UNIVERSITATEA "POLITEHNICA" DIN TIMIŞOARA
FACULTATEA DE ELECTRONICĂ ŞI TELECOMUNICATII

CONTRIBUŢII LA ESTIMAREA MIŞCĂRII BAZATĂ PE REGIUNI PENTRU COMPRESIA SECVENŢELOR DE IMAGINI

BIBLIOTECA CENTRALĂ
UNIVERSITATEA "POLITEHNICA"
TIMIŞOARA

TEZĂ DE DOCTORAT

Conducător ştiinţific,
Prof.dr.ing. Corneliu I. TOMA

Doctorand,
Ing. ALEXA Florin

1999

BUPT

CUVÂNT ÎNAINTE

În aceste rânduri de început aş dori să exprim mulţumirile mele domnului prof. dr. ing. Corneliu I. TOMA, conducătorul ştiinţific al tezei, pentru sprijinul şi îndrumarea constantă şi mai ales pentru felul în care m-a încurajat în momentele dificile de căutări ce au apărut pe tot parcursul realizării acesteia. Discuţiile şi încurajările primite m-au făcut să înţeleg că întotdeauna există posibilitatea perfecţionării şi să conştientizez importanţa prezentării rezultatelor cercetărilor într-o formă cât mai accesibilă. Pot să afirm aici cu toată convingerea că această lucrare nu ar fi fost posibilă fără sprijinul oferit de Domnia sa.

Sincere mulţumiri domnului profesor Zoran BOJKOVIĆ, de la Universitatea din Belgrad, pentru sprijinul şi încrederea acordată. Schimbul de idei şi sugestiile privind organizarea tezei precum şi a rezultatelor obţinute au condus la accelerarea ritmului de prelucrare a datelor necesare prezentei teze de doctorat. Încrederea acordată s-a materializat în publicarea împreună a trei lucrări ştiinţifice la manifestări prestigioase în domeniul prelucrării imaginilor. Sugestiile oferite pe parcursul elaborării tezei mi-au fost de un mare folos, contribuind la îmbunătăţirea variantei finale a acesteia.

Mulţumesc domnului prof.dr.ing. Mircea PETRESCU, de la Universitatea "Politehnica" din Bucureşti, pentru ajutorul acordat şi pentru acceptul de a deveni referent al tezei.

Mulţumiri speciale domnilor prof. dr. ing. Ioan NAFORNIŢĂ şi prof. dr. ing. Alimpie IGNEA, pentru sugestiile oferite pe parcursul elaborării acestei teze, pentru consecvenţa cu care au urmărit finalizarea acesteia, precum şi pentru sprijinul acordat în activitatea de zi cu zi în cadrul colectivului Departamentului de Comunicaţii, Facultatea de Electronică şi Telecomunicaţii.

Sincere mulţumiri domnului conf. dr. ing. Vasile GUI pentru faptul că mi-a deschis drumul spre fascinatul domeniu al prelucrărilor de imagini şi pentru sugestiile valoroase şi extrem de utile privitoare la conţinutul tezei. Discuţiile purtate pe parcursul elaborării tezei mi-au confirmat unele din rezultatele obţinute şi au contribuit hotărâtor la tema abordată.

O listă lungă de persoane au avut o contribuţie mai mare sau mai mică la finalizarea acestei lucrări, şi de aceea pentru a nu nedreptăţi pe cineva, voi mulţumi aici tuturor colegilor din Departamentul de Comunicaţii şi Facultatea de

BUPT

Electronică şi Telecomunicaţii care m-au ajutat prin crearea unei atmosfere colegiale de lucru şi pentru încurajările permanente.

O menţiune specială revine, de asemenea, familiei mele, care a ştiut să înţeleagă importanţa acestui moment, a avut încredere în mine şi mi-a acordat tot sprijinul necesar.

Timişoara,

Decembrie 1998

BUPT

CUPRINS

1. INTRODUCERE **9**

2. ESTIMAREA MIŞCĂRII ÎN SECVENŢELE DE IMAGINI **14**

2.1 Introducere în analiza mişcării 14
2.2 Percepţia şi modelarea secvenţelor de imagini 17
2.2.1 Percepţia secvenţelor de imagini 17
2.2.2 Modelarea mişcării în vederea analizei secvenţelor de imagini 19
2.2.2.1 Mişcarea corpului rigid 20
2.2.2.2 Modelarea mişcării suprafeţelor plane 22
2.2.2.3 Alegerea celui mai adecvat model 24
2.3 Estimarea mişcării 25
2.3.1 Tehnici de estimare bazate pe gradient 26
2.3.2 Tehnici de estimare bazate pe algoritmi recursivi pe pixel 34
2.3.3 Algoritmi de corespondenţă a blocurilor 38
2.3.3.1 Căutarea logaritmică bidimensională 41
2.3.3.2 Căutare prin creşterea preciziei 41
2.3.3.3 Căutarea în direcţii conjugate 42
2.3.3.4 Un nou algoritm de căutare rapidă în patru paşi 42
2.3.4 Tehnici de estimare bazate pe reprezentarea în frecvenţă 47
2.3.4.1 Intercorelaţia în domeniul transformat 47
2.3.4.2 Corespondenţa fazei 48
2.3.5 Estimarea mişcării în prezenţa discontinuităţilor 49
2.3.6 Metode de estimare ierarhice 51
2.3.7 Algoritmi de estimare bazaţi pe regiuni 54
2.4 Concluzii 55

3. CODAREA SECVENŢELOR DE IMAGINI **57**

3.1 Codarea imaginilor 57
3.1.1 Fundamente 57
3.1.2 Entropia semnalului video 60
3.2 Tehnici de codare a semnalelor de imagine 63
3.2.1 Codarea predictivă cu compensarea mişcării 65
3.2.2 Codarea hibridă : compensarea mişcării şi transformările ortogonale ... 68
3.2.3 Codarea multirezoluţie cu compensarea mişcării 70
3.2.3.1 Codare predictivă cu estimarea mişcării în imaginea originală 72

BUPT

3.2.3.2 Codare predictivă cu estimarea mişcării în fiecare subbandă 73
3.2.3.3 Codarea hibridă subbandă cu determinări multiple 74
3.2.4 Codarea bazată pe model 75
3.2.5 Codarea orientată pe regiuni 77
3.3 Concluzii 80

4. SEGMENTAREA IMAGINILOR 83

4.1 Introducere 83
4.2 Extragerea contururilor 87
4.2.1 Detecţia punctelor 87
4.2.2 Detecţia liniilor 87
4.2.3 Detecţia muchiilor 88
4.3 Segmentarea bazată pe prag 96
4.3.1 Segmentare prin determinarea pragului global 96
4.3.2 Segmentare prin detecţia pragului optim 97
4.3.3 Segmentare bazată pe caracteristicile muchiilor 98
4.3.4 Segmentare cu căutare automată a pragului (ISODATA) 101
4.4 Segmentare orientată pe regiuni 103
4.4.1 Definirea problemei 103
4.4.2 Creşterea regiunilor 104
4.4.3 Algoritmi de tip divide şi contopeşte 106
4.4.4 Segmentarea morfologică 108
4.4.5 Algoritm de segmentare de tip divide şi contopeşte adaptiv [ALE 98] 110
4.4.5.1 Divizarea 115
4.4.5.2 Contopirea 119
4.5 Concluzii 122

5. ESTIMAREA MIŞCĂRII BAZATĂ PE REGIUNI ÎN SECVENŢELE DE IMAGINI 124

5.1 Introducere 124
5.2 Tehnici de estimare bazate pe segmentare 127
5.2.1 Estimarea mişcării utilizând modelul parametric 127
5.2.2 Estimarea mişcării bazată pe segmentarea spaţio-temporală 129
5.2.3 Urmărirea 131
5.2.3.1 Urmărirea obiectelor din scenă 132
5.2.3.2 Algoritmi de filtrare Kalman 132
5.3 Algoritmul propus de autor pentru estimarea mişcării în vederea codării bazate pe regiuni a secvenţelor de imagini 133
5.3.1 Etichetarea 136
5.3.2 Calculul parametrilor regiunilor 136
5.3.3 Reprezentarea şi codarea conturului 138
5.3.4 Calculul vectorilor de mişcare 141

