt variabile, efortul de eliminare a influențelor care apar ca o consecință va fi considerabil.

Presupunem că am ales varianta de inspecție vizuală și că ne interesează rezoluția imaginii afișate pe monitor. Alegem un câmp de vedere de 4,75 mm pe orizontală pentru ca partea utilă să ocupe cea mai mare parte a imaginii afișate pe monitor. Dintre formații CCD disponibile presupunem că vom folosi o cameră de 1/2" (6,4 mm pe orizontală). Pentru că, așa cum am precizat deja, dorim să umplem imaginea de pe monitor cu imaginea corespunzătoare câmpului de vedere, cu cât alegem un senzor de dimensiune mai mare cu atât vom avea nevoie de lentile cu magnificare mai mare. Acum se poate calcula magnificarea primară, magnificarea secundară (să presupunem un monitor cu diagonală de 13") și magnificarea sistemului. Cum magnificarea sistemului corespunde magnificării lentilelor, se va folosi această valoare pentru a alege un obiectiv adecvat. Este de preferat un obiectiv care să permită și un zoom (mărirea câmpului de vedere) sau pentru a avea posibilitatea să vizualizez o porțiune mai mare din obiectul care conține detaliul care ne interesează.

Rezoluția camerelor CCD este frecvent specificată în linii TV pe orizontală și pe verticală. Dacă ne referim la catalogul Edmund Scientific atunci putem găsi camera analogică alb-negru Sony XC-75 cu un senzor 1/2" CCD care oferă 570 TVL[H] ceea ce corespunde la o rezoluție de 16,8 μm. Dacă acceptăm ca rezoluția sistemului este limitată de rezoluția camerei atunci putem calcula rezoluția sistemului folosind în formulă magnificarea primară a lentilelor. Rezoluția în linii TV ia în considerare mărimea senzorului, deci pentru același număr de linii TV se obține o aceeași rezoluție a sistemului dacă și câmpul de vedere este același. Astfel camerele Sony XC-73 și XC-75 vor oferi aceeași rezoluție pentru că ambele au 570 TVL[H] deși prima are un senzor mai mic. În schimb camera numerică Pulnix TM-1200 oferă 55,5 perechi de linii/mm (18,0 μm) rezoluția camerei pe orizontală pentru un senzor de 9,07 mm. Deși rezoluția camerei Pulnix este mai slabă decât a camerei Sony, pentru că folosește un senzor mai mare, totuși camera Pulnix poate crește rezoluția sistemului (9,4 μm).

Pentru că am exemplificat cu o aplicație de inspecție vizuală în care se urmărește afișarea pe monitor a unei imagini de cât mai bună rezoluție, putem lua în calcul obiective care includ reglaje ale diafragmei (iris) astfel încât să se influențeze și adâncimea câmpului de vedere. Pentru aplicațiile automate de vedere artificială la care se exclude orice intervenție on-line a operatorului uman, toate reglajele obiectivelor și ale camerelor trebuie făcute off-line, înainte de startarea aplicației propriu-zise, în faza de calibrare. Dacă se intervine ulterior asupra acestor reglaje, atunci în programul care controlează aplicația vor trebui schimbate multe valori determinante în faza de calibrare, în special valori de prag (thresholds). Spre exemplu, dacă detaliul de interes ar trebui extras în mod automat din imagine, atunci în acest proces trebuie folosite câteva praguri relative la nivelele de gri din imagine. Aceste praguri depind de nivelul de gri al mesei pe care este aşeat dispozitivul mecanic și de distribuția nivelelor de gri de pe suprafața dispozitivului. O reglare a diafragmei obiectivului în timpul aplicației, chiar dacă ar crește calitatea imaginii la o apreciere vizuală a operatorului, va schimba nivelele de gri din imagine ceea ce va conduce la rezultate finale imprevizibile. O nouă calibrare se impune după un astfel de reglaj care modifică oricare parametru al sistemului (câmp de vedere, magnificare primară, distanță de lucru, focalizare etc.).

Marea problemă a aplicațiilor de vedere artificială, și în particular a celor de inspecție, o constituie iluminarea scenelor. Prin natura lor obiectele inspectate presupun suprafețe foarte reflectante. În imagine vor apărea străluciri și umbre care vor pune mari probleme detectoarelor de contururi și regiuni implementate în programul sistemului. Este recomandabil să se insiste pe asigurarea unor condiții favorabile de iluminare decât pe încercarea de a elimina prin program efectele nedorite. Becul clasic cu incandescență este soluția cea mai simplă. Este o soluție ieftină iar intensitatea luminoasă poate fi ușor controlată. Deși este într-o oarecare măsură o aproximare a sursei de lumină punctiforme, becul clasic asigură o iluminare direcțională care produce umbre, și deci probleme în procesarea ulterioară a imaginii. Un alt dezavantaj al becului cu incandescență, insesizabil de către ochiul uman, este emisia de radiație infraroșie care poate crea probleme anumitor camere de luat vederi. Pentru cele mai multe aplicații este utilă o iluminare difuză, nedirecțională și care deci să nu producă umbre. Sursele fluorescente produc o astfel de iluminare și acoperă o zonă destul de întinsă ca suprafață. Acolo unde este posibil, se poate așeza obiectul pe o masă translucidă sub care este plasată o sursă de lumină ("back-lighting"). Se obține astfel ușor o segmentare, o extragere a conturului obiectului. Dacă în aplicație apar obiecte în mișcare, este utilă "înghețarea" imaginii prin iluminarea stroboscopică și achiziția, sincronizată cu iluminarea, a imaginii. Alternativ se poate folosi un timp de expunere foarte scurt pentru cameră ("shutter speed"). În acest caz perioada de timp pe care se face integrarea sau medierea luminii incidente, este foarte mică și deci surprinde o imagine staționară a unei scene chiar în mișcare. Un alt aspect neplăcut introdus de sursele de lumină este fluctuația lor ("flickery"), care nu este sesizată de om, al cărui sistem de vedere integrează puternic. O soluție pentru înlăturarea acestor fluctuații este controlul sursei de alimentare, de exemplu folosirea surselor de curent continuu. Multe sisteme de iluminare se

degradează în timp ceea ce face utilă măcar o calibrare la începutul fiecărei aplicații, dacă o calibrare dinamică este prea complicată. În acest fel se pot alege corespunzător pragurile ("thresholds") implicate în procesare. O alternativă este reglarea efectivă, periodică a intensității luminoase a sursei la o valoare constantă, aceeași de-a lungul timpului. Pentru exemplul nostru, poate fi aleasă, din catalogul firmei Edmund Scientific, o sursă axială difuză de lumină care să minimizeze zonele de strălucire de pe suprafața obiectelor și să eliminate umbrele. Pentru o imagine cu contrast foarte bun, ne putem aștepta la o eroare de câțiva pixeli, de exemplu 3. Această eroare se poate dubla pentru o imagine cu contrast slab.

Auratețea măsurătorii poate fi calculată acum cu formulele anterioare și ea va depinde esențial și de numărul total de pixeli de la nivelul senzorului, respectiv al plăcii de achiziție și prelucrare de imagini.

- Magnificarea primară $PMAG = 6,4 \text{ mm} / 4,75 \text{ mm} = 1,35X$
- Magnificarea sistemului $= 1,35 \times (330 \text{ mm} / 8 \text{ mm}) = 55,7X$
- Rezoluția camerei $= 570 \times (1,33 / (2 \times 6,4 \text{ mm})) = 59,4 \text{ perechi de linii/mm} \Rightarrow 16,8 \mu\text{m}$
- Rezoluția sistemului $= 16,8 \mu\text{m} / 1,35X = 12,4 \mu\text{m}$
- Auratețea de măsurare $= (3 \times 4,75 \times 1000) / 640 = 22,3 \mu\text{m}$

In concluzie, pentru exemplul nostru, se poate alege camera CCD Sony XC-75, cu obiectivul VZM 300 și cu o sursă de lumină difuză. Se va obține o rezoluție a sistemului de $12,4 \mu\text{m}$. Auratețea măsurării va fi de $22,3 \mu\text{m}$, ceea ce este suficient pentru cerințele formulate. Dacă se alege camera Pulnix TM 1020 atunci acuratețea măsurătorii va fi de $14 \mu\text{m}$.

Exemplul anterior trebuie considerat ca o variantă teoretică susținută în special de producătorii și distribuitorii de echipamente. Rezoluția și acuratețea astfel calculate sunt ceea ce putem obține în cel mai bun caz posibil cu respectivul sistem. În practică, este recomandabil ca proiectantul sistemului de vedere artificială să introducă un coeficient propriu de neîncredere pentru că surse de zgomot și perturbații se găsesc pe întreg lanțul de la sursa de lumină, la lentile, senzori, electronica camerei, până la cabluri, conectori și componentele plăcii de achiziție și prelucrare de imagini. În aceste condiții nu considerăm exagerat să nu ne bazăm pe zecimile de μm rezultate din calculele de mai sus!

Curs 5

2.5 Plăci de achiziție și prelucrare de imagini

2.5.1 Considerante generale

Plăcile de achiziție de imagini pentru aplicații de vedere artificială (nu și cele pentru multimedia) sunt identificate, în general, prin termenul de frame-grabber. Numele nu definește complet o astfel de placă deoarece nu se achiziționează ("culege") o singură imagine ci se convertește mai multe cadre ("frame"), la viteze echivalente cu zeci de cadre pe secundă (tradițional 25-30), ceea ce este denumită, în general, achiziție în timp real. Placa de achiziție primește semnalul analogic de la camera de luat vederi, îl convertește sub formă numerică (figura 2.17) și depune această informație într-o zonă de memorie accesibilă calculatorului. Cuantificarea în amplitudine oferă nivelele de gri, eșantionarea spațială precizează poziția poziția respectivei valori în cadrul unei linii a imaginii. Imaginele sunt alcătuite din cadre, dacă se lucrează întrețesut, cadrele din linii, iar linii din pixeli (elementul fundamental din care este alcătuită o imagine, și căruia i se asociază un nivel de gri și o poziție).

În mod uzual standardele specifică anumite intervale pentru parametrii care caracterizează semnalul oferit de camerele de luat vederi. Se poate întâmpla ca două semnale furnizate de două camere conform standardelor RS-170 sau NTSC, spre exemplu, să nu fie perfect identice. Este important ca placa de achiziție să aibă posibilitatea ca prin program să se controleze tactul pixelilor ("clock pixel") pentru a se adapta acestor diferențe amintite mai sus. Acest tact determină intervalul de timp după care se face eșantionarea. Dacă frecvența semnalului oferit de cameră, pe de o parte, și aceea după care se face conversia analog-numeric la nivelul plăcii de achiziție, pe de altă parte, nu coincid se va înregistra o pierdere a calității imaginii cu care se lucrează. O astfel de facilitate va permite și achiziția de semnale de la surse nestandard. Deasemenea, dacă camera are aceste posibilități, este util ca și placa de achiziție să poată controla funcții ca viteza de achiziție ("capture speed"), amplificarea ("gain"), translatarea semnalului ("offset"), prin intermediul calculatorului.

Chiar dacă o cameră numerică oferă un semnal numeric și nu unul analogic, care să fie nevoie să fie convertit, totuși la ora actuală și multe dintre aceste camere au nevoie de interfețe cu calculatorul. Producătorii tradiționali de camere de luat vederi fac eforturi să realizeze camere de luat vederi care să ofere imagini de o cât mai bună calitate și nu tehnologii de cuplare pe magistrala calculatoarelor, sisteme de operare sau software de aplicații. Se poate considera că este preferabil ca plăcile de conversie a imaginii, din forma analogică în forma numerică, să fie achiziționate de la firmele specializate în domeniu.

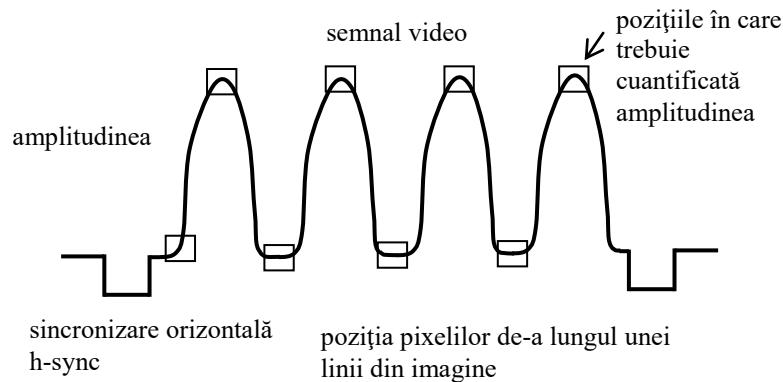


Figura 2.17 Eșantionarea și cuantificarea semnalului imagine

Plăcile specializate pentru editarea / prezentarea de imagini multimedia (identificate frecvent cu termenul de video blaster) oferă posibilități de conversie a semnalelor TV (video complex) standardizate (NTSC sau PAL, spre exemplu), uneori inclusiv a semnalului audio, într-un semnal standardizat pentru editarea video (TIFF, BMP, JPEG, GIF, etc). Rezoluția standardizată pentru sursele de semnal video este impusă aici la valori care să nu producă abateri ("jitter") observabile cu ochiul liber. Se tinde către utilizare de compresii, de exemplu JPEG sau MPEG, care să realizeze un echilibru între volumul de memorie necesar și păstrarea calității imaginii. Plăcile de achiziție pentru aplicații științifice, medicale, industriale au o mai mare flexibilitate în cuplarea cu sursele de semnal și o mai mare fidelitate. Nu se pune accent și multe dintre aceste plăci nu oferă posibilități de editare video, achiziție de semnal sonor sau conversii între diferențele formate, inclusiv compresii, pentru memorarea de imagini.

Există multe criterii care pot influența prețul unei plăci de achiziție și prelucrare de imagini. Timpul și efortul de proiectare cresc considerabil dacă se dorește obținerea unei înalte fidelități a imaginii, un zgomot redus și o rată de transfer mare. Din alt punct de vedere contează ceea ce firma oferă în prețul furniturii. La un preț mic s-ar putea să trebuiască plătite separat interfețele software, cablurile și/sau programele de aplicație sau dezvoltare de aplicații. La un preț mai mare, parțial sau total, aceste facilități pot fi incluse în furnitura de bază. Plăcile de achiziție și prelucrare de imagini pentru aplicații științifice și industriale se vând într-un număr mai mic, iar recuperarea investiției se face mai greu. În aceste condiții, reproiectarea lor are loc după un timp mai lung în raport cu plăcile similare dedicate aplicațiilor multimedia. La acest tip de plăci vânzarea se face într-un număr mare, iar recuperarea investițiilor, inclusiv de proiectare, se poate realiza mai rapid și, ca urmare, produsele noi apar mult mai frecvent. În cazul plăcilor de achiziție și prelucrare de imagini pentru aplicații științifice și industriale se impun condiții foarte stricte în ceea ce privește stabilitatea parametrilor în funcție de condițiile de lucru, care pot fi unele grele, și, respectiv, fiabilitatea, care poate fi esențială pentru procesul la care sunt cuplate plăcile. Din această cauză, prețul unui frame-grabber este mai mare, iar variantele noi apar mai rar. În lipsa unor resurse financiare corespunzătoare, integratorii de sistem recurg la utilizarea unor sisteme ieftine pentru a implementa aplicații dedicate. De exemplu, se poate imagina utilizarea unui TV-tuner pentru a achiziționa imaginea de la un microscop. Dacă se dorește, în acest caz, realizarea unor măsurători pe respectivele imagini, precizia și repăratibilitatea nu vor putea să fie garantate. Sau, un alt exemplu, ne putem imagina comparativ ce consecințe are defectarea unui sistem de achiziție și prelucrare de imagini care susține conducerea unei linii de fabricație din industria automobilelor în raport cu defectarea unui sistem similar utilizat într-un echipament multimedia de uz comun.

Există diferențe între metodele utilizate de diferiți producători. O diferență majoră este maniera în care se realizează secvențierea temporală a conversiei, timing. Fiecare semnal video are un semnal de sincronizare orizontală, hsync. Prin detectarea acestui semnal sistemul poate determina când să înceapă conversia, când s-o sfărșească și cu ce viteză să facă această conversie. Capacitatea

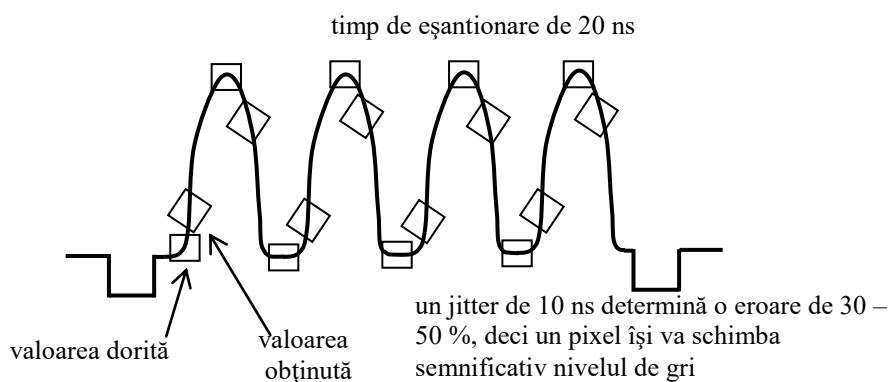


Figura 2.18 Efectul jitterului asupra conversiei

de a detecta și de areacționa la acest semnal de sincronizare este decisivă în obținerea unei imagini curate. Phase Lock Loop (PLL) este un mecanism tradițional în acest sens. PLL crează un clock, un semnal de ceas, ca o frecvență de referință. Acuratețea acestui semnal este măsurată prin fenomenul de "jitter" asociat unui pixel. În funcție de acest parametru se diferențiază frame-grabberele. De exemplu (figura 2.18) cele mai multe PLL implementate sunt limitate la un jitter inferior (deci mai mare) pragului de +/- 10 nanosecunde (uzual se obțin chiar valori de +/- 20 nanosecunde). Dacă perioada de eșantionare este de 80 nanosecunde (cam cât este durata pentru un semnal NTSC) atunci o variație de 10 nanosecunde este foarte mare. Dar pentru sistemele de înaltă rezoluție, unde perioada de eșantionare este de 20 nanosecunde, acest jitter este catastrofal. De asemenea dacă un PLL se sincronizează o singură dată pentru un întreg cadru atunci jitterul se va suma cu fiecare nouă linie din cadru. O urmare a acestui fapt va fi aceea că linii drepte din imagine vor tinde să se curbeze. Din această cauză în unele frame grabbere se utilizează un circuit numeric de ceas și o variantă modificată de PLL. Astfel se poate face resincronizare la fiecare linie din cadru astfel ca erorile se elimină la începutul fiecarei linii. Se poate ajunge astfel ca jitterul asociat unui pixel să fie de +/- 2 nanosecunde, sau chiar mai mic.

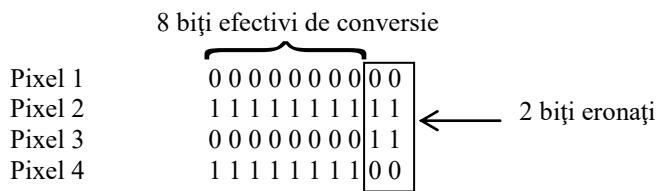


Figura 2.19 Conversie pe 10 biți cu 2 biți eronați

Principial fiecare bit în plus pe care se face conversia aduce informație în plus despre nivelul de gri asociat pixelului (dublează acest număr de nivele). Multe frame-grabbere lucrează pe 8 biți (deci cu 256 de nivele de gri în mod monocrom). Există producători care oferă conversii pe 10 biți. Nu întotdeauna se poate beneficia de acest avantaj (figura 2.19). Alți producători oferă conversii pe 12 sau 16 biți. De asemenea se poate afirma că jitterul și raportul semnal/zgomot pot conduce la o diferență între numărul de biți pe care se face conversia și numărul de biți efectiv folosiți. Un număr de biți care nu este multiplu de 8 (octet) va determina un consum suplimentar de timp pentru stocarea și regăsirea informației în memorie. Un volum mai mare de informație va necesita un timp mai mare pentru prelucrarea ei. Trebuie făcut un balanț între precizia câștigată și timpul consumat pentru a alege varianta optimă.

Zgomotul este unul din factorii importanți care afectează calitatea unei imagini. Zgomotul poate reduce volumul și acuratețea informației achiziționate și el este introdus cel mai frecvent de interferențele magnetice și electronice care sunt prezente în apropierea echipamentelor sau componentelor calculatorului gazdă. Cea mai defavorabilă situație se întâlnește când zgomotul afectează chiar circuitele care alcătuiesc frame grabberul. Aici este un punct în care diferențele produse se deosebesc tehnologic. Calitatea proiectării și fabricației afectează volumul zgomotului introdus. O măsură a zgomotului introdus este raportul semnal/zgomot. Dacă semnalul are 40 dB și 2 dB reprezintă zgomotul atunci raportul semnal/zgomot este 38 dB (rezolvare logaritmică!). Cu cât acest raport semnal/zgomot este mai mare cu atât performanțele plăcii sunt mai bune. Raportul este de fapt un număr care reprezintă diferența dintre zgomot și maximul semnalului, deci în esență semnalul util. Ca urmare după ce semnalul este convertit numeric unii dintre biți vor fi eronați, începând cu LSB (Least Significant Bit). Dacă se lucrează pe 8 biți și zgomotul afectează ultimii 2 biți atunci se poate utiliza numai informația din 6 biți. Astfel numărul de biți efectivi, diferența dintre numărul de biți pe care se face conversia și numărul de biți eronați, devine mai important decât numărul de biți pe care se face conversia.

Achiziționarea unui frame-grabber pune și problema posibilității de cuplare pe calculatorul gazdă. Este greu de găsit o magistrală de calculator care să fie efectiv un standard industrial. Pentru calculatoarele personale această standardizare este mai ușor de realizat. La calculatoarele personale schimbarea acestui tip de magistrală se face mai des decât la calculatoarele industriale, din motive similare cu acelea prezentate într-un paragraf anterior referitoare la plăcile de achiziție și prelucrare de imagini. În aceste condiții, pentru a permite utilizarea plăcilor specializate de achiziție și prelucrare de semnale, calculatoarele personale au păstrat multă vreme și câte un slot sau două dintr-o versiune mai veche de magistrală de comunicație. O serie de parametri ai magistralei de comunicație influențează performanțele sistemului global: lărgimea de bandă de frecvență, viteza, opțiunile de autoconfigurare (plug & play) și independența de platforma software în care se dezvoltă aplicația.

Dacă se dorește achiziționarea unui frame grabber pentru a avea un singur sistem de achiziție și prelucrare de imagini atunci trebuie căutat unul care să se cupleze cu o anumită cameră de luat vederi, pe de o parte, iar pe de altă parte cu un anumit sistem de programe software. Sunt preferabile în această situație firmele care furnizează cât mai multe drivere pentru cuplarea cu multiple sisteme. Dacă se caută un frame grabber pentru a dezvolta un sistem care să fie ulterior revândut în mai multe exemplare atunci trebuie privite mult mai critic performanțele legate de jitter și raportul semnal/zgomot. Tot aici devine deosebit de importantă posibilitatea de a dezvolta aplicații software pe acel frame grabber. Multe firme oferă biblioteci de programe sursă pentru exemple de aplicații, suport software pentru dezvoltarea de aplicații și conexiuni telefonice directe care să permită actualizări ale programelor sau consultanță ulterioară cumpărării produsului.

2.5.2 Structura unei plăci de achiziție și prelucrare de imagini

Structura unei plăci de achiziție și prelucrare de imagini poate include următoarele componente (figura 2.20), unele optionale:

- Multiplexor de intrare,
- Convertor analog - numeric,
- LUT (Look Up Table) de intrare,
- Generator de semnale de sincronizare,
- Plan video,
- Plan grafic,
- Procesor de semnal,
- LUT (Look Up Table) de ieșire,
- Convertor numeric - analog.

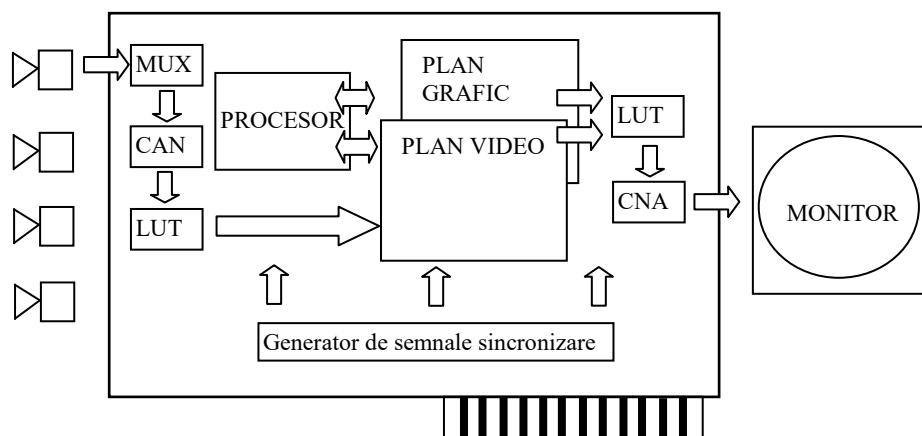


Figura 2.20 O structură de frame grabber

Multiplexorul permite cuplarea mai multor camere, oferind însă un singur semnal la un moment dat. Dacă viteza de prelucrare globală a sistemului este suficient de mare atunci, printr-un sistem de împărțire a timpului, se vor prelucra succesiv un număr de cadre pe secundă provenind de la fiecare cameră în parte. Soluțiile uzuale presupun 2, 4 sau 8 intrări de la tot atâtea camere diferite.

Semnalul selectat de multiplexor este preluat de **convertorul analog – numeric** care furnizează astfel nivelul de gri al pixelului curent. Dacă se lucrează cu imagini color vor exista trei convertoare pentru cele trei culori de bază RGB. Pentru camerele numerice conversia se face la nivelul camerei și atunci convertorul de pe placă de achiziție de imagini, la care se cuplează astfel de camere, nu mai este necesar. Dacă luăm ca exemplu o imagine cu nivele de gri (figura 2.21), atunci conversia se face între un prag inferior LOW, asociat cu negru, și un prag superior HIGH, asociat cu alb. Între aceste valori se repartizează cele 2^n nivele de gri, dacă se lucrează pe n biți. În mod curent numărul de biți pe care se face conversia nu poate fi schimbat pentru o placă de achiziție și prelucrare de imagini. Cea mai folosită soluție este repartiția uniformă a nivelelor de gri în interiorul intervalului de conversie. Nivelele de gri din imaginea reală presupun un cel mai închis nivel de gri G_{min} , care nu este întotdeauna negrul, și un cel mai deschis nivel de gri G_{max} , care nu este întotdeauna albul. Aceste valori se pot determina ușor din histograma imaginii.

Pentru a folosi ideal gama dinamică a senzorului, este necesar ca intervalul de conversie [LOW, HIGH] să fie identic cu intervalul nivelelor de gri din imagine [G_{min} , G_{max}]. Dacă imaginea nu este statică, este practic imposibil ca această identitate să se păstreze prin reglarea dinamică a pragurilor convertorului analog – numeric. Dacă $[G_{min}, G_{max}] \subset [LOW, HIGH]$ atunci toate nivelurile de gri furnizate de senzor vor apărea în imaginea numerică, dar convertorul va lucra și pentru valori care nu apar în semnalul imagine [LOW, G_{min}] și [G_{max} , HIGH]. În acest mod va avea loc o pierdere a preciziei de reprezentare a nivelelor de gri și deci o pierdere de contrast în raport cu situația ideală. Spre exemplu, dacă se fixează pragurile convertorului analog – numeric la $LOW = 0,2$ V și respectiv $HIGH = 1$ V, pentru o conversie pe 10 biți vom avea:

$$u_{g1} = (1000 \text{ mV} - 200 \text{ mV})/2^{10} = 800 \text{ mV}/1024 = 780 \text{ mV/nivel}$$

Dacă imaginea nu are niveli de gri decât în plaja $[G_{min}, G_{max}]$ cu $G_{min} = 0,4$ V și $G_{max} = 0,8$ V, atunci prin reglarea convertorului analog – numeric la $LOW = 0,4$ V și $HIGH = 0,8$ V atunci vom:

$u_{g2} = (800 \text{ mV} - 400 \text{ mV})/2^{10} = 400 \text{ mV}/1024 = 390 \text{ mV/nivel}$, deci o precizie dublă de reprezentare a nivelelor de gri.

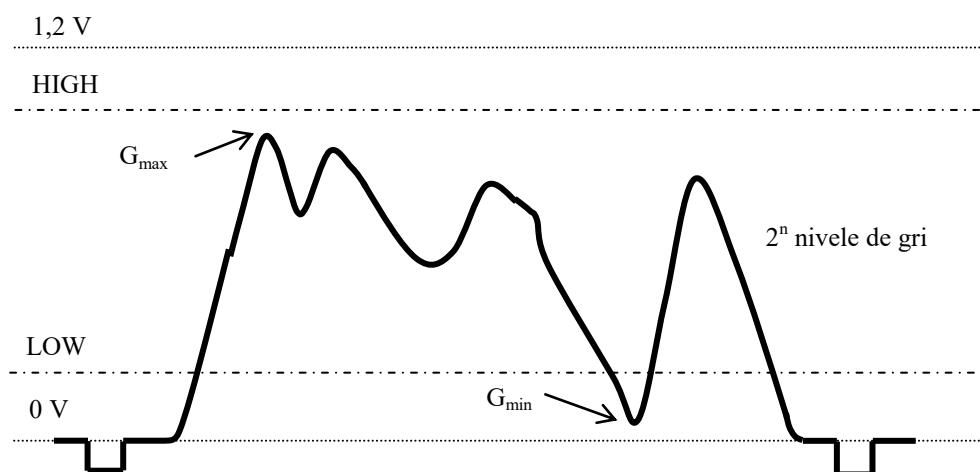


Figura 2.21 Pragurile CAN

Dacă $[G_{\min}, G_{\max}] \not\subset [\text{LOW}, \text{HIGH}]$ atunci există nivele de gri furnizate de senzor care nu sunt corect evidențiate în imaginea numerică. Nivelele de gri din plaja $[G_{\min}, \text{LOW}]$ vor fi toate forțate la negru și/sau nivelele de gri din plaja $[\text{HIGH}, G_{\max}]$ vor fi forțate toate la alb. Se va produce astfel o pierdere de rezoluție în reprezentarea nivelelor de gri extreme, cele mai închise și cele mai deschise.

În general prin reglarea pragurilor, inferior și superior, de conversie se obțin următoarele rezultate:

- prin scăderea pragului LOW imaginea devine mai deschisă, anumiți pixeli negri vor căpăta nivele de gri închise, dar diferite de negru,
- prin creșterea pragului LOW imaginea devine mai închisă, anumiți pixeli care aveau nivele de gri închise vor deveni negri,
- prin scăderea pragului HIGH imaginea devine mai deschisă, anumiți pixeli care aveau nivele de gri deschise vor deveni albi,
- prin creșterea pragului HIGH imagine devine mai închisă, anumiți pixeli albi vor căpăta nivele de gri deschise, dar diferite de alb.

Conversia analog – numerică este influențată și de tactul de conversie, care are valori de ordinul megahertzilor (zeci de cadre pe secundă, zeci de mii de linii pe secundă, sute sau mii de pixeli pe o linie). Valori mai mari ale tactului de pixeli “îngustează” pixelul și imaginea afișată cu același număr de coloane, iar valori mai mici ale acestui tact “lățesc” pixelul și imaginea afișată cu același număr de coloane.

LUT - ul de intrare este, din punct de vedere logic, o tabelă de conversie a nivelelor de gri sau colorilor. Din punct de vedere fizic, LUT - ul este o zonă de memorie. Valoarea numerică N oferită de convertorul analog - numeric este interpretată ca adresă în tabela de conversie (figura 2.22). Conținutul locației respective de memorie este noua valoare numerică $I(i,j)$ – nivelul de gri asociat pixelului situat în linia i coloana j în imaginea numerică - care o înlocuiește pe aceea oferită de convertorul analog – numeric. Este necesar, în acest caz, ca adresa și datele să fie reprezentate pe același număr de biți. Dacă se lucrează cu imagini color vor fi folosite 3 LUT – uri, câte unul pentru fiecare culoare de bază RGB.

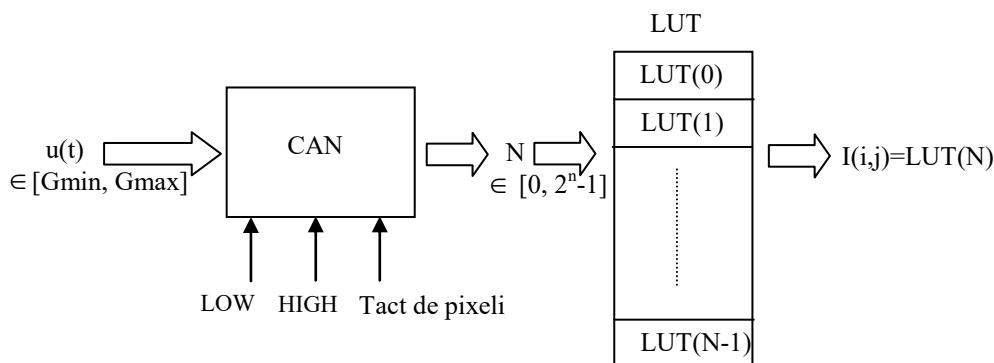


Figura 2.22 Interconectarea LUT-ului

LUT - ul oferă o modalitate foarte eficientă și rapidă de transformare a nivelelor de gri ale imaginilor numerice. LUT- ul realizează, din punct de vedere logic, o funcție $f(N) = M$ care transformă un nivel de intrare N în alt nivel de gri M , nivelul de gri de ieșire. Această transformare se face în timp real într-o memorie de 2^n-1 cuvinte de câte n biți. Programarea acestei tabele de conversie permite utilizarea tuturor nivelurilor de gri. Pentru a lăsa imaginea neschimbată tabela este programată în funcție identitate (figura 2.23 a): $LUT(N) = N$. Pentru a obține o imagine negativată funcția programată este (figura 2.23 b): $LUT(N) = (2^n-1) - N$.