BUPT

5.3.5 Codarea datelor obţinute prin estimarea mişcării 147
5.3.6 Refacerea imaginii 150
5.4 Concluzii 152

6. REZULTATE EXPERIMENTALE 154

6.1 Introducere 154
6.2 Imaginile utilizate 154
6.3 Segmentarea imaginilor 156
6.4 Estimarea mişcării 160
6.4.1 Rezultate ale estimării mişcării pentru imagini sintetice 160
6.4.2 Rezultate ale estimării mişcării în imagini reale de referinţă 165
6.5 Imaginile de test reconstruite 173
6.6 Concluzii 175

7. CONCLUZII 177

7.1 Contribuţii personale 177
7.2 Posibilităţi de dezvoltare ulterioară 179

ANEXE 182

A.1 Modele de mişcare 182
A.1.1 Echivalenţă între viteză şi deplasare în estimarea mişcării 182
A.1.2 Modele parametrice ale suprafeţei plane 182
A.2 Aprecierea calităţii imaginii 183
A.2.1 Aprecierea calităţii prin măsurători cantitative 183
A.2.2 Aprecierea subiectivă a calităţii 186

BIBLIOGRAFIE 188

BUPT

INTRODUCERE

Ultimii ani au constituit o etapă cu o dezvoltare spectaculoasă a domeniului comunicațiilor. Au fost introduse foarte multe servicii telematice şi de televiziune ce au dus la creşterea numărului de imagini transmise la diferite niveluri. Progresele înregistrate în electronică, în tehnologiile software şi, nu în ultimul rând, corelarea eforturilor de cercetare ale specialiştilor din domenii distincte ca electronica, grafica computerizată, vederea artificială etc., au făcut ca azi să se vorbească tot mai mult despre: supermagistrale de informații, aplicații multimedia supersofisticate, echipamente performante de comunicații interpersonale şi multe altele. Traficul de informație pe rețelele de calculatoare existente prin intermediul serviciilor Internet ca: E-Mail, Telnet, FTP, Usenet, WWW, Gopher etc., a crescut continuu putând spune că se marchează startul spre o noua eră: *era informației*. Toate acestea demonstrează clar că industria de telecomunicații capătă o nouă formă şi ocupă o poziție tot mai importantă în cadrul societății moderne.

În acest context compresia digitală a imaginilor, utilizând tehnici performante, a cunoscut o dezvoltare impresionantă în decursul ultimilor decenii. S-au elaborat şi propus foarte multe standarde atât pentru secvențele de imagini: H-261, H-263, MPEG-1, MPEG-2, MPEG-4, MPEG-7, cât şi pentru imagini statice: JPEG, JPEG-2000, multe dintre ele constituind elementul cheie al noilor tehnologii informatice în diverse aplicații: multimedia, efecte speciale în televiziune şi cinematografie, grafică pe calculator, sisteme de videoconferință, teledetecție, medicină etc. Cele mai multe dintre aceste servicii necesită transmiterea unei cantități impresionante de date pe canale foarte înguste.

Codarea la rate foarte scăzute de biți, necesară transmisiei pe canalele de comunicații existente, implică rate de transmisie sub 64 kb/s pentru semnalul video. Sursele video aferente acestor aplicații, având 8 biți/ eşantion pentru fiecare culoare, conțin în mod obişnuit aproximativ 1,25 Mocteți pentru un cadru de imagine color. Cum frecvența cadrelor este de 25 Hz (30 Hz), rezultă o rată de date, fără implicarea unei compresii, de peste 10 Mb/s. În aceste condiții compresia semnalelor trebuie să fie cuprinsă între 150:1 si 1000:1.

Problema de mai sus poate fi exemplificată în video-telefonie, care presupune transmiterea imaginilor prin rețelele telefonice uzuale. După cum se ştie, rețeaua telefonică disponibilă este folosită în special pentru transmiterea semnalului vocal. Informația video are o bandă mult mai largă comparativ cu cea vocală. În condițiile

BUPT

în care se adoptă formatul CIF ("Common Intermediate Format" - 352×288), rata de biţi pentru o secvenţă este de aproximativ 37 de milioane b/sec. Dacă încercăm transmiterea unui astfel de semnal prin reţeaua telefonică, presupunând capacitatea canalului extinsă la 16 kb/sec şi că utilizam 8 kb/sec pentru informaţia video, atunci rata de compresie trebuie să fie de peste 4666:1!

În tabelul 1.1 sunt prezentate compresiile necesare pentru obţinerea unei rate de 10 kb/s (specifică aplicaţiilor de codare la rate foarte scăzute), pornind de la câteva formate video standard.

Tabelul 1.1 Compresia necesară pentru o rata de 10 kb/s pentru diferite formate de imagine

Frecvenţa cadrelor [Hz]	CCIR 601 (720 × 576)	CIF (352 × 288)	QCIF (176 × 144)
7,5	4979 : 1	915 : 1	229 : 1
10	6637 : 1	1216 : 1	304 : 1
15	9952 : 1	1824 : 1	456 : 1
25	16586 : 1	3040 : 1	760 : 1
30	19.904 : 1	3648 : 1	912 : 1

În general, este imposibilă compresia unui semnal TV cu un asemenea factor şi totodată păstrarea unei calităţi bune a imaginii reconstituite. Din acest motiv, există restricţii impuse în astfel de aplicaţii. Pentru exemplificare se consideră tot semnalul pentru videotelefonie. Scenele tipice din videotelefonie au următoarele caracteristici:

- *conţinut fix al scenei*: Scena tipică este imaginea capului şi umerilor vorbitorului. Datorită cunoaşterii a-priori a obiectelor din scenă se pot folosi unele elemente din aceasta, cum ar fi forma 3-D a feţei umane.
- *mişcarea limitată din scenă*: Mişcarea este în principal determinată de mişcarea vorbitorului, iar camera este in general fixă (această situaţie nu este valabilă pentru videotelefonia mobilă, dar şi în aceste cazuri camera are deplasări limitate: scalare, vibraţie). Mişcarea vorbitorului cuprinde o mişcare globală (capul şi umerii), precum şi o mişcare locală determinată de schimbarea expresiilor feţei. Datorită inerţiei corpului uman, mişcarea globală este relativ

BUPT

lentă şi poate fi descrisă utilizând doar câţiva biţi pe cadru. În acest fel vor putea fi utilizaţi mai mulţi biţi pentru descrierea expresiilor faciale.

- *particularităţi de percepţie vizuală:* Comunicaţiile video nu necesită de obicei întreaga rezoluţie cerută de televiziunea radiodifuzată sau de formatul *CIF*, scopul fiind acela de a putea reda corect aspectul emoţional al imaginii vorbitorilor. Din acest motiv, se utilizează un format de imagine cu rezoluţie mai scăzută. Un astfel de format, utilizat în mod curent este *QCIF*, care are următorii parametri: rezoluţia semnalului de luminanţă 176×144, rezoluţia semnalului de crominanţă 88×72, frecvenţa cadrelor 10 Hz sau chiar 6 Hz.

Combinând astfel cunoaşterea scenei, redundanţa spaţio-temporală, rezoluţia scăzută etc., se pot obţine rate de compresie foarte ridicate pentru semnalul video.

Teza elaborată de autor se înscrie în această problematică, constituindu-se într-o contribuţie la înţelegerea şi aplicarea celor mai importante tehnologii şi metode specifice noilor algoritmi utilizaţi pentru estimarea mişcării în secvenţele de imagini. Teza urmăreşte realizarea unei abordări sistematice a domeniului atât din punct de vedere teoretic, cât şi din punct de vedere al implementării practice a algoritmilor de estimare a mişcării. Întreaga abordare urmăreşte prezentarea contribuţiilor autorului în ceea ce priveşte elaborarea de noi algoritmi de estimare a mişcării şi în special pentru codarea bazată pe regiuni a secvenţelor de imagini. În centrul preocupărilor autorului a stat abordarea codării secvenţelor de imagini din punct de vedere al sistemului vizual uman.