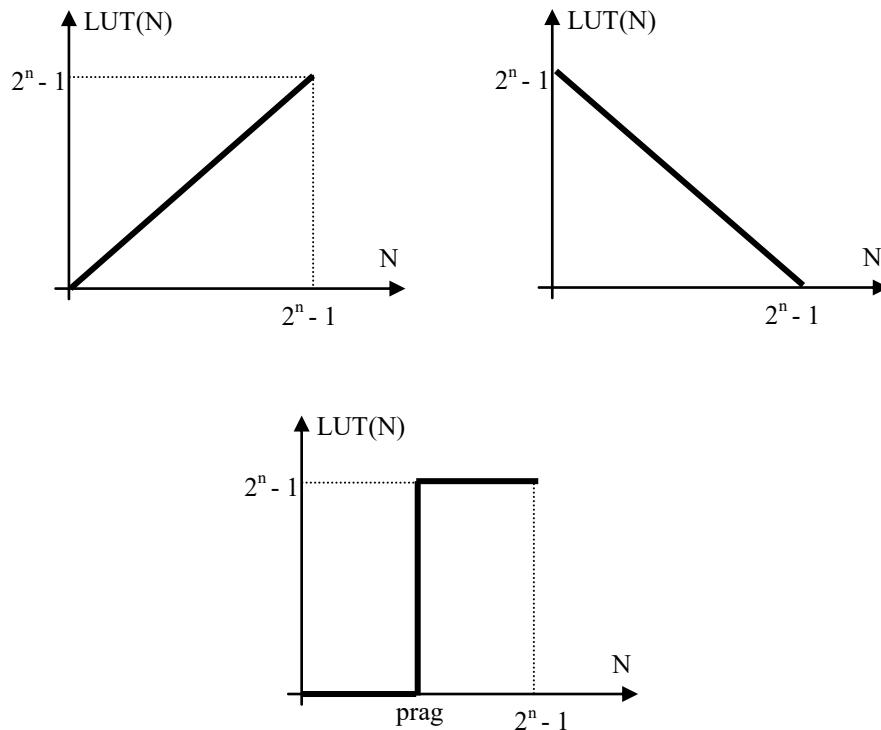


Figura 2.23 Modalități de programare a LUT-ului

Pentru a obține o imagine cu doar două nivele de gri, alb și negru, adică pentru a binariza imaginea (în sensul că ea poate fi atunci reprezentată cu doar două valori 1 și 0) se alege o funcție de tip prag (figura 2.23 c):

$$LUT(N) = \begin{cases} 0 & \text{pentru } N \leq \text{prag} \\ 2^n - 1 & \text{pentru } N > \text{prag} \end{cases}$$

Valoarea de prag utilă poate fi aleasă din analiza histogramei, dacă aceasta este bimodală.

O imagine poate fi corectată (dacă este prea luminoasă sau prea întunecată) prin modificarea distribuției de nivele de gri, utilizând LUT-ul (se spune că se modifică dinamica imaginii) prin alegerea unor funcții de transformare mai complicate, formate din porțiuni liniare de pante diferite sau chiar neliniare. Luminozitatea și contrastul unei imagini pot fi controlate cu ajutorul LUT-ului. Creșterea luminozității (figura 2.24) presupune creșterea nivelelor de gri în fiecare pixel și ca urmare se obține printr-o translatare în sus a funcției implementate în LUT. Analog, scăderea luminozității presupune scăderea nivelelor de gri în fiecare pixel și ca urmare se obține printr-o translatare în jos a funcției implementate în LUT. Luând în considerare că intervalul nivelelor de gri este limitat la $[0, 2^n - 1]$, se poate observa că nivelurile de gri care prin translația în sus ar trebui să depășească valoarea $2^n - 1$ vor fi limitate la această valoare, deci la alb. Similar, nivelurile de gri care prin translația în jos ar trebui să scadă sub valoarea 0 vor fi limitate la această valoare, deci la negru. În ambele cazuri are loc o pierdere de rezoluție în zona nivelelor de gri deschise, respectiv închise.

Creșterea contrastului (figura 2.25) presupune creșterea diferenței dintre nivelurile de gri asociate pixelilor și ca urmare se obține printr-o creștere a pantei funcției implementate în LUT. Dacă, spre exemplu, inițial $LUT(10) = 10$ și $LUT(20) = 20$ atunci diferența dintre cele două nivele de gri este $D_1 = 10$. Dacă se crește panta funcției din LUT astfel încât $LUT(10) = 15$ și $LUT(20) = 30$ atunci diferența dintre cele două nivele de gri este $D_2 = 15 > D_1$. Analog, scăderea contrastului presupune

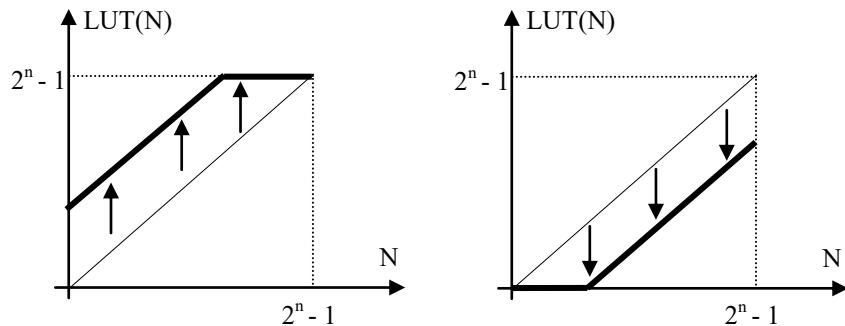


Figura 2.24 Controlul luminozității prin LUT

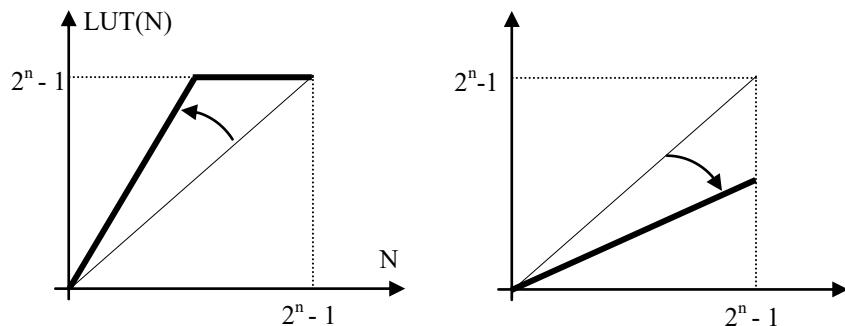


Figura 2.25 Controlul contrastului prin LUT

scăderea diferenței dintre nivelele de gri și ca urmare se obține printr-o scădere a pantei funcției implementate în LUT. Luând în considerare că intervalul nivelelor de gri este limitat la $[0, 2^n - 1]$, se poate observa că nivelele de gri care prin creșterea pantei ar trebui să depășească valoarea $2^n - 1$ vor fi limitate la această valoare, deci la alb. Are loc o pierdere de rezoluție în zona nivelelor de gri deschise.

Dacă are loc o programare a LUT-ului ca în figura următoare (figura 2.26), atunci se poate spune că în zonele din imagine cu pixeli care au nivele de gri în intervalul $[0, P_1]$ are loc o scădere a rezoluției prin forțarea acestor nivele la negru. În zonele din imagine cu pixeli care au nivele de gri în intervalul $[P_1, P_2]$ sunt loc de creștere a contrastului. În zonele din imagine cu pixeli care au nivele de gri în intervalul $[P_2, 2^N - 1]$ sunt loc de scădere a rezoluției prin forțarea acestor nivele la alb. Este de presupus că o astfel de soluție este convenabilă numai pentru imaginile care au $[G_{min}, G_{max}] \subset [P_1, P_2]$.

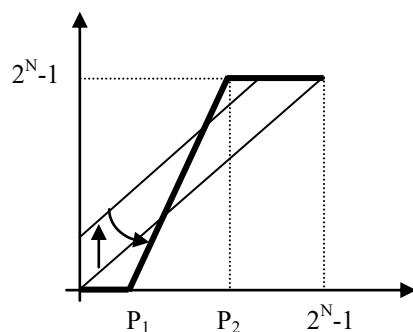


Figura 2.26

Modificările nivelelor de gri rezultate în urma analizei histogramelor sunt implementate tot cu ajutorul LUT – ului. În acest mod se pot modifica nivelele de gri în funcție de conținutul efectiv al imaginii, după ce aceasta a fost analizată prin calculul histogramei. Histograma este o funcție care asociază fiecărui nivel de gri numărul de pixeli care au acel nivel într-o imagine, ceea ce dă o informație utilă despre conținutul și distribuția informației din imagine.

Există diferite tehnici de redistribuire a nivelelor de gri pe baza analizei histogramei imaginii dintre care amintim aici următoarele trei:

- Punerea la scară: după o analiză a histogramei, nivelele de gri sunt configurate pentru a utiliza întreaga dinamică a semnalului de ieșire al senzorului. De exemplu, dacă cel mai mic nivel de gri corespunde, în imaginea reală, la P_1 , iar cel mai mare la P_2 , atunci LUT-ul va fi programat ca în figura anterioară.
- Egalizarea histogramei: teoretic imaginea cu cel mai bun contrast este aceea care utilizează egal toate nivelurile de gri. Într-o astfel de imagine, funcția de repartiție, adică totalul pixelilor mai întunecăți decât nivelul de gri curent (cu valori asociate mai mici decât aceea curentă), este o dreaptă. După egalizarea histogramei, vor fi realocate nivelurile de gri astfel încât funcția de repartiție să tindă, pe cât posibil adică atât cât permite conținutul imaginii reale, la o dreaptă.
- Reglajul contrastului: metoda prezentată anterior are inconvenientul că pentru imagini cu un fond omogen de mare suprafață, această corecție duce la un contrast prea mare pentru o zonă de gri limitată, ceea ce face vizibil zgromotul electronic. Reglajul după contrast aduce un remediu pentru că se face o corecție combinată între o aducere la scară și egalizarea histogramei.

Dacă la nivelul convertorului analog – numeric este puțin probabil să putem schimba numărul de biți și deci numărul de nivele de gri cu care se lucrează, acest lucru, în sensul micșorării numărului de nivele de gri, se poate face la nivelul LUT-ului. Soluția este să se programeze o funcție în trepte (figura 2.27), egale (pentru o distribuție liniară a nivelelor de gri) sau inegale (pentru o distribuție neliniară a nivelelor de gri)

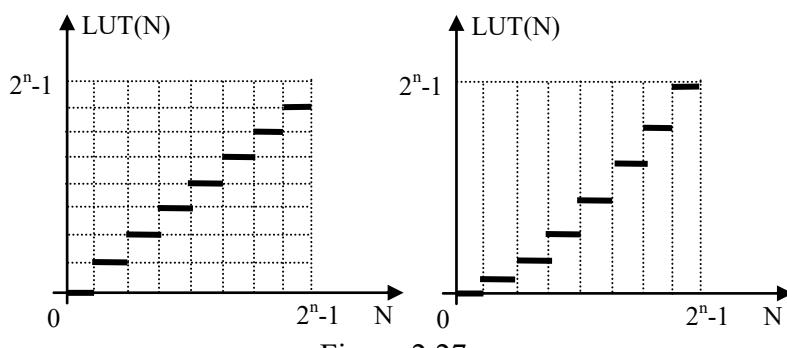


Figura 2.27

Valorile numerice furnizate de LUT sunt depuse în memoria video a plăcii, identificat aici prin termenul de plan video. Mărimea acestei memorii determină formatul maxim al imaginilor cu care poate lucra sistemul. Plăcile specializate pentru editarea / prezentarea de imagini lucrează cu formatele standardizate pentru sistemele de calcul, în raportul de 3/4 între dimensiuni: 640 x 480, 800 x 600, 1024 x 768, 1600 x 1200 etc. Plăcile de achiziție pentru aplicații științifice, medicale, industriale lucrează, în general, cu imagini pătrate 256 x 256, 512 x 512 sau 2048 x 2048. Dacă în primul caz pixelii au asociate culori reprezentate pe 3 sau 4 octeți, 24 sau 32 biți, pentru cel de al doilea caz cel mai frecvent se lucrează cu nivelurile de gri reprezentate pe 8 sau 10 biți sau cu culori reprezentate pe 3 x 8 biți. Cu cât memoria video este mai mare, cu atât mai multe imagini pot fi memorate pe placă. Cu cât există memorie video mai mare, cu atât crește posibilitatea să se realizeze prelucrarea imaginii la nivelul plăcii. Dacă nu există memorie video suficientă sau nu

există un procesor specializat la nivelul plăcii, atunci singura variantă este aceea a transferului imaginilor în memoria calculatorului gazdă și prelucrarea lor acolo.

Planul grafic este tot o memorie video, dar care conține o imagine generată și nu achiziționată și care se suprapune la afișare (este încrustată) peste imaginea achiziționată, memorată în planul video. În acest mod utilizatorul poate crea o imagine care să includă texte și primitive grafice care, eventual, să expliciteze elemente din imaginea achiziționată. Plasarea planului grafic într-o memorie separată permite păstrarea informației aferente lui deși planul video își împrospătează conținutul de câteva zeci de ori pe secundă.

LUT-urile de ieșire funcționează similar cu cele de intrare, dar valorile conținute în locațiile tableei vor înlocui pe acele din planul video. Noile valori vor fi convertite de convertoroarele numeric – analogic și afișate ulterior pe un monitor. Modificarea LUT-urilor de intrare schimbă informația din planul video și implicit imaginea de pe monitor. Modificarea LUT-urilor de ieșire nu schimbă decât imaginea afișată, dar lasă neschimbat conținutul planului video.

Convertoarele numeric – analogice furnizează semnalele standardizate necesare pentru a fi afișate pe monitoarele video. Pentru aplicațiile de achiziție și prelucrare, inclusiv recunoaștere, automate a imaginilor, imaginea afișată pe monitor influențază mai puțin performanța globală a sistemului.

Procesorul de semnal poate să fie un procesor obișnuit care este programat să execute prelucrările dorite asupra imaginilor sau un procesor specializat care execută hard, cu viteză crescută, aceste prelucrări. Suportul hard pentru prelucrările la nivel de pixel poate asigura o creștere semnificativă a performanțelor globale ale sistemului. La o imagine de 1024 x 1024 simpla aplicare a unui filtru sub forma unei măști 3 x 3 necesită, dacă ignorăm alte operații decât cele aritmetice, efectuarea a $9.446.400$ înmulțiri și $1.049.600$ împărțiri. Pentru timp real ar fi necesară prelucrarea a 25 sau 30 de imagini în fiecare secundă. Evident că o astfel de filtrare nu este decât una dintre prelucrările necesare. Pentru a putea partaja sarcina realizării acestor prelucrări între mai multe plăci identice sau nu, este nevoie ca procesoarele respective să prezinte facilități, pe de o parte de comunicație de date de mare viteză pentru schimbul de date între procesele care se desfășoară în paralel, și, pe de altă parte, de sincronizare pentru a putea coordona aceste procese între ele.

Curs 6

3. Formarea imaginii numerice

Există două aspecte ale procesului prin care fluxul luminos emis de scena reală trece prin lentilă și este proiectat pe planul senzorial.

3.1 Transformarea optică

Mai întâi se pune problema dacă distanțele dintre obiect și lentilă, dintre lentilă și planul imaginii, verifică legile specifice ale opticii, deci dacă imaginea este bine focalizată. În acest caz planul senzorial va coincide cu planul imaginii. În caz contrar între cele două plane va exista o distanță nenulă care va introduce o estompare, o neclaritate în imagine (efect flou).

Studiile optice ale fenomenului au arătat că dependența între imaginea ideală I_i , corespunzătoare unei focalizări perfecte, și imaginea reală I_r , corespunzătoare unei distanțe d între planul imagine și planul senzorial, se poate exprima sub formă liniară:

$$I_r(x,y) = h(x,y) * I_i(x,y)$$

unde $*$ reprezintă o conoluție. h este răspunsul impulsional al sistemului și depinde de natura și caracteristicile lentilei, respectiv ale luminii și sursei de lumină. Pentru o lumină albă și coerentă se poate face o aproximare gaussiană a răspunsului impulsional:

$$h(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad \text{unde dacă } d \rightarrow 0 \text{ atunci } \sigma \rightarrow 0 \text{ și } h \rightarrow 1.$$

Scena și iluminarea obiectelor de interes din ea sunt foarte importante pentru rezultatul final, care este clasificarea / recunoașterea acestor obiecte. Deși alegerea soluției de iluminare este clar dependentă de aplicație, totuși unele aspecte generale pot fi evidențiate și au fost facute într-un paragraf anterior.

3.2 Transformarea geometrică

Al doilea aspect, dintre cele două amintite anterior, este proiecția perspectivă care aduce scena 3D în imaginea 2D. Una dintre deosebirile dintre grafica pe calculator și vederea artificială, este legătura mult mai directă cu realitatea, legătură pe care o presupune al doilea domeniu enumerat. Imaginea reprezintă aici chiar scena reală, inclusiv zgomotele, iar scopul final este folosirea informațiilor obținute din această imagine în acțiuni relative tot la scena reală (de exemplu manipularea unor obiecte). Fie și numai din aceste considerente se desprinde necesitatea găsirii legăturii dintre coordonatele din imaginile numerice disponibile, și coordonatele dintr-un sistem de referință legat de scena reală.

În cea mai simplă aproximare camera este o lentilă sferică cu centrul optic în C și cu distanța până la planul imaginii f . P' este proiecția punctului P , din spațiu, în planul imaginii ca în figura 3.1. Avem atunci dezavantajul că imaginile apar răsturnate. Pentru a înlătura acest dezavantaj, plasăm prin convenție, planul imaginii în fața lentilei ca în figura 3.2.

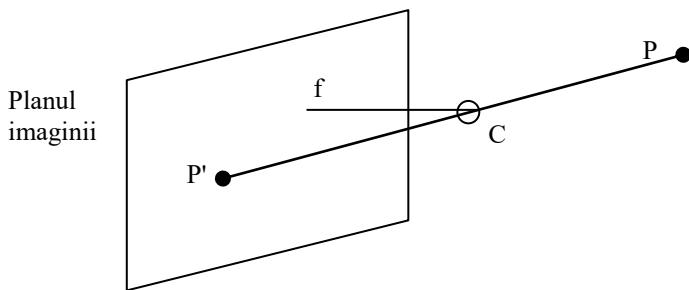


Figura 3.1 Formarea imaginii

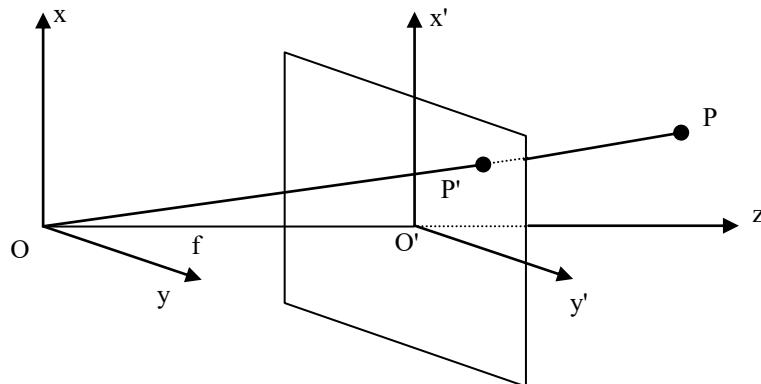


Figura 3.2 Mutarea planului imaginii

Apare aici o nedeterminare în sensul că toate punctele de pe dreapta PC au aceeași proiecție P'. Această nedeterminare nu va putea fi rezolvată decât prin prezența unei două camere de luat vederi: vederea stereo, după modelul vederii binoculare umane.

Problema se poate pune în două moduri: să se găsească coordonatele proiecției P' dacă se cunosc cele ale punctului real P, și invers, să se găsească locul geometric al punctelor P dacă se cunosc coordonatele proiecției P'.

Transformarea $P \rightarrow P'$ se numește transformarea perspectivă directă, iar transformarea $P' \rightarrow P$ se numește transformarea perspectivă inversă. În acest model axa optică a camerei punctuale ("pin-hole") este aliniată cu axa z și planul imaginii se găsește la distanța $z = f$, f distanța focală. Folosind simple asemănări geometrice se poate scrie:

$$\begin{cases} x' = \frac{x \cdot f}{z} \\ y' = \frac{y \cdot f}{z} \\ z' = f \end{cases}$$

Observație: Față de figură, originea sistemului O' este translatată pe axa z în O .

Se observă că transformarea perspectivă directă este simplă, dar neliniară. Pentru a obține o reprezentare liniară, apelăm la coordonatele omogene; se introduce un coeficient de scară w într-un spațiu 4D. Dacă $v = (x, y, z)^t$ în 3D atunci în coordonate omogene $v' = (wx, wy, wz, w)^t$. Pentru un punct din 3D nu există o singură reprezentare în coordonate omogene. Dacă considerăm, spre exemplu, raportul între z și f ca parametru atunci avem:

$$\lambda = \frac{z}{f} \quad \text{și} \quad w = \frac{1}{\lambda}$$

Și deci forma liniară pentru transformarea perspectivă este:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} wx \\ wy \\ wz \\ w \end{bmatrix} \quad \text{și} \quad \begin{bmatrix} \lambda x' \\ \lambda y' \\ \lambda \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{f} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Din considerente practice, este util să reprezentăm coordonatele în raport cu un sistem de referință global. Coordonatele (x, y, z) se vor raporta acum la un astfel de sistem. Sistemul de referință legat de cameră își păstrează originea în centrul lentilei, iar planul imaginii se află tot la distanța f față de această origine în lungul axei z' . În acest caz, în transformarea perspectivă intervin translații și rotații în jurul axelor de coordonate ale sistemelor de referință. Dacă translația se face cu vectorul T_x, T_y, T_z , iar rotațiile se reprezintă în secvență R_x, R_y, R_z , respectiv cu unghiurile α, β, γ în jurul axelor de coordonate (figura 3.3), atunci transformarea perspectivă ia forma matriceală următoare:

$$\begin{bmatrix} \lambda x' \\ \lambda y' \\ \lambda \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{f} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & T_x \\ 0 & 1 & 0 & T_y \\ 0 & 0 & 1 & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

$$\cdot \begin{bmatrix} \cos \gamma & \sin \gamma & 0 & 0 \\ -\sin \gamma & \cos \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos \beta & 0 & -\sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Sau mai compact, trecerea de la coordonate obiect la coordonate imagine se scrie:

$$P' = A_p \cdot A_t \cdot R_z \cdot R_y \cdot R_x \cdot P$$

Calculăm invers $P(P')$ și trebuie să găsim transformata perspectivă inversă și deci locul geometric din spațiu, o dreaptă pe care se găsesc punctele P care se proiectează toate în P' . Dacă ar interveni o a doua cameră de luat vederi, atunci la intersecția celor două locuri geometrice s-ar putea localiza punctul P în spațiu, cunoscând proiecțiile lui în cele două imagini și parametrii camerelor de luat vederi.

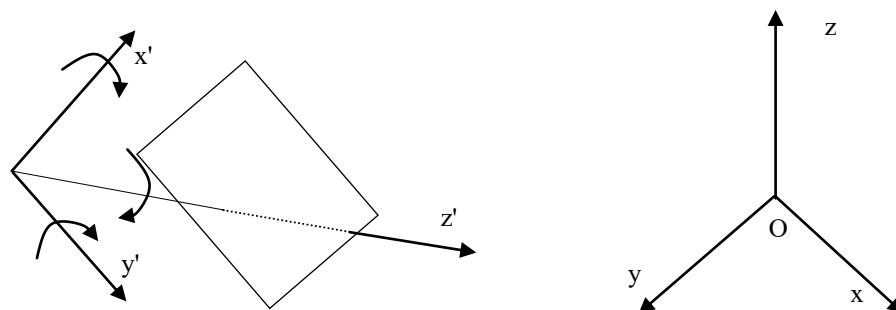


Figura 3.3 Raportarea la un sistem de coordonate global

3.3 Calibrarea camerelor de luat vederi

O etapă preliminară în aplicațiile de vedere artificială este tocmai determinarea localizării camerei (sau camerelor) de luat vederi, a parametrilor din transformările perspective asociate, în vederea controlării poziționărilor în spațiul de lucru. Modelul camerei de luat vederi exprimat sub formă matriceală este folosit pentru a transforma coordonatele obiect în coordonate imagine. Determinarea acestui model se numește calibrarea camerei de luat vederi.

Folosim relațiile deja găsite:

$$C_1 \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda x' \\ \lambda y' \\ \lambda \\ 1 \end{bmatrix} \quad \text{sau} \quad C \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda x' \\ \lambda y' \\ \lambda' \\ 1 \end{bmatrix}$$

unde $C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix}$ este matricea omogenă a transformării.

Dacă înlocuim pe λ găsim următoarele două ecuații:

$$\begin{cases} c_{11}x + c_{12}y + c_{13}z + c_{14} - x'(c_{31}x + c_{32}y + c_{33}z + c_{34}) = 0 \\ c_{21}x + c_{22}y + c_{23}z + c_{24} - y'(c_{31}x + c_{32}y + c_{33}z + c_{34}) = 0 \end{cases}$$

Ne punem problema principală a găsirii matricei omogene a transformării, a calibrării camerei de luat vederi. Dacă se stabilește o corespondență, deci dacă se știu coordonatele x , y , z pentru P din scenă și x' , y' pentru punctul P din imagine, atunci avem două ecuații (cele de mai sus) și 12 necunoscute (coeficienții matricei C). Mai mult, dacă găsim 6 astfel de corespondențe, atunci vom avea 12 ecuații pentru 12 necunoscute c_{ij} . Datorită omogenității matricei C , se poate considera $c_{34}=1$ și rămân numai 11 necunoscute, deci sistemul este supradeterminat. Atunci cele două ecuații se rescriu astfel:

$$\begin{aligned} c_{11}x + c_{12}y + c_{13}z + c_{14} + c_{21}0 + c_{22}0 + c_{23}0 + c_{24}0 - x'c_{31}x - x'c_{32}y - x'c_{33}z - x'c_{34} &= x' \\ c_{11}0 + c_{12}0 + c_{13}0 + c_{14}0 + c_{21}x + c_{22}y + c_{23}z + c_{24} - y'c_{31}x - y'c_{32}y - y'c_{33}z - y'c_{34} &= y' \end{aligned}$$

Punând sistemul sub formă matriceală se poate da o soluție de eroare pătratică minimă folosind pseudo-inversa:

$$C^T = [c_{11} \ c_{12} \ c_{13} \ c_{21} \ c_{22} \ c_{23} \ c_{31} \ c_{32} \ c_{33}]$$

$$V^T = [x'^1 \ y'^1 \ x'^2 \ y'^2 \ x'^3 \ y'^3 \ x'^4 \ y'^4 \ x'^5 \ y'^5 \ x'^6 \ y'^6]$$

unde indicele numeric superior corespunde numărului punctului de calibrare considerat.

$$C = (U^T U)^{-1} U^T V$$

unde:

$$U = \begin{bmatrix} x^1 & y^1 & z^1 & 1 & 0 & \dots & -x'^1x^1 & -x'^1y^1 & -x'^1z^1 \\ 0 & 0 & 0 & 0 & x^1 & \dots & -y'^1x^1 & -y'^1y^1 & -y'^1z^1 \\ x^2 & y^2 & z^2 & 1 & 0 & \dots & -x'^2x^2 & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & x^6 & y^6 & z^6 & -y'^6x^6 & -y'^6y^6 & -y'^6z^6 \end{bmatrix}$$

S-a considerat că se fac 6 corespondențe între puncte din scena reală și puncte din imagine. Se poate aprecia că este util să se ia în considerare mai mult de 6 corespondențe, pentru că prin supradeterminarea sistemului să se poată crește precizia de determinare a modelului camerei de luat vederi. Deși simplă la o primă apreciere, determinarea cu precizie a coordonatelor pentru cele șase puncte din scena reală este dificilă în practică. Determinarea cu precizie de 1 mm a coordonatelor x, y, z ale unuia din vârfurile unei corp poliedral, de dimensiuni de ordinul centimetrilor așezat pe o masă, față de un sistem de coordonate atașat unuia din colțurile mesei, nu este în practică o problemă simplă. Se introduc erori semnificative prin aproximarea direcțiilor paralele și perpendiculare. Nici determinarea coordonatelor în planul imaginii nu este lipsită de probleme. Semnalăm numai fie și problema determinării centrului imaginii, care nu este aproape niciodată chiar în centrul imaginii. Centrul imaginii este punctul în care axa lentilei înțeapă dreptunghiul senzorial în care se formează imaginea. Dacă am schimba focalizarea obiectivului, atunci centrul imaginii ar fi singurul punct din imagine care rămâne pe loc. Diverse metode de calibrare propun diferite modalități de alegere a mulțimii de puncte, minim 6, de calibrare. Spre exemplu o soluție este trasarea unei grile rectangulare de calibrare cu determinarea coordonatelor punctelor de intersecție din grilă. Pentru aplicațiile de robotică, se preferă să se programeze robotul astfel ca el însuși să marcheze punctele de calibrare. Metoda are avantajul că pentru tandemul sistem de vedere - robot, un anumit punct va avea aceleași coordonate chiar dacă ele pot dифeri eventual de cele măsurate de un alt treilea sistem. Dacă avem matricea de transformare a camerei, putem determina coordonatele punctelor din scenă dacă le cunoaștem pe cele din imagine.

$$\begin{cases} c_{11}x + c_{12}y + c_{13}z + c_{14} = x'(c_{31}x + c_{32}y + c_{33}z + c_{34}) \\ c_{21}x + c_{22}y + c_{23}z + c_{24} = y'(c_{31}x + c_{32}y + c_{33}z + c_{34}) \end{cases}$$

se obține:

$$\begin{cases} (c_{11} - x'c_{31})x + (c_{12} - x'c_{32})y + (c_{13} - x'c_{33})z + (c_{14} - x'c_{34}) = 0 \\ (c_{21} - y'c_{31})x + (c_{22} - y'c_{32})y + (c_{23} - y'c_{33})z + (c_{24} - y'c_{34}) = 0 \end{cases}$$

sau cu o nouă notație suplimentară:

$$\begin{cases} a_1x + b_1y + c_1z + d_1 = 0 \\ a_2x + b_2y + c_2z + d_2 = 0 \end{cases}$$

Acste două ecuații sunt ecuațiile a două plane; intersecția acestor două plane determină o dreaptă pe care se află punctul real de coordonate (x,y,z) care s-a proiectat în imagine în punctul (x',y'). Prin rezolvare se găsesc coordonatele x și y, dar nu și z-ul.

$$\begin{cases} x = \frac{z(b_1c_2 - b_2c_1) + (b_1d_2 - b_2d_1)}{a_1b_2 - a_2b_1} \\ y = \frac{z(c_1b_2 - c_2b_1) + (d_1a_2 - d_2a_1)}{a_1b_2 - a_2b_1} \end{cases}$$

Cum se obține cea de a treia dimensiune ? Adică distanța de la punctul din scenă la camera de luat vederi. În unele aplicații această distanță este constantă și cunoscută. Este cazul, spre exemplu al camerei fixe montate deasupra unei benzi transportoare. Dar situația nu se poate generaliza. În alte aplicații, se folosește aşa numita tehnica a iluminării structurate. În acest caz se cunosc poziția și direcția sursei de iluminare, și poziția camerei de luat vederi, iar punctul iluminat este ușor identificabil în scenă (de exemplu o iluminare punctiformă). Din aceste informații se poate deduce adâncimea, cea de a treia dimensiune prin metoda triangulației active.

O metodă binecunoscută este și vederea stereo, inspirată din binocularitatea vederii biologice. Dacă avem două camere de luat vederi, aşa cum am sugerat deja, și cunoaștem pentru fiecare modelul transformării, atunci în loc să facem intersecția a două plane, vom avea intersecția a patru plane (figura 3.4), care conduc la intersecția a două drepte ceea ce dă un punct, adică localizează în spațiul 3D punctul care a creat cele două puncte corespondente în cele două imagini:

$$\begin{cases} a_1x + b_1y + c_1z + d_1 = 0 \\ a_2x + b_2y + c_2z + d_2 = 0 \\ p_1x + q_1y + r_1z + s_1 = 0 \\ p_2x + q_2y + r_2z + s_2 = 0 \end{cases}$$

unde a_i , b_i , ..., s_i , $i=1, 2$. sunt funcții liniare de C_{1ij} și C_{2ij} , adică coeficienții modelelor camerelor și de x_1' , y_1' , x_2' și y_2' , adică coordonatele proiecțiilor punctului din scenă în cele două imagini, respectiv stânga și drepta. Avem acum trei necunoscute x , y , z , dar patru ecuații, și deci se poate calcula o soluție de eroare pătratică minimă folosind o pseudo-inversă ca mai înainte. Aici problema cea mai dificilă nu este calculul matematic, iar problemele legate de erorile de măsură țin, așa cum am arătat, de faza de determinare a modelelor camerelor. Principala problemă care trebuie rezolvată, aici este punerea în corespondență a punctelor: care sunt, cu precizie, cele două puncte P_1' și P_2' din cele două imagini care corespund punctului real P ale cărui coordonate se doresc a fi calculate. În acest caz, sistemul de coordonate de referință este legat de sistemul celor două camere (poate fi unul dintre ele sau fixat la mijloc). Dacă poziția și orientarea relativă a celor două camere, ca și parametrii lor intrinseci nu se schimbă atunci calibrarea rămâne valabilă.

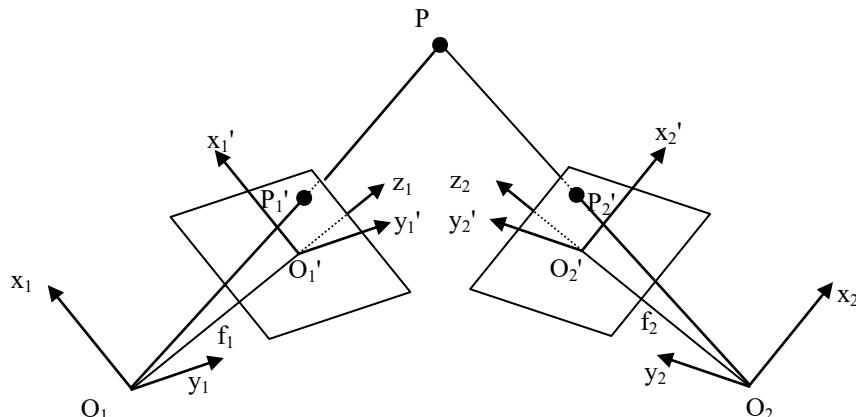


Figura 3.4 Videre binoculară

Ca o scurtă concluzie, se poate afirma că în problema calibrării camerelor de luat vederi, prin metode pasive (n-am menționat nimic despre laser, de exemplu), aparatul matematic folosit este bine stăpânit și nu foarte complicat. Calibrarea se face off-line și deci nu se pun probleme de viteză de calcul sau necesar de memorie, care oricum nu au valori restrictive. Problemele sunt legate de obținerea unei precizii bune și imune la zgomotele și perturbațiile care apar în timpul aplicației propriu-zise. Orice schimbare în poziția, orientarea sau parametrii intrinseci ai camerei de luat vederi anulează rezultatele calibrării, matricele se modifică și o nouă calibrare se impune.

Nu sunt lipsite de interes două observații. Deși se cunosc multe și diverse abordări ale problemei, totuși în literatura de specialitate apar în continuare noi metode de rezolvare a problemei. Inevitabil, sistemul senzorial de vedere umană este luat ca model și referință, deși se poate trage concluzia că datorită diferenței de performanță a celor două sisteme (sistemul biologic fiind de departe mai bine dotat) nu este poate, cel puțin deocamdată, cea mai bună idee. Chiar și aşa sistemul de vedere umană este mai degrabă performant din punct de vedere al aprecierii calitative, decât din punct de vedere al măsurătorii efective, al aprecierii cantitative. Încercați, și veți constata că vă este mult mai ușor să apreciați vizual care obiect este mai apropiat de dumneavoastră (deși și aici sunt posibile iluzii) decât să apreciați în centimetri, cu o eroare de câteva procente, distanța în milimetri sau centimetri până la un obiect.

Este interesant să analizăm cum se calibrează sistemul de vedere uman. Sistemul de vedere uman se bazează pe doi ochi, doi senzori care formează un sistem de vedere stereo. Creierul suprapune în timp real cele două imagini, adică rezolvă corespondența stereo împerechind punctele din cele două imagini care corespund aceluiași punct din spațiu. Se realizează în acest fel corespondența “ochi – mâna”, adică omul știe unde se află ceea ce vede și ca urmare poate localiza corect obiectele și le poate eventual manipula. Această operație este facilitată de poziția fixă a ochilor în craniu și de poziționarea și orientarea controlată a craniului față de scheletul care include și mâinile (“manipulatoarele”). În acest mod corespondența dintre punctele din imagine și cele din realitate este învățată prin experiență și ea nu se schimbă. Dacă încercați să manipulați obiectele folosind o imagine provenind de la o oglindă sau una provenind de la o cameră de luat vederi puteți avea dificultăți pentru că noua corespondență nu a fost învățată, ea trebuie căutată pe loc, în timp real. Dacă cele două camere de luat vederi (cei doi ochi) și-ar schimba permanent poziția și/sau orientarea și/sau parametrii intrinseci atunci corespondența dintre punctele din imagine și punctele din spațiul de operare s-ar schimba permanent.

3.4 Conversia analog-numerică

Scopul acestei operații este transformarea imaginii proiectate de lentilă pe suprafața senzorului I_r într-o imagine numerică I_n .

3.4.1 Eșantionare

O eșantionare ideală constă în multiplicarea imaginii I_r cu un impuls Dirac multiplicat spațial ca în figura 3.5.

$$I_n = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} I_r(i \cdot \Delta x, j \cdot \Delta y) \delta(x - i \cdot \Delta x, y - j \cdot \Delta y)$$

unde pașii de eșantionare $\Delta x, \Delta y$ verifică teorema lui Shannon.

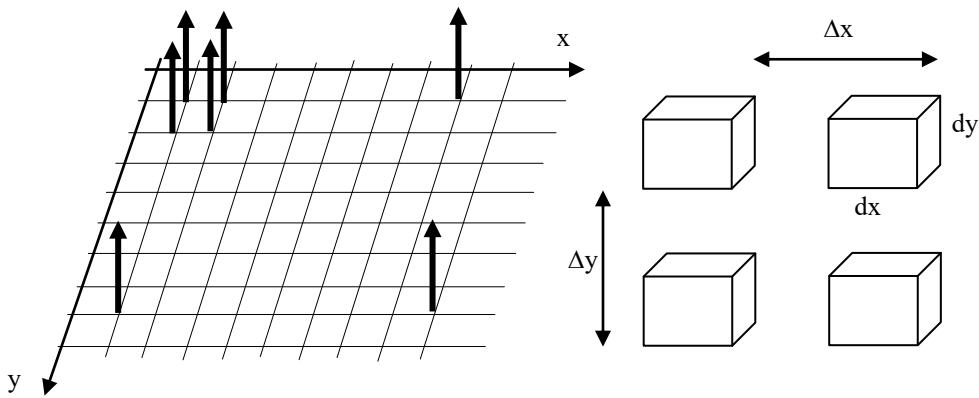


Figura 3.5 Eșantionarea imaginilor

Referindu-ne la principiul de funcționare al camerelor CCD eșantionarea în fiecare punct corespunde la integrarea imaginii formate pe fiecare zonă de dimensiune $dx \cdot dy$ (dimensiunea pixelului):

$$I_n(i \cdot \Delta x, j \cdot \Delta y) = \int_{i \cdot \Delta x - \frac{dx}{2}}^{i \cdot \Delta x + \frac{dx}{2}} \int_{j \cdot \Delta y - \frac{dy}{2}}^{j \cdot \Delta y + \frac{dy}{2}} I_r(u, v) \cdot du \cdot dv$$

Dacă presupunem că se face o singură integrare într-o unitate de timp, atunci putem extinde la infinit limitele de integrare:

$$I_n(i \cdot \Delta x, j \cdot \Delta y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_r(u, v) \cdot \Pi\left(\frac{dx}{2}, \frac{dy}{2}\right) \cdot (i \cdot \Delta x - u, j \cdot \Delta y - v) \cdot du \cdot dv$$

unde $\Pi\left(\frac{dx}{2}, \frac{dy}{2}\right)$ este o funcție de tip palier definită pe zona $dx \cdot dy$ (figura 3.5).

Funcția continuă imagine numerică I_n se poate obține atunci ca un produs de conoluție între I_r și Π .

$$I_n(x, y) = (I_r * \Pi\left(\frac{dx}{2}, \frac{dy}{2}\right))(x, y) \cdot \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \delta(x - i \cdot \Delta x, y - j \cdot \Delta y)$$

Deci eșantionarea imaginii I_r revine la o conoluție cu o funcție de tip palier, urmată de o eșantionare cu impulsuri Dirac. Se poate arăta că prin interpolare se poate reconstrui imaginea inițială: la o eșantionare ideală corespunzând o imagine reconstruită ideală.

Eșantionarea spațială este urmată de o eșantionare în frecvență, prin care valorile eșantionate sunt proiectate într-un interval discret $[0, 2^m - 1]$, unde 2^m este numărul de nivele de gri. În principiu se pune problema unei cuantificări optimale prin minimizarea, de tipul celor mai mici pătrate, a distorsiunilor între valoarea reală și valoarea cuantificată.

Pentru conversia analog – numerică a imaginii se mai folosesc termenii digitizare sau discretizare: a eșantiona imaginea pentru a discretiza coordonatele (x, y) și a cuantifica nivelele de gri pentru a discretiza amplitudinea într-un punct.

În aplicațiile de vedere artificială, ușual se lucrează cu 2^m nivele de gri cu $m = 8, 10, 12$ și 16 și cu imagini de dimensiune $N \times N$ cu $N=2^n$ cu $n= 9, 10, 11$ (adică $512 \times 512, 1024 \times 1024$ pixeli, 2048×2048 pixeli). Deci numărul de biți necesari pentru a memora o imagine este $b = N \times N \times m$. De exemplu o imagine de 512×512 pixeli cu 256 nivele de gri (deci pe 8 biți) ocupă $512 \times 512 \times 8 = 512 \times 512$ octeți = 256 kB și analog o imagine color de 1024×1024 pixeli cu $2^8 \times 2^8 \times 2^8 = 2^{24} = 16$ megaculatori ocupă 3 MB.

Pot fi identificați trei factori care limitează rezoluția efectivă a imaginii și fidelitatea imaginilor numerice:

- frecvența de eșantionare a senzorului,
- lățimea de bandă a semnalului video,
- frecvența de eșantionare a convertorului analog-numerice.

Cât de mică trebuie să fie zona de eșantionare pentru a putea lucra cu varianta discretizată a imaginii, cu imaginea numerică în loc de imaginea continuă, analogică? Teorema de eșantionare a lui Shannon stipulează, în esență, că dacă un semnal de bandă limitată BHz trebuie eșantionat, atunci frecvența utilizată în acest proces trebuie să fie mai mare sau egală cu dublul frecvenței de tăiere. Dacă un senzor liniar are 756 de fotodiode dispuse liniar pe direcția orizontală, atunci el va fi capabil să reprezinte o variație în intensitatea luminoasă care are o frecvență spațială de 378 de ciclii pe unitatea de distanță. Eșantionarea unei imagini este o operație analoagă care se obține prin decuparea unui semnal bidimensional, după o rețea pătrată pe x și pe y. Eșantionarea nu depinde numai de frecvența aleasă pentru impulsurile Dirac. Cu cât dimensiunea unei celule, CCD sau CMOS, este mai mai mare, cu atât se introduce un prefiltraj al imaginii (o conoluție a imaginii sau o multiplicare cu o funcție de transfer) mai important. Pentru standardul CCIR frecvența pe linii este 15625 Hz. În plus, dacă lărgimea de bandă nominală a semnalului video este 5,0 MHz, adică un semnal poate transmite o imagine video cu cinci milioane de variații periodice în nivelul (strălucirea) semnalului, atunci rezultă un maxim de $5 \times 10^6 / 15625 = 320$ variații periodice pe linie, deci frecvența spațială maximă care poate fi reprezentată corect de un semnal video este de 320 ciclii pe linie.

Curs 7

3.4.2 Poziția pixelilor în imagine

Modelul imaginii numerice introdus anterior, a arătat că imaginea se reduce la o matrice de pixeli codificați pe m biți și ca urmare acoperind 2^m nivele de gri sau culori. Un pixel reprezentat prin intensitatea sau nivelul de gri sau culoarea, considerate atribute ale punctului $P(i,j)$, are patru puncte vecine în pozițiile: $(i+1,j)$, $(i-1,j)$, $(i,j+1)$, $(i,j-1)$. Putem nota acest ansamblu de pixeli (4 vecini) prin $N_4(P)$ și îi putem identifica prin punctele cardinale cărora le corespund. Pe diagonală mai putem identifica 4 vecini în pozițiile: $(i+1,j+1)$, $(i+1,j-1)$, $(i-1,j-1)$, $(i-1,j+1)$. Acestea îl putem nota cu $N_D(P)$. Dacă facem reuniunea $N_8 = N_4(P) \cup N_D(P)$ obținem o mulțime de 8 pixeli care se numesc 8-vecinii pixelului curent (figura 3.6).

$P(x-1,y-1)$	$P(x-1,y)$	$P(x-1,y+1)$
$P(x,y-1)$	$P(x,y)$	$P(x,y+1)$
$P(x+1,y-1)$	$P(x+1,y)$	$P(x+1,y+1)$

Figura 3.6 8-vecinătate

Asociind poziției și atributul unui pixel, de exemplu nivelul de gri, se poate introduce noțiunea de conectivitate relativ la doi pixeli $P_1(x_1,y_1)$ și $P_2(x_2,y_2)$ cu nivelele de gri g_1 respectiv g_2 . Putem defini 3 tipuri de conectivitate impunând condiția $g_1, g_2 \in [a, b]$ în următorul sens:

- 4-conectivitate: P_1, P_2 sunt 4-conectați dacă $P_1 \in N_4(P_2)$,
- 8-conectivitate: P_1, P_2 sunt 8-conectați dacă $P_1 \in N_8(P_2)$,
- conectivitate mixtă: se obține din modificarea 8-conectivității prin eliminarea conectivităților multiple care sunt surse de erori. P_1, P_2 sunt mixt - conectați dacă $P_1 \in N_4(P_2)$ sau dacă $P_1 \in N_D(P_2)$ și $N_4(P_1) \cap N_4(P_2) = \emptyset$ (un exemplu în figura 3.7)

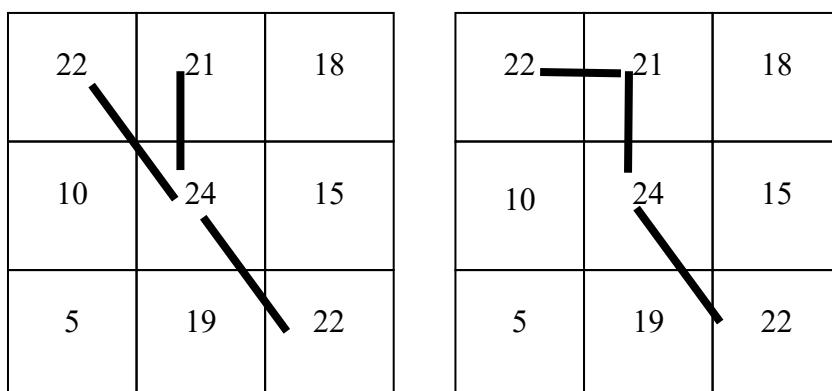


Figura 3.7 Exemplu de conectivitate mixte pentru $a=20$ și $b=24$

Putem calcula distanțele între doi pixeli $P_1(x_1, y_1)$ și $P_2(x_2, y_2)$, dintr-o imagine numerică cu nivele de gri, folosind metrice de următoarele forme:

- distanța euclidiană: $d(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

- distanța Manhattan: $d(P_1, P_2) = |x_1 - x_2| + |y_1 - y_2|$

- distanța de tip şah: $d(P_1, P_2) = \max(|x_1 - x_2|, |y_1 - y_2|)$

Așa cum s-a arătat, raportul dimensiunilor, orizontală și verticală, pentru imaginile TV tradiționale este 4/3. Pentru sistemele de vedere artificială care presupun inspecție și măsurare este utilă o imagine pătrată cu un raport de 1/1. Dacă se folosește un semnal video cu raportul 4/3, iar imaginea analogică este digitizată pentru a obține o imagine numerică pătrată, atunci va avea loc o deformare a distanțelor pe cele două direcții perpendiculare, distanța efectivă dintre pixelii vecini pe orizontală devenind de 4/3 ori mai mare decât aceea pe verticală. Dacă nu am lăsat în considerare această deformare, am putea considera că distanța pe diagonală între doi pixeli vecini ar fi $\sqrt{1^2 + 1^2} = \sqrt{2}$. În

realitate această distanță este $\sqrt{1 + \left(\frac{4}{3}\right)^2} = \sqrt{\frac{25}{9}} = \frac{5}{3}$. Mai mult, la un standard TV nu toate liniile

din semnal sunt afișate. Astfel la standardul CCIR numai 576 din cele 625 de linii sunt afișate efectiv în timp ce restul sunt folosite pentru altceva. Dacă se lucrează cu o imagine numerică de dimensiune 512x512 atunci se va introduce o nouă distorsionare. Noul raport între dimensiuni va fi

$$\frac{4}{3} \cdot \frac{512}{576} = \frac{4 \cdot 192}{512} = \frac{768}{512} = \frac{3}{2}$$
 ceea ce conduce la următoarele distanțe între doi pixeli vecini: 1 pentru

verticală, $3/2$ pe orizontală și $\frac{\sqrt{13}}{2}$ pe diagonală.

4. Detectia contururilor

4.1 Introducere

Extragerea contururilor este o etapă necesară pentru a realiza caracterizarea regiunilor din imagine în vederea identificării obiectelor din lumea reală. În plus această operație facilitează o reducere importantă a volumului de informație de la sute de mii sau milioane de pixeli la mii de puncte de contur.

O etapă preliminară destinată creșterii calității imaginii este pretratarea ei. Se poate interveni la nivelul convertorului analog numeric pentru o mai bună folosire a dinamicii camerei de luat vederi, (se controlează limitele domeniul convertit astfel încât acesta să se suprapună pe cât posibil peste domeniul real al nivelelor de gri din imagine) și se poate controla LUT-ul (Look Up Table) de intrare (pentru a se adapta sistemul de achiziție la modul de variație al nivelelor de gri). Trebuie subliniat faptul că nu întotdeauna aprecierea calității imaginii poate fi corect făcută cu ochiul liber. Cu alte cuvinte, nu imaginea pe care sistemul de vedere uman o apreciază drept cea mai bună va conduce cu necesitate la cele mai bune contururi. Intervin aici particularitățile sistemului de vedere artificială. Histogramele imaginilor sunt un instrument util în controlul efectului pe care diferitele operații de pretratare le au asupra imaginilor numerice.

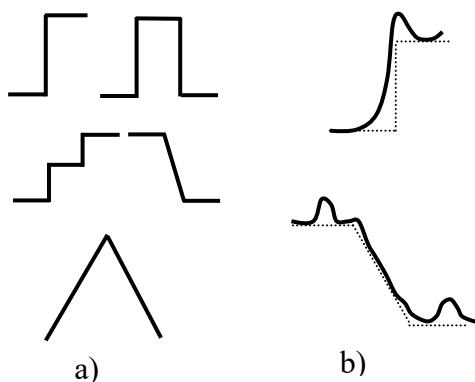


Figura 4.1 Tipuri de variație pentru muchii

Vom privi contururile ca pe granițele regiunilor care alcătuiesc imaginile obiectelor. Segmentarea este operația prin care se pun în evidență regiunile din imagine. Dacă suprafețele din scenă ar primi lumina numai direct de la sursa de lumină, atunci muchiile ar corespunde la variații ale intensității luminoase de tip treaptă, impuls, scară, rampă, sau creastă. Dacă se ia în considerare faptul că pe de o parte, suprafețele reflectă lumină unele spre altele (reflexii), iar pe de altă parte această reflexie se poate face într-o direcție unică sau în toate direcțiile, atunci pentru muchii pot apărea profile de variație de tipuri mult diferite de cele ideale (figura 4.1).

Studiind metodele de segmentare prezentate în literatura de specialitate le putem împărți în două mari categorii:

- **Metode care pun în evidență regiunile ca agregări de pixeli cu nivele de gri sau culori apropriate. Contururile se vor obține, eventual, într-un al doilea pas ca frontiere ale regiunilor.**

Există abordări care folosesc histogramele imaginilor pentru a obține informații sintetice asupra repartiției nivelelor de gri, și aplică detecția cu praguri pe aceste histograme pentru a pune în evidență regiunile din imagine. Este clasic exemplul histogramei bimodale la care prin aplicarea unui singur prag se izolează obiectul de fondul imaginii. În experimentele practice s-a evidențiat faptul că una dintre cele mai dificile probleme este tocmai alegerea pragurilor implicate în diferite procese de prelucrare a imaginilor. În cazul de mai sus, influența alegerii pragului se traduce în distribuția pixelilor albi și negri din imagine, și deci implicit în mărimea (aria) regiunii reprezentate de obiect în imagine. Dacă se dorește numai localizarea aproximativă în spațiul de operare (de exemplu poziția centrului de greutate al unui obiect), sau numai determinarea tipului formei plane din imagine, atunci metoda este deosebit de directă și în consecință se poate aplica cu succes în aplicații practice. Influența alegerii pragului pe histogramă asupra poziției centrului de greutate este mai mică decât influența asupra lungimii conturului și mult mai mică decât influența asupra ariei. Când este vorba de obiecte poliedrale, metoda devine restrictivă și își păstrează aplicabilitatea doar dacă se folosesc metode bazate pe conturul absolut exterior al obiectului, și nu se iau în considerare toate muchiile obiectului, inclusiv cele interioare, acelea care nu-l separă de fondul imaginii.

Alte abordări de acest tip fie pleacă de la o regiune maximală din imaginea pe care o divizează succesiv, fie pleacă de la regiuni minime cărora le cresc dimensiunea prin agregare. În ambele abordări, respectiv ascendentă și descendentă, criteriul folosit este nivelul de gri, iar în final se ajunge la regiunile reale din imagine.

- **Metode care pun în evidență inițial contururile. Regiunile vor fi atunci zonele din imagine delimitate de contururi.**

Metoda generală care s-a impus presupune trei pași succesivi:

- Se filtrează trece - jos imaginea și se elimină zgomotele care sunt asociate frecvențelor înalte. Operația este în principiu o mediere, care introduce și o pierdere a detaliilor și o estompare a contururilor, care sunt și ele asociate tot frecvențelor înalte.
- Se filtrează trece - sus imaginea. Operația este în principiu o derivarea (derivata este mare acolo unde variația este mare).
- Se aplică o detecție cu prag pe imaginea derivată pentru a pune în evidență punctele cu derivată mare, care corespund la o variație importantă a nivelului de gri și care sunt declarate puncte de contur.

Problema alegerii pragurilor revine: praguri prea mici conduc la contururi prea groase care necesită o etapă suplimentară de subțiere a contururilor, iar praguri prea înalte conduc la contururi eventual întrerupte care necesită o etapă suplimentară de închidere a contururilor. Sunt interesante metodele de segmentare simple și deci rapide, care să se preteze la aplicații practice, industriale. Aceste aplicații, pe de o parte cer un răspuns în timp scurt, iar pe de altă parte oferă condiții relativ stabile și controlabile de lucru.

4.2 Histogramele imaginilor și prelucrarea lor

4.2.1 Conceptul general de histogramă

Fie o imagine digitizată pe 2^m nivele de gri, și care are $n = N \times N = N^2$ pixeli. Procesarea de imagini este des folosită pentru a crește calitatea imaginii în sensul de-a fi mai ușor înțeleasă de observatorul uman, sau pentru a permite extragerea unor caracteristici importante de către un sistem de vedere artificială. În acest sens este frecvent necesar să se identifice un grup de pixeli legați între ei printr-o caracteristică comună. Informații utile pot fi extrase și dintr-o reprezentare făcută dintr-un punct de vedere mai general, cum este repartitia nivelelor de gri care conduce la histograma acestor nivele. Se măsoară și se reprezintă, pe ordonată, frecvența de apariție a fiecărui nivel de gri (figura 4.2). Prin generalitatea reprezentării se pierd relațiile geometrice dintre pixeli, iar corespondența histogramă-imagine nu este reciprocă.

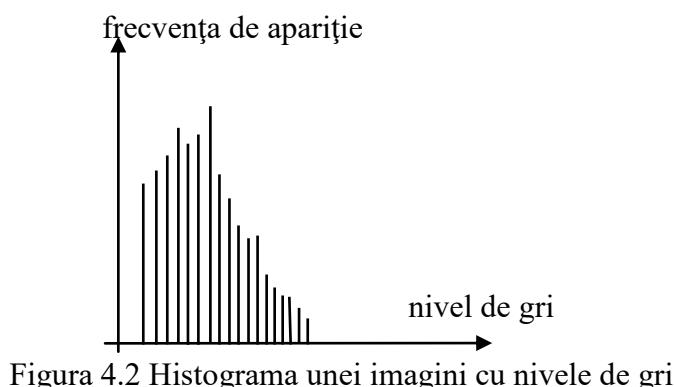


Figura 4.2 Histograma unei imagini cu nivele de gri

Totuși această reprezentare are cel puțin două avantaje:

- conține informații despre natura fundamentală a imaginii, și în anumite privințe despre calitatea ei.
- constituie baza de definire a unor prelucrări pentru transformarea imaginii. Una dintre informațiile utile conținute în histogramă este aceea despre contrastul imaginii (și deci despre contururi).

Dacă histograma este plată, relativ uniformă (figura 4.3 a), putem trage concluzia că s-a realizat o bună utilizare a nivelelor de gri disponibile, și deci este de așteptat o bună observabilitate a detaliilor din imagine. O histogramă cu un maxim într-o extremitate a intervalului de nivele de gri și o lipsă a nivelelor de gri dintr-o extremitate a intervalului (figura 4.3 b) va semnala un contrast slab

și deci în consecință o pierdere a detaliilor din imagine în zonele predominant albe sau negre din imagine. O similară lipsă de contrast poate fi semnalată și în cazul unei distribuții relativ uniforme pe întreaga scară de nivele de gri, dar cu un vârf proeminent într-o extremitate (figura 4.4 a).

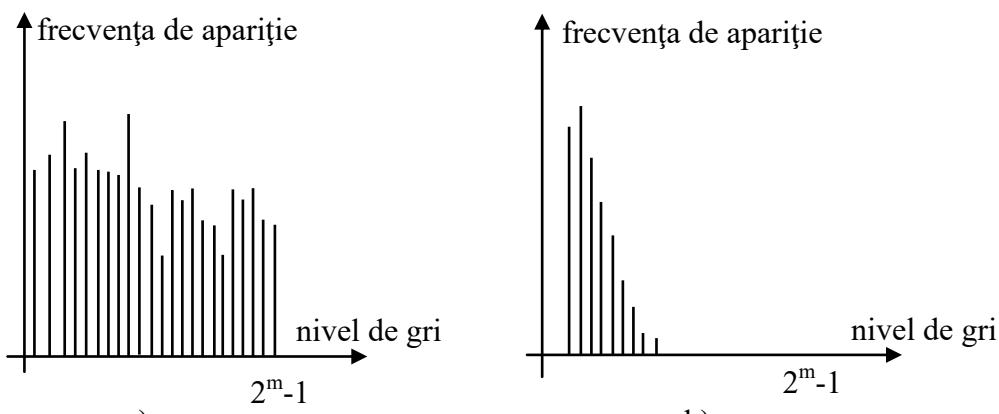


Figura 4.3 Exemple de histograme a) histogramă plată b) histogramă unei imagini care nu folosește toată gama de nivele de gri

O cale de control a calității imaginii ar fi reachiziționarea ei numerică după modificarea anumitor parametrii pe lanțul de achiziție, în sensul creșterii calității imaginii, și recalcularea histogramei pentru controlul efectului obținut. Chiar și în situațiile în care nu interesează foarte mult calitatea reprezentării detaliilor, histograma poate fi un instrument util. De exemplu, o histogramă bimodală (figura 4.4 b), semnalează prezența unei zone (obiect) de culoare deschisă și a unei zone (fond) de culoare închisă. Această situație este des întâlnită în aplicațiile practice, de exemplu prezența unui reper pe o bandă transportatoare. În astfel de cazuri poate fi utilă reprezentarea imaginii pe un singur bit (binarizarea imaginii), doar două nivele de gri: alb și negru. Trebuie găsit pragul care să delimitizeze cele două nivele de gri pentru a programa corect convertorul analog – numeric implicat în achiziție. Histograma oferă un ajutor important în această alegere. Problema nu este la fel de simplă dacă în aplicația respectivă contează în detaliu forma și mărimea obiectelor. În acest caz alegerea pragului cere o precizie mult mai mare.

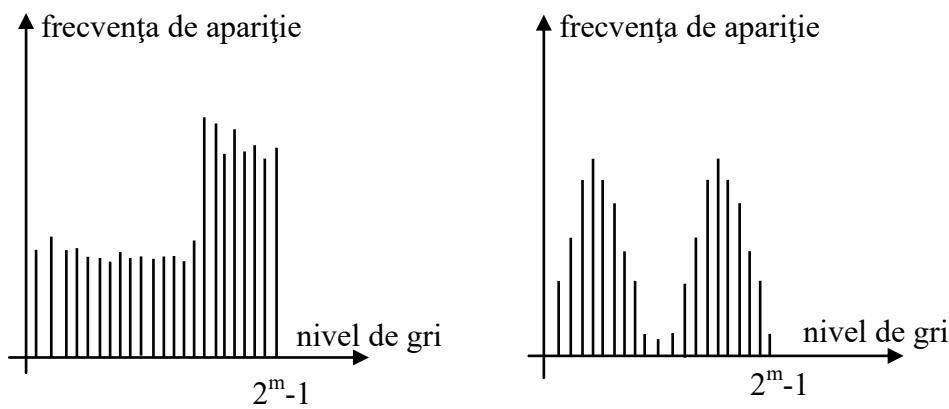


Figura 4.4 Exemple de histograme a) histogramă cu un maxim într-o extremitate a intervalului de nivele de gri b) histogramă bimodală

4.2.2 Tehnici de modificare a histogramei

Fie $I_n = f(x,y)$ o imagine numerică, $h(g)$ histograma ei, g nivelele de gri. Fie Φ o transformare a nivelelor de gri, $g \xrightarrow{\Phi} g'$, care conduce la $f'(x,y)$ noua imagine, $h'(g')$ noua histogramă, g' noile nivele de gri.

- Translatarea liniară a nivelelor de gri. Să luăm exemplul foarte simplu al următoarei histograme bimodale. Din cele prezentate anterior înțelegem că este cazul unui obiect de culoare deschisă pe un fond întunecat. Dacă această imagine trebuie afișată pe un display care are o neliniaritate pentru nivelele de gri în plaja (g_1, g_2) și dacă interesează numai forma și mărimea obiectului, nu și culoarea lui, atunci o simplă translatare a nivelelor de gri aduce vârful de pe histograma inițială într-o zonă în care displayul funcționează corect (figura 4.5). Atunci Φ , funcția de transformare, poate fi reprezentată ca o translație cu θ .

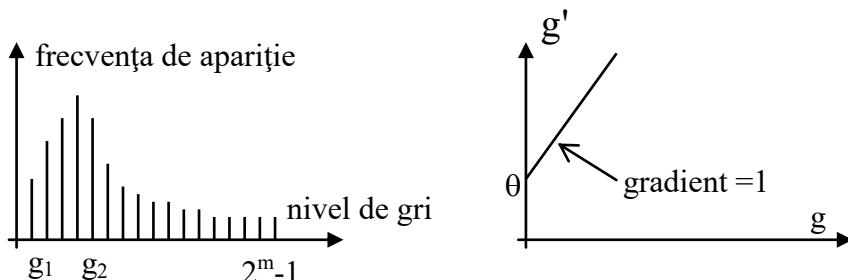


Figura 4.5 Translatarea nivelelor de gri

- Punerea la scară. După o analiză a histogramei, nivelele de gri sunt configurate pentru a utiliza întreaga dinamică a semnalului de ieșire al senzorului. Pentru a folosi ideal gama dinamică a senzorului, este necesar ca intervalul de conversie [LOW, HIGH] să fie identic cu intervalul nivelelor de gri din imagine [Gmin, Gmax]. Aceste valori se pot determina ușor din histograma imaginii. Dacă imaginea nu este statică, este practic imposibil ca această identitate să se păstreze prin reglarea dinamică a pragurilor convertorului analog – numeric.

- Aplatizarea histogramei (egalizarea). Teoretic imaginea cu cel mai bun contrast este aceea care utilizează egal toate nivelele de gri. Într-o astfel de imagine, funcția de repartitie, adică totalul pixelilor mai întunecați decât nivelul de gri curent (cu valori asociate mai mici decât aceea curentă), este o dreaptă. După egalizarea histogramei, vor fi realocate nivelele de gri astfel încât funcția de repartitie să tindă, pe cât posibil adică atât cât permite conținutul imaginii reale, la o dreaptă. Un exemplu de funcție care implică o astfel de schimbare pe histograma imaginii este următoarea:

$$g_k' = \frac{2^m - 1}{N^2} \sum_{j=0}^k h(g_j)$$

Pe histogramă, efectul este unul de aplatizare (figura 4.6), cu valorile asociate nivelelor de gri mai uniform distribuite (în special dacă se vizualizează anvelopa). Pe imagine încep să apară și detaliile dintr-o imagine cu contrast inițial mai slab. Se întărește contrastul, se îmbunătățește dinamica imaginii, prin folosirea tuturor nivelelor de gri. În același timp imaginea poate căpăta un aspect dur, iar zgomotele sunt și ele amplificate. Funcția descrisă este cunoscută și ca "Funcția cumulativă de distribuție" (histograma este o funcție densitate de probabilitate). În aplicațiile practice, datorită discretizării, rezultatele nu sunt la fel de bune ca în cazul ideal.

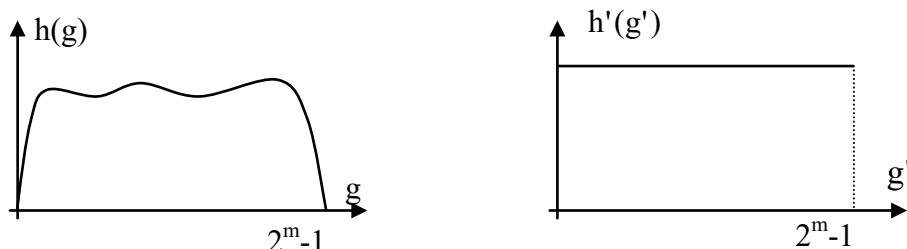


Figura 4.6 Egalizarea histogramei

Transformările de mai sus nu sunt singurele moduri de a transforma histograma. Egalizarea histogramei poate avea inconvenientul că pentru imagini cu un fond omogen de mare suprafață, această corecție duce la un contrast prea mare pentru o zonă de gri limitată, ceea ce face vizibil zgometul electronic. Reglajul după contrast aduce un remediu pentru că se face o corecție combinată între o aducere la scară și egalizarea histogramei. Alternativ pot fi făcute transformări exponențiale sau hiperbolice, spre exemplu. Se mai procedează curent la sumare sau scădere de imagini. Histograma sumei este egală cu produsul de convoluție al histogramelor imaginilor inițiale. Se mai pot aplica transformări punctuale asupra pixelilor individuali aşa cum este cazul LUT-ului.