Astfel s-au evaluat din punct de vedere cantitativ, precum şi din punct de vedere calitativ conţinutul vizual al acestor imagini. Principala sarcină a constituit-o determinarea informaţiei minime şi de strictă necesitate necesară pentru înţelegerea completă a mesajului transmis sau memorat, modelul percepţiei fiind ales implicit sau explicit funcţie de algoritmul utilizat, deoarece în toate aplicaţiile observatorul uman este analizatorul final al calităţii serviciilor oferite. Plecând de la algoritmul de estimare a mişcării bazat pe corespondenţa blocurilor, autorul tezei de doctorat propune un nou algoritm de estimare a mişcării, în care imaginea este reprezentată pe regiuni. Acest algoritm se înscrie în tendinţa actuală de obţinere a unor rate cât mai scăzute pentru transmiterea informaţiei de imagine.

Teza este structurată astfel:

- Capitolul 2 examinează principalele strategii utilizate în analiza secvenţelor de imagini, precum şi legătura dintre ele. Autorul propune o analogie între nivelele de cunoaştere şi complexitate rezultate din analiza secvenţelor de imagini şi nivelele utilizate în procesul de codare al acestora, ceea ce conduce la ideea utilizării informaţiilor semantice din imagine pentru obţinerea de rezultate superioare în procesul de codare.

  În paragraful 2.2 este introdus conceptul de percepţie a imaginilor, realizându-se totodată şi prezentarea principalelor caracteristici ale sistemului vizual uman care, mai apoi, vor fi luate în consideraţie la analiza secvenţelor de

BUPT

imagini. Sunt introduse din bibliografie conceptele de modelare a imaginii şi a mişcării.

Pentru metodele bazate pe blocuri, autorul prezintă un algoritm de corespondenţă rapidă a blocurilor propus în lucrarea "A New Block Matching Algorithm for Motion Estimation" având rezultate apropiate cu cel cu căutare completă (full search) şi mai bune decât algoritmii rapizi propuşi în literatură.

Analiza stadiului actual al metodelor de codare a condus autorul spre concluzia că, pentru obţinerea performanţelor cerute de aplicaţiile actuale, soluţia de estimare a mişcării o constituie dezvoltarea metodelor bazate pe conţinutul semantic al imaginilor.

- Capitolul 3 prezintă tehnicile de codare a secvenţelor de imagini. Motivată de aplicaţiile imediate (video-telefonie, video-conferinţe, televiziune digitală, multimedia, etc.), codarea la rate foarte scăzute reprezintă principalul scop al celor mai multe din cercetările actuale. În capitol este prezentată o clasificare a principalelor tehnici de codare cunoscute având la bază o schemă de codare în trei paşi: transformarea spaţiului imaginii, codarea cu pierderi şi codarea în scopul adaptării la transmisia pe canal. În final este prezentată o schemă de codare bazată pe regiuni.

- Capitolul 4 tratează segmentarea imaginilor ca un prim pas al estimării mişcării, cu un rol esenţial la obţinerea reprezentării imaginii sub formă de primitive semantice lipsite de ambiguitate (contururi, regiuni).

Autorul prezintă o scurtă introducere în segmentarea imaginilor cu evidenţierea principalelor caracteristici ale imaginilor (discontinuităţi, similarităţi) precum şi metodele de evidenţiere ale acestora. După prezentarea metodelor şi algoritmilor de extragere a contururilor (gradient, laplacian) şi regiunilor (pe baza pragurilor, ISODATA adaptiv, creşterea regiunilor), autorul prezintă un algoritm original de segmentare adaptivă a imaginilor, algoritm ce se bazează pe proprietăţile sistemului vizual uman

- informaţia referitoare la contur şi muchii este apreciată foarte mult de sistemul vizual uman, fiind responsabila principală în ceea ce priveşte percepţia;
- textura are o importanţă diminuată, fiind asociată cu informaţia suplimentară, influenţând percepţia numai atunci când este asociată cu informaţia de contur.

- Capitolul 5 prezintă algoritmul propus de autor în vederea estimării mişcării pentru codarea bazată pe regiuni a secvenţelor de imagini. În introducerea capitolului, se realizează o prezentare sintetică a principalilor algoritmi propuşi în literatură bazaţi pe reprezentarea semantică a imaginilor. Algoritmul propus se bazează pe segmentarea statică a imaginilor şi căutarea corespondenţelor între cadrele succesive din imagine. Pe baza corespondenţelor se realizează o

BUPT

atribuire adaptivă a modelului de mişcare fiecărei regiuni, model ce este ap utilizat la codarea secvenţei. Pentru regiunile ce nu au o corespondenţă imaginea anterioară se utilizează o metodă de codare cu pierderi pe baz conturului, a centrului de greutate şi a modelului de aproximare a textu utilizat la segmentare.

Atribuirea adaptivă a modelului de mişcare pentru fiecare regiune bazează pe presupunerea că scena cuprinde numai obiecte rigide (corpu nedeformabile) şi există limite maxime ale mişcării între imaginile succesive

Pentru estimarea mişcării autorul propune un set de trăsături caracteristic cu ajutorul cărora este posibilă determinarea rotaţiei şi translaţiei regiunilor. lângă implementarea calculului caracteristicilor regiunilor pe baza momente care conduce la identificarea unui vector de mişcare şi pentru fundal, autor propune o metodă nouă de determinare a translaţiei şi rotaţiei. Aceasta bazează pe determinarea recursivă a centrului de greutate al regiuni determinarea unghiului de rotaţie prin utilizarea unor puncte oarecar aparţinând conturului regiunii. În acest scop, autorul propune o teoremă presupune că în imagine există numai corpuri rigide. Metoda este bazată identificarea regiunii şi a punctelor determinate de intersecţia unui cerc avâ raza cuprinsă între distanţa minimă şi maximă a centrului de greutate faţă conturul regiunii. În acest fel se obţin puncte ce caracterizează regiunea şi baza cărora, prin minimizarea unui criteriu de eroare introdus de autor, determină unghiul de rotaţie al regiunii.

Rezultatele obţinute sunt apoi utilizate pentru codarea secvenţei imagini. Codarea se bazează pe informaţia de mişcare şi tabelul corespondenţă pentru regiunile care au o corespondenţă în imaginea preceden şi pe contur, centrul de greutate şi modelul de aproximare al texturii pent regiunile considerate noi. Pentru fiecare regiune se transmit numai ac parametri de mişcare ce caracterizează cel mai bine mişcarea regiunii.

- Capitolul 6 prezintă rezultatele implementării pe diferite clase de secvenţe imagini. Aceste rezultate sunt evaluate comparativ prin intermediul rat biţi/pixel şi a raportului semnal/zgomot, precum şi prin intermediul apreciere subiective a calităţii imaginii obţinute. Concluzia este că se obţine un raport compresie mare în condiţiile unei calităţi bune a imaginii reconstruite. D analiza raportului semnal/zgomot şi a calităţii imaginii rezultă că raport semnal/zgomot nu este relevant în cazul codării cu pierderi, întrucât în urm refacerii imaginile arată mai bine decât ne-am fi aşteptat din punct de vedere percepţiei vizuale, datorită reprezentării pe regiuni a acestora.
- Capitolul 7 prezintă concluziile cu evidenţierea principalelor contribuţii a autorului şi sugestii privind posibilele direcţii de continuare a cercetărilor

BUPT

Capitolul 2

ESTIMAREA MIŞCĂRII ÎN SECVENŢELE DE IMAGINI

2.1 Introducere în analiza mişcării

Unele din etapele de prelucrare a imaginii încearcă să exploateze corelaţia intercadre în scopul codării. Se poate distinge, în mod esenţial, una din prelucrări ca fiind detecţia mişcării pentru estimarea şi compensarea acesteia.