De exemplu, pentru egalizarea de histogramă (figura 4.7) la o imagine de $n = 64 \times 64 = 4096$ pixeli, $2^3 = 8$ nivele de gri, datele sunt prezentate în tabelul 4.1, cu titlu de exemplificare.

r_k	n_k	$p_1(r_k)$	s_k	s_k'	s_k''	n_k'	$p_2(s_k'')$
$r_0 = 0$	790	0,19	0,19	1/7	1/7	790	0,19
$r_1 = 1/7$	1023	0,25	0,44	3/7	3/7	1023	0,25
$r_2 = 2/7$	850	0,21	0,65	5/7	5/7	850	0,21
$r_3 = 3/7$	656	0,16	0,81	6/7	6/7	985	0,24
$r_4 = 4/7$	329	0,08	0,89	6/7			
$r_5 = 5/7$	245	0,06	0,95	1	1	448	0,11
$r_6 = 6/7$	122	0,03	0,98	1			
$r_7 = 1$	81	0,02	1	1			

Tabelul 4.1 Exemplu de aplatizare a unei histogramme

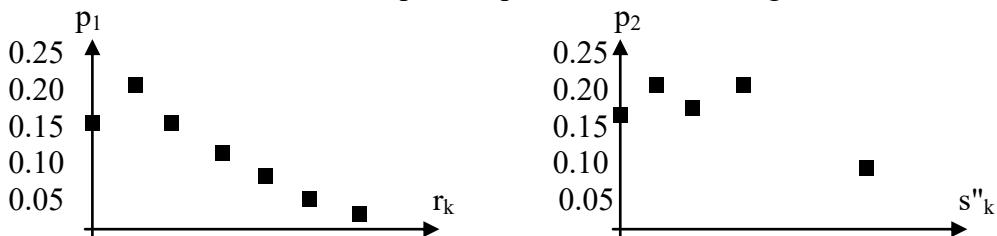


Figura 4.7 Rezultatul aplatizării

r_k - nivele de gri; n_k - număr de pixeli; $p_1(r_k)$ - densitate de probabilitate; s_k - funcția cumulată s_k' - cele mai apropiate nivele de gri; s_k'' - eliminare valori dublate; n_k' - redistribuția pixelilor

$p_2(s_k'')$ - noua densitate de probabilitate; $p_1(r_k) = \frac{n_k}{n}$; $s_k = \sum_{i=0}^k p_1(r_i)$

Este util ca transformările asupra histogramelor să se poată face hard, prin componente specializate, deci mai rapid. Mecanismul pe care îl oferă în acest sens plăcile de achiziție și prelucrare de imagini (frame-grabber) este cunoscut sub numele de LUT (Look Up Table). LUT-ul este o zonă de memorie adresată de valoarea numerică corespunzătoare nivelului de gri oferit, pentru fiecare pixel, de convertorul analog-numeric Valoarea de la respectiva locație din LUT, este noul nivel de gri care se asociază pixelului și care va fi memorată în planul video, deci în memoria imagine. Are loc astfel o transformare pixel cu pixel a histogramei.

Dacă histograma normalizată este $H(g) = \frac{h(g)}{n}$ unde $n = N \times M$ reprezintă numărul total de pixeli

din imagine, $h(g)$ este histograma (o funcție care reprezintă numărul de pixeli pentru fiecare nivel de gri g), $G = 2^m$ este numărul total de nivele de gri, atunci se pot defini mărimi sintetice care caracterizează o histogramă și deci au legătură cu calitatea sau conținutul imaginii, cum sunt, de exemplu, următoarele:

- media $\bar{g} = \sum_{g=0}^{G-1} g H(g)$

- varianță $\sigma_g = \sqrt{\sum_{g=0}^{G-1} (g - \bar{g})^2 H(g)}$

- energia $g_N = \sum_{g=0}^{G-1} (H(g))^2$

Curs 8

4.3 Filtrarea imaginilor numerice

4.3.1 Filtrarea în frecvență

Vom prezenta principal filtre liniare cu defazaj nul, și care atenuează în măsuri egale părțile reală respectiv imaginată ale componentelor semnalului.

Filtrarea se scrie:

- $g(x,y) = f(x,y) \otimes h(x,y)$ pentru planul imaginii (poziție – nivel de gri) ca o conoluție cu răspunsul h în impuls al filtrului.
- $G(u,v) = F(u,v) \cdot H(u,v)$ pentru planul frecvențelor ca o înmulțire cu funcția de transfer $H(u,v)$ a filtrului.

Pentru că aceste două operații sunt absolut echivalente, este suficient să se găsească funcțiile de transfer corespunzătoare.

Trebuie remarcat că o astfel de filtrare presupune trei pași succesivi:

- transformarea 2D $f(x,y) \rightarrow F(u,v)$,
- filtrare în planul frecvențelor $G(u,v) = F(u,v) \cdot H(u,v)$
- respectiv transformarea 2D inversă pentru regăsirea imaginii filtrate $G(u,v) \rightarrow g(x,y)$.

Fără un echipament (hardware) specializat, timpul de calcul pentru o astfel de filtrare este mare.

Filtrare trece - jos (FTJ)

Ca rezultat al unei astfel de filtrări imaginea va deveni mai estompată (flou), în sensul că se estompează contururile, care corespund la frecvențele înalte din imagine.

- **Filtrul ideal:** $H(u,v) = \begin{cases} 1 & \text{pentru } D(u,v) \leq D_0 \\ 0 & \text{pentru } D(u,v) > D_0 \end{cases}$

Filtrul prezintă simetrie radială în jurul originii. $D(u,v) = \sqrt{u^2 + v^2}$ este distanța de la punctul (u,v) la origine, iar D_0 este pragul de tăiere. Frecvențele atenuate corespund zonei exterioare unui cerc de raza D_0 . Numai frecvențele care corespund la puncte din interiorul cercului trec neatenuate. Într-un plan, pentru $v = 0$, distanța devine $D_0 = D(u_0,0)$, unde u_0 este frecvența de tăiere, mărime utilă pentru aprecierea caracteristicilor filtrelor în raport cu filtrul ideal (figura 4.8).

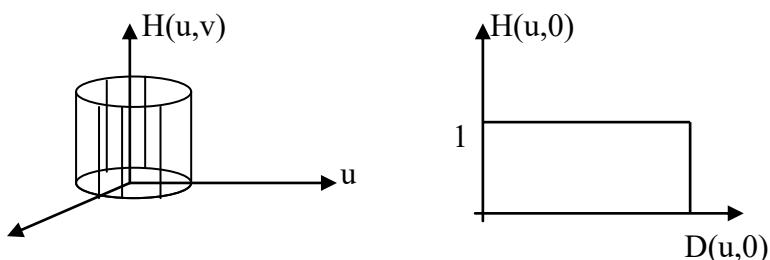


Figura 4.8 Filtru trece - jos

În planul frecvențelor, energia este aditivă, și deci se poate vorbi de zone de energie, zone circulare în care energia conținută este proporțională cu raza. Energia imaginii se concentrează spre centrul planului frecvențelor: $R_1 = 10\%R \Rightarrow E_1 = 90\%E$; $R_2 = 53\%R \Rightarrow E_2 = 99,5\%E$. Deci eficacitatea filtrelor trece-jos este mare, pentru că este suficient să înlăturăm 10% din energia totală a unei imagini pentru ca marginile obiectelor legate de frecvențele înalte, să dispară și să lase locul unei imagini estompeate (flou).

- Filtrul Butterworth:

$$H(u, v) = \frac{1}{1 + \left(\frac{D(u, v)}{D_0}\right)^{2n}} \quad \text{sau} \quad H(u, v) = \frac{1}{1 + (\sqrt{2} - 1) \cdot \left[\frac{D(u, v)}{D_0}\right]^{2n}}$$

Alegerea lui n schimbă ordinul filtrului, în mod curent $n = 1,2$. Acest filtru este un filtru trece - jos fără discontinuități brutale între frecvențele atenuate și cele neatenuate. Efectul de estompare este mai redus ca la filtrul ideal, și nu introduce efect de dublare a contururilor (ringing) pe imagine.

- Filtrul exponențial: $H(u, v) = e^{\left[\frac{-D(u, v)}{D_0}\right]^n}$, unde n este ordinul filtrului.

Prin acest tip de filtrare se introduc efecte mici de estompare și de dublare a contururilor.

- Filtrul trapezoidal:

$$H(u, v) = \begin{cases} 1 & \text{pentru } D \leq D_0 \\ \frac{D(u, v) - D_1}{D_0 - D_1} & \text{pentru } D_0 < D \leq D_1 \\ 0 & \text{pentru } D > D_1 \end{cases}$$

În general D_0 corespunde la 90% din energie. D_1 este de 2 până la 5 ori mai mare decât D_0 . Acest filtru introduce un efect de dublare a contururilor, dar mai puțin pronunțat decât la cazul ideal.

Filtrare trece - sus (FTS)

Prin aplicarea filtrării trece - sus pe imagini se produce o întărire, accentuare, a detaliilor și contururilor regiunilor dintr-o imagine inițial ștearsă. Această filtrare nu convine eliminării zgomotelor compuse din impulsuri abrupte și aleatoare.

- Filtrul ideal:

$$H(u, v) = \begin{cases} 0 & \text{pentru } D \leq D_0 \\ 1 & \text{pentru } D > D_0 \end{cases}$$

- Filtrul Butterworth:

$$H(u, v) = \frac{1}{1 + \left(\frac{D_0}{D}\right)^{2n}} \quad \text{sau} \quad H(u, v) = \frac{1}{1 + (\sqrt{2} - 1) \cdot \left(\frac{D_0}{D}\right)^{2n}}$$

La $D = D_0$, $H(u, v)$ ia valoarea $\frac{1}{\sqrt{2}}$ din valoarea maximă.

- Filtrul exponențial:

$$H(u, v) = e^{\left(\frac{-D_0}{D}\right)^n} \quad \text{sau} \quad H(u, v) = e^{\left(\ln \frac{1}{\sqrt{2}}\right) \cdot \left(\frac{D_0}{D}\right)^n}$$

La $D = D_0$ $H(u, v)$ ia valoarea $\frac{1}{\sqrt{2}}$ din valoarea maximă.

- Filtrul trapezoidal:

$$H(u, v) = \begin{cases} 0 & \text{pentru } D \leq D_1 \\ \frac{D(u, v) - D_1}{D_0 - D_1} & \text{pentru } D_1 < D \leq D_0 \\ 1 & \text{pentru } D > D_0 \end{cases}$$

Prin filtrări trece-sus se obțin accentuări ale contururilor, dar se poate ajunge și la exagerări datorate atenuării frecvențelor joase (imaginea devine dură). Pentru ameliorare, o tehnică ar fi adunarea unei constante, aproape 1, la funcția de transfer. Apare atunci o amplificare prea mare a frecvențelor înalte. O tehnică alternativă presupune o redistribuire a nivelelor de gri înainte de filtrare adică egalizarea histogramei.

4.3.2 Filtrarea în domeniul spațial

Acest tip de filtrare se referă la operații directe pe aggregate de pixeli și la atributul principal: nivelul de gri. Scopul urmărit este atenuarea zgomotului și înlăturarea defectelor și perturbațiilor conținute în imaginea inițială. Se ajunge astfel la operații de netezire analoage cu filtrările trece - jos.

- Medierea vecinilor

Această tehnică este cea mai directă pentru a restabili o imagine denaturată în contururile sale, distorsionată de un zgomot de înaltă frecvență. Nivelul de gri în fiecare pixel se înlocuiește cu media nivelelor de gri pentru cei n-vecini (4 - vecini sau 8 - vecini, cel mai frecvent, figura 4.9).

Deci relația folosită este: $g'(i, j) = \frac{1}{n} \sum_{(i,j) \in E} g(i, j)$, unde E este ansamblul n - vecinilor

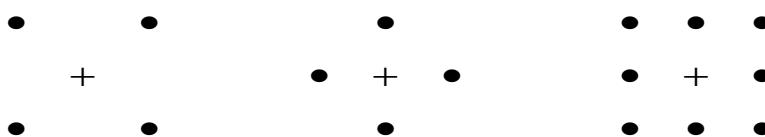


Figura 4.9 Vecinătăți pentru mediere

- Filtrul median

Tehnica medierii vecinilor are drept defect principal crearea efectului de estompare (flou) la marginile obiectelor sau în punctele cu mare gradient de intensitate. Acest efect secundar este puternic diminuat prin tehnică filtrări mediene, care înlocuiește nivelul de gri al fiecărui pixel printr-o combinație a nivelelor de gri ale pixelilor vecini. Considerând o fereastră din imagine conținând $N = 2p+1$ pixeli, N impar, mai întâi se ordonează valorile $\{g_i\}$ ale nivelelor de gri astfel încât la sfârșit să se obțină $g_1 \leq g_2 \leq \dots \leq g_N$. Apoi printr-o operație liniară asupra acestei secvențe ordonate, se efectuează următoarea sumă: $g' = \sum_{i=1}^N c_i g_i$ care este noua valoarea care va fi afectată pixelului central, unde c_i sunt coeficienți constanti de ponderare.

Filtrul median standard se obține dacă se alege $c_{p+1} = 1$, iar toți ceilalți coeficienți sunt nuli. Se poate înțelege intuitiv că filtrul median forțează pixelii de intensitate luminoasă ieșită din comun, în raport cu vecinătatea lor, să fie netezite, adică să ia valori apropriate de ale vecinilor. Acest filtru este un filtru trece - jos care elimină zgomotul de tip ninsoare, dar care antrenează o ușoară pierdere de rezoluție.

- Medierea pe imagini

Tehnica medierii pe imagini constă în a considera ansamblul pixelilor dintr-o imagine și nu pe fiecare în parte, separat. Trebuie să dispunem deci de mai multe imagini, luate riguros din același punct de vedere, ale unei aceleiași scene statice.

Sistemul opto - electric de preluare a imaginii poate fi același sau altul pentru fiecare imagine în parte, ceea ce face ca și zgomotul să fie specific fiecărei imagini în parte. În aceste condiții, dacă notăm $m_i(x,y)$ fiecare imagine și $b_i(x,y)$ zgomotul care îi este atașat, se poate considera că semnalul imaginii nealterate $f(x,y)$ satisfacă relația:

$$m_i(x,y) = f(x,y) + b_i(x,y), \text{ pentru } i=1, 2, 3.., n.$$

Pentru a atenua zgomotul și pentru a extrage din aggregatele de informație valoarea cea mai apropiată de $f(x,y)$ se face adunarea după i , până la n :

$$\sum_{i=1}^n m_i(x,y) = nf(x,y) + \sum_{i=1}^n b_i(x,y).$$

Ipoteza că pentru sistemele opto - electronice zgomotul nu este de loc corelat (b_i nu sunt corelate între ele) este foarte apropiată de realitate și în plus zgomotele nu au nici o șansă să se repete ciclic, deci media fiecărui zgomot b_i este nulă. Deci putem scrie:

$$\bar{b}_i = 0; \quad \bar{b}_i b_j = 0;$$

$$\overline{m(x,y)} = \frac{1}{n} \sum_{j=1}^n (f(x,y) + b_i(x,y))$$

$$\overline{m(x,y)} = f(x,y) + \sum_{i=1}^n b_i(x,y) = f(x,y) + b$$

unde b este zgomotul final după mediere. Pentru o imagine obținută prin mediere abaterea medie se calculează ca mai jos ceea ce duce la concluzia că prin medierea pe n imagini se obține o îmbunătățire a raportului semnal/zgomot de \sqrt{n} ori. Luând un număr mare de imagini și mediindu-le, se poate obține o imagine cât mai apropiată de imaginea ideală, fără zgomot. În practică, numărul de sumări nu poate fi crescut foarte mult pentru că datorită unei erori inevitabile de suprapunere a imaginilor rezultatele pot fi compromise.

$$\begin{aligned} \sigma^2 &= \bar{b}_i^2 - \bar{b}_i^2 = \bar{b}_i^2 \\ \sigma^2 &= \bar{b}^2 - \bar{b}^2 = \left(\frac{1}{n} \sum_{j=1}^n b_j \right)^2 \\ &= \frac{1}{n^2} \left(\sum_{i=1}^n b_i^2 + \sum_{i \neq j} b_i b_j \right) = \frac{1}{n^2} \left(\sum_{i=1}^n b_i^2 \right) + \frac{1}{n^2} \left(\sum_{i \neq j} b_i b_j \right) = \\ &= \frac{1}{n^2} \left(\sum_{i=1}^n b_i^2 \right) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n} \Rightarrow \sigma_{\text{med}} = \frac{\sigma}{\sqrt{n}} \end{aligned}$$

- Extragerea fondului

Scăderea imaginilor numerice se referă la o operație care efectuează diferența între valorile numerice, nivelele de gri, asociate punctelor corespunzătoare ca poziție din cele două imagini; deci imaginile se scad punct cu punct. Dacă spre exemplu se consideră ca referință imaginea fără nici un obiect în scenă, atunci prin scăderea dintr-o altă imagine care conține un obiect, acesta din urmă poate fi localizat cu precizie. Decalibrarea fotometrică este una dintre cele mai importante aplicații pentru extragerea fondului (background subtraction). În anumite circumstanțe camerele pot oferi răspunsuri neliniare în funcție de iluminarea câmpului vizual ceea ce poate duce într-o oarecare măsură la o strălucire mai accentuată a centrului imaginii în raport cu zonele periferice. Acest fenomen este foarte dezavantajos pentru unele prelucrări asupra imaginilor cum ar fi, spre exemplu, detecția cu prag pe histogramă. O soluție ar fi înlocuirea lentilelor. Altfel se poate face o modelare a neliniarității răspunsului lentilei luând o imagine a unei suprafețe umbrite uniform, detectarea minimului pentru nivelele de gri din această imagine, extragerea acestei valori din nivelul de gri al fiecărui punct și obținerea astfel a imaginii de referință a fondului. Aceasta imagine va fi scăzută succesiv din toate imaginile obținute prin achiziție și va contribui la eliminare efectului nedorit amintit anterior.

4.3.3 Utilizarea măștilor de convoluție

Pentru realizarea rapidă a operațiilor de mediere și derivare a imaginilor, o abordare foarte folosită este utilizarea măștilor de convoluție.

Înloc de convoluția sub formă continuă:

$$g(x) = f(x) \circ h(x) = \int_{-\infty}^{\infty} f(\alpha) \cdot h(x - \alpha) \cdot d\alpha$$

putem scrie pentru un semnal discret 1D convoluția:

$$g_l = f_l \cdot h_l = \sum_{i=-\infty}^{+\infty} f_i \cdot h_{l-i}$$

Dacă semnalele sunt simetrice în raport cu originea $x = 0$ și finite de dimensiune $2m+1$, iar răspunsul impulsional al filtrului este finit și are $2k+1$ elemente atunci:

$$g_l = \sum_{i=\inf(m,l-k)}^{\sup(-m,l-k)} f_i \cdot h_{l-i},$$

mai mult dacă $m >> k$ atunci

$$g_l = \sum_{i=l-k}^{l+k} f_i \cdot h_{l-i}$$

Se observă că sub semnul Σ apar produse între componentele vectorilor:

$$f_i = [f_{l-k} \quad \dots \quad f_l \quad \dots \quad f_{l+k}] \text{ și } h_{l-i} = [h_k \quad \dots \quad h_o \quad \dots \quad h_{-k}]$$

Dacă considerăm acum un semnal $f(x,y)$, deci facem trecerea la 2D, care corespunde unei imagini presupuse pătrate și digitizate cu $N = 2m+1$ linii și tot atâtea coloane, atunci funcția $h(x,y)$ este o fereastră de dimensiunea $2k+1$. Folosind indicii i și i' pentru linii și coloane, produsul $g_{ii'}$ va fi o imagine pătrată de dimensiune $2(m-k)+1$ linii și coloane:

$$g_{ii'} = \sum_{i=n-k}^{l+k} \sum_{j=l'-k}^{l'+k} f_{ij} \cdot h_{l-i, l'-j}.$$

Pentru $k \ll m$ această formulă este adevarată peste tot, în afară de marginile imaginii. Calculele scad proporțional cu k . Se observă analogia între conoluția pentru filtrare și plimbarea unei ferestre pe imagine. Dacă avem $k=1$, atunci fereastra are dimensiunea de $3x3$, caz în care pentru imaginile uzuale, de exemplu de $512x512$ pixeli, se obțin performanțe, tempi de calcul, mai bune decât cele pentru filtrarea în domeniul frecvență pentru transformata Fourier 2D.

Anterior am presupus atât imaginea cât și filtrul ca fiind pătrate. Dacă luăm $k = 3 \ll m$ atunci filtrul apare ca o fereastră cu care se poate baleia, parcurge imaginea. Formula arată că pentru toți pixelii din imagine și 8 - vecinii lor se multiplică nivelelor lor gri cu valorile corespunzătoare filtrului (elementele măștii corespunzătoare). Produsele sunt adunate într-o sumă care este noua valoare a nivelului de gri asociat pixelului central. Apar limitări de calcul la linile și coloanele extreme. Mască poate fi asimilată, ca formă, cu o matrice $|M|$ cu care trebuie parcursă o porțiune $|I|$ din imagine, centrată pe pixelul curent (figura 4.10). Se definește această variantă de produs matriceal \otimes prin:

$$|I| \oplus |M| = \sum_{i=1}^3 \sum_{j=1}^3 g_{ij} \cdot m_{ij}$$

Dacă valorile g_{ij} și m_{ij} ale lui I , respectiv M sunt așezate vectorial în \vec{I} respectiv \vec{M} atunci:

$$P_s = \vec{I} \cdot \vec{M} = g_1 \cdot m_1 + g_2 \cdot m_2 + \dots + g_9 \cdot m_9.$$

În general ponderile din măști au valori întregi ceea ce conduce la necesitatea introducerii unui factor de scară global cu scopul de a păstra noile valori în aceeași plajă ca și valorile inițiale și deci să permită afișarea rezultatelor: $|I| \oplus |M| = f_s \cdot \sum_{i=1}^3 \sum_{j=1}^3 g_{ij} \cdot m'_{ij}$. Fără factorul de scară, am introduce în calcul nivele de gri între negru și alb, iar rezultatele ar rezulta și în afara acestui interval.

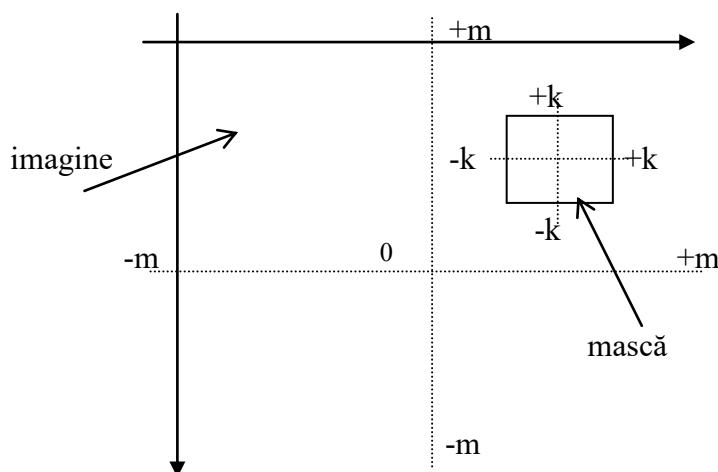


Figura 4.10 Parcurgerea imaginii cu o mască

Mediere

Lucrând pe imagini numerice, integrarea care este asociată cu filtrarea trece – jos este o mediere, sau mai simplu o sumare. Măștile asociate acestui tip de filtrare au numai ponderi pozitive. În cazul cel mai general, filtrarea trece - jos care trebuie să preceadă orice extragere de contur, are ca scop eliminarea zgomotului și deci metoda aplicată depinde de natura acestuia. Bazându-ne pe repetabilitatea și controlul parametrilor în aplicațiile industriale, putem aprecia că principalul efect așteptat aici de la această operație este o uniformizare a diferențelor zone din imagine și eliminarea variațiilor punctuale exagerate ale nivelor de gri. Pot fi utilizați operatori de mediere pe vecini, pentru 8-vecinătăți, de următorul tip:

$$M_1 = \frac{1}{16} \cdot \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad M_2 = \frac{1}{64} \cdot \begin{bmatrix} 7 & 7 & 7 \\ 7 & 8 & 7 \\ 7 & 7 & 7 \end{bmatrix} \quad M_3 = \frac{1}{5} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Se observă că efectul primei măști este mai slab decât efectul celei de a doua pentru că ponderea pixelului central este mai mare ($4/16 > 8/64$). Filtrarea trece jos are efect mai slab atunci când noua valoare de gri din pixelul central depinde mai mult de valoare de gri veche din poziția respectivă decât de valorile de gri ale vecinilor. Si reciproc, efectul este mai puternic atunci când noua valoare de gri din pixelul central depinde mai mult de valorile de gri ale vecinilor decât de valoarea de gri veche din poziția respectivă. La limită, cea mai slabă filtrare este aceea care alocă ponderi 0 vecinilor și deci pixelul central are pondere maximă, iar cea mai puternică filtrare este realizată când pixelul central are ponderea 0 în timp ce vecinii săi au pondere maximă.

Mai general, un exemplu sub formă parametrizată depinzând de factorul x:

$$M_4 = \frac{1}{(x+2)^2} \cdot \begin{bmatrix} 1 & x & 1 \\ x & x^2 & x \\ 1 & x & 1 \end{bmatrix}$$

Se caută și obținerea unor forme avantajoase din punct de vedere al aplicării într-un algoritm de calcul implementabil pe calculator (se împarte mai repede la 4 decât la 3). Factorul de scară a fost introdus pentru a păstra nivelele de gri în aceeași plajă cu cele din imaginea inițială.

Un alt filtru trece-jos utilizat face, de exemplu, o mediere în cruce următorul tip, la care este evident că pentru economie de calcul nu se va implementa în program chiar baleierea imaginii cu o mască de dimensiune 5x5 (nu are rost să mai inmulțim cu zero, este suficient să facem 9 adunări). Un operator similar de mediere în cruce se poate construi și sub formă de mască 3x3. Filtrarea trece jos se poate aplica repetat de un anumit număr de ori, astfel încât să se eliminate zgomotele nedorite, dar trebuie luat în calcul că în acest fel se elimină și detaliile din imagine.

$$M_5 = \frac{1}{9} \cdot \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Curs 9

4.3.3 Utilizarea măștilor de conoluție (continuare)

Derivare

Lucrând pe imagini numerice, filtrarea trece – sus care este asociată cu operația de derivare se realizează numai prin diferențe între nivelele de gri. Măștile asociate au și ponderi pozitive și ponderi negative. Derivarea imaginii rămâne în esență o operație de accentuare a detaliilor dintr-o imagine numerică.

Din punct de vedere teoretic există două modalități de a detecta contururile (în special cele de tip treaptă):

- Calculul gradientului în fiecare punct al imaginii.

$$\vec{G} = (G_x, G_y), \quad G_x = \frac{\partial I_r(x, y)}{\partial x}, \quad G_y = \frac{\partial I_r(x, y)}{\partial y}$$

În această variantă se pot obține informații despre direcția conturului (gradientul este perpendicular pe contur). Compararea valorilor normei gradientului cu un prag ales va selecta punctele de pe contur (figura 4.11). Măștile asociate au o simetrie în raport cu o anumită direcție, direcția muchiilor detectate.

- Calculul operatorului Laplace în fiecare punct al imaginii.

$$\nabla^2 I_r(x, y) = \frac{\partial^2 I_r(x, y)}{\partial x^2} + \frac{\partial^2 I_r(x, y)}{\partial y^2}$$

În această variantă nu se obțin informații legate de direcția conturului. Detectarea trecerilor prin zero ale acestui operator va permite selectarea punctelor de pe contur (figura 4.11). Măștile asociate au o simetrie în jurul centrului măștii și nu una direcțională. Ca dezavantaj se poate menționa sensibilitatea la zgomot a acestui operator.

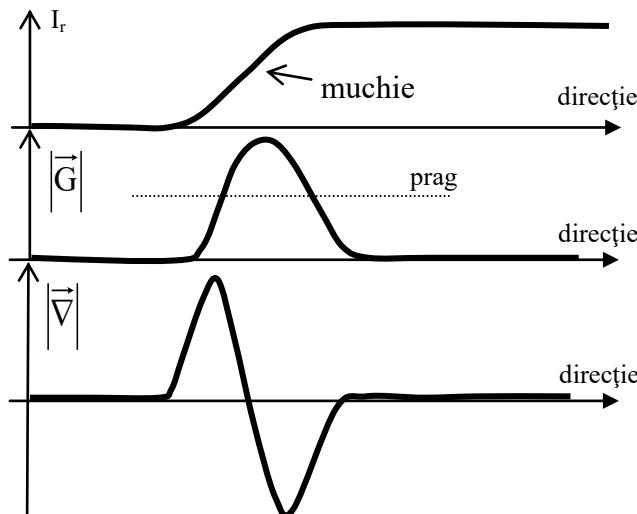


Figura 4.11 Variațiile intensității luminoase și derivatelor corespunzătoare unei muchii.

Considerând un semnal imagine $I_r(x,y)$ continuu și urmărind o coloană a sa cu $y = \text{constant}$, derivata parțială trăduce variația în intensitate luminoasă pe verticală:

$$\frac{\partial I_r}{\partial x} = \lim_{h \rightarrow 0} \frac{I_r(x,y) - I_r(x+h,y)}{h}.$$

Similar mergând pe linii, cu $x = \text{constant}$, se calculează și $\frac{\partial I_r}{\partial y}$ deci gradientul intensității luminoase în (x,y) este:

$$\vec{G} = \begin{vmatrix} \frac{\partial I_r}{\partial x} \\ \frac{\partial I_r}{\partial y} \end{vmatrix}$$

\vec{G} este orientat în direcția variației maximale ale lui I_r . Norma gradientului se scrie:

$$|\vec{G}| = \sqrt{\left(\frac{\partial I_r}{\partial x}\right)^2 + \left(\frac{\partial I_r}{\partial y}\right)^2}$$

și este valoarea maximă a variației $f(x,y)$ pe unitatea de distanță (amplitudinea gradientului).

Metoda permite calculul direcției gradientului: $\Theta = \arctg\left(\frac{G_y}{G_x}\right)$

Imaginea numerică I_n este alcătuită din pixeli cu nivele de gri $g(i,j)$, care corespund la intensități luminoase $I_n(x,y)$, despărțite de distanțe unitare și deci se poate scrie:

$$|\vec{G}| = \sqrt{[g(i,j) - g(i+1,j)]^2 + [g(i,j) - g(i,j+1)]^2}$$

sau

$$|\vec{G}| = |g(i,j) - g(i+1,j)| + |g(i,j) - g(i,j+1)|$$

Aceasta ultimă formulă convine unui calcul numeric rapid în binar.

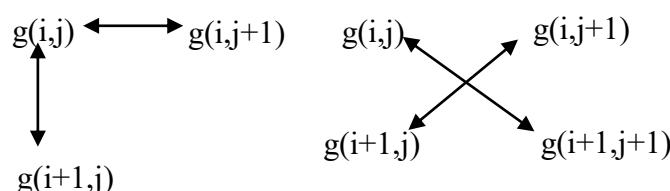


Figura 4.12 Modalități de calcul rapid al gradientului

Este clasică aproximarea Roberts care face apel la diferențele încrucișate (figura 4.12) și este utilă pentru că detectează mai bine conturul atunci când gradientul variază puțin. În această aproximare se scrie :

$$|\vec{G}| = |g(i,j) - g(i+1,j+1)| + |g(i+1,j) - g(i,j+1)|$$

Contururile într-o imagine se pot extrage comparând $|G|$ cu un prag T :

$$|\vec{G}| \geq T \Rightarrow \text{punct pe contur},$$

$$|\vec{G}| < T \Rightarrow \text{punct care nu se află pe contur}.$$

Conturul poate apărea ca o curbă cu grosime variabilă de mai mulți pixeli dacă T este ales prea mare. Devine util în multe aplicații să se aducă conturul la grosime de un pixel și ca urmare este necesară o operație suplimentară de subțiere a conturului. Dacă se alege un prag T prea mic, atunci este posibilă obținerea unui contur cu întreruperi și ca urmare este necesară o operație suplimentară de închidere a conturului (vezi figura 4.11).

Dacă continuăm derivarea la ordinul doi pentru a calcula expresia $\frac{\partial^2 I_r}{\partial x^2} + \frac{\partial^2 I_r}{\partial y^2}$ găsim, pentru cazul imaginii numerice I_n cu nivelele de gri $g(i,j)$, modulul operatorului Laplace:

$$\begin{aligned} |L| &= |g(i+1,j) - 2g(i,j) + g(i-1,j) + g(i,j+1) - 2g(i,j) + g(i,j-1)| = \\ &= |g(i+1,j) + g(i-1,j) + g(i,j+1) + g(i,j-1) - 4g(i,j)| \end{aligned}$$

Prin detectarea trecerilor prin zero ale valorii laplaceanului, se decide care dintre pixeli aparțin conturului (vezi figura 4.11). Acest operator amplifică zgomotul imaginii, ceea ce necesită prelucrări suplimentare, și este izotrop deci nu conține informații utile despre direcția conturului. Matriceal ultimii doi operatori se scriu ca măști astfel:

$$\text{Gradient: } \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{Laplacean: } \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

Ca soluție alternativă încă mai simplă, se pot deriva imaginile folosind următoarea formulă de tip gradient:

$$G = \max \{|g(i,j) - g(i+1,j)|, |g(i,j) - g(i,j+1)|, |g(i,j) - g(i+1,j+1)|, |g(i+1,j) - g(i,j+1)|\}$$

Avantajele obținute în acest caz sunt: volumul relativ redus de calcul, obținerea direcției conturului (codificat Freeman) dacă se iau în calcul și semnele diferențelor din formula de mai sus, relativa independență a pragului în raport cu tipul imaginii, ceea ce evită căutarea lui dinamică și obținerea unui contur în general subțire cu poziția determinată de poziția pixelului din dreapta sus dintr-o vecinătate de patru pixeli.

Gradintul se scrie sub formă de măști 3x3 sub diferite forme, dintre care se pot exemplifica:

$$\begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{sau} \quad \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{pentru detectie de muchii verticale,}$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \end{bmatrix} \quad \text{sau} \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \quad \text{pentru detecție de muchii orizontale.}$$

Laplaceanul poate fi scris sub formă de diferite măști, dintre care se pot exemplifica cu următoarele:

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} \quad \begin{bmatrix} -1 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & -1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Cu cât diferența dintre suma ponderilor pozitive și suma ponderilor negative este mai mare cu atât derivarea este mai puternică pentru că introduce o amplificare mai mare a diferențelor dintre nivelele de gri.

Pentru detecția de drepte (pixelii de pe dreapta au nivele de gri diferite de ale regiunilor pe care dreapta le separă, iar dreptele sunt subțiri) se pot da exemple de măști direcționale de următoarele tipuri:

pentru detecție de drepte orizontale: $\begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix}$

sau pentru detecția de drepte inclinate la 45° : $\begin{bmatrix} -1 & -1 & 2 \\ -1 & 2 & -1 \\ 2 & -1 & -1 \end{bmatrix}$

Foarte cunoscute și utilizate sunt măștile gradient de tip Sobel și Prewitt:

$$S_x = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad S_y = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad \text{cu} \quad G = \sqrt{S_x^2 + S_y^2}$$

$$R_x = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad R_y = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \quad \text{cu} \quad G = \sqrt{R_x^2 + R_y^2}$$

care oferă un raport performanță/complexitate foarte bun. Această metodă permite și calculul direcției Θ a pseudogradientului: $\Theta = \arctg(G_y / G_x)$

O formă particulară capătă gradientul în reprezentarea Kirsch. Pseudogradientul este dat aici de combinațiile:

$$|\vec{G}| = \max \left\{ 1, \max_{i=1}^7 (5S_i - 3T_i) \right\}$$

unde $S_i = m_i g_i + m_{i+1} g_{i+1} + m_{i+2} g_{i+2}$ și $T_i = m_{i+3} g_{i+3} + m_{i+4} g_{i+4} + m_{i+5} g_{i+5} + m_{i+6} g_{i+6} + m_{i+7} g_{i+7}$

Se consideră pixelii în sens orar în jurul pixelului curent, și se alege valoarea maximă găsită după calculul tuturor acestor posibile permutări circulare. De exemplu se pot alege: $m_1 = m_2 = m_3 = -1$, $m_4 = m_6 = m_7 = m_8 = m_9 = 1$, $m_5 = -2$ sau $m_1 = m_2 = m_3 = 5$, $m_4 = m_6 = m_7 = m_8 = m_9 = -3$, $m_5 = 0$.

Pentru detecția direcțională a muchiilor, după notația Freeman (figura 4.13), se pot utiliza măștile detectoare de contur de următorul tip:

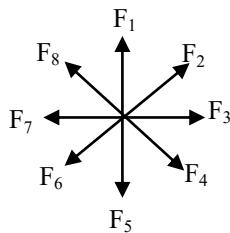


Figura 4.13 Direcțiile Freeman

$$\begin{aligned} F_1 &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{bmatrix} & F_2 &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -1 & -1 \end{bmatrix} & F_3 &= \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & -1 \end{bmatrix} \\ F_4 &= \begin{bmatrix} 1 & -1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & 1 \end{bmatrix} & F_5 &= \begin{bmatrix} -1 & -1 & -1 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix} & F_6 &= \begin{bmatrix} -1 & -1 & 1 \\ -1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \\ F_7 &= \begin{bmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{bmatrix} & F_8 &= \begin{bmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{bmatrix} \end{aligned}$$

Apartenența la contur se decide comparând valoarea $I'_n = \max_{i=1}^8 \{F_i \otimes I_n\}$ cu un prag prestabilit. Indicele care corespunde valorii maxime indică direcția muchiei pe care se plasează pixelul respectiv.

Pentru a valorifica similitudinea cu măștile atenuatoare de contur M_4 prezentate în paragraful anterior se pot defini măști amplificatoare de contur parametrizate de următorul tip. și din natura nedirecțională a măștii de mai jos reiese că ea este o variantă de operator Laplace.

$$M_6 = \frac{1}{(x-2)^2} \cdot \begin{bmatrix} 1 & -x & 1 \\ -x & x^2 & -x \\ 1 & -x & 1 \end{bmatrix}$$

Se pot folosi și măști detectoare de contur de dimensiune 5×5 de mai multe tipuri, ca de exemplu:

$$M_7 = \frac{1}{9} \cdot \begin{bmatrix} 0 & 2 & 0 & -2 & 0 \\ 2 & 1 & 0 & -1 & -2 \\ 2 & 1 & 0 & -1 & -2 \\ 2 & 1 & 0 & -1 & -2 \\ 0 & 2 & 0 & -2 & 0 \end{bmatrix}$$

dar se apreciază că în majoritatea cazurilor, plusul de precizie obținut nu compensează efortul de calcul suplimentar. Sistemele industriale de prelucrare a imaginii oferă astfel de operatori ajungând să se extragă contururile în cadență video (viteză mare).

Măștile în care ponderile au același semn presupun operații de adunare ceea ce corespunde la o mediere, o integrare a imaginii, deci la un filtru trece-jos care poate produce o eliminare a zgomotului din imagine. Măștile în care apar ponderi cu semne contrare presupun operații de scădere ceea ce corespunde la o derivare a imaginii, deci la un filtru trece-sus care poate pune în evidență frecvențele înalte corespunzătoare schimbărilor în nivelul de gri. Cu aceste măști se poate detecta apartenența unui pixel la muchie: prin comparație cu un prag – pentru derivata de ordinul 1 sau prin detecția trecerilor prin zero pentru derivata de ordinul 2.

Se dovedește utilă o filtrare trece-jos inițială urmată de o extragere a contururilor printr-o filtrare trece-sus și detecția cu prag corespunzătoare. Filtrarea trece-jos se poate realiza prin aplicarea succesivă repetată a diferitelor măști de acest tip. Alegerea numărului de aplicări successive se face, după experiență, în funcție de tipul imaginii și zgomotului, astfel încât să se îmbunătățească raportul semnal zgomot: o singură parcurgere poate fi insuficientă, dar foarte multe pot estompa complet detaliile din imagine. Filtrarea trece-jos de tip Gauss, deci convoluția cu funcție de tip Gauss, ar putea fi probabil cea mai des folosită. În aplicațiile industriale, din păcate utilizarea ei este limitată de viteza de calcul care scade datorită dimensiunii măștilor. Există însă plăci specializate care realizează această funcție. O abordare foarte utilizată este acea bazată pe filtrarea trece-jos cu un mediator de tip Gauss și pe un extractor de muchii de tip Laplace.

Funcția Gauss este definită în maniera cunoscută:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

unde σ determină, într-o relație directă lărgimea, întinderea funcției. În mască se trec valorile numerice corespunzătoare acestei funcții. Problema care se pune este cât de mare alegem masca, ca dimensiune. Atunci când x și y devin foarte mari, funcția tinde către zero dar nu se anulează. O alegere poate fi aceea în care se alege amplitudinea maximă pentru punctul central (de ex. 1000, prin normalizare, deci prin împărțire cu suma elementelor din mască) și se rețin valorile diferite de zero pentru a alcătui masca. Pentru $\sigma = 3$ (pixeli) se ajunge la o mască de dimensiune 23x23 pentru a reprezenta corect o funcție bidimensională Gauss. Alegând parametrul σ se alege efectul filtrării: pentru valori mici imaginea va păstra detaliile, iar pentru valori mari imaginea va pierde detaliile și va păstra numai structurile importante.

4.4 Urmărirea contururilor

Așa cum s-a afirmat la începutul capitolului, extragerea contururilor se poate face prin depistarea pixelilor în care au loc schimbări importante ale nivelerelor de gri prin aplicarea unui prag pe derivată. Această metodă poate conduce fie la contururi groase fie la contururi întrerupte. O soluție alternativă, care conduce la contururi subțiri și fără întreruperi, este parcurgerea contururilor. Prin urmărirea contururilor, se poate face și o extragere de contur pe de o parte dar, pe de altă parte se pot obține și informații legate de relațiile dintre regiunile pe care aceste contururi le separă pentru că le sunt frontiere. Din literatura de specialitate se pot selecta, adapta și experimenta mai multe metode de parcurgere a contururilor.

O primă metodă generală se poate descrie astfel:

- Se alege din imaginea gradient, derivată imaginii inițiale, pixelul cu cea mai mare valoare. Acesta este punctul de start.
- Într-o vecinătate 3×3 a acestui pixel se alege pixelul cu valoarea cea mai mare a gradientului ca al doilea punct al conturului (figura 4.14).
- Începând de aici se caută în vecinătăți 3×3 , centrate pe punctul curent, cel mai bun candidat, situat pe direcția locală a conturului, pentru a fi adăugat la contur.
- Algoritmul funcționează atâtă timp cât mai găsim în vecinătate puncte cu un gradient mai mare ca un prag prestabilit, sau am regăsit punctul de plecare și conturul s-a închis.
- Se scot din imagine punctele de contur detectate și se reaplică algoritmul pentru punctele rămase.
- Procedura se oprește când nu mai găsim puncte cu un gradient mai mare decât un prag (prestabilit la valoarea minimă așteptată pentru un punct de pe contur) pentru modulul gradientului.



Figura 4.14 Exemplu cu direcțiile posibile de căutare a continuării conturului

La folosirea gradientului (o derivare pe două direcții perpendiculare) se poate obține direcția conturului: perpendiculară pe direcția gradientului. Pentru aplicațiile la care nu este nevoie de informații direcționale, se folosește ca operator izotrop operatorul Laplace $\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$. În acest caz se determină trecerile prin zero ale laplaceanului.

Se poate face o adaptare la 8 – conectivitate dintr-o imagine cu nivele de gri (figura 4.15):

- Dacă P_i este punctul curent, atunci vecinul său pe contur va fi căutat pe direcția $\text{Arg}(\vec{\text{grad}}_{P_i} f) \pm \frac{\pi}{2}$.
- Dacă P_j este un vecin astfel determinat, atunci se verifică dacă modulul gradientului său este mai mare decât un prag T .
- Dacă DA atunci P_j este reținut ca punct de pe contur.
- Dacă NU se determină dacă P_j aparține sau nu obiectului comparându-i nivelul de gri cu nivelele de gri ale vecinilor lui.
- Se căută P_k adiacent lui P_i pe direcția perpendiculară la $\text{Arg}(\vec{\text{grad}}_{P_i} f) \pm \frac{\pi}{4}$ și se reia procedura.

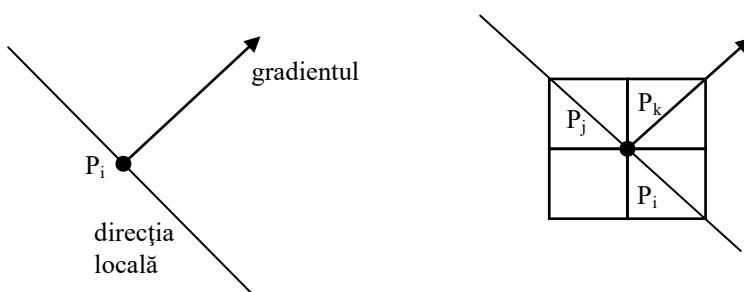


Figura 4.15 Urmărirea conturului

Conturul obiectelor dintr-o imagine binară (obținută de exemplu prin binarizare cu prag pe histogramă) poate fi extras și el printr-o procedură de urmărire (figura 4.16).

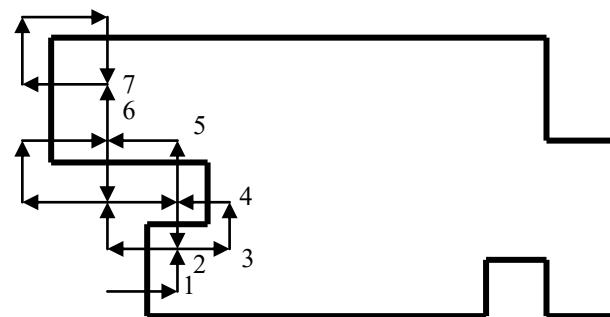


Figura 4.16 Parcurgerea conturului pe imagini binare

- Se baleiază imaginea de jos în sus și, linie cu linie, până ce se găsește un pixel de pe obiect. Fie acesta pixelul 1. Apoi, ținând cont de o 4-conectivitate ușuală se trece la pasul următor.
- Când se găsește un pixel pe obiect se schimbă direcția spre stânga.
- Când se găsește un pixel care aparține fondului, se schimbă direcția spre dreapta.
- Se oprește procesul când se regăsește punctul de start.

Dintre punctele de contur, o importanță deosebită o au punctele de bifurcare: punctele în care se întâlnesc frontierele regiunilor adiacente, minim trei (figura 4.17). Un punct de contur care are un singur vecin pe contur sau care are exact doi vecini puncte de contur, dar aceștia sunt și vecini imediați între ei, este un punct de capăt, și a fost avut în vedere la metodele de închidere a contururilor intrerupte. În cazul contururilor subțiri (de un pixel grosime) un pixel de contur care are doi vecini pe contur, dar aceștia nu sunt și vecini între ei, este un punct obișnuit, de continuitate al frontierei. Într-o reprezentare 8-conectată pentru o imagine numerică, un punct de contur în care se întâlnesc mai multe frontiere este dificil de reprezentat. Rezolvarea problemei detecției acestor puncte presupune abordări euristică (în special pentru rezolvarea situațiilor particulare care se pot întâlni și care nu au fost eliminate prin restricțiile inițiale care particularizează lumea studiată), ca de fapt majoritatea implementărilor algoritmilor de prelucrare și recunoaștere de imagini. Elementul de bază este desigur numărul vecinilor pixelului analizat, vecini căutați totuși, în cvasi-totalitatea aplicațiilor, în vecinătăți 3x3.

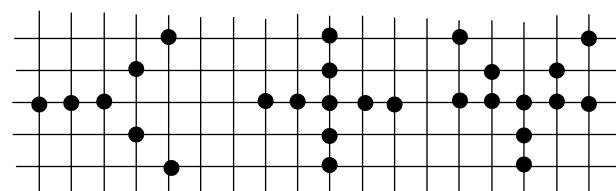


Figura 4.17 Bifurcații cu probleme

Curs 10

5. Segmentarea imaginii

5.1 Prelucrarea contururilor

5.1.1 Închiderea contururilor

Dacă la extragerea conturului s-au aplicat detectoare cu prag bazate pe calculul gradientului, atunci apare problema, semnalată deja într-un capitol anterior, a alegerii pragului. Un prag prea înalt va conduce la contururi întrerupte, în timp ce un prag prea coborât va determina obținerea de contururi prea groase. Alte prelucrări ulterioare ale imaginilor impun utilizarea de contururi închise. Ca urmare se impune implementarea unui algoritm de închidere a contururilor întrerupte. Metoda folosește atât informație din imaginea cu nivele de gri inițială, cât și informație din imaginea care conține numai conturul detectat. Principal se încearcă găsirea punctelor de capăt ale conturului, și căutarea printre vecinii lui al celui mai bun pixel candidat care să fie adăugat la contur. Procedeul se repetă până când se întâlnesc din nou conturul, sau se depășește o lungime a lanțului de pixeli adăugați, de la care nu se mai poate aștepta o închidere a conturului.

Un pixel de capăt, în accepțiunea pe care i-o atribuim în contextul de față, este un pixel care îndeplinește una din următoarele două condiții: are un singur vecin care face parte din contur, sau are doi vecini situați pe contur, dar atunci oricare doi dintre acești trei pixeli sunt vecini între ei. Încercarea de a determina pixelii de capăt pe contururi groase conduce la dificultăți majore, pentru că se obține un număr prea mare de pixeli de capăt sau, mai rău, între aceștia nu se găsesc unii dintre adevărații pixeli de capăt. Se impune subțierea acestor contururi și aducerea lor la grosime de un pixel, într-o etapă anterioară de prelucrare a imaginii. După ce s-au găsit pixelii de capăt, se testează cei 8 vecini ai săi din imagine, pentru a găsi cea mai bună continuare a conturului. Varianta cea mai simplă pe care o putem folosi, este de a calcula gradientul în pixeli candidați și de a-l alege pe acela cu gradientul cel mai mare. O variantă mai elaborată, presupune o parcurgere arborescentă: se calculează suma gradientilor pentru toate căile posibile din arbore. O cale unește pixelul de capăt cu pixelul de pe cel mai departat nivel al arborelui. Găsirea căii cu cea mai mare sumă a gradientului, va indica pixelul care trebuie adăugat conturului: primul pixel din lanțul căii respective.

În concluzie algoritmul general este următorul:

```

FOR (fiecare pixel din imagine) DO
IF (pixelul se află pe contur) THEH
IF (pixelul este pixel de capăt) THEN
    BEGIN
        (initializează contorul lungimii lanțului de închidere)
    REPEAT
        (determină mulțimea pixelilor probabili)
        (alege pixelul pentru continuare dintre pixelii candidați)
        (pixel current ← pixelul ales)
    UNTIL (pixelul current este pixel de capăt) ȘI (contor < prag)
    END;
  
```

5.1.2 Subțierea contururilor

Din mai multe motive care depind de modul în care s-a făcut achiziția imaginii și de prelucrările ulterioare, (deci condițiile de iluminare, rezoluție, acumulare de zgomote, alegerea pragurilor pe histogramă sau pentru extractoarele de contur) se poate întâmpla ca imaginea binară (alb - negru), obținută din imaginea inițială cu nivele de gri, să conțină contururi mai groase decât este util (pixeli mai mulți decât unul pe grosime). Devine necesar pentru multe aplicații să subțiem conturul, păstrând proprietățile legate de forma obiectului. Se cunosc mai mulți algoritmi de subțiere, dar majoritatea se bazează pe algoritmul principal dezvoltat de Sherman care poate fi ilustrat în organigramă următoare (figura 5.1).

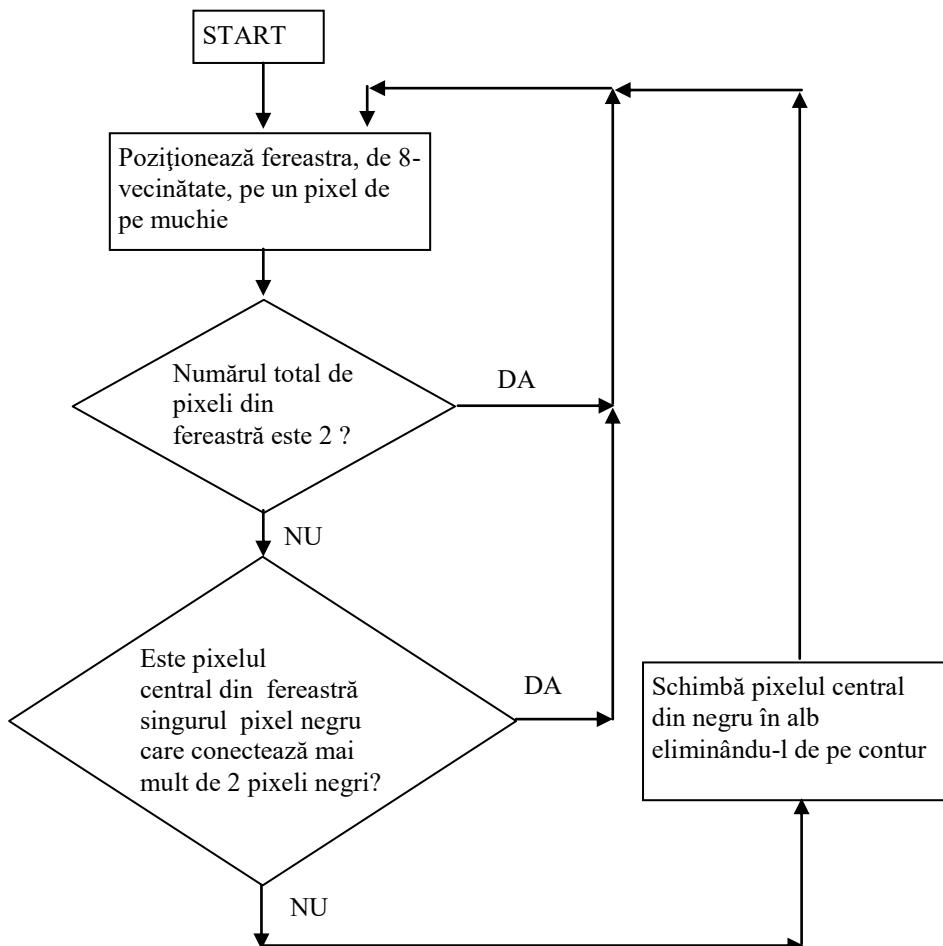


Figura 5.1 Algoritmul general de subțiere a conturului

Pixelul central este eliminat cu condiția să nu erodeze sfârșitul unui contur neînchis, sau să nu creeze o discontinuitate în contur, a cărui lățime este de doar un pixel (figura 5.2). Un algoritm de subțiere de contur se bazează pe depistarea celor două condiții care nu permit îndepărțarea unui pixel de pe contur prin depistarea următoarelor situații (figura 5.3) și a celorlalte șase situații analoage obținute prin rotiri cu câte 90°.

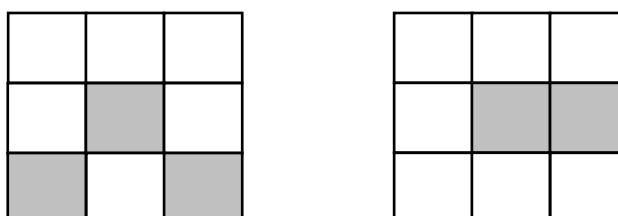


Figura 5.2 Pixelii care nu se pot înălța de pe contur

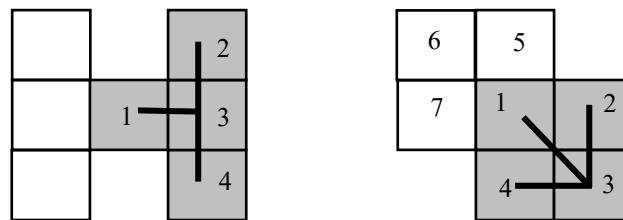


Figura 5.3 Pixelii 1 pot fi înlátaři de pe contur

Depistarea se face realizând următoarea funcție logică:

$$F_s = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot \bar{f}_5 \cdot \bar{f}_6 \cdot \bar{f}_7$$

sau

$$F_s = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot \overline{f_5 + f_6 + f_7}$$

În continuare se aplică următoarea regulă simplă: dacă $F_s = 1$ atunci pixelul central se forțează în 0, iar în caz contrar pixelul central se lasă neschimbăt. Se repetă iterativ atât timp cât apare cel puțin o schimbare la parcurgerea unei imagini complete. Punctele 5, 6, 7 servesc pentru a testa dacă pixelul nu aparține deja conturului, și dacă este cazul ca el să fie păstrat. Se folosesc opt măști diferite pentru a nu prejudicia nici o direcție.

Obținerea unor contururi (granițele regiunilor) subțiri poate fi înlocuită cu utilizarea scheletului regiunii. Să precizăm ce se înțelege prin schelet. Dacă regiunea R are conturul C și P este un pixel din regiune, atunci se calculează distanța cea mai mică de la pixel la contur. Scheletul S este mulțimea punctelor P care au cel puțin două puncte de pe conturul C cu distanțe egale până la P. Câteva exemple sunt prezentate în continuare (figura 5.5).

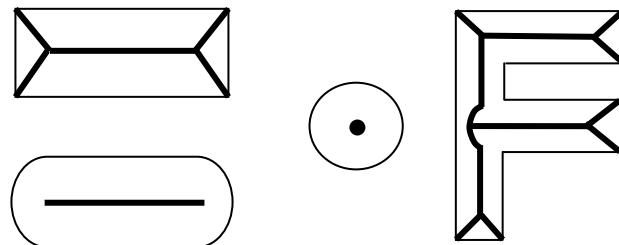


Figura 5.5 Exemple de schelete

5.2 Aproximarea poligonală a contururilor

O metodă relativ simplă de reducere a volumului de informații asociate unei regiuni, este exprimarea sub formă analitică a frontierelor acelei regiuni. Pentru că timpul de răspuns al implementării unei metode numerice este un criteriu de selecție esențial, se poate recurge cu succes la utilizarea unei aproximări poligonale a frontierelor regiunilor din imaginile numerice.

Tinând cont de forma discretă sub care se prezintă frontierele regiunilor în imaginile numerice, se poate observa ușor că cea mai precisă aproximare poligonală este, la limită, aceea care transformă toate punctele de pe muchie în vârfuri și laturile poligonului în segmentele din codul lanț asociat frontierei. Practic rămâne valabilă necesitatea reducerii volumului de informație asociată grafului de descriere, și atunci se au în vedere diferite metode pentru abordarea celei de-a treia variante.

Metodele se diferențiază după criteriul pe care îl folosesc la introducerea unor noi vârfuri poligonale, între două vârfuri deja existente, la parcurgerea frontierelor (figura 5.6):

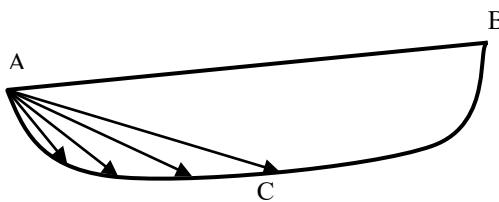


Figura 5.6 Introducerea de noi vârfuri pe contur

Se pot impune valori limită pentru distanța de la segmentul poligonal căutat pentru aproximare până la frontieră sau pentru eroarea medie pătratică introdusă prin această operație. Depășirea acestor limite în procesul de parcurgere al frontierei determină introducerea unui nou segment și deci a unui nou vârf. Vârful se poate introduce acolo unde se atinge un maxim pentru distanța de la segment la frontieră (figura 5.7) sau în poziția care împarte aria dintre segment și frontieră în două părți egale (figura 5.8).

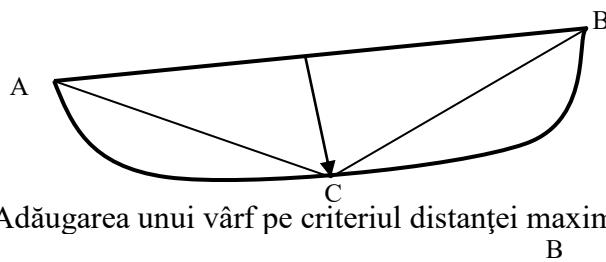


Figura 5.7 Adăugarea unui vârf pe criteriul distanței maxime

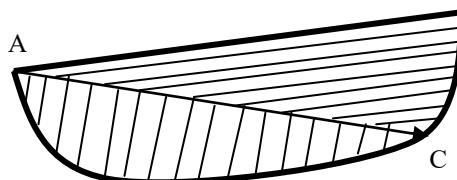


Figura 5.8 Adăugarea unui vârf pe criteriul împărțirii ariei

Trebuie definit un criteriu după care să se decidă continuarea sau oprirea procesului de aproximare poligonală. Acest criteriu poate fi distanța de la segment la frontieră, sau aria cuprinsă între segment și frontieră, sau numărul total de segmente. În toate aceste variante se impune definirea prealabilă a unor valori de prag care să controleze parametrii respectivi. În urma implementărilor practice ale multiplelor variante de aproximare poligonală pentru frontierele regiunilor din imaginile numerice, se ajunge la concluzia că aspectul critic al oricărei metode îl constituie metoda de calcul a distanțelor. Erorile introduse de natura discretă și rezoluția limitată ale imaginilor numerice, sunt importante și încercarea de a le controla prin metode de rafinare mai precise, poate conduce la creșterea importantă a timpului de calcul. În acest sens se poate renunța la a mai folosi norma euclidiană standard pentru calculul distanței între două puncte P_1 și P_2 datorită inconvenientelor legate de calculul radicalului, eroare și timp de calcul: $\|P_1P_2\| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. Se poate folosi norma: $\|P_1P_2\| = |x_1 - x_2| + |y_1 - y_2|$.

Pentru metoda bazată pe distanța minimă de la frontieră la segmentul poligonal, algoritmul presupune următorii pași (figura 5.9):

- Calculează distanța dintre arcul AB și segmentul AB ca maximul distanțelor $P_kP'_k$, unde P'_k este proiecția lui P_k pe segmentul AB.
- Dacă $P_kP'_k < pd$, prag anterior stabilit, atunci AB este segmentul căutat.
- Dacă $P_kP'_k \geq pd$ atunci se introduce C în poziția P_k corespunzătoare distanței maxime și reia procedeul pentru segmentele AC și CB.

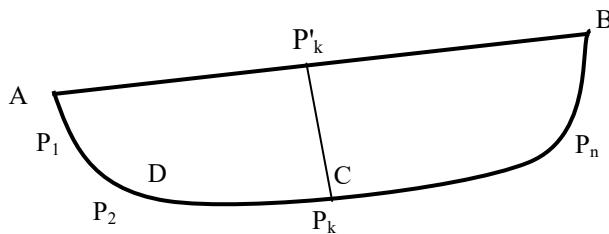


Figura 5.9 Algoritmul de adăugarea unui vârf pe criteriul distanței maxime

Într-o altă variantă de abordare, C determinat ca mai sus este privit numai ca vârf temporar: se caută un nou candidat pe arcul CA și dacă se găsește D, atunci în ipoteza că AD este acceptat ca segment al aproximării (numai există vârfuri între A și D), următorul arc analizat este DB. La punctele 2 și 3 ale algoritmului de mai sus se poate lua o valoare absolută pentru pragul p_d , sau una relativă la lungimea, în pixeli, a arcului AB.

Se poate utiliza și metoda bazată pe criteriul de arie avându-se în vedere că este o metodă rapidă și care induce erori relativ mici în aplicațiile practice. Această metodă presupune următorii pași:

- Calculează aria S_k marginată de arcul AP_k și segmentul AP_k .
- Dacă $S_k \geq p_a$, prag anterior stabilit, atunci $C \equiv P_{k-1}$ și reia algoritmul pentru segmentul CB.
- Dacă $S_k < p_a$ atunci calculează S_{k+1} prin incrementarea lui S_k și reia de la punctul 1.
- Oprește procesul când s-a întâlnit punctul B, punctul final.

5.3 Transformarea Hough

O problemă întâlnită frecvent la extragerea caracteristicilor obiectelor din imagini, este detecția dreptelor sau curbelor care aparțin contururilor obiectelor, adică a curbelor cărora aparțin punctele declarate anterior ca puncte de contur. Aceste drepte și curbe aparțin unei familii C cu expresia generală $f(X_i, a) = 0$, unde X este un vector care reprezintă punctul M_i din spațiu, iar a un vector de parametri care caracterizează pe C.

Fiind dat un ansamblu de puncte discrete M_i care reprezintă conturul obiectului, se caută să se determine dacă unul sau mai multe submulțimi de puncte M_i fac parte dintr-o curbă C_a , unde a rămâne de determinat (figura 5.10). Dacă se încearcă testarea celor n puncte din M_i , două câte două, se ajunge la un număr prea mare de iterații, cel puțin superior lui n^2 . Hough propune o metodă diferită pentru testarea punctelor (x, y) . Principiul metodei constă în a calcula pentru fiecare punct M_i (x_i, y_i) aparținând conturului obiectului ansamblul parametrilor a care verifică $f(x_i, y_i, a) = 0$. Aceasta revine la a evalua în spațiul parametrilor a, suprafața reprezentată de toți parametrii tuturor curbelor care trec prin M_i (x_i, y_i) .

Căutarea dreptelor în plan este din punct de vedere cronologic prima aplicație a transformantei Hough. Presupunem că imaginea originală a fost derivată și s-a binarizat imaginea gradient. Fie un pixel alb $P(x, y)$ dintr-o astfel de imagine, cu fondul negru. Ecuația unei drepte care trece prin acest punct este $y = ax + b$. Aceasta ecuație definește în spațiul (a, b) , spațiul parametrilor, o dreaptă.

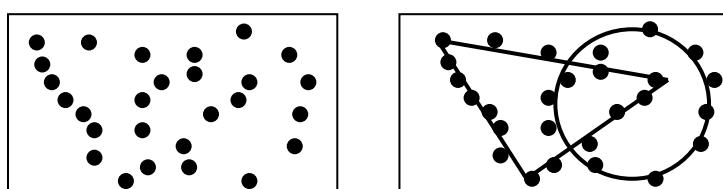


Figura 5.10 Identificarea curbelor din imagini prin metoda Hough

Fiecare punct de pe această dreaptă îi corespunde câte o dreaptă în planul (x,y) (figura 5.11), această ultimă dreaptă trece prin $P(x,y)$. Și reciproc, la fiecare punct din (x,y) îi corespunde câte o dreaptă în planul (a,b) .

Căutarea unei drepte în planul imagine se reduce la căutarea unui punct în planul parametrilor. Procedura poate fi prezentată astfel:

- Pentru fiecare punct din planul parametrilor (a,b) se definește câte un contor $C(a,b)$,
- Pentru fiecare punct alb $P(x,y)$ al imaginii gradient se incrementează $C(a,b)$ corespunzător lui $D(a,b)$ care verifică ecuația $y = ax + b$.

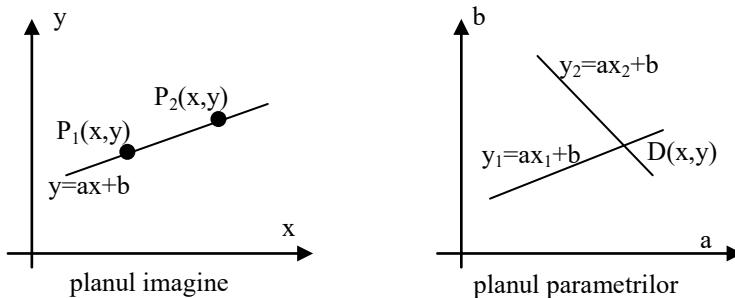


Figura 5.11 Corespondența variabile - parametrii

Dacă există într-adevăr o dreaptă $y = ax + b$ în imagine, atunci $C(a,b)$ corespunde la o valoare mare la sfârșitul procedurii. Identificarea dreptelor prezente în imagine revine la a identifica contoarele C cu valori ridicate. Dacă am presupune că matricea contoarelor este o imagine numerică pe care s-o afișăm, atunci ea trebuie binarizată, iar grafic imaginea va prezenta dreptele cu alb pe fond negru. De observat că pentru aplicarea practică, în cazul dreptelor verticale de exemplu, apar parametrii infiniti ceea ce pune probleme de implementare. Se poate aborda atunci problema parametrică după cum urmează, (figura 5.12).