Dacă se acceptă că semnalul secvenţelor de imagini este nestaţionar în spaţiu şi timp, atunci pentru o imagine dată, funcţia densităţii de probabilitate a luminanţei nu este uniformă, ea modificându-se atât în suprafaţa imaginii cât şi în timpul secvenţei (tabelul 3.2 – paragraful 3.1.1). Pentru a elimina cât mai mult redundanţa între cadre, este necesară analiza conţinutului dinamic al imaginii (prezenţa mişcării în imagine). Această analiză poate fi realizată la diferite niveluri de cunoaştere şi complexitate. În continuare se prezintă pe scurt principalele niveluri de analiză şi de utilizare a rezultatelor obţinute, urmând ca în următoarele paragrafe să se prezinte pe larg procesul de estimare al mişcării, în vederea codării secvenţelor de imagini.

Detecţia mişcării – Scopul acestui proces este de a se identifica zonele staţionare şi cele în mişcare dintr-o secvenţă de imagini, printr-o serie de observaţii bazate în principal pe diferenţele dintre cadre. Rezultatul îl constituie o decizie binară a prezenţei/absenţei mişcării prezentată sub forma unei hărţi spaţiale a obiectelor mobile. Datele obţinute sunt insuficiente pentru eliminarea ambiguităţilor de tipul: suprapunerea obiectelor, apariţia de noi obiecte prin descoperire, suprafeţe interioare ce aparţin unui obiect în mişcare.

Estimarea mişcării – Reprezintă o etapă superioară în analiza mişcării, cu scopul principal de a măsura mişcarea percepută prin detecţie (viteza aparentă). Există numeroşi algoritmi pentru măsurarea câmpului vectorilor de mişcare, ale cărui principale caracteristici sunt:

- densitatea – care este mare, dacă măsurarea este realizată pentru fiecare pixel şi mică, dacă este determinată numai pentru trăsăturile de bază;
- tipul de primitive utilizate pentru reprezentare: pixel (estimarea mişcării locale), contur (estimare unidimensională), regiune (estimare bidimensională), obiect (estimare tridimensională).

După această identificare, care implică definirea modulelor de mişcare, măsurările cantitative sunt definite cu un nivel de precizie ce poate fi mai mic decât un pixel. În acest caz, precizia de estimare a vectorului de mişcare poate fi 1/8 sau chiar 1/16 pixel/cadru.

BUPT

Segmentarea mişcării – Este importantă analiza scenei din punct de vedere al regiunilor omogene. Spre deosebire de algoritmii de detecţie a mişcării, care conduc spre ambiguităţi, algoritmii de segmentare bazaţi pe mişcare asigură fragmentarea scenei în regiuni având mişcări distincte şi identificabile.

Cele două niveluri de analiză ale mişcării, detecţia şi segmentarea, ridică una din cele mai dificile probleme, şi anume, optimizarea lor simultană. Problema este cunoscută sub denumirea de "Chicken and Egg problem" [TZI 94]. Aceasta poate fi sintetizată prin următoarele: obţinerea unei bune segmentări, în zone omogene din punct de vedere al mişcării, este determinată de obţinerea mai întâi a parametrilor de mişcare; dar algoritmii de mişcare pot fi efectiv aplicaţi numai pentru regiuni a căror reprezentare are un model continuu (în acest caz este vorba despre modelul mişcării), lucru ce implică o segmentare efectivă în regiuni omogene. Datorită acestui fapt au apărut algoritmi iterativi sau multiplexaţi care se bazează pe estimarea şi segmentarea simultană a mişcării.

Interpretarea mişcării – Poate fi interesantă din punct de vedere al codării semantice, analiza scenelor utilizând modele pentru mişcare. Aceasta presupune analiza secvenţei de imagini din punct de vedere calitativ al mişcării şi anume: obiect 1 - rotaţie, obiect 2 - modificare dimensiuni.

Interpretarea de înalt nivel se bazează pe etapele de analiză inferioare ale mişcării: detecţie-segmentare. Acest tip de prelucrare este esenţial atunci când într-un sistem de comunicaţii este necesară îmbunătăţirea performanţelor de compresie şi diagnostic, utilizând aceeaşi algoritmi de prelucrare. Una dintre aplicaţii o constituie sistemele de supraveghere activă din cadrul reţelelor de comunicaţii video.

Ţinând cont de posibilele niveluri de analiză a mişcării, se pot defini corespondenţele între analiza mişcării şi codarea secvenţelor de imagini prezentate în tabelul 2.1.

Tabelul 2.1 Corespondenţele între algoritmii de estimare ai mişcării şi codarea secvenţelor de imagini

Analiza mişcării	Codare
Detecţie	Transmisie condiţionată
Estimare	Compensare mişcare
Estimare-segmentare	Codare spaţio-temporală
Interpretare	Codare semantică

Actualizare condiţionată – Detectorul de mişcare realizează o localizare binară (da sau nu) a zonelor mobile din imagine, iar pentru imaginile reconstituite se vor reactualiza numai acele zone în care a fost detectată o mişcare faţă de cadrul anterior. Reactualizarea temporală este, deci, condiţionată de detecţia mişcării. În aceste condiţii imaginea ce se transmite este de fapt imaginea diferenţă între cadre:

BUPT

$$I_D(x,y;t) = I(x,y;t) - I(x,y;t-1) \qquad (2.1)$$

unde $I(x,y;t-1)$ reprezintă imaginea anterioară atât la codare cât şi la decodare.

Compensarea mişcării – Compensarea mişcării pentru o imagine implică o analiză prealabilă a mişcării (estimarea vectorului de mişcare al fiecărui punct) şi utilizarea datelor obţinute pentru crearea unei imagini virtuale care, în fiecare punct, bloc sau regiune, funcţie de densitatea de informaţie de estimare a mişcării, caută informaţia:

$$I_{t-1} = I(x - D_x, y - D_y; t-1) \qquad (2.2)$$

unde $\vec{D}=(D_x, D_y)$ reprezintă deplasarea estimată, între momentele t şi $t-1$, pentru pixelul având coordonatele (x,y), cu referinţă în imaginea de la momentul t. Compensarea mişcării permite astfel utilizarea corelaţiei intercadre în concordanţă cu mişcarea estimată, pentru reducerea redundanţei temporale din secvenţa de imagini. Semnalul util ce se transmite în acest caz este:

$$I_{DD}(x,y;t) = I(x,y;t) - I(x - D_x, y - D_y; t-1) \qquad (2.3)$$

Codarea orientată pe obiecte – În acest caz scena este descrisă ca o reuniune de regiuni omogene din punct de vedere al mişcării (se pot adăuga criterii privind modele ale suprafeţei şi intensităţii). Semnalul obţinut cuprinde evoluţia acestor regiuni, incluzând noţiuni de apariţie/dispariţie, deformare şi contopire a obiectelor. Prelucrarea datelor obţinute din analiza mişcării presupune:

- codarea structurilor dinamice obţinute prin segmentare;
- codarea parametrilor modelului de mişcare;
- codarea texturii.

Codarea semantică – Acest tip de prelucrare necesită existenţa unor modele de înalt nivel pentru forma şi luminanţa regiunilor şi a imaginilor prelucrate. Etapa de interpretare rezultată prin analiza mişcării generează o descriere calitativă compactă, echivalentă cu utilizarea unei secvenţe de elemente de limbaj de nivel înalt. Evident, acest tip de codare poate fi folosit numai în aplicaţii restrânse, în care se pot însera cunoştinţe apriorice numeroase sau când acestea pot fi identificate printr-o învăţare anterioară. Împreună cu metodele semantice, se pot utiliza tehnicile predictive.

În completarea metodelor de codare descrise anterior (care se bazează pe mişcare), putem menţiona două care nu se bazează numai pe mişcare:

- cuantizarea vectorială adaptivă, care cuprinde situaţii nestaţionare în timp şi a căror principală caracteristică este îmbunătăţirea bibliotecii de coduri [GER 92];
- metode de intrapolare/extrapolare cu utilizarea mişcării, având ipoteze bazate pe modelul traiectoriei de mişcare estimate pentru o secvenţă de câteva cadre (aşa numitul "grup de imagini" – GOP, Group Of Picture – utilizat în standardul MPEG).

BUPT

2.2 Percepţia şi modelarea secvenţelor de imagini

În acest paragraf se face o trecere în revistă a principalelor caracteristici, ce sunt luate în consideraţie în ceea ce priveşte percepţia, atât din punctul de vedere al sistemului vizual uman, cât şi al vederii artificiale. Este clar că aceste observaţii depind de caracteristicile semnalului analizat (contrast, bandă de frecvenţe a semnalului, mişcare), dar şi de sistemul de percepţie, care poate fi un senzor (vedere artificială) sau sistem vizual uman.