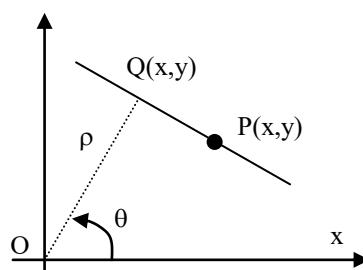


Figura 5.12 Abordarea parametrică a metodei Hough

$$\begin{aligned} \overline{QP} \perp \overline{OQ} &\Rightarrow \overline{QP} \cdot \overline{OQ} = 0 \Rightarrow (\rho \cos \theta \cdot \vec{i} + \rho \sin \theta \cdot \vec{j}) \cdot ((x - X) \cdot \vec{i} + (y - Y) \cdot \vec{j}) = 0 \\ \rho((x - X) \cos \theta - (y - Y) \sin \theta) &= 0 \Rightarrow x \cos \theta + y \sin \theta = X \cos \theta + Y \sin \theta \\ X = \rho \cos \theta, \quad Y = \rho \sin \theta, \quad \rho &= \sqrt{x^2 + y^2} \end{aligned}$$

Planul dual este deci planul (ρ, θ) care rămâne mărginit $0 \leq \theta \leq \pi$ și $-\rho_{\text{lim}} \leq \rho \leq \rho_{\text{lim}}$. La fiecare punct (x,y) corespunde o curbă în (ρ, θ) . La fiecare punct (ρ, θ) corespunde o dreaptă în (x,y) (figura 5.13).

Totuși transformata Hough necesită timp de calcul mare. O metodă de îmbunătățire a timpului de calcul, este asocierea metodei cu calculul gradientului, și deci calculul contoarelor numai pentru o zonă de puncte indicate de direcția gradientului (figura 5.14).

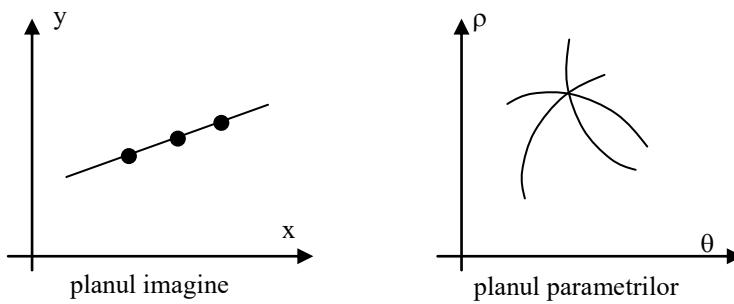


Figura 5.13 Abordarea Hough parametrică

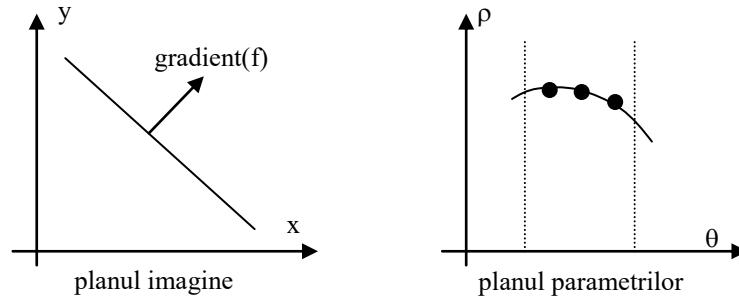


Figura 5.14 Selecția zonelor folosind direcțiile indicate de gradient

Căutarea cercurilor în plan, poate fi abordată într-o manieră similară. Mai mult, transformata Hough permite găsirea oricărui tip de curbă într-o imagine. De exemplu, pentru detecția cercurilor de ecuație: $(x-x_c)^2+(y-y_c)^2=R^2$ spațiul dual al parametrilor este acum spațiul 3D: (x_c, y_c, R) . La orice punct (x, y) din planul imagine îi corespunde o suprafață pătratică în (x_c, y_c, R) . La orice punct (x_c, y_c, R) îi corespunde un cerc în planul (x, y) . Dacă se căută în planul imagine numai cercuri de o anumită rază R_0 , atunci spațiul dual devine planul (x_c, y_c) și corespondența se face între puncte din acest plan și cercuri din planul imagine. Se poate face o generalizare a transformatei pentru spațiul tridimensional.

În concluzie, transformata Hough are două avantaje majore în căutarea unor curbe în imagini: nu cere decât o descriere elementară a obiectelor din imagine, adică extragerea contururilor prin segmentare, iar în plus permite determinarea tuturor aparițiilor parțiale dintr-un contur într-o singură trecere. În schimb apar și trei inconveniente: calcul este laborios, contoarele necesită spațiu de memorie destul de mare, alegerea pasului de cuantificare este dificilă: pasul mic este precis dar dă sensibilitate la zgomote, pasul grosier egalizează valorile contoarelor și împiedică detecția. Există probleme de continuitate a segmentelor de dreaptă cu aceeași direcție. Dacă două drepte sunt paralele și foarte apropiate, ele vor fi greu de detectat. Este utilă introducerea de praguri pentru valorile contoarelor și de probabilități de apariție a dreptelor sau cercurilor.

5.4 Etichetarea regiunilor

Problema care se pune este aceea de a identifica mulțimile de pixeli conexe din imagine numite regiuni. Această identificare presupune în final atribuirea a câte unei etichete pentru fiecare din aceste regiuni. Pentru izolare obiectelor din imagine, această etapă este absolut necesară.

Cu titlu de exemplificare prezentăm numi o metodă particulară foarte simplă adaptată la imagini binare (alb – negru):

- Se face o binarizare a imaginii, de exemplu cu prag pe histogramă, și se izolează obiectele, negre sau albe, pe fondul imaginii, alb respectiv negru (figura 5.15 a).

- Se baleiază imaginea de la stânga la dreapta și de sus în jos cu un şablon de următorul tip (figura 5.15 b) unde se precizează următoarele semnificații: U – sus, L – stânga, C – pixel curent, 0 – pixel de pe fond, 1 – pixel pe obiect. Se aplică următorul algoritm:

```

K = 1
IF f(C) = 0 THEN alege următorul pixel
ELSE
BEGIN
    IF (f(U) = 1 & f(L) = 0) THEN label(C) = label(U)
    IF (f(U) = 0 & f(L) = 1) THEN label(C) = label(L)
    IF (f(U) = 1 & f(L) = 1) THEN
        BEGIN
            label(C) = label(L)
            label(L) = label(U)
        END
    IF (f(U) = 0 & f(L) = 0) THEN
        BEGIN
            label(C) = k;
            k = k+1
        END
    END.

```

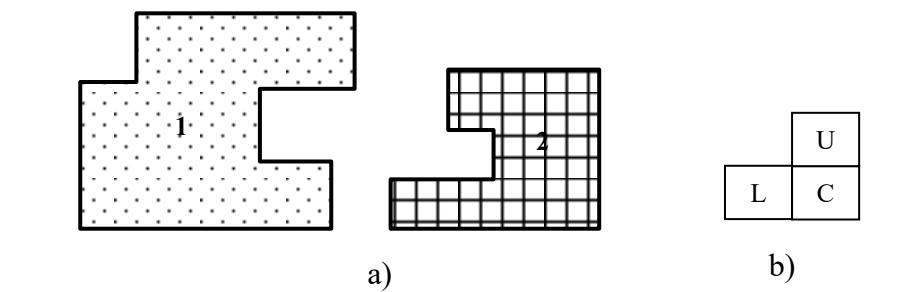


Figura 5.15 a) Etichetarea regiunilor b) Şablonul cu care se parcurge

5.5 Dilatare și eroziune

Astfel de transformări, din grupul transformărilor morfologice, vor fi prezentate spre exemplificare numai pentru imagini binare, alb-negru unde "1" este asociat cu alb, iar "0" este asociat cu negru. În această ipoteză, dilatarea elimină punctele negre dintre cele albe. Un obiect (regiune din imagine) alb pe un fond negru într-o imagine binară va fi astfel dilatat, mărit. Eroziunea elimină punctele albe dintre cele negre. Un obiect alb (regiune din imagine) pe un fond negru într-o imagine binară va fi astfel erodat, micșorat.

Procesul de dilatare se desfășoară atunci după cum urmează. Punctele negre ("0") care au ca vecini și puncte albe ("1") sunt schimbate, atribuite obiectului, deci sunt "albite", și deci crește suprafața obiectului. Pe ferestre de dimensiune $(2N + 1 \times 2N + 1)$ se face funcția logică SAU între toți pixelii, mai puțin pixelul central. Dacă rezultatul este "1" atunci pixelul central este forțat în "1". Dacă rezultatul este "0" pixelul central rămâne neschimbat. Procesul de eroziune se desfășoară în mod complementar după cum urmează. Punctele albe ("1") care au ca vecini puncte negre ("0") sunt schimbate, atribuite fondului, deci "înnegrite", crește suprafața fondului. Pe ferestre de dimensiune $(2N + 1 \times 2N + 1)$ se face funcția logică SI între toți pixelii mai puțin pixelul central. Dacă rezultatul

este "1", atunci pixelul central rămâne neschimbat. Dacă rezultatul este "0" pixelul central devine zero.

Prin dilatare și eroziune se modifică mărimea regiunilor din imaginile binare. Pentru a elimina acest efect, se aplică succesiv un număr de dilatări urmat de un număr egal de eroziuni, caz în care zgomotele izolate, insulele de dimensiuni mici datorate probabil zgomotelor, se elimină fără a modifica dimensiunea obiectelor (regiunilor) de interes din imagine. Abordarea variantei complementare cu obiect negru pe fond alb se face într-o manieră absolut similară, și putem păstra termenii de dilatare pentru mărirea ariei obiectului respectiv de eroziune pentru micșorarea acestei arii (figura 5.16). Operația de dilatare poate fi asociată cu un pas din procesul de subțiere, cu diferența că aici nu se pun condiții asupra păstrării vreunei conectivități.

inițial	dilatare	eroziune
0 0 0 0 0	0 0 0 0 0	0 0 0 0 0
0 0 0 0 0	0 1 1 1 0	0 0 0 0 0
0 0 1 0 0	0 1 1 1 0	0 0 0 0 0
0 0 0 0 0	0 1 1 1 0	0 0 0 0 0
0 0 0 0 0	0 0 0 0 0	0 0 0 0 0

inițial	dilatare	eroziune
1 1 1 1 1	1 1 1 1 1	1 1 1 1 1
1 1 1 1 1	1 1 1 1 1	1 0 0 0 1
1 1 0 1 1	1 1 1 1 1	1 0 0 0 1
1 1 1 1 1	1 1 1 1 1	1 0 0 0 1
1 1 1 1 1	1 1 1 1 1	1 1 1 1 1

Figura 5.16 Exemple simple de eroziuni și dilatări

O aplicație interesantă și simplă pentru aceste două operații, este detectarea defectelor de pe cablajele imprimante: prin eroziuni se poate pune în evidență o zonă de traseu care nu are grosimea necesară, iar prin dilatări se poate evidenția o zonă în care traseele riscă să se atingă pentru că nu s-a respectat distanța impusă tehnologic (figura 5.21). Trebuie amintit că grupul transformărilor morfologice este mult mai complex, și că aceste două transformări prezentate mai sus sunt numai cele mai simple dintre ele, selectate cu rol exemplificator.

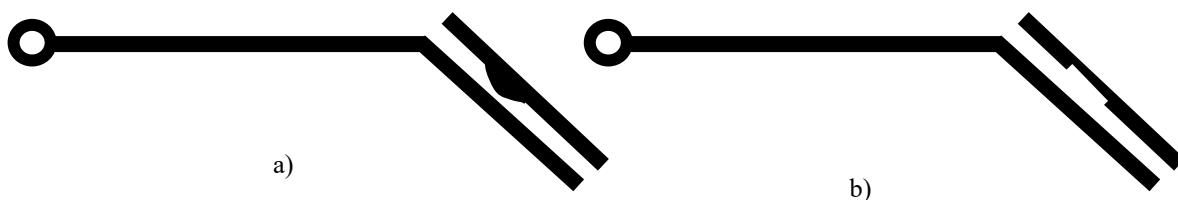


Figura 5.17 a) Cele două trasee se vor atrinde după un număr de dilatari mai mic decât pragul impus b) Unul din trasee se va îintrerupe după un număr de eroziuni mai mic decât pragul impus

Curs 11

6. Descrierea obiectelor

6.1 Descriptori de contur

6.1.1 Lanț numeric (codul lanț)

În contextul acestui capitol termenul “formă” (shape) are avea semnificația regiune din imagine. Conturul este granița sau frontiera regiunii. Într-un capitol viitor termenul “formă” (pattern) va fi folosit pentru a indica un vector de caracteristici. Pentru a sublinia această diferență de semnificație vom folosi acum termenul obiect.

Această metodă este utilă pentru descrierea obiectelor mărginite de frontiere, alcătuite din segmente de dreaptă de lungimi și direcții cunoscute. Codul lanț se bazează pe 8 - conectivitate (figura 6.1) sau 4 - conectivitate și formează un cod numeric plecând de la notația circulară Freeman a direcțiilor.

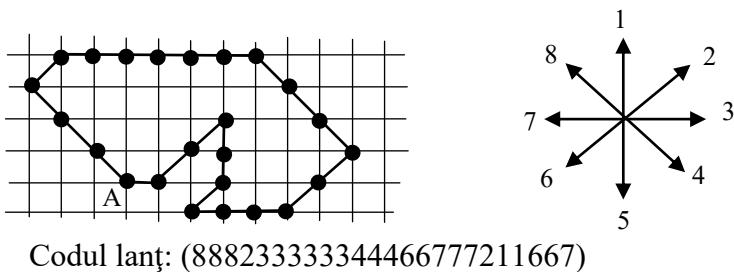


Figura 6.1 Exemplu de cod lanț într-o 8 - conectivitate

Codul lanț depinde de punctul de start, de mărimea și orientarea obiectului în imagine. Translatarea obiectului în imagine nu modifică codul lanț decât în măsura în care pasul de digitizare este prea mare, și ca urmare contează poziția obiectului relativ la grila de digitizare. Codul lanț se poate normaliza relativ la punctul de start, tratându-l ca secvență circulară a numerelor asociate direcțiilor, și redefinind acest punct astfel încât secvența rezultată să aibă amplitudinea minimă. Pentru orientare, se poate normaliza codul lanț făcând să apară prima diferență în locul codului lanț propriu-zis. Pentru prima diferență calculul se face în sens trigonometric contorizând numărul de direcții care separă două elemente adiacente ale codului lanț (figura 6.2). Codul diferențelor se scrie în inel, prima direcție se compară cu ultima. Un cod lanț are mai multe prime diferențe depinzând de punctul de start, punctul de start în parcurgerea conturului. Numărul obiectului este prin definiție prima diferență cu amplitudinea cea mai mică. Ordinul n, pentru un număr de obiect, este definit ca numărul de cifre conținut în codul lanț. Un contur închis, într-o 4 - conectivitate, este caracterizat de un număr de obiect par. Valoarea ordinului limitează numărul de obiecte diferite posibile.

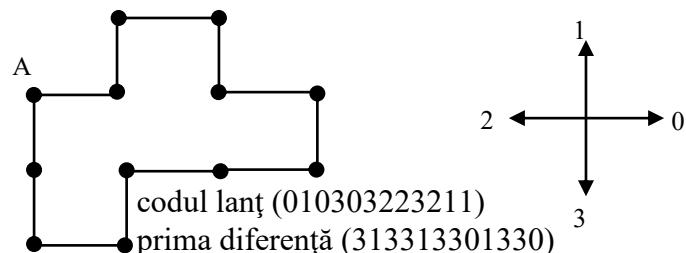


Figura 6.2 Exemplu de primă diferență în 4 - conectivitate

	Obiect 1	Obiect 2	Obiect 3	Obiect 4	Obiect 5
Ordin	4	6	4	8	8
Cod lanț	0321	003221	00332211	00323211	00032221
Diferență	3333	303303	30303030	30331330	30033003
Numărul obiectului	3333	033033	03030303	03313303	00330033

Figura 6.3 Exemple pentru numerele obiectelor

Codul numărul obiectului nu depinde de rotație, dar este dependent de orientarea grilei de conectivitate (figura 6.3). Normalizarea definiției grilei în raport cu obiectul se poate face alegând, de exemplu, axa principală ca linia dreaptă cea mai lungă, iar axa secundară, perpendiculară pe prima și limitată de contur. Normalizările propuse nu sunt singurele posibile.

În general rezultatele așteptate în urma acestor normalizări sunt exacte în măsura în care și contururile sunt ele însele invariante la rotație și la mărime (un același obiect apare în imagine mai mare sau mai mic, în funcție de distanța până la cameră), ceea ce este dificil de obținut din simplul motiv că imaginea trebuie considerată numai în nodurile rețelei rectangulare de digitizare. Cu toate acestea, codul lanț, prin simplitate, se face util într-un număr de aplicații. Codul lanț reduce dimensiunea reprezentării, dar oferă și avantaje utilizabile pentru măsurarea unor caracteristici fizice ale obiectelor care apar în aplicații industriale: aria, perimetrul, grosimea.

6.1.2 Descriptori Fourier

Considerăm conturul ca o curbă închisă reprezentată în planul complex. Fiecare punct de coordonate (x,y) de pe contur îi corespunde un punct $z = x + iy$ în planul complex. Parcurgând curba în sens trigonometric cu viteză uniformă, se poate da o reprezentare parametrică în t : $z(t) = x(t) + iy(t)$. Viteza de parcursare trebuie să conducă la o perioadă 2π . Parcurgerea se poate face repetat, și deci se asigură periodicitatea și posibilitatea de a dezvolta pe $z(t)$ într-o serie Fourier convergentă. Dacă se discretizează $z(t)$ cu un pas uniform și se obține $z(k)$, varianta discretă a semnalului, cu N componente atunci se poate scrie transformata Fourier discretă astfel:

$$z(k) = \sum_{n=0}^N a_n e^{-nk \frac{2\pi i}{N}}, \text{ unde } a_n \text{ corespund celor mai mici } N \text{ frecvențe după relația: } a_n = \frac{1}{N} \sum_{k=0}^N z(k) e^{nk \frac{2\pi i}{N}}.$$

Această serie este reprezentarea Fourier clasică a conturului obiectului. Se impune o normalizare a acestei reprezentări pentru a elimina efectul mărimii, punctului de start, poziției și orientării obiectului, și se face astfel reprezentarea aptă de a fi utilizată într-un proces de identificare de obiecte. Normalizarea la translație se face prin forțarea lui a_0 la zero. Normalizarea în raport cu mărimea obiectului se face prin împărțirea tuturor coeficienților prin $|a_1|$, care în condițiile de mai sus este cel mai mare coeficient. Punctul de start și orientarea conturului obiectului afectează numai faza coeficienților. Fiind doi factori implicați pentru normalizare, trebuie luați în considerare doi coeficienți, cel puțin. Se încearcă aducerea la zero a fazelor celor doi coeficienți cu amplitudinile

cele mai mari. Fie aceştia a_1 şi a_k . Se multiplică toţi coeficienţii a_i prin $e^{\frac{[(i-k)\beta+(l-i)\beta]}{k-1}}$, unde α şi β sunt fazele celor doi coeficienţi amintişi. Totuşi condiţia de fază nulă pentru cei doi coeficienţi este îndeplinită pentru $|k-1|$ combinaţii diferite de orientări şi puncte de start. Dacă avem $k = 2$ totul este în regulă, dar aceasta nu se întâmplă în general şi atunci normalizarea trebuie continuată prin implicarea unui al treilea coeficient.

6.1.3 Funcţia unghiulară cumulativă

Conturul închis al obiectului este reprezentat parametric sub forma $\{x(t), y(t)\}$, unde t este lungimea măsurată de la punctul de start până la punctul curent, pe curbă, şi deci $0 \leq t \leq L$, cu L lungimea curbei.

Funcţia unghiulară cumulativă se defineşte ca: $\Phi(t) = \Theta(t) - \Theta(0)$, unde $\Theta(t)$ este direcţia unghiulară în punctul curent t . $\Theta(t)$ este unghiul pe care tangentă la curbă, în punctul t , îl face cu axa absciselor. Pentru a nu lua în calcul informaţia legată de mărimea obiectului L , ci o periodicitate 2π , se poate folosi o altă formulare: $\Phi^*(t) = \Phi\left(\frac{Lt}{2\pi}\right) + t$ pe $[0, 2\pi]$, care asigură invariантă nu numai la mărime ci şi la translaţie şi orientare. Se observă că pentru un cerc $\Phi^*(t) = 0$, ceea ce poate indica faptul că funcţia măsoară diferenţa dintre forma respectivă şi o formă circulară. Toate curbele plane şi închise identice ca formă şi punct de start se transformă prin acest procedeu în aceeaşi funcţie $\Phi^*(t)$.

Fie V_0, V_1, \dots, V_{m-1} vîrfurile conturului poligonal. Dacă distanţa $V_{i-1}V_i$ este Δl_i , atunci schimbarea unghiulară corespunzătoare vîrfului i este $\Delta\Phi_i$. Folosind schimbarea de variabilă din relaţia precedentă, se constată că pentru calculul numeric nu este convenabilă utilizarea $\Phi(\lambda)$ şi $d\lambda$, unde $\lambda = \frac{L \cdot t}{2 \cdot \pi}$, pentru că acestea sunt mărimi greu accesibile din punct de vedere practic. Mai accesibile sunt, pentru un contur poligonal, lungimea arcului $l_k = \sum_{i=1}^k \Delta l_i$ şi schimbarea de orientare cumulativă $\Phi_k = \sum_{i=1}^k \Delta\Phi_i$, pâna la, respectiv într-un vîrf al conturului poligonal.

Se poate considera acum situaţia ilustrată în figura 6.6, calculând schimbările de direcţie în punctele successive de pe contur după formula: $\Phi(t) = \sum_{i=1}^k \Delta\Phi_i$, unde $\sum_{i=1}^k \Delta l_i \leq l \leq \sum_{i=1}^{k+1} \Delta l_i$,

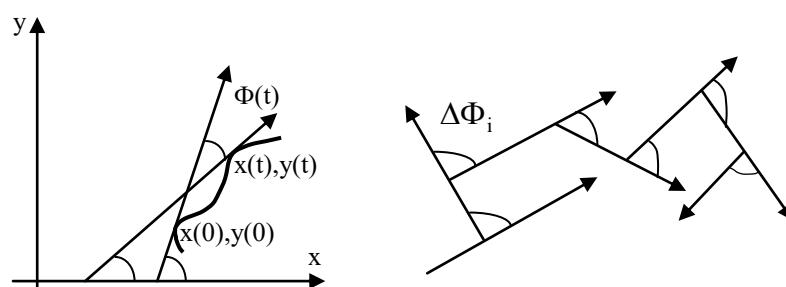


Figura 6.6 Funcţia cumulativă de schimbare de direcţie

Funcția Φ^* poate fi descompusă într-o serie Fourier astfel: $\Phi^*(t) = \mu_0 + \sum_{k=1}^{\infty} (a_k \cos kt + b_k \sin kt)$ și

varianta polară: $\Phi^*(t) = \mu_0 + \sum_{k=1}^{\infty} A_k \cos(kt - \alpha_k)$, unde (A_k, α_k) sunt coordonatele polare ale punctului (a_k, b_k) .

Fiecare figuri îi corespunde o regiune plană în imagine. Fiecare regiune are un contur care poate fi aproimat poligonal. Fiecarui contur poligonal i se poate asocia o funcție Φ^* care poate fi dezvoltată în serie Fourier. Se obține astfel o formă asociată figurii inițiale și care este un vector alcătuit din coeficienții (A_k, α_k) .

Reținând o mulțime rezonabilă (8-16) dintre acești coeficienți (A_k, α_k) , se poate caracteriza obiectul respectiv în vederea identificării. Mai puțini coeficienți înseamnă pierdere de informație, alterarea obiectului, iar prea mulți înseamnă un timp de calcul mai mare. Dacă Φ^* asigură invarianța la mărimea, poziția și orientarea obiectului, rămâne de rezolvat invarianța la schimbarea punctului de start. Pentru aceasta se poate roti obiectul până ce punctul de start se suprapune, sau este cel mai aproape posibil de axa principală de simetrie a obiectului. Acești descriptori asigură și reconstrucția obiectului inițial.

6.1.4 Curbura

Problema care se pune este cum definim curbura pentru o reprezentare discretă. Curbura într-un punct al unei curbe plane se poate defini ca raportul între variația instantanea a pantei tangentei la curbă în punctul respectiv și variația arcului de curbă corespunzător: $k(s) = \frac{d\Psi(s)}{ds}$. Inversul curburii este raza unui cerc cu aceeași curbură ca și curba în punctul considerat. Pentru o reprezentare parametrică și normalizată la intervalul $[0,1]$ a curbei, $\{x(t), y(t)\}$, curbura se poate scrie astfel:

$$k = \frac{y''}{(1 + (y')^2)^{3/2}} \quad \text{unde} \quad y' = \frac{dy}{dx}, y'' = \frac{d^2y}{dx^2}.$$

Dacă $\dot{x} = \frac{dx}{dt}$, $\ddot{x} = \frac{d^2x}{dt^2}$, $\dot{y} = \frac{dy}{dt}$, $\ddot{y} = \frac{d^2y}{dt^2}$ atunci $y' = \frac{\dot{y}}{\dot{x}}$, $y'' = \frac{\ddot{x}\dot{y} - \dot{x}\ddot{y}}{\dot{x}^2}$. Curbura se poate scrie

$$\text{acum: } k = \frac{\ddot{x}\dot{y} - \dot{x}\ddot{y}}{\sqrt{\left(\frac{\dot{x}^2 + \dot{y}^2}{\dot{x}^2}\right)^3}}.$$

Aceeași problemă se pune aici că la oricare descriptor: se poate reconstrui forma dându-se descriptorii? Răspunsul este că și demonstrația se poate face cu ajutorul geometriei diferențiale: curbele similare ca formă au și curburi similare. Problemele care apar sunt date de natura discretă a informației dintr-o imagine: panta tangentei nu poate varia infinitezimal ci doar cu 45 de grade! Mai mult, derivatele fac curbura sensibilă la zgromot. Se poate apela la o variantă de calcul în care reprezentarea parametrică este modificată printr-o convoluție cu un nucleu de tip Gauss și pentru $x(t)$ și pentru $y(t)$. O valoare de $\sigma = 5$ și calculul curburii pe zone de dimensiune 5σ , spre exemplu, asigură un echilibru între păstrarea detaliilor (convoluția Gauss are efect de netezire) și suprimarea influenței zgromotului.

Din modul de calcul parametrizat de-a lungul curbei se elimină efectul translatării sau rotirii formei. Pentru normalizarea legată de poziția punctului de start, se alege un punct cât mai apropiat de axa principală de simetrie, aşa cum s-a mai amintit. Pentru normalizarea legată de mărimea curbei se face împărțirea cu valoarea maximă a curburii.

6.1.5 Semnătura polară

După obținerea codului lanț al conturului, se calculează coordonatele centrului de greutate G și în continuare distanța de la fiecare punct al conturului la centrul de greutate în funcție de direcție. Se determină astfel semnătura polară care este o caracteristică pe baza căreia se poate face identificarea obiectelor. Un rol important în obținerea unui rezultat corect îl are precizia cu care se determină coordonatele centrului de greutate. Dacă acest centru de greutate se calculează numai folosind punctele de pe contur, atunci eventualele erori în aprecierea lungimii efective pe contur pot fi importante.

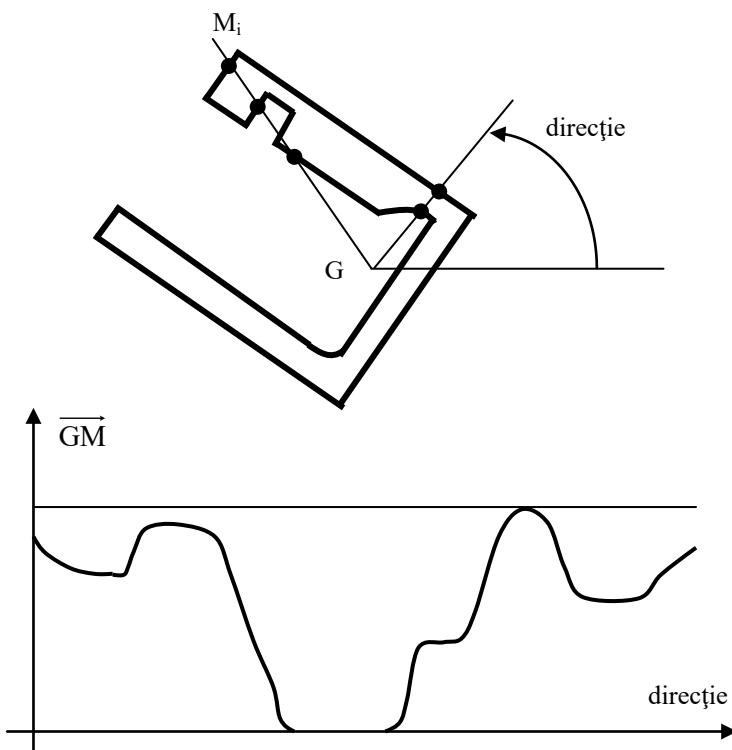


Figura 6.7 Semnătura polară cu pas unghiular constant

Se poate parcurge conturul în două moduri: fie cu un pas de lungime constantă, de-a lungul conturului, fie cu un pas unghiular constant în jurul centrului de greutate. Prima variantă are dezavantajul unui număr variabil de distanțe în vectorul semnăturii, și avantajul că permite reconstruirea formei, dacă se dă semnătura. A doua variantă are dezavantajele imposibilității reconstrucției formei plecând de la semnătura polară și al unor probleme legate de precizia de calcul și de existența intersecțiilor multiple. Tot această metodă prezintă însă avantajul că oferă o semnătură polară de lungime cunoscută aprioric și controlabilă.

Vom descrie în continuare varianta cu pas unghiular constant. Vectorul \$GM\$ baleiază conturul cu un pas unghiular $\Delta\alpha = \frac{2\pi}{N}$ unde \$N\$ este numărul de valori ales pentru semnătură. Pentru fiecare poziție a vectorului \$GM\$ se pot obține mai multe puncte de intersecție, lungimile obținute se combină, se

calculează minimul, media sau maximul acestor valori și se va obține semnătura polară GM(a), funcție periodică de perioadă 2π (figura 6.7).

Se pune problema imunizării semnăturii polare la mărimea obiectului, la poziția și la orientarea obiectului în imaginea numerică. Coordonatele centrului de greutate se calculează cu formulele:

$$X_G = \frac{\sum_{i=1}^N \sum_{j=1}^M x_i \cdot f(i,j)}{\sum_{i=1}^N \sum_{j=1}^M f(i,j)} \quad Y_G = \frac{\sum_{i=1}^N \sum_{j=1}^M y_j \cdot f(i,j)}{\sum_{i=1}^N \sum_{j=1}^M f(i,j)}, \text{ unde } I(i,j) \text{ este } 1 \text{ dacă punctul aparține}$$

obiectului și 0 în caz contrar, iar i și j sunt indicii de linie și coloană din imagine. Raportarea la aceste coordonate înălțură efectul translației obiectului în imagine.

Pentru înălțurarea efectului factorului de scară care face ca obiectul să apară mai mare sau mai mic în funcție de distanța de la care este privit (nu se iau în considerare deformările neasemenea date de o eventuală direcție de privire care nu este perpendiculară pe planul formei) se poate recurge la o metodă foarte simplă de normalizare cu distanța maximă. Efectul orientării diferite a obiectului în imagine (este evident că în fazele de învățare și de clasificare obiectul va avea orientări diferite) a fost înălțurat prin aplicarea unei rotații circulare a semnăturii, până ce o pseudo-axă de simetrie este adusă într-o poziție prestabilită. O primă soluție ar fi să se compare semnătura obiectului examinat cu toate semnăturile învățate pentru piesele din lumea cunoscută. De remarcat că există $N \times N$ posibilități de a compara două semnături de dimensiune N , datorită orientărilor diferite (comparare circulară). O altă variantă este bazată pe estimarea axei de simetrie maximală care trece prin G . Direcția ei, Θ_p , este aceea pentru care se minimizează funcția:

$$x(\Theta_j) = \sum_{i=1}^N [GM(\alpha_i + \Theta_j) - GM(-\alpha_i + \Theta_j)] \quad j = 0, 1, \dots, N-1$$

Aceasta este direcția s cu proprietatea că $\Delta_s < \Delta_j \forall j = 0, 1, 2 \dots N-1$. Determinarea axei de simetrie maximale nu este suficientă pentru determinarea orientării piesei, eroarea fiind de $\pm \pi$. Comparările care se fac trebuie să distingă între următoarele patru situații (figura 6.8). Pentru imunizarea la orientare, semnătura polară se rotește circular astfel încât direcția Θ_p să se afle pe direcția 180°.

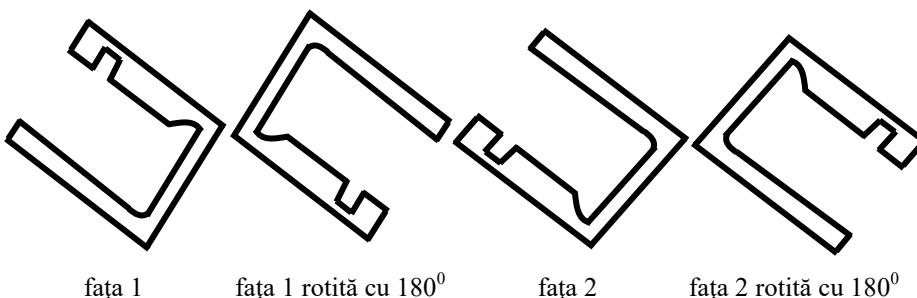


Figura 6.8 Situațiile care conduc la aceeași semnătură

Precizia obținută este proporțională cu N , uzual se lucrează cu o semnătură de 360 de grade, deci cu o discretizare a direcției din grad în grad. $N = 360 \{d(k), k = 0, 1, 2 \dots N-1\}$. Determinarea orientării și identificării feței vizibile a unei piese, corespunde la o complexitate mai mică decât aceea cerută de compararea circulară. Varianta cu pas de lungime constantă se utilizează mai ales acolo unde se întâlnesc forme cu multiple concavități, și în acest caz se acordă o atenție particulară ponderării distanțelor pe contur, în funcție de poziția relativă a pixelilor vecini, pentru a minimiza erorile de calcul.

6.1.6 Histograma de intersecție

O altă metodă de descriere a obiectelor care poate fi avută în vedere se bazează pe construirea histogramelor de intersecție dintre conturul formei și un set de drepte generate aleator (figura 6.9). Orientarea, poziția și mărimea formelor respective se presupun normalizezate anterior.

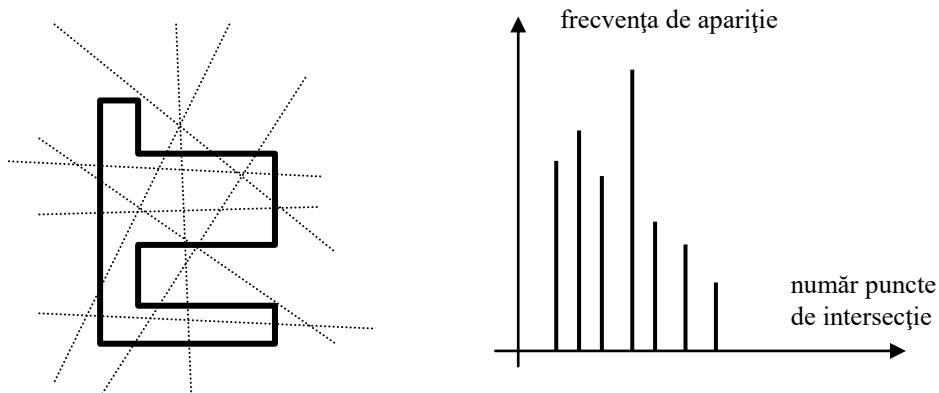


Figura 6.9 Histograma de intersecție

Se generează aleator, atât în faza de antrenament cât și în faza de clasificare seturi de căte N (de exemplu N=256) drepte aleatoare și se contorizează pentru fiecare în parte numărul de puncte de intersecție cu conturul formelor respective. Histograma, care reprezintă probabilitatea ca o dreaptă să aibă un anumit număr de puncte de intersecție cu conturul, se poate folosi drept caracteristică, devine o semnătură, pentru obiectul respectiv în procesul de clasificare.

6.2 Descriptori de suprafață

6.2.1 Descriptori topologici

Descriptorii topologici sunt descriptori legați de proprietățile topologice ale obiectului, și care sunt invariante la rotație, translație și omotetie. Ei acceptă chiar și deformări plastice, dar nu și pe cele de tip pliere sau rupere. Proprietățile prezentate în continuare sunt insensibile la distanța între puncte, dar nu și la conexitatea punctelor: numărul de găuri H dintr-o suprafață omogenă, respectiv numărul de componente conexe C. O componentă conexă a unui ansamblu este un subansamblu de mărime maximă, astfel încât două puncte ale ei să poată fi legate între ele de o dreaptă sau o curbă continuă care rămâne în întregime în interiorul subansamblului. Regiunea din figura 6.10 are trei componente conexe reprezentate de trei zone independente.

Relația topologică care definește descriptorul clasic cunoscut sub numele de numărul lui Euler se scrie: $E = C - H$. Formula lui Euler (figura 6.11) este în acest context: $S - Q + F = C - H$, unde S este numărul de vârfuri, Q este numărul de laturi, F este numărul de fețe.

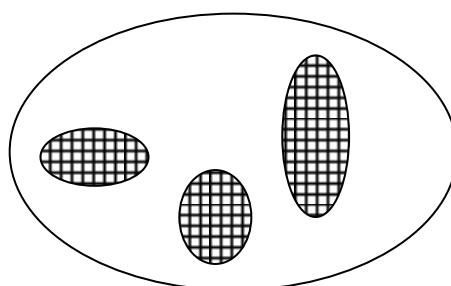


Figura 6.10 Regiuni conexe

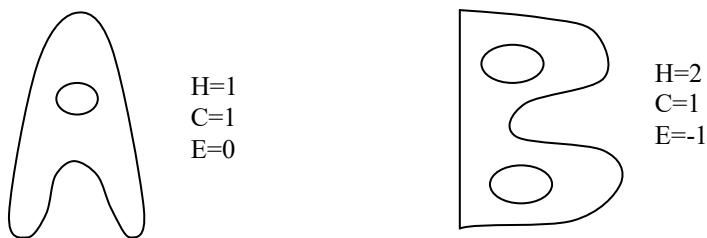


Figura 6.11 Exemple de calcul pentru numărul lui Euler

6.2.2 Descriptori geometrici

Acești descriptori sunt cei mai simpli descriptori de suprafață pentru că ei iau în considerare numai atribute metrice: perimetru, suprafață, raza cercului înscris etc. Numai cu titlu de exemplu putem menționa:

- excentricitatea: $\frac{L}{L'} = \frac{\text{axa_mare}}{\text{axa_mică}}$

- compactitatea: $\frac{p^2}{A} = \frac{(\text{perimetru})^2}{\text{aria}}$

- finețea, inversa compactății: $\frac{A}{p^2}$

- raportul dintre diametrul cercului înscris sau circumscris și perimetru: $\frac{\Phi}{p}$

Pentru obiectele din figura 6.12 se pot aprecia următoarele valori pentru descriptorii amintiți mai sus:

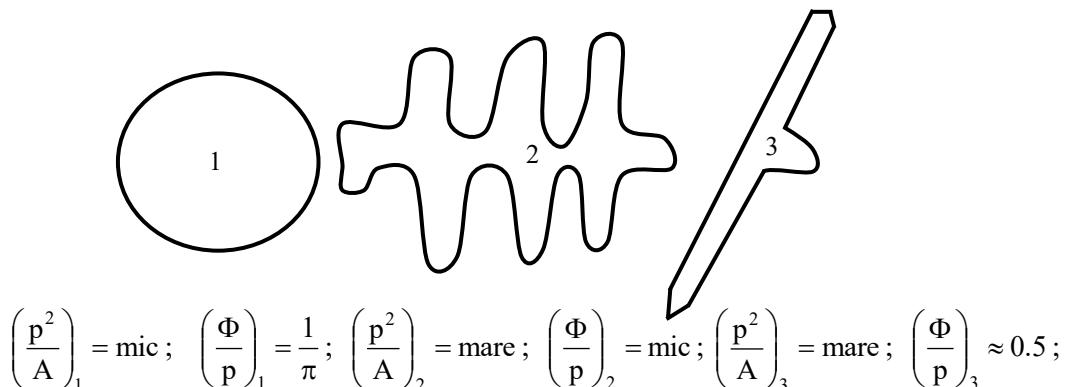


Figura 6.12 Exemple de calcul pentru descriptori geometrici

6.2.3 Momente

Momentele sunt descriptori care iau în considerare organizarea interioară a punctelor unei suprafete. Ele pot fi aduse la forme invariante la translație, rotație sau variație omotetică.

Momentul de ordinul $(p+q)$ al unei funcții $f(x,y)$ se scrie: $m_{pq} = \iint_{-\infty}^{\infty} x^p y^q f(x,y) dx dy$, pentru $p, q = 0, 1, 2, \dots$. Aplicată la imagini discrete formula devine: $m_{pq} = \sum_i \sum_j x_i^p y_j^q f(i,j)$.

Pentru cazul imaginii binare negre pe fond alb, adică: $f(i,j) = 1$ dacă $P(i,j) \in$ obiectului:

- momentul de ordin zero reprezintă aria obiectului din imagine: $m_{00} = \sum_i \sum_j f(i,j) = A$,
- momentele de ordin 1 se calculează după formulele: $m_{10} = \sum_i \sum_j x_i f(i,j)$, $m_{01} = \sum_i \sum_j y_j f(i,j)$,
- coordonatelor centrului de greutate al obiectului din imagine: $X_G = \frac{m_{10}}{m_{00}}$ $y_G = \frac{m_{01}}{m_{00}}$.

Momentele centrale de ordin $(p+q)$ sunt date de expresia: $\mu_{pq} = \sum_i \sum_j (x_i - x_G)^p (y_j - y_G)^q f(i,j)$.

Momentele centrale sunt invariante la translație, pentru că centrul de greutate se deplasează în același timp cu obiectul. Momentele centrale de ordin 2 sunt definite astfel: $\mu_{20} = \sum_i \sum_j (x_i - x_G)^2 f(i,j)$, $\mu_{02} = \sum_i \sum_j (y_j - y_G)^2 f(i,j)$, $\mu_{11} = \sum_i \sum_j (x_i - x_G)(y_j - y_G) f(i,j)$. Aceste momente permit să se găsească axele de inerție ale obiectelor. Mai indicăm, pentru exemplificare, câteva momente importante:

$$\begin{aligned}\mu_{10} &= \sum_i \sum_j (x_i - x_G)^1 (y_j - y_G)^0 f(i,j) = m_{10} - \frac{m_{10}}{m_{00}} m_{00} = 0 \\ \mu_{00} &= m_{00}; \quad \mu_{11} = m_{11} + y_G m_{10} \\ \mu_{10} &= 0; \quad \mu_{30} = m_{30} - 3x_G m_{20} + 2m_{10} x_g^2 \\ \mu_{01} &= 0; \quad \mu_{03} = m_{03} - 3y_G m_{02} + 2m_{01} y_G^2 \\ \mu_{20} &= m_{20} - x_G m_{10}; \quad \mu_{12} = m_{12} - 2y_G m_{11} - x_G m_{02} + 2m_{10} y_G^2 \\ \mu_{02} &= m_{02} - y_G m_{01}; \quad \mu_{21} = m_{21} - 2x_G m_{11} - y_G m_{20} + 2m_{01} x_G^2\end{aligned}$$

Momentele de mai sus pot fi normalizate la schimbarea mărimeii formei înlocuind momentele centrale μ_{pq} cu variantele lor normalizate η_{pq} . Momentele centrale normate se definesc prin:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad \text{cu} \quad \left\{ \begin{array}{l} p, q = 2, 3, \dots \\ \gamma = \frac{p+q}{2} + 1 \end{array} \right.$$

Următoarele momente sunt exemple de momente invariante la translație și rotație:

$$\begin{aligned}\Phi_1 &= \eta_{20} + \eta_{02}; \quad \Phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2; \quad \Phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2; \\ \Phi_4 &= (\eta_{30} - \eta_{12})^2 + (\eta_{21} + \eta_{03})^2; \quad \Phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) + \\ &+ (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2); \\ \Phi_6 &= (\eta_{20} - \eta_{02})(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ \Phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) - \\ &- (\eta_{30} - 3\eta_{12})(\eta_{21} - \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2)\end{aligned}$$

În această manieră se generează momentele invariante generalizate (invariante la translație, rotație și omotetie, adică la grupul similitudinilor afine). S-au realizat aplicații în care, de exemplu, un avion a fost descris în vederea recunoașterii în zbor prin mărimele celor șapte momente generalizate; tot momentele au fost folosite pentru a ajuta la găsirea unei soluții la problema orientării unui obiect în spațiu. Există variante de momente 3D.

Curs 12

6.2.4 Textura

În familia atributelor suprafețelor textura este un descriptor care permite, în anumite circumstanțe, identificarea obiectelor. Textura nu are o definiție formală dar ea descrie cantitativ atributele de suprafață ca: regularitate, finețe, granulație, rugozitate. Distingem o primă familie a texturii obiectelor naturale la scară mare ca: vederile aeriene ale pădurilor, întinderilor de ape (lacuri, mări), câmpuri cu monocultură, și respectiv până la cele de scară mică: radiografiile medicale ale ţesuturilor organice, fotografiile tăieturilor prin arbori, straturilor geologice. Într-o altă categorie se poate include textura obiectelor artificiale: stofe cu motive repetitive, ziduri de piatră, acoperișuri de țiglă, piese industriale (suprafețele lor) înainte și după prelucrare.

Un instrument util în descrierea texturilor este spectrul energiei Fourier. Energiea totală a transformării Fourier discrete $F(u,v)$ în domeniul frecvență NxN se scrie:

$$E_T = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} |F(u,v)|^2 .$$

S-a stabilit că variațiile progresive, lente ale nivelor de gri sunt asociate frecvențelor joase u și v , iar variațiile brutale sunt corespunzătoare frecvențelor înalte. Deci o textură regulată de variație spațială lină a nivelor de gri prezintă un spectru energetic important în domeniul frecvențelor joase, iar o textură grosieră cu variații brutale ale intensității luminoase oferă un spectru întărit în domeniul frecvențelor înalte. Deci se poate imagina o metodă de diferențiere a texturilor prin calculul spectrului energiei în interiorul inelelor cu centrul în originea planului frecvențelor (figura 6.13). Energia unui inel de rază interioară r_1 și de rază exterioară r_2 este dată de expresia:

$$E(r_1, r_2) = \sum_{u=r_1}^{r_2} \sum_{v=r_1}^{r_2} |F(u,v)|^2 , \quad r_1^2 < u^2 + v^2 < r_2^2 , \quad 0 < r_1, r_2 < N - 1 .$$

Inelele energetice apropiate de centrul O corespund la texturile fine, în vreme ce inelele depărtate indică texturi grozioase. Spectrul energetic poate fi calculat și într-un domeniu de integrabilitate limitat de două direcții Θ_1 și Θ_2 în planul frecvențelor. Atunci atributul energetic ia forma:

$$E(\Theta_1, \Theta_2) = \sum_u \sum_v |F(u,v)|^2 , \text{ cu } \Theta_1 < \operatorname{arctg} \left| \frac{u}{v} \right| < \Theta_2 .$$

Energia $E(\Theta_1, \Theta_2)$ trece printr-un maxim de fiecare dată când o zonă limitată de direcțiile Θ_1 și Θ_2 surprinde orientări remarcabile ale texturii, care sunt de exemplu linii cu gradient puternic al intensității luminoase.

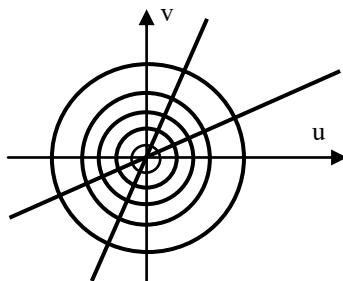


Figura 6.13 Energia în spațiu de frecvențe

Un alt instrument pentru caracterizarea texturilor, de data aceasta în domeniul spațial (linii, coloane, pixeli, culori), sunt matricele de detectare a perechilor. Se iau în considerare poziția relativă a pixelilor și intensitățile pixelilor conecți. De exemplu în varianta cea mai simplă, acești pixeli sunt considerați în perechi, câte doi. Se stabilește dinainte un criteriu de distanță și intensitate. Doi pixeli, o pereche, sunt selectați dacă se află la distanță d și se găsesc pe o dreaptă de direcție Θ . Pentru a ține cont de intensitate, punctele M și N sunt caracterizate de nivelelor lor de gri i și j . Perechile M_i, M_j sunt contorizate de numerele $P(d, \Theta, i, j)$ care și ele sunt așezate într-o matrice $|A_{ij}| = a_{ij}$ după nivelele de gri. Pentru normalizare $P(d, \Theta, i, j)$ sunt împărțite prin $n = P(d, \Theta)$ și aranjate într-o matrice similară $|C| = c_{ij}$. Termenii c_{ij} evaluează probabilitatea de existență a perechilor de puncte care satisfac parametrii impuși pentru direcție, distanță și nivele de gri. $|C|$ este matricea de incidență. Exemplu: Fie o imagine de 5×5 pixeli cu 4 nivele de gri ca în figura 6.14. Se caută perechi de pixeli orientate pe direcția D' la distanță d , pasul rețelei din 8-conectivitate. $n=16$, $\Theta = 3\pi/4$. Pentru un timp de calcul rezonabil, pentru că dimensiunea lui C depinde de numărul de nivele de gri, acesta trebuie să rămână în jurul lui 8. C depinde de $P(d, \Theta)$, iar aici cei mai mulți sunt pixelii de intensitate 2 dispusi pe direcția $3\pi/4$.

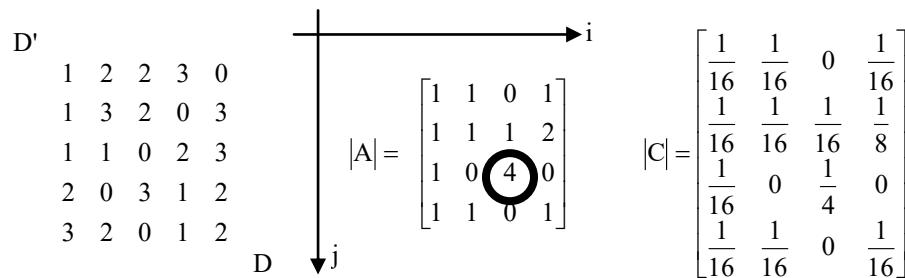


Figura 6.14 Exemplu de contorizare a perechilor

S-au propus și alți descriptori dintre care exemplificăm:

- maximul probabilității: $|c_{ij}|$, detectează motivul cel mai frecvent.
- eterogenitate: $\sum_i \sum_j c_{ij}^2$, o imagine cu textura uniformă are o eterogenitate aproape nulă.
- entropie: $\sum_i \sum_j c_{ij} \log c_{ij}$, măsoară aspectele aleatoare ale texturii.
- momentele diferențelor de gri: $\sum_i \sum_j (i - j)^2 c_{ij}$, acest descriptor scade când valorile mari c_{ij} sunt pe diagonala matricei C .
- contrast: $\sum_{m=0}^{N-1} m^2 \sum_i c_{ij}$ $m = |i - j|$, acest descriptor este o sumă ponderată a probabilităților, valoarea sa este mare în cazul texturilor cu contrast mare pentru că atunci ecarturile mari de nivele de gri duc la ponderi importante.
- descriptorul Gallaway: Se caută alinieri de pixeli cu nivele de gri foarte apropriate, alese într-o bandă foarte îngustă $j = i \pm 1$. Lungimea λ a aliniamentelor este parametrul cel mai important pentru că în general o textură grosieră are aliniamente lungi, iar într-o textură fină aliniamentele de pixeli sunt mai numeroase și mai scurte. Se calculează q numere $P_q(i, j, \Theta)$ pentru lungimi λ_q de nivele de gri apropriate de i în direcția Θ ; $q=1, 2, \dots, L$. K este numărul total de nivele de gri $K=2^2$, dacă sunt m biți/pixel. L este numărul de lungimi λ_q căutate. Dacă banda îngustă presupune chiar $i = j$ se pot defini: $G_1 = \frac{1}{N} \sum_{i=1}^K \sum_{q=1}^L \frac{P_q(i, \theta)}{q^2}$, $G_2 = \frac{1}{N} \sum_{i=1}^K \sum_{q=1}^L q^2 P_q(i, \theta)$, unde $N = \sum_{i=1}^K \sum_{q=1}^L P_q(i, \theta)$. Dacă G_1 este mare textura are un număr important de aliniamente scurte în direcția θ . Dacă G_2 este mare textura se face pe aliniamente lungi.

6.3 Procesare pentru normalizare

În procesul de clasificare/recunoaștere este util ca mărimea (scara), poziția (translația) sau orientarea (rotirea) sub care apar obiectele în imagine să nu infueneze decizia. Unii dintre descriptorii folosiți pot avea intrinsec, din modul cum sunt definiți, această invarianță. În alte cazuri este nevoie de prelucrări speciale asupra acestor descriptori pentru a obține invarianță dorită. În alte aplicații, pe lângă identificarea unui obiect, este nevoie și de localizarea lui și de determinarea atitudinii lui spațiale, atunci informațiile legate de mărime, poziție și orientare devin utile.

Spre exemplu (figura 6.15) dacă apare un caracter alfanumeric presupus într-o reprezentare de imagine binară, alb și negru, prezentat în diferite ipostaze din punct de vedere al mărimii, poziției și orientării. Pentru o aplicație cum ar fi citirea unor texte pe niște etichete de identificare, în toate exemplele din figură trebuie identificat același obiect în fiecare din cele cinci imagini diferite. Ce se întâmplă dacă descriptorul folosit este sensibil la mărimea, poziția sau orientarea obiectului în imagine? În acest caz, în afara celor menționate mai sus și subliniate în cursul prezentării anterioare a diferenților descriptori, o soluție poate fi o prelucrare a imaginii care să o normalizeze în raport cu aceste transformări, anterior extragerii de caracteristici.

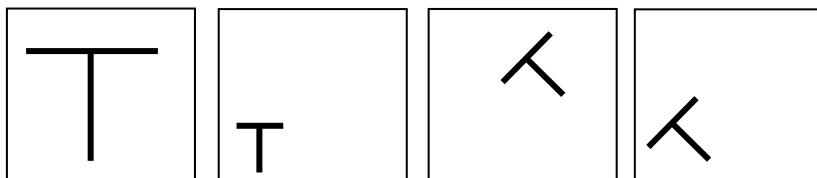


Figura 6.15 Exemple de obiect în poziții, cu orientări și mărimi diferite

Pentru **normalizarea la poziție** în literatura de specialitate se amintește că un mijloc simplu de a normaliza un obiect în legătură cu poziția în interiorul granițelor unei imagini, acela de a redefini axele de coordonate astfel încât să se alinieze extremitățile obiectului la puncte particulare din imagine, spre exemplu la o margine a imaginii (figura 6.16). Problema majoră este atunci "care extremități ?!" pentru că în această fază obiectul nu este identificat. O abordare alternativă care ține mai bine seama de proprietățile obiectului este să se centreze obiectul de interes în matricea imagine. Aceasta metodă presupune calcularea coordonatelor centrului de greutate după formula și translatarea originii axelor de coordonate în acest punct sau într-o poziție relativă la el dar fixă față de acest punct.

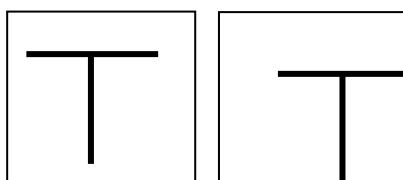


Figura 6.16 Propunere de normalizare în raport cu poziția

Pentru **normalizarea relativă la orientarea obiectului**, dar și pentru a putea găsi o formă care se potrivește cel mai bine cu un model, este utilă și rotirea unui obiect într-o imagine până când acesta este adus într-o poziție prestabilită. Pentru aceasta trebuie recalculate pozițiile pe care le ocupă în imagine punctele de pe obiect. Pentru o rotire a obiectului cu un unghi Θ în sens trigonometric (figura 6.17), se folosesc formulele similare de la grafica pe calculator sau de la robotică:

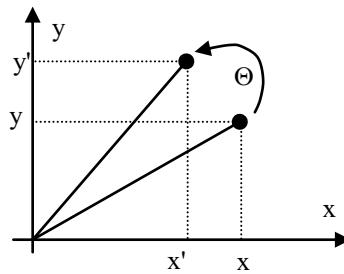


Figura 6.17 Calculul coordonatelor la rotirea unui obiect

$$\begin{cases} x' = x \cos \Theta - y \sin \Theta \\ y' = x \sin \Theta + y \cos \Theta \end{cases} \text{ sau matriceal: } \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \Theta & -\sin \Theta \\ \sin \Theta & \cos \Theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Desigur că problema mai complicată este tocmai determinarea poziției prestabilite amintite și deci a unghiului Θ . În general se recurge la soluții euristică (bazate pe experiență). De exemplu, se rotește obiectul până când devin orizontale sau verticale axa de simetrie maximă sau axa determinată de centrul de greutate și cel mai depărtat punct de pe obiect în raport cu acest centru. Unghiul Θ este tocmai o măsură a poziției inițiale a obiectului.

7. Clasificarea formelor

Metodele uzuale de clasificare fac apel la un număr de atribute globale care permit discriminarea între obiecte și decizia apartenenței lor la diferite clase de obiecte aprioric cunoscute. Când numărul modelelor crește și când erorile care le influențează sporesc în importanță, trebuie căutate alte metode care să folosească eventual fragmente de siluete ale formelor. În acest caz atributele globale pot fi înlocuite de informații oferite de fiecare punct de contur, de exemplu: transformata Hough care cauță segmente de dreaptă, curbe în general în imagini. Atributele sunt înlocuite de primitive locale: vârfuri, colțuri, puncte de curbură singulare sau primitive extinse: părți din curbe sau suprafețe sau forme particulare de contur sau suprafețe.

Dacă ne referim la o imagine 2D sub formă matricială numerică, de dimensiune $N \times N$, atunci procesul de clasificare a formelor poate fi prezentat ca în figura 7.1. Facem precizarea că în aici **forma** este un **vector de caracteristici** și deci recunoașterea de forme (pattern recognition) are un înțeles mai larg decât recunoașterea de figuri provenite din imagini grafice. Confuzia poate fi alimentată de sensurile multiple ale cuvântului forma din limba română.

$x(t)$ este un semnal care aparține la una din cele k clase cunoscute aprioric: C_1, C_2, \dots, C_k . Prin eșantionare se obține un vector numeric: $\{x(m)\} = \{x(1), x(2), \dots, x(N)\}$. Vectorul $z^t = \{z_1, z_2, \dots, z_d\}$ va fi numit modelul sau semnătura obiectului, adică forma. Clasificatorul este un dispozitiv de luat decizii și care a fost antrenat să clasifice un semnal de intrare $x(t)$ ca aparținând la una din clasele C_i , $i = 1, 2, \dots, k$.

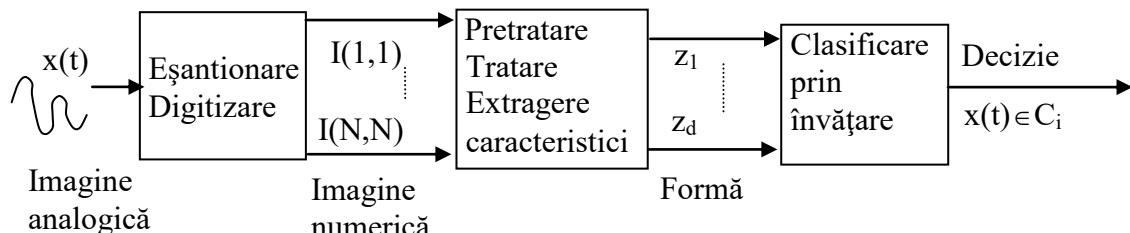


Figura 7.1 Clasificarea formelor provenite din imagini

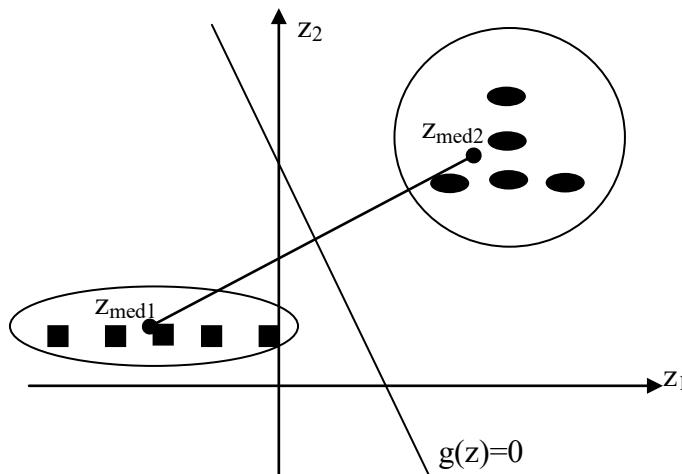


Figura 7.2 Clasificarea pentru două clase de obiecte

7.1 Conceptul de învățare și clasificatori

7.1.1 Clasificatori liniari

Să presupunem că dorim să antrenăm un clasificator astfel încât să fie capabil să clasifice automat un model z care aparține sau la C_1 sau la C_2 . Setul de antrenament (adică o mulțime a cărei clasificare este cunoscută pentru fiecare element) constă în următoarele mulțimi de modele cu două componente: z_{ij} al j -lea element al clasei C_i , $i = 1, 2$. Modelele z_{ij} care aparțin la C_1 sau C_2 sunt localizate în spațiul 2-dimensional al caracteristicilor. Fie z_{med1} și z_{med2} vectorii medii asociati cu cele două clase respective. Deci din figura 7.2 reiese că o graniță rezonabilă de decizie, adică o dreaptă în spațiul 2-dimensional care să separe cele două clase, este mediatoarea segmentului care unește punctele corespunzătoare vectorilor medii introdusi anterior. Considerăm un punct oarecare z pe această graniță de decizie. Dacă ea este mediatoarea segmentului $z_{med1}z_{med2}$ atunci:

$$\|z - z_{med1}\|^2 = \|z - z_{med2}\|^2 \Rightarrow (z_{med1} - z_{med2})^t z = \frac{1}{2} \{\|z_{med1}\|^2 - \|z_{med2}\|^2\}$$

Înlocuind obținem ecuația corespunzătoare graniței de decizie. Cantitatea $\frac{1}{2} \{\|z_{med1}\|^2 - \|z_{med2}\|^2\}$ se numește pragul clasificatorului.

Deci se poate lua următoarea decizie simplă, cu ajutorul funcției discriminant:

- Dacă $g(z) > 0 \Rightarrow z \in C_1$
- Dacă $g(z) < 0 \Rightarrow z \in C_2$

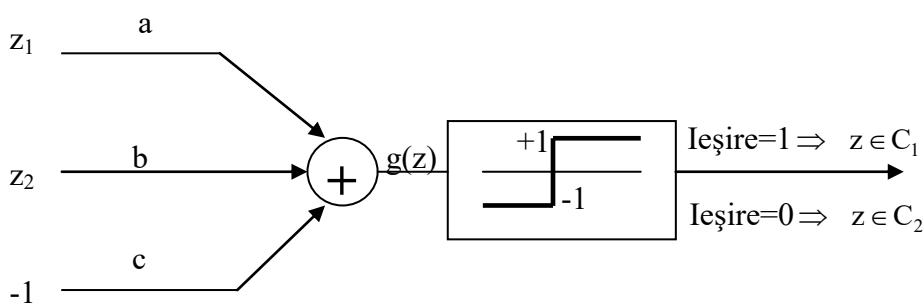


Figura 7.3 TLU pentru 2 clase și 2 caracteristici

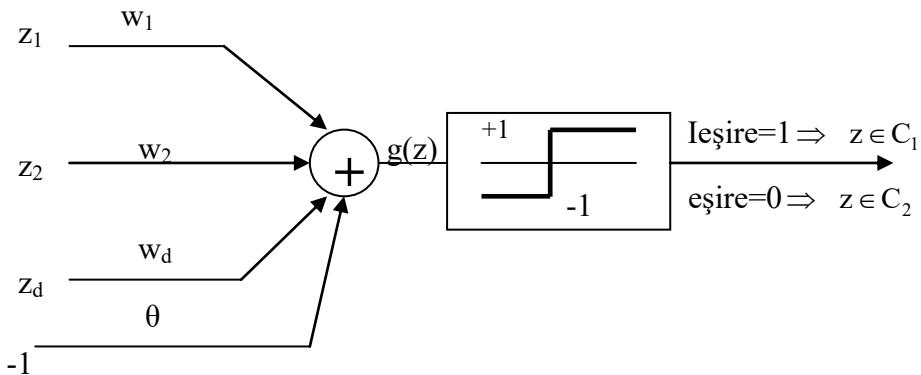


Figura 7.4 TLU pentru 2 clase și d caracteristici

Clasificatorul de mai sus este numit "unitate logică cu prag" (TLU - Threshold Logic Unit) și este liniar (figura 7.3). În concluzie, un set de forme este folosit pentru antrenamentul clasificatorului calculându-se funcția discriminant $g(z)$, după calculul căreia clasificatorul devine capabil să decidă apartenența unei forme la o clasă pe criteriul distanței minime. Se poate face trecerea spre forme caracterizate de d atribute $\mathbf{z}^t = [z_1, z_2 \dots z_d]$ (figura 7.4). Atunci funcția discriminant are forma:

$$g(z) = w_1 z_1 + w_2 z_2 + \dots + w_d z_d - \Theta$$

unde w_i sunt ponderile sau parametrii clasificatorului, Θ este pragul, iar w_i și Θ sunt obținuți din procesul de învățare. În acest caz, granița este un hiperplan $g(z) = 0$ în spațiul d-dimensional al caracteristicilor.

Pentru $d = 3$ hiperplanul se reduce la un plan. Generalizarea poate continua cu clasificarea cu 3 clase de forme (figura 7.5). Clasificatorul de distanță minimă este proiectat să opereze după următoarea regulă de decizie: o formă dată $\mathbf{z} \in C_i$ dacă \mathbf{z} este mai aproape de \mathbf{z}_{medi} , $i=1,2,3$. Fie D_i distanța de la \mathbf{z} la \mathbf{z}_{medi} , $i = 1, 2, 3$ (figura 7.6). Atunci avem: $D_i^2 = \|\mathbf{z} - \mathbf{z}_{\text{medi}}\|^2 = (\mathbf{z} - \mathbf{z}_{\text{medi}})^t (\mathbf{z} - \mathbf{z}_{\text{medi}})$, $i=1,2,3$. Simplificând: $D_i^2 = \|\mathbf{z}\|^2 - 2 \{ \mathbf{z}_{\text{medi}}^t \mathbf{z} - \frac{1}{2} \|\mathbf{z}_{\text{medi}}\|^2 \}$ $i=1,2,3$. D_i^2 este minimă dacă valoarea care se scade, aceea dintre accolade, este maximă. Deci clasificatorul este descris de funcția: $2 \{ \mathbf{z}_{\text{medi}}^t \mathbf{z} - \frac{1}{2} \|\mathbf{z}_{\text{medi}}\|^2 \}$, $i=1, 2, 3$.

Următorul pas este clasificarea a 3 clase cunoscute, dar cu d caracteristici. Iar cel mai general k clase cu d caracteristici (figura 7.7).