2.2.1 Percepţia secvenţelor de imagini

O analogie simplă poate aproxima sistemul vizual uman cu o cameră. Aproximarea poate fi făcută în cazul unei descompuneri funcţionale şi structurale brute, principalele elemente rezultate din această descompunere fiind:

- senzorul de intrare,
- suprafaţa de recepţie - retina pentru sistemul vizual uman şi matricea pentru senzorul CCD,
- convertorul de informaţie.

La o analiză atentă a mecanismelor biologice şi psiho-chimice, această analogie nu este valabilă. Motivul principal este că, în cazul sistemului vizual uman, nu este posibilă izolarea elementelor funcţionale. Desigur, se pot face diferite interpretări, dar, spre deosebire de cameră, elementele biologice sunt interdependente, operând continuu în timp şi aproape continuu în spaţiu.

Codarea secvenţelor de imagine nu poate ignora aceste rezultate, deoarece în ultimă instanţă aprecierea calităţii imaginii reconstituite este realizată de ochiul uman. Studiile privitoare la percepţie în sistemul psiho-vizual încearcă definirea următoarelor modele [NAK 85]:

- modele descriptive a informaţiilor percepute – acestea pot fi modele structurale, statistice (mai rar utilizate) sau în frecvenţă;
- modele geometrice – acestea sunt foarte asemănătoare cu cele utilizate în vederea artificială, unul dintre cele mai utilizate fiind modelul de proiecţie al perspectivei;
- modele ale funcţiilor de transfer - pentru celulele elementare de recepţie precum şi pentru cele de recepţie senzorială;
- modele de sensibilitate temporală şi spaţială – care sunt echivalente cu determinarea pragurilor de vizibilitate, pornindu-se de la premisa liniarităţii şi a separabilităţii spaţiale [ANA 93];
- modele pentru efectul de mascare – admiţând că se generează informaţii vizuale diferite, funcţie de contextul în care sunt obţinute, atât din punct de vedere spaţial cât şi temporal. Această observaţie fundamentală pentru codarea imaginii este utilizată în cazul codării adaptive intracadre pentru alegerea predictorului sau adaptarea cuantizorului.

BUPT

S-a demonstrat că sistemul vizual uman este sensibil la câteva trăsături ale obiectelor, cum ar fi: graniţele dintre obiecte sau contururile obiectelor, colţurile (punctele de curbură maximă), regiunile (zonele cu textură omogenă) [MAR 82].

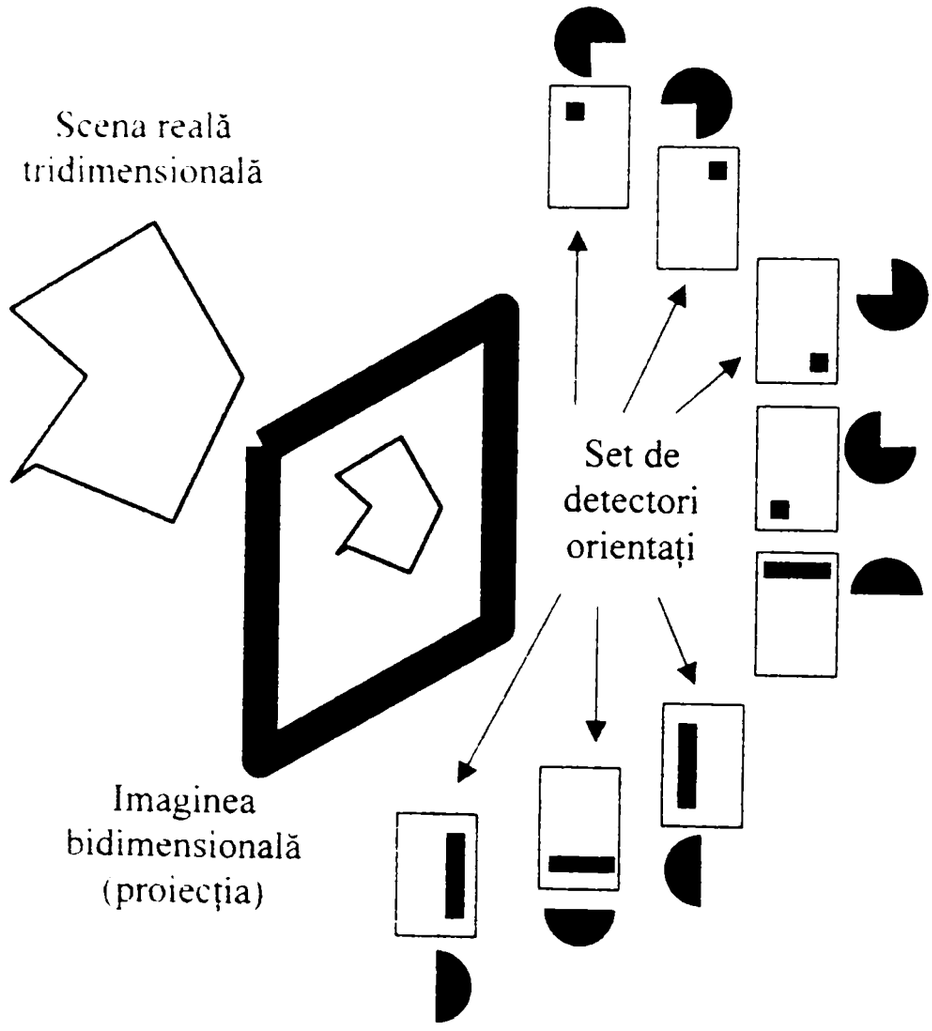


Fig 2.1 *Set de detectori orientaţi pentru trăsături*

Ullman [ULL 79] înaintea lui Marr, a extins conceptul referitor la modelare şi pentru mişcarea aparentă. Această observaţie este similară cu cea făcută de psihofizicieni, referitoare la existenţa unor câmpuri receptive ce fac posibilă, prin excitaţia sau inhibiţia unor detectoare elementare, diagnosticarea prezenţei sau absenţei unor trăsături ale obiectelor. În figura 1.1 se prezintă un set de detectori orientaţi pentru trăsături.

Un alt nivel de modelare este acela al conexiunilor, pentru care există următoarele niveluri:

a) Celula elementară – care reprezintă câmpul de recepţie orientat. Dacă se injectează un stimul staţionar, atunci suprafeţele excitate şi inhibate formează un detector de caracteristici vizuale.

BUPT

b) Coloana acordată – compusă din celule elementare care detectează aceeaşi caracteristică. Noţiunea de acord introdusă în modelare este echivalentă cu introducerea unui filtru trece bandă.
c) Supercoloana – introdusă în neuropsihologie, care este obţinută prin compunerea unui set de coloane acordate şi care face posibilă distingerea între stimuli.

Watson propune modelarea sistemului vizual uman ca un filtru liniar spaţio-temporal, având următorul model analitic [WAT 86]:

$$h(t) = \xi[h_1(t) - \zeta h_2(t)] \tag{2.4}$$

în care

- $h_1(t) = u(t)(\tau\,(n-1)!)^{-1}\left(\frac{t}{\tau}\right)e^{-\frac{t}{\tau}}$, reprezintă răspunsul unei cascade de filtre trece bandă, cu transformată Fourier $H_1(f) = (2i\pi\, f\tau + 1)^{-n_1}$
- $h_2(t)$ – este un filtru identic dar cu o constantă de timp diferită K_τ şi lungime n_2
- ξ - factorul de amplificare;
- ζ - coeficient de ponderare care permite introducerea atenuărilor la frecvenţe joase.