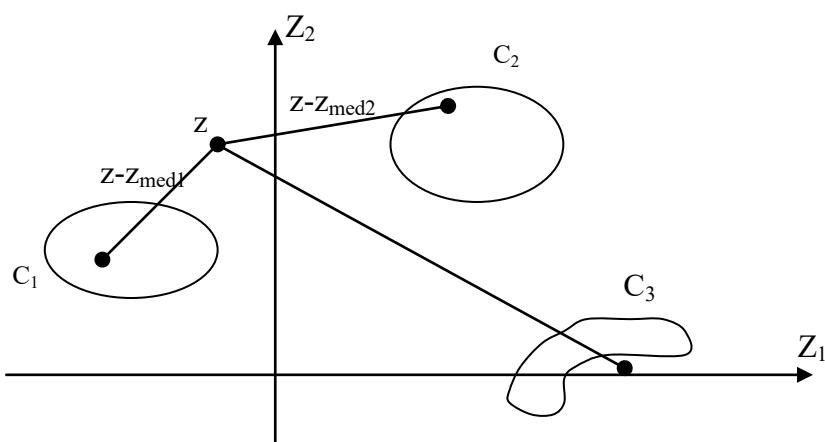


Figura 7.5 Clasificarea pentru 3 clase

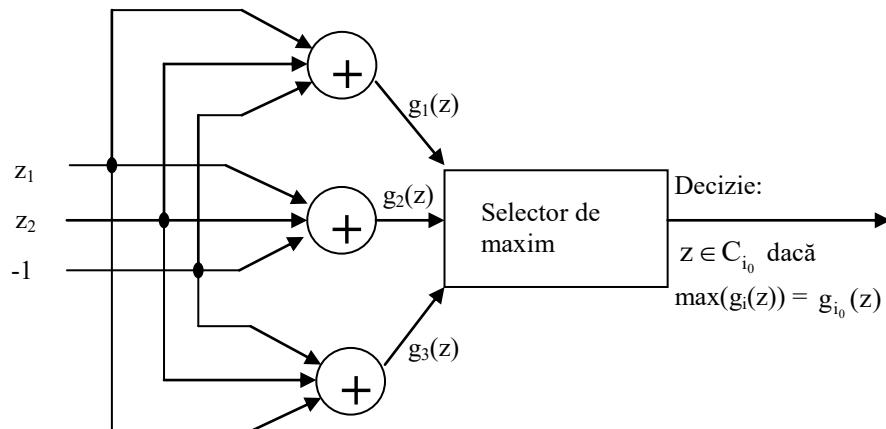
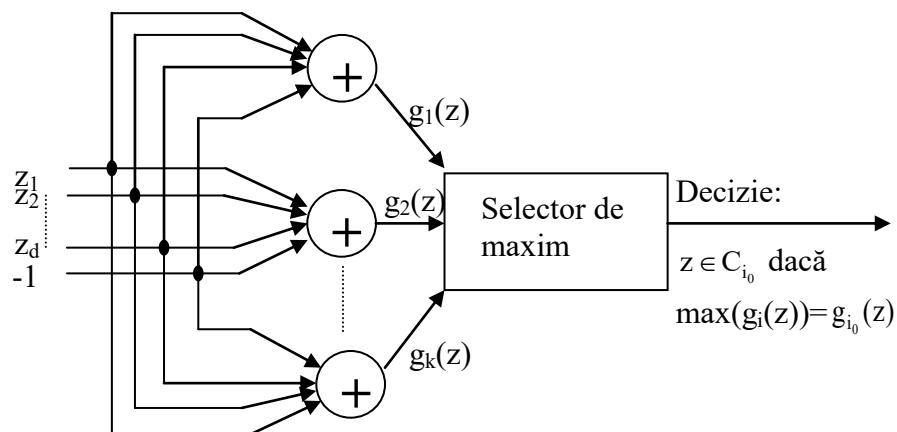


Figura 7.6 TLU pentru 3 clase și 2 caracteristici

Figura 7.7 TLU pentru k clase și d caracteristici

Acste metode de clasificare pot fi implementate și testate practic într-o aplicație exemplificatoare aşa cum este precizat succint în continuare. Clasele C_0, C_1, \dots, C_9 ($k = 10$) să corespundă cu cele 10 cifre din notația arabă: 0, 1, 2 ... 9. Caracterele grafice corespunzătoare cifrelor pot fi digitizate pe 64×64 pixeli. Drept caracteristici se pot alege coeficienții unei transformări ortogonale (TFD, TWH...), în număr de 16, $d = 16$, pentru funcția cumulativă de schimbare a direcției pe contur. Mulțimea formelor de antrenament poate cuprinde, de exemplu, 100 de caractere grafice pentru fiecare clasă. Rezultatele obținute în procesul de clasificare sunt cu câteva procente mai slabe decât cele obținute pe mulțimea formelor de antrenament, deci în procesul de învățare. Numărul caracteristicilor necesare pentru obținerea unei bune clasificări crește în ordine pentru TFD, TWH, HT, IT, (TWD - Transformarea Walsh-Hadamard, HT-Transformare Haar, IT - Transformarea Identică). Se poate repeta experimentul pentru alte tipuri de caracteristici, dar pentru aceleași clase de forme: histograma intersecției cu un set de drepte aleatoare, semnătura polară cu pas unghiular constant, momente.

Curs 13

7.1.2 Tehnica celor mai mici pătrate

În prezentarea clasificatorilor de distanță minimă s-a presupus că inevitabil clasele de forme se grupează în jurul valorilor medii, formând clustere în spațiul caracteristicilor z_{medi} , $i = 1, 2, \dots, k$. În multe aplicații, această presupunere nu poate fi considerată rezonabilă. În astfel de situații, clasificatorul trebuie mai întâi să mapeze formele într-un spațiu de decizie în care formele C_i sunt făcute să se grupeze în jurul unor puncte prestabile v_i , $i=1, 2, \dots, k$.

Transformarea matriceală A care oferă această posibilitate de trecere din spațiul caracteristicilor la spațiul de decizie, este în general aleasa astfel încât abaterea medie pătratică totală să fie minimă. Pentru a clasifica o formă dată, ea este mai întâi mapată într-un spațiu de decizie și atunci clasificată ca aparținând la C_{i_0} dacă este mai apropiată de v_{i_0} . Acest tip de clasificator va fi numit clasificator de distanță minimă folosind cele mai mici pătrate.

Considerăm o mulțime de dimensiune N_i de forme z_{ij} , $j = 1, 2, \dots, N_i$ care trebuie mapate în punctul specificat din spațiul k -dimensional, notat $v_i^t = [v_1, v_2, \dots, v_k]$. Căutăm transformarea A care mapează $\{z_{ij}\}$ în v_i astfel încât eroarea medie pătratică totală introdusă de mapare să fie minimă. Să presupunem că mappingul lui z_{ij} este în realitate L_{ij} . Vectorul corespunzător de eroare este: $\varepsilon_{ij} = L_{ij} - v_i = A \cdot z_{ij} - v_i$. Eroarea pătratică medie totală la maparea tuturor punctelor z_{ij} în sau mai

precis în jurul lui v_i , este: $\varepsilon = \frac{1}{N_i} \sum_{j=1}^{N_i} \|\varepsilon_{ij}\|^2$. Folosind un calcul matriceal bazat pe reguli de următorul tip:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \nabla_A P = \begin{bmatrix} \frac{\delta P}{\delta a_{11}} & \frac{\delta P}{\delta a_{12}} \\ \frac{\delta P}{\delta a_{21}} & \frac{\delta P}{\delta a_{22}} \end{bmatrix}$$

$$\nabla_A (x^t \cdot A^t \cdot A \cdot x) = 2A \cdot (x \cdot x^t) \quad \nabla_A (x^t \cdot A^t \cdot v) = v \cdot x^t \quad \nabla_A (v^t \cdot v) = 0$$

unde ∇_A operator de derivare, x, v vectori, iar A matrice.

A trebuie ales astfel încât eroarea să fie minimă deci $\nabla_A \varepsilon = 0$.

$$\varepsilon = \frac{1}{N_i} \sum_{j=1}^{N_i} \{z_{ij}^t \cdot A^t \cdot A \cdot z_{ij} - z_{ij}^t \cdot A^t \cdot v_i + \|v_i\|^2\}$$

$$\frac{1}{N_i} \sum_{j=1}^{N_i} \Delta_A (z_{ij}^t \cdot A^t \cdot Az_{ij}) - \frac{2}{N_i} \sum_{j=1}^{N_i} \Delta_A (z_{ij}^t \cdot A^t \cdot v_i) + \Delta_A (\|v_i\|^2) = 0$$

$$\frac{1}{N_i} \sum_{j=1}^{N_i} 2A(z_{ij} \cdot z_{ij}^t) - \frac{2}{N_i} \sum_{j=1}^{N_i} v_i \cdot z_{ij}^t = 0$$

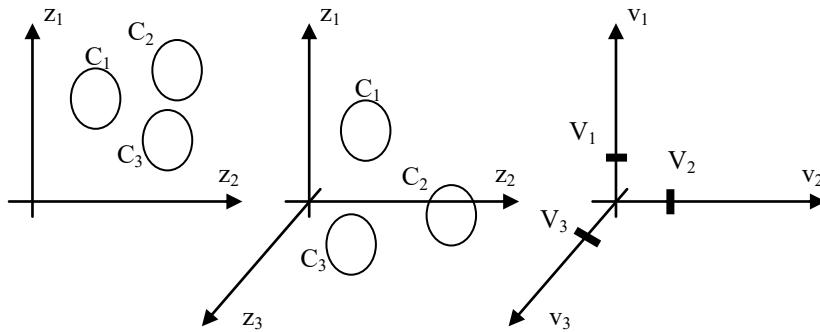


Figura 7.8 Maparea claselor în tehnica celor mai mici pătrate

De unde se poate calcula A după formula: $A = \left[\sum_{j=1}^{N_i} v_i \cdot z_{ij}^t \right] \left[\sum_{j=1}^{N_i} z_{ij} \cdot z_{ij}^t \right]^{-1}$

Noțiunea de spațiu majorat al caracteristicilor este convenabilă pentru dezvoltarea unui clasificator de distanță prin metoda celor mai mici pătrate. Din definirea funcției discriminant se poate scrie: $g(z) = w^t \cdot \hat{z}$, unde $w^t = [w_1, w_2, \dots, w_d, \theta]$, $\hat{z}^t = [z_1, z_2, \dots, z_d, -1] = [z^t, -1]$. Se observă că \hat{z} este ușor de obținut dintr-o formă dată z prin adăugarea unei componente, (-1). Spațiul $(d + 1)$ -dimensional al formelor \hat{z} se numește spațiu majorat al caracteristicilor.

Dacă pentru trei clase C_i , $i = 1, 2, 3$, folosim exemplul anterior și adăugăm (-1) așa cum am arătat mai sus atunci rezultatul se poate ilustra grafic ca în figura 7.8.

$$\hat{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \quad v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Căutăm o transformare care să mapeze $\hat{z} \in C_k$, $k = 1, 2, 3$, în v_k , $k = 1, 2, 3$, în sensul celor mai mici pătrate. Alegem exact versorii celor trei axe ca în figura 7.8. Fie probabilitățile apriorice P_k corespunzătoare pentru C_k , $k = 1, 2, 3$. După metoda pe care am mai folosit-o anterior căutăm A:

$$A \left[\frac{P_1}{N_1} \sum_{j=1}^{N_1} \hat{z}_{1j} \cdot \hat{z}_{1j}^t + \frac{P_2}{N_2} \sum_{j=1}^{N_2} \hat{z}_{2j} \cdot \hat{z}_{2j}^t + \frac{P_3}{N_3} \sum_{j=1}^{N_3} \hat{z}_{3j} \cdot \hat{z}_{3j}^t \right] = \left[\frac{P_1}{N_1} \sum_{j=1}^{N_1} v_1 \cdot \hat{z}_{1j}^t + \frac{P_2}{N_2} \sum_{j=1}^{N_2} v_2 \cdot \hat{z}_{2j}^t + \frac{P_3}{N_3} \sum_{j=1}^{N_3} v_3 \cdot \hat{z}_{3j}^t \right]$$

Calculând pe A găsim: $A = S_{\hat{z}\hat{z}}^{-1} \cdot S_{v\hat{z}}$, unde $S_{v\hat{z}} = \sum_{i=1}^3 \sum_{j=1}^{N_i} \frac{P_i}{N_i} (v_i \cdot \hat{z}_{ij}^t)$ este matricea de corelație încrucișată iar $S_{\hat{z}\hat{z}} = \sum_{i=1}^3 \sum_{j=1}^{N_i} \frac{P_i}{N_i} (\hat{z}_{ij} \cdot \hat{z}_{ij}^t)$ este matricea de autocorelație, iar A este o matrice de dimensiune $[3 \times (d+1)]$.

Pentru a clasifica o formă $z^t = [z_1, z_2, \dots, z_d, -1]$ clasificatorul calculează prima dată $L = A \cdot \hat{z}$ și deci mapează \hat{z} în spațiul deciziilor. În continuare se folosește următoarea regulă de decizie folosind distanță minimă: dacă L este apropiat de v_{i_0} , atunci z este clasificat în C_{i_0} .

Acum distanța pe care clasificatorul trebuie să o calculeze este: $D_i^2 = \|L - V_i\|^2$, $i = 1, 2, 3$. $D_i^2 = \|L\|^2 - 2V_i^T \cdot L + \|V_i\|^2$. Dar am ales $\|V_i\|^2 = 1$ și deci D_i^2 este minim când $V_i^T L$ este maxim. Este preferabil să calculăm în loc de D_i^2 pe $d_i = V_i^T L$, $i = 1, 2, 3$. Înlocuind mai departe:

$$V_1^T = [1 \ 0 \ 0], \ V_2^T = [0 \ 1 \ 0], \ V_3^T = [0 \ 0 \ 1], \ L = A\hat{z} \Rightarrow$$

$$\begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot A \cdot \hat{z} = A \cdot \hat{z}$$

Dacă $A = [a_{ij}]$, $i = 1, 2, 3$. și $j = 1, 2, \dots, d+1$. Atunci

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1d} & \theta_1 \\ a_{21} & a_{22} & \dots & a_{2d} & \theta_2 \\ a_{31} & a_{32} & \dots & a_{3d} & \theta_3 \end{bmatrix}$$

Deci d este funcție discriminant care specifică clasificatorul. Este clar acum că specifiicatorul este complet definit de matricea A care este obținută din mulțimea formelor de antrenament. Pentru a clasifica o anumită formă, clasificatorul calculează trei numere $g_1(z)$, $g_2(z)$, $g_3(z)$. Dacă $\max_{i=1,2,3} (g_i(z)) = g_{i_0}$ atunci forma este asignată clasei C_{i_0} (figura 7.7). Rezultatele clasificării chiar pentru mulțimea de antrenament pot fi prezentate sub forma unei matrice de confuzie: elementele din afara diagonalei reprezintă numărul de erori în timpul clasificării. f_{ij} , $i \neq j$, elementul din linia i și coloana j , este numărul de forme din C_i atribuite greșit la C_j .

Orice clasificator care implementează granițe liniare este împins la erori în cazul în care clasele de forme nu sunt liniar separabile. Pentru k clase trecerea este directă. Formele C_i sunt mapate pe versorii din spațiul k -dimensional v_i , $V_i^T = [00\dots 1\dots 00]$ unde 1 se află în poziția a i a, $i = 1, 2, \dots, k$. Vom obține o matrice A , de dimensiune $k \times (d+1)$, și funcția discriminant a clasificatorului în mod absolut similar cu aceea din cazul cu 3 clase, cu singura deosebire că la limitele superioare de sumare se înlocuiește 3 cu k .

Procedura de învățare poate fi ușor extinsă la implementarea granițelor cuadratici. Reținem totuși ca alternativă și folosirea granițelor liniare pe bucăți (piece wise) cu rezultate mai bune decât cele liniare pur și simplu. Pentru granițe cuadratici clasificatorul va fi în schimb mai complicat. Funcția discriminant care specifică un clasificator quadratic este:

$$g_i(z) = \sum_{j=1}^d w_{jj} \cdot z_j^2 + \sum_{j=1}^{d-1} \sum_{k=j+1}^d w_{jk} \cdot z_j \cdot z_k + \sum_{j=1}^d w_j \cdot z_j - \theta_i$$

Se observă prezența termenilor pătratici; $g_i(z)$ are $\frac{(d+1)(+2)}{2}$ ponderi după cum urmează: d coeficienți pentru z_j^2 adică w_{jj} , d coeficienți pentru z_j adică w_j , $\frac{d(d-1)}{2}$ coeficienți pentru $z_j z_k$ adică w_{jk} , un termen care este valoarea de prag θ_i , și care nu este coeficient. Acești parametri sunt obținuți în procesul de învățare. Pentru a explica implementarea lor definim mai întâi vectorul G , q -dimensional, cu componente f_1, f_2, \dots, f_q funcții de z_i , $i = 1, 2, \dots, d$. Primele d componente ale lui

G sunt $z_1^2, z_2^2, \dots, z_d^2$, următoarele $\frac{d(d-1)}{2}$ componente sunt $z_1z_2, z_1z_3, \dots, z_{d-1}z_d$, iar ultimele d componente sunt z_1, z_2, \dots, z_d . Numărul lor total este $\frac{d(d+3)}{2}$. Notăm această corespondență cu $G = G(z)$, unde $G(z)$ este o transformare unu la unu. Deci pentru fiecare formă $z^t = [z_1, z_2, \dots, z_d]$ din spațiul d -dimensional există un unic vector $G^t = [f_1, f_2, \dots, f_q]$ în spațiul q -dimensional. Această corespondență biunivocă face posibilă scrierea lui $g_i(z)$ ca o funcție liniară de componentele lui G cu rezultatul că fiecare funcție discriminant quadratică a lui z_i există o funcție discriminant liniară a lui G (figura 7.9).

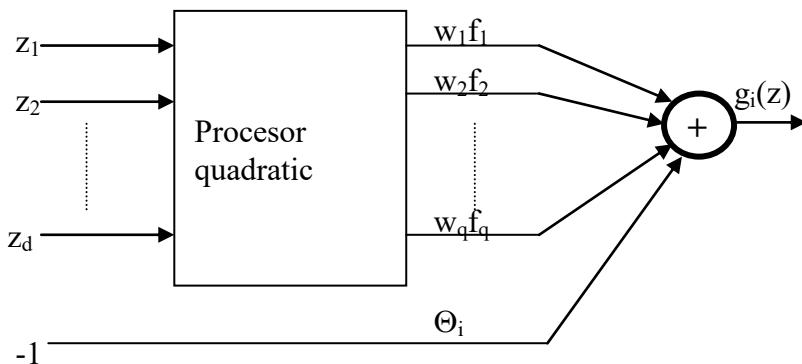


Figura 7.9 Clasificator quadratic

$$g_i(z) = \bar{w}_1 f_1 + \bar{w}_2 f_2 + \dots + \bar{w}_q f_q - \theta_i$$

O altă metodă de clasificare a formelor în clase cunoscute, presupune evaluarea distanțelor, în spațiul caracteristicilor, până la toate punctele care corespund formelor de antrenament și alegerea a k dintre vecini și anume cei mai apropiati. Apartenența acestora la clasele de forme (sau a majorității lor) implică apartenența formei de clasificat.

Metodele anterioare sunt identificate prin sigla 1 - NN (1 - Nearest Neighbor) pentru că se caută un prototip, cel mai apropiat (prototipul este z_{med}); iar metoda amintită ulterior prin sigla K-NN (K - Nearest Neighbor).

7.2 Clasificări statistice

În faza de instruire, clasificatorul asociază fiecărei forme S_j un vector de atrbute, de caracteristici, (pattern) X_j . Perechile (S_j, X_j) , sau modelele, sunt afectate de imperfecțiunile de la segmentare, celor datorate zgomotelor sistemului de achiziție a imaginilor, condițiilor de iluminare, erorilor de calcul, imperfecțiunilor de normalizare etc.. Apare deci o dispersie a atrbutelor. Mai mult, aceste atrbute diferă și datorită micilor deosebiri dintre formele aceleiași clase. În aceste condiții, devine utilă o abordare statistică a clasificării formelor.

Se definesc atunci clasele C_j , $j = 1, 2, \dots, N$, ca ansambluri de obiecte similare cu vectorii de atrbute grupați în mulțimea Y . Când variația vectorilor X_j are o dispersie mare, se face apel la calculul cu probabilități. Fie:

- $p(C_j)$ probabilitatea de apariție a clasei C_j .
- $p(X)$ funcția densitate de probabilitate de apariție a vectorului X .

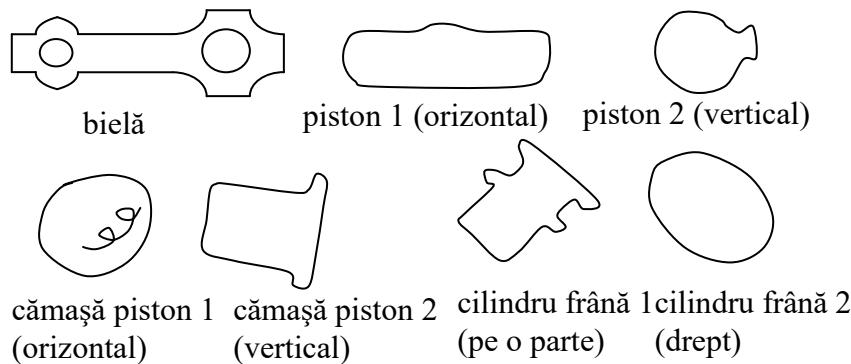


Figura 7.10 Exemple de clase de obiecte

- $p(X/C_j)$ funcția densitate de probabilitate condiționată de apariție a lui X știind că X îi corespunde lui C_j .
- $p(C_j/X)$ funcția densitate de probabilitate aposteriori de apariție a clasei C_j știind valoarea vectorului X (probabilitatea ca X să corespundă respectivei clase)

La Stanford Research Institute, cu mulți ani în urmă, a fost implementată o aplicație industrială, devenită clasică, în care apar piese de patru tipuri: biele, pistoane, cămași de pistoane și cilindri de frână (figura 7.10). Datorită faptului că pozițiile în care pot apărea aceste piese pe o bandă transportoare sunt diferite, (de exemplu un piston poate apărea vertical sau orizontal în raport cu banda transportoare) se definesc șapte clase de forme. Presupunem că există posibilitatea de a ști de unde provin aceste piese și deci să prediciționăm numărul relativ în care ele pot apărea (dacă există șanse ca vreuna din clase să apară mai frecvent decât celelalte). Putem presupune că datorită numărului de componente necesare uzual pentru o mașină, (câte patru biele, pistoane, cămași de cilindru și cilindri de frână) probabilitățile de apariție ale claselor de obiecte sunt egale. În schimb pentru clasele de forme avute în vedere putem face o diferențiere în sensul că, de exemplu, poziția orizontală pentru un piston pe banda transportoare este mult mai probabilă. Experimentele probabilistice pot impune o alegere de următorul tip: $p(C_1) = 25\%$, $p(C_2) = 20\%$; $p(C_3) = 5\%$, $p(C_4) = 5\%$; $p(C_5) = 20\%$, $p(C_6) = 5\%$; $p(C_7) = 20\%$.

Pentru alegerea caracteristicilor claselor de forme, trebuie parcursă o etapă preliminară de apreciere a puterii de disociere pe care respectiva caracteristică o are în raport cu clasele analizate. Dacă se dorește luarea deciziei pe baza unui arbore de decizii, atunci această etapă presupune următorii pași:

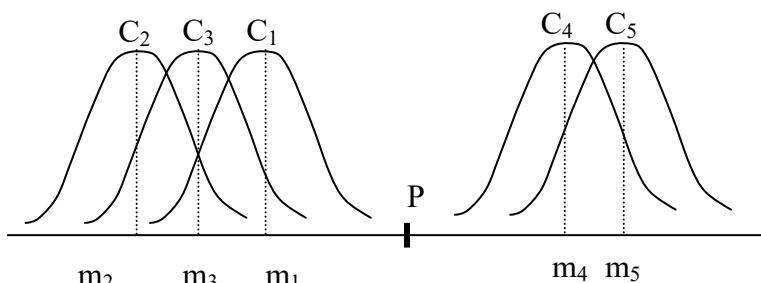


Figura 7.11 Selecția statistică a caracteristicilor

- se măsoară distribuțiile de probabilitate pentru caracteristicile X_1, X_2, \dots folosind o mulțime de forme de antrenament cât mai mare pentru fiecare clasă de forme.
- se calculează valorile medii $m_i(j)$ pentru cea de a i-a caracteristică X_i pentru obiectele din cea de a j-a clasă de obiecte.
- pentru fiecare caracteristică X_i se determină un prag de decizie P în mijlocul intervalului care separă toate obiectele în două grupe distincte (de exemplu ca în figura 7.11 unde dacă valoarea

măsurată pentru X_i este mai mare ca pragul, atunci obiectul este mai degrabă din C_4 sau C_5 decât din C_1, C_2 sau C_3 .

- se ordonează caracteristicile după mărimea intervalului de separare care l-a dat, în fiecare caz, pe P. Cu cât intervalul de separare este mai mare, cu atât încrederea în decizia care se bazează pe acest prag este mai importantă.

- se poate face o selecție a caracteristicilor utilizabile după mărimea intervalului de separare (se aleg numai acele caracteristici cu un astfel de interval suficient de mare).

Dacă nu se dorește implementarea unui clasificator Bayes, atunci probabilitățile $p(X_i)$ sunt necesare pentru a determina prin calcul probabilitatea $p(C_i/X)$ folosită în decizia de clasificare. Considerăm totuși că o etapă intermediară de tipul celei descrise anterior pentru varianta cu arbore de decizie, rămâne utilă chiar și în acest caz, pentru a putea rejecta caracteristicile care nu disting între diferitele clase de forme.

Algoritmul de decizie poate cuprinde atunci următorii pași:

- evaluarea probabilităților condiționate în raport cu o caracteristică măsurabilă X .
- obținerea valorii măsurate pentru aceeași caracteristică dar pentru obiectul de clasificat.
- dându-se X se determină probabilitatea ca forma să aparțină la una din clasele de obiecte considerate.
- repetă pasul anterior pentru toate clasele.
- decide că X este din C_i după regula: $p(C_i/X) > p(C_j/X) \quad \forall i \neq j$

Pentru determinarea practică a lui $p(C_i/X)$ ar trebui să se facă un foarte mare număr de clasificări de antrenament anterioare, pe forme cunoscute, în condiții greu de respectat (de exemplu să se prezinte clasificatorului un număr mare de forme de diferite tipuri, toate cu aceeași valoare măsurată pentru o caracteristică și să se măsoare probabilitatea de apartenență la o clasă).

Se poate folosi teorema lui Bayes: $p(C_i/X) = \frac{p(X/C_i) \cdot p(C_i)}{p(X)}$, unde $p(X) = \sum_i p(X/C_i) \cdot p(C_i)$ poate fi ignorat pentru că nu influențează valorile probabilităților relative (este un factor de scară). Deci regula de decizie devine: $p(X/C_i) \cdot p(C_i) > p(X/C_j) \cdot p(C_j), \forall i \neq j \Rightarrow X \in C_i$.

Clasificantul Bayes prezentat mai sus este folosit ca referință în aprecierea altor clasificatori (deși rămâne o aproximare a clasificatorului ideal). Dacă un obiect dintr-o clasă C_i a fost clasificat eronat ca aparținând la o clasă C_j , atunci această eroare implică un cost asociat L_{ij} . În inspecția automată printr-un proces de fabricație, putem aprecia că acceptarea unui rebut are un cost mult mai mare decât respingerea eronată a unui produs bun. De exemplu, dacă se inspectează un traseu electric pe o placă cu componente electronice, este mai mic costul semnalării unui defect inexistent (se pierde din greșală o placă) decât cel al nesenzării defectului (respectivul defect poate produce defectarea și altor componente cu care placă se couplează electric). Dar, se poate considera pentru alte aplicații

varianta simplificatoare (unde δ_{ij} este simbolul lui Kronecker): $L_{ij} = \begin{cases} 0 & i = j = 1 - \delta_{ij} \\ 1 & i \neq j \end{cases}$

Cum X poate să aparțină în principiu oricareia din clasele cunoscute, fiecare din clase implicând un risc de clasificare greșită, atunci riscul condițional asociat unei decizii X din C_j este calculabil după cum vom arata în continuare. Se definește riscul de asignare a lui X la clasa C_j , riscul mediu condițional, astfel: $r_j(X) = \sum_{i=1}^M L_{ij} \cdot p(C_i/X)$, unde M este numărul de clase de obiecte considerate.

Clasificatorul calculează $r_1(X), r_2(X), \dots$ și alege varianta având riscul cel mai mic. Utilizând

formula lui Bayes: $p(C_i/X) \cdot p(X) = p(C_i) \cdot p(X/C_i)$, atunci formula riscului se transformă în:

$$r_j(X) = \frac{1}{p(X)} \sum_{i=1}^M L_{ij}(X/C_i) \cdot p(C_i), \text{ unde din motivele amintite mai sus } p(X) \text{ poate fi neglijat:}$$

$$r_j(X) = \sum_{i=1}^M L_{ij} p(X/C_i) \cdot p(C_i) - \text{care este formula clasificatorului Bayes de risc minim.}$$

Dacă $M = 2$ distingem cazurile următoare:

- pentru strategia 1, adică dacă decidem că X este din C_1 , riscul este:

$$r_1(X) = L_{11} \cdot p(X/C_1) \cdot p(C_1) + L_{21} \cdot p(X/C_2) \cdot p(C_2)$$

- pentru strategia 2, adică dacă decidem că X este din C_2 , riscul este:

$$r_2(X) = L_{12} \cdot p(X/C_1) \cdot p(C_1) + L_{22} \cdot p(X/C_2) \cdot p(C_2)$$

Dacă $r_1(X) < r_2(X)$ atunci X este asignat la clasa C_1 .

Se poate scrie în continuare:

$$(L_{21} - L_{22}) \cdot p(X/C_2) \cdot p(C_2) < (L_{12} - L_{11}) \cdot p(X/C_1) \cdot p(C_1)$$

$$\frac{p(X/C_1)}{p(X/C_2)} > \frac{p(C_2) \cdot (L_{21} - L_{22})}{p(C_1) \cdot (L_{12} - L_{11})}$$

pentru că în general se admite că $L_{ij} > L_{ii}$ atunci $L_{12}(X) = \frac{p(X/C_1)}{p(X/C_2)}$ se numește raport de adevărată

asemănare. Pragul $S_{12} = \frac{p(C_2) \cdot (L_{21} - L_{22})}{p(C_1) \cdot (L_{12} - L_{11})}$ permite luarea deciziilor de tipul:

- $L_{12}(X) > S_{12} \Rightarrow X \in C_1$
- $L_{12}(X) < S_{12} \Rightarrow X \in C_2$

Extinderea pentru mai multe clase se face pornind de la ideea că $L_{ii} = 0$, riscul pentru decizie corectă, și L_{ij} ia o valoare unică în caz de decizie incorrectă, iar riscul deciziilor incorecte e normalizat la 1.

$$\Rightarrow L_{ij} = 1 - \delta_{ij} \quad \delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

Atunci: $r_j(X) = \sum_{i=1}^M (1 - \delta_{ij}) \cdot p(X/C_i) p(C_i) = p(X) - p(X/C_j) \cdot p(C_j)$, X este asignat la clasa C_i dacă este îndeplinită condiția:

$$\begin{aligned} p(X) - p(X/C_i) \cdot p(C_i) &< p(X) - p(X/C_j) \cdot p(C_j) \\ p(X/C_i) \cdot p(C_i) &> p(X/C_j) \cdot p(C_j) \end{aligned}$$

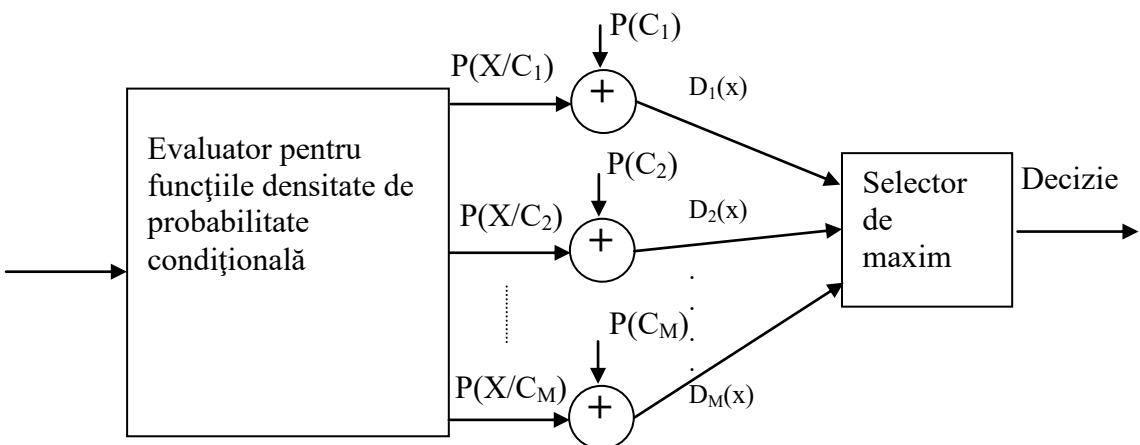


Figura 7.12 Clasificator statistic

Deci sarcina de calcul presupune găsirea valorilor $D_j(X) = p(X/C_j) \cdot p(C_j)$ și alegerea $X \in C_i$ dacă $D_i(X) > D_j(X) \quad \forall i, j = 1, 2, \dots, M; i \neq j$. Utilizarea clasificatorului Bayes (risc minim la decizie corectă și risc egal pentru toate deciziile incorecte) este foarte frecventă, dar se bazează pe cunoașterea probabilităților apriorice pentru fiecare clasă și a costurilor deciziilor (figura 7.12). Să presupunem că $D_1(X)$ este maximul și $D_k(X)$ valoarea cea mai mare următoare. Dacă diferența dintre cele două valori este relativ mare, atunci putem avea încredere în decizia luată, dacă această diferență este relativ mică (de exemplu sub 5%) încrederea în decizie scade foarte mult. Se introduce noțiunea de marjă de rejecție sau nivel de încredere, care se calculează ca diferența prezentată anterior și numai depășirea unui anumit prag pentru această valoare permite luarea deciziei. Un prag se poate impune nu numai la diferența între maxime, dar de asemenea și la valorile absolute ale maximului. Se poate condiționa decizia și de faptul că valoarea maximului depășește un prag impus. Dacă determinarea probabilităților cerute nu se poate face convenabil, ca timp sau performanțe, printr-o fază experimentală, atunci se poate pleca de la o distribuție de probabilități aprioric cunoscută și definită analitic.