2.2.2 Modelarea mişcării în vederea analizei secvenţelor de imagini

Pentru analiza secvenţelor de imagini este necesară definirea unui model de mişcare. Anandan introduce următoarele categorii de modele de mişcare [SEZ 93]:

- Modele neparametrice – care se bazează pe câmpul local al mişcării ("optical flow" [HOR 81]) şi care se aplică în condiţiile estimării mişcării pentru fiecare pixel (estimarea unui câmp al vectorilor de mişcare dens), fără a fi necesare ipoteze cu privire la apartenenţa acestora la anumite regiuni din imagine. Dezavantajul acestor modele îl constituie necesitatea impunerii unor anumite condiţii referitoare la imagine (uniformitatea locală a semnalului de luminanţă).
- Modele cvasiparametrice – care se bazează pe reprezentarea mişcării printr-un set de parametri, utilizând pe lângă modelarea corpului rigid şi modelarea locală a câmpului vectorilor de mişcare.
- Modele parametrice – se bazează pe reprezentarea mişcării globale printr-un set de parametri. Aceste modele au avantajul unei mai mari precizii datorate contribuţiei tuturor pixelilor din regiune la estimarea mişcării.

Odată cu introducerea tehnicilor de codare din cea de-a doua generaţie [KUN 85], estimarea mişcării nu s-a mai bazat pe reprezentarea stohastică a imaginii, ci pe o reprezentare semantică a acesteia. Aceasta presupune descrierea imaginii prin obiectele prezente în scena reală. În acest fel, aplicarea modelului parametric la descrierea mişcării este mult mai apropiată de realitate.

BUPT

Aşa cum am văzut la codarea secvenţelor de imagini, este necesară transmiterea, pe lângă semnalul principal, a informaţiei de mişcare. În acest context, câmpurile de mişcare dense obţinute pentru modelele neparametrice sau evasiparametrice necesită o cantitate de informaţie suplimentară mare faţă de modelul parametric, care asigură o reprezentare compactă a informaţiei de mişcare şi care tinde, deci, la micşorarea informaţiei suplimentare.

Se poate concluziona că, modelele parametrice ce derivă din proiecţia mişcării tridimensionale a scenei sunt preferate deoarece contribuie la creşterea factorului de compresie şi la reprezentarea mai compactă a imaginii. În plus, deoarece mişcarea tridimensională poate fi descrisă printr-un model de mişcare rigid, rezultă că aceasta poate fi reprezentată prin proiecţia ortografică [TZI 94].

2.2.2.1 Mişcarea corpului rigid

Prin proiecţia ortografică se realizează reprezentarea mişcării tridimensionale într-un plan, care în cazul real este realizată de camera de televiziune. Dacă considerăm că vectorul $\vec{R} = [X, Y, Z]^T$ reprezintă coordonatele carteziene ale scenei reale în raport cu camera şi $\vec{r} = [x, y]^T$ coordonatele imaginii plane (astfel încât Z este axa perpendiculară pe imaginea plană care defineşte axa optică, figura 2.2), atunci $\vec{r}$ poate fi exprimat în funcţie de vectorul $\vec{R}$ astfel [HOR 86]:

$$x = \frac{X}{Z} \qquad y = \frac{Y}{Z} \tag{2.5}$$

în care distanţa focală a fost normalizată, fără pierderi în ceea ce priveşte generalitatea transformării.

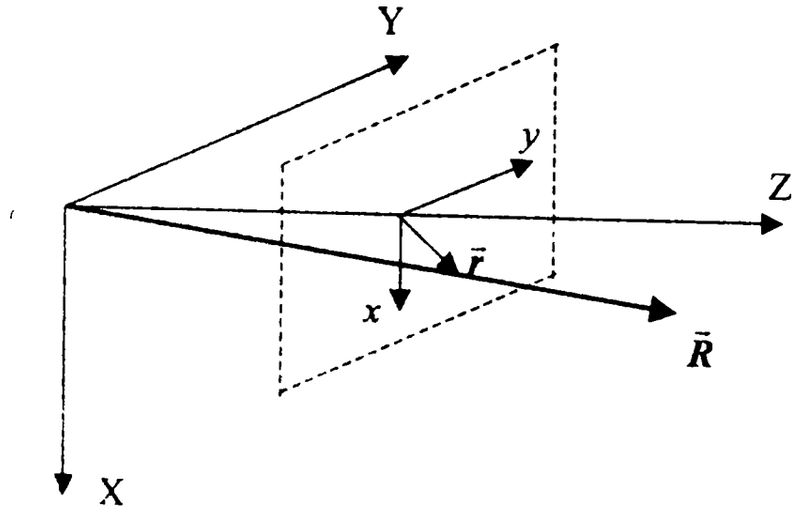


Fig 2.2 *Proiecţia plană a scenei reale tridimensionale*

BUPT

În cele mai multe cazuri se presupune că scena este compusă din corpuri rigide, ipoteză motivată de dificultatea de a considera corpuri nerigide. Cu toate acestea au fost propuşi şi algoritmi de estimare a mişcării corpurilor nerigide [PEN 91]. În plus, această ipoteză este justificată şi de faptul că se pot aproxima corpurile nerigide cu corpuri rigide atunci când frecvenţa de eşantionare temporală este suficient de mare, iar diferenţa dintre două cadre succesive este suficient de mică. Această presupunere se va considera valabilă pe durata întregului capitol fără a introduce distorsiuni sau erori suplimentare.

Pentru analiza mişcării plane a corpului rigid tridimensional pot fi formulate două abordări distincte şi anume:

1. din punct de vedere al vitezei instantanee – în care mişcarea corpului rigid este analizată prin viteza instantanee dată de relaţia [HOR 86]:

$$\vec{V} = \vec{\omega} \times \vec{R} + \vec{T} \tag{2.6}$$

unde:

$\vec{R}$ - componenta temporală

$\vec{\omega} = [\omega_x, \omega_y, \omega_z]^T$ - viteza unghiulară instantanee

$\vec{T} = [T_x, T_y, T_z]^T$ - viteza de translaţie instantanee

Dacă ţinem cont de relaţiile (2.5) şi (2.6), viteza instantanee a imaginii plane poate fi exprimată cu ajutorul relaţiei:

$$\vec{V} = \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} -\omega_x xy + \omega_y (x+1) - \omega_y y + \frac{T_x - T_z x}{z} \\ -\omega_x (y+1) + \omega_y xy - \omega_z x + \frac{T_y - T_z y}{z} \end{pmatrix} \tag{2.7}$$

2. din punct de vedere al deplasării – în care se analizează mişcarea prin deplasarea corpului rigid dată de relaţia:

$$\vec{R}' = \Omega \vec{R} + \vec{D} \tag{2.8}$$

unde: $\vec{R}$ şi $\vec{R}'$ reprezintă poziţiile la momentul t şi respectiv t+1, Ω matricea ce descrie mişcarea de rotaţie, iar $\vec{D} = [D_x, D_y, D_z]^T$ reprezintă translaţia. Din analiza ecuaţiei (1.8) putem concluziona că Ω şi D ne arată diferenţa în orientare şi poziţie de-a lungul intervalului de timp. În mod similar, prin utilizarea ecuaţiilor (2.5) şi (2.8) obţinem pentru reprezentarea mişcării plane:

BUPT

$$\vec{r} = \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \dfrac{x - \Omega_z y + \Omega_y + \dfrac{D_x}{z}}{1 - \Omega_y x + \Omega_x y + \dfrac{D_z}{z}} \\ \dfrac{\Omega_z x + y - \Omega_x + \dfrac{D_y}{z}}{1 - \Omega_y x + \Omega_x y + \dfrac{D_z}{z}} \end{pmatrix} \tag{2.9}$$

Pentru acest caz matricea Ω a fost aproximată ca în ecuaţia (2.10), presupunând unghiuri de rotaţie foarte mici:

$$\Omega = \begin{pmatrix} 1 & -\Omega_z & \Omega_y \\ \Omega_z & 1 & \Omega_x \\ -\Omega_y & \Omega_x & 1 \end{pmatrix} \tag{2.10}$$

Ecuaţiile (2.7) şi (2.9) definesc mişcarea bidimensională a imaginii plane rezultate prin proiecţia mişcării tridimensionale a corpului rigid în termeni ai vitezei instantanee, ecuaţia (2.7) şi ai deplasării, ecuaţia (2.9) [ADI 85]. Aceste ecuaţii corespund modelului cvasiparametric, putându-se arăta că cele două formulări sunt echivalente [SEZ 93] şi Anexa 1.

2.2.2.2 Modelarea mişcării suprafeţelor plane

Dependenţa modelului de profunzimea scenei Z în ecuaţiile (2.7) şi (2.9) determină două probleme:

- dificultatea de estimare precisă a hărţii de profunzime (adâncime a scenei);
- creşterea cantităţii de informaţie necesară pentru a fi transmisă.

Pentru a elimina aceste probleme, se impune ca model al scenei o suprafaţă plană mozaicată, care poate aproxima foarte bine corpurile rigide tridimensionale. Această simplificare elimină dependenţa de harta de profunzime pentru modelul mişcării şi conduce la un model complet parametric. În cadrul acestui model o suprafaţă plană este definită astfel:

$$k_x X + k_y Y + k_z Z = 1 \tag{2.11}$$

sau, în mod echivalent:

$$k_x x + k_y y + k_z = \frac{1}{z} \tag{2.12}$$

Dacă dorim formularea în termeni ai vitezei instantanee, introducem ecuaţia (2.12) în ecuaţia (2.7) şi obţinem:

BUPT

$$\begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} a_1 + a_2 x + a_3 y + a_7 x^2 + a_8 xy \\ a_4 + a_5 x + a_6 y + a_7 xy + a_8 x^2 \end{pmatrix} \tag{2.13}$$

Cei opt parametri ai modelului, definiţi de ecuaţia (2.13), specifică vitezele instantanee în imaginea plană corespunzătoare suprafeţelor plane în mişcare, din perspectiva proiecţiei şi sunt prezentaţi în Anexa 1. În mod similar, dacă dorim formularea în termeni ai deplasării, obţinem:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \frac{a_1 + a_2 x + a_3 y}{a_7 x + a_8 y + a_9} \\ \frac{a_4 + a_5 x + a_6 y}{a_7 x + a_8 y + a_9} \end{pmatrix} \tag{2.14}$$

Pentru a obţine un model cu 8 parametri, se realizează normalizarea prin termenul a_9, parametrii fiind prezentaţi în Anexa 1.

Atunci când distanţa faţă de scenă este mult mai mare decât variaţia distanţei dintre obiectele scenei, proiecţia scenei tridimensionale are o aproximaţie mai bună dacă este utilizată proiecţia ortografică (proiecţia ortografică presupune că razele de lumină sunt paralele cu axa optică) [HOR 86]. Cu alte cuvinte, această simplificare este valabilă atunci când profunzimea scenei este mai mică decât distanţa acesteia faţă de cameră.

Considerând o proiecţie ortografică, termenii de ordinul doi ai ecuaţiei (2.13) pot fi neglijaţi, obţinându-se un model parametric optim:

$$\begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} a_1 + a_2 x + a_3 y \\ a_4 + a_5 x + a_6 y \end{pmatrix} \tag{2.15}$$

Modelul definit de ecuaţia (2.15) este admis pentru reprezentarea mişcării suprafeţelor plane obţinute printr-o proiecţie ortografică. În condiţiile unei proiecţii date, se poate obţine un model similar prin dezvoltarea în serie Taylor de ordinul I a ecuaţiei (2.14). Modelul obţinut mai poartă denumirea şi de model afin [TZI 94]. Este util de reţinut faptul că, modelul afin poate fi echivalat cu un vector de translaţie, doi factori de scalare şi două unghiuri de rotaţie [WU 90]. Modelul afin poate fi simplificat, rezultând modelul de translaţie, relaţia (2.16), care este cel mai simplu şi cel mai des utilizat, fiind obţinut prin anularea parametrilor a_2, a_3, a_5, a_6.

BUPT

$$\begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_4 \end{pmatrix} \tag{2.16}$$

Ecuaţiile (2.13) – (2.16) corespund modelului parametric.

2.2.2.3 Alegerea celui mai adecvat model

Având posibilitatea de alegere a mai multor modele, problema alegerii adaptive în cadrul unei aplicaţii a celui mai adecvat model rămâne una foarte complicată. O altă dificultate constă în lipsa unei măsuri pentru evaluarea performanţelor modelului, care este strâns legat de estimarea mişcării. Din aceste motive, precum şi datorită existenţei mai multor modele, alegerea sa se face ţinând cont de aplicaţia propriuzisă.

Aplicaţiile de codare video, în contrast cu cele de analiză a mişcarii, necesită transmiterea parametrilor de mişcare către decodor. În aceste condiţii, costul transmiterii unui câmp local al mişcării dens ("optical flow" - descris în paragraful 2.3.1) devine foarte mare, limitând alegerea la modelul parametric. Este evident că un model parametric mai complex poate reprezenta mişcări mai complexe, făcând posibilă descrierea mişcării unei regiuni cu o mai mare precizie. Cu toate acestea, procesul de estimare tinde să devină tot mai dificil şi, în particular, mai sensibil la zgomot. În plus, estimarea mai multor parametri necesită o complexitate de calcul mai mare şi implicit creşterea cantităţii de informaţie suplimentară transmisă.

Standardele actuale de codare MPEG 1 [LeG 91], MPEG 2 [LeG 92], H 261 şi H 263 [BOJ 97] se bazează pe codarea cu transformare, care are la bază reprezentarea imaginii prin blocuri de pixeli. În acest context, metodele de estimare a mişcarii bazate pe corespondenţa blocurilor ("block matching") s-au dovedit a fi mai eficiente [MUS 85], [DUF 95]. Aceste tehnici divid imaginea în blocuri de pixeli de formă pătrată, atribuind fiecăruia un set de parametri de mişcare, complexitatea modelului depinzând de dimensiunea blocurilor. Dacă regiunea descrisă de un set de parametri este mică, mişcarea poate fi reprezentată eficient printr-un model simplu. Cu toate că au fost propuse tehnici de corespondenţă a blocurilor generalizate ce permit utilizarea unor modele mai complexe [SEF 93], cel mai utilizat model rămâne cel translaţional descris de ecuaţia (2.16).

Au fost propuse tehnici diferenţiale de estimare a mişcării bazate pe blocuri, care constau din modele de complexităţi variabile ce pot fi aplicate fiecărui bloc. Dificultatea implementării acestora constă în determinarea performanţelor modelelor respective, pentru definirea unui criteriu de a alege unul optimal pentru fiecare bloc în parte [NIC 91].

În cazul estimării mişcării bazate pe segmentare, un model al mişcării este aplicat unui obiect al scenei, care reprezintă uzual o regiune de formă oarecare. În

BUPT

acest context, un model translaţional simplu apare ca fiind ineficient, fiind necesar un model parametric complex. În particular un model afin, ecuaţia (2.15), arată a fi un bun compromis, şi a fost adoptat în multe aplicaţii [BOU 93], [DUF 95b], [PAR 94].

2.3 Estimarea mişcării

În acest paragraf se face o trecere în revistă a principalelor abordări în estimarea mişcării: metode bazate pe gradient, metode bazate pe căutare şi metode bazate pe reprezentarea în frecvenţă. În timp ce primele două tehnici de estimare a mişcării se bazează pe imaginea originală, cea de-a treia realizează transformarea imaginii într-un nou domeniu, obţinând estimarea pentru coeficienţii transformării.

Estimarea mişcării este o problemă ce a fost abordată în foarte multe teme de cercetare ale ultimilor 20 de ani. Există foarte multe articole de sinteză ce prezintă principalele abordări în domeniu [MUS 85], [AGG 88], [VEG 89], [BAR 94], [DUF 95]. În continuare se prezintă câteva dintre cele mai uzuale abordări ale problemei, care să ne conducă la noul algoritm propus, care are ca principală aplicaţie codarea la rate foarte scăzute a secvenţelor de imagini.

Există o distincţie între *câmpul real al mişcării*, care reprezintă proiecţia mişcării relative tridimensionale şi *câmpul local al mişcării*, care descrie variaţia în spaţiu şi timp a structurii nivelurilor de gri. Ideal, cele două caracterizări ale mişcării sunt identice, dar cu toate acestea, de cele mai multe ori, diferenţele obţinute sunt semnificative. Aceasta poate fi rezultatul procesului de formare al imaginii (zgomotul introdus de circuitele electronice, modificările de iluminare) sau al caracteristicilor scenei (cel mai uzual exemplu reprezentându-l iluzia creată prin mişcarea de rotaţie a unei roţi care se deplasează). Este clar că în codarea video este posibilă numai estimarea traseului optic (câmpul local al mişcării), cu toate că cea mai bună compensare a mişcării se obţine prin cunoaşterea mişcării reale.

Estimarea vitezei sau a deplasării se poate face fie într-un domeniu bidimensional (pentru blocuri, regiuni, întreaga imagine), fie pentru un domeniu unidimensional (pentru contururi sau muchii). Cele două abordări (bidimensională şi unidimensională) pot fi utilizate împreună, analiza mişcării contururilor fiind urmată de analiza mişcării regiunii. Ambele metode au un rezultat similar în cazul unei secvenţe de imagini, dar nu este rezolvată problema estimării locale.

Algoritmii de estimare a mişcării se pot clasifica în patru mari grupe:

- tehnici bazate pe gradient [HOR 81], [NAG 83],
- tehnici recursive bazate pe algoritmi recursivi pe pixel [NET 91], [CAF 81], [WAL 84]
- tehnici de corespondenţă a blocurilor (block matching) [JAI 81], [KOG 81],
- tehnici bazate pe reprezentarea în frecvenţă [HAS 74], [HEE 88], [FLE 90]

BUPT

2.3.1 Tehnici de estimare bazate pe gradient

Aşa cum am văzut în paragraful anterior şi am definit scena reală în paragraful 4.1, mişcarea percepută într-o secvenţă nu este identică cu proiecţia câmpului tridimensional al vitezelor. De aceea mişcarea aparentă rezultă dintr-un proces complex ce implică pe lângă mişcarea reală şi următoarele aspecte: iluminările scenei respective, proprietăţile de reflexie a obiectelor componente, forma suprafeţelor precum şi caracteristici ale senzorului de imagine utilizat pentru achiziţie. De asemenea este necesară o distincţie între mişcarea reală (câmpul bidimensional al vitezelor) şi mişcarea aparentă (traseul optic).

Pe de altă parte, variaţia intensităţii luminoase în timp şi în spaţiu depinde atât de mişcarea obiectului, cât şi de: caracteristicile scenei tridimensionale, componenţa suprafeţelor obiectului şi proprietăţile fotometrice ale acestor suprafeţe. Bazându-ne pe optica geometrică şi fotometrie, se poate stabili că intensitatea luminoasă sesizată în imagine este egală cu radiaţia emisă de suprafaţa obiectului către senzor, cu condiţia ca sistemul de achiziţie optică să fie corect calibrat radiometric [VER 89].

Dacă se consideră semnalul video ca o funcţie de trei variabile $I(x,y;t)$ (două spaţiale x, y şi una temporală t) şi se presupune că modificarea semnalului de luminanţă $I(x,y;t)$ se datorează numai mişcării structurii spaţiale a acestuia, atunci se poate defini deplasarea $(\delta x, \delta y)$ pixelului aflat în poziţia (x,y) între momentele t şi $t+1$, sub următoarea formă:

$$I(x,y;t) = I(x+\delta x, y+\delta y; t+\delta t) \tag{2.17}$$

Prin dezvoltarea termenului din stânga într-o serie Taylor pentru $(x,y;t)$ obţinem ecuaţia:

$$I(x+\delta x, y+\delta y; t+\delta t) = I(x,y;t) + \delta x\frac{\partial I}{\partial x} + \delta y\frac{\partial I}{\partial y} + \delta t\frac{\partial I}{\partial t} + e(x,y;t) \tag{2.18}$$

în care termenul $e(x,y;t)$ conţine termenii de ordin superior ai dezvoltării în serie Taylor. Prin divizarea cu δt şi considerând limita când acesta tinde la zero, se obţine ecuaţia traseului optic:

$$\frac{\partial I}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial I}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial I}{\partial t} = \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \tag{2.19}$$

unde u şi v reprezintă vitezele pentru axa x, respectiv axa y, sau aşa cum este prezentată în [HOR 81]:

$$(\nabla I(x,y;t))^T \mathbf{V}(x,y;t) + \left.\frac{\partial I}{\partial t}\right|_{(x,y;t)} = 0 \tag{2.20}$$

BUPT

unde V reprezintă viteza imaginii iar ∇I reprezintă vectorul gradientului spaţial al intensităţii:

$$\nabla I(x,y;t) = \left[\left.\frac{\partial I}{\partial x}\right|_{(x,y;t)} \quad \left.\frac{\partial I}{\partial y}\right|_{(x,y;t)} \right]^T \tag{2.21}$$

Presupunând că pentru suprafaţa considerată, lumina emisă nu este dependentă de direcţie şi scena este luminată uniform, Verry şi Poggio [VER 89], au obţinut următoarea ecuaţie pentru exprimarea vitezei $\left(\frac{\delta x}{\delta t}, \frac{\delta y}{\delta t}\right)$:

$$\frac{\partial I}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial I}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial I}{\partial t} = \rho \overline{I}\overline{\Omega} \times \overline{N}(a,b,t) \tag{2.22}$$

unde:

- *(a, b)* – sunt coordonatele intrinseci ale punctului de pe suprafaţă;
- ρ - este coeficientul de reflexie al suprafeţei (presupunând că nu este variabil spaţio-temporal);
- $\overline{I}$ - este vectorul ce indică direcţia şi intensitatea iluminării;
- $\overline{N}$ - este vectorul unitar, normal la suprafaţă în punctul (a,b);
- $\overline{\Omega}$ este viteza unghiulară a punctului;
- $\frac{\partial I}{\partial t}$ - este gradientul temporal, pe care în continuare il vom nota I_t.

Ecuaţia (2.22) reprezintă generalizarea ecuaţiei traseului optic (ecuaţiile (2.19) şi (2.20)). Acestea se pot obţine din ecuaţia 1.22 punând condiţia independenţei de caracteristicile geometrice şi fotometrice ale imaginii. Astfel partea stângă a ecuaţiei realizează legătura între vectorul de viteză bidimensional şi gradientul spaţio-temporal al intensităţii din imagine. În mod similar putem spune că partea dreaptă a ecuaţiei depinde în totalitate de caracteristicile tridimensionale, care nu sunt uşor accesibile.

Aşa cum s-a observat, traiectoria optică (traseul optic) şi vectorul viteză sunt diferite, lucru evidenţiat şi din compararea ecuaţiilor (2.19) şi (2.22). Astfel, din analiza celor două ecuaţii, se poate obţine accentuarea diferenţelor dintre mişcarea reală şi mişcarea aparentă. Un exemplu concludent îl constitue o sferă ce se roteşte în jurul axei proprii [HOR 86]. Ţinând cont de ipotezele anterioare, gradientul temporal este nul, iar traseul optic exprimat de ecuaţia (2.19) este de asemenea nul. In realitate mişcarea nu este nulă iar cel de-al doilea termen al ecuaţiei (2.22) nu este nul.

Schunk [SCH 84] a arătat că ecuaţia (2.20) este valabilă în cazul funcţiilor de intensitate diferenţiabile şi deci suficient de flexibilă pentru a fi compatibilă cu discontinuităţile prezente în intensitatea imaginii (muchiile). În plus ecuaţia (2.20)

BUPT

conţine două necunoscute: cele două componente ale vectorului viteză V, care într-un model continuu reprezintă soluţia dorită, descriind evoluţia conturului de intensitate constantă $I(x,y,t)=c$.

De fapt, această ecuaţie permite numai calculul componentei traseului optic în direcţia vectorului gradient $\nabla I = (I_x, I_y)$ la momentul t în care $I_x = \frac{\partial I}{\partial x}$ şi $I_y = \frac{\partial I}{\partial y}$ Presupunând că $\|\nabla I\| \neq 0$, prin rezolvarea ecuaţiei (2.20) obţinem:

$$\vec{V} = -\frac{I_t \nabla I}{\|\nabla I\|^2} \tag{2.23}$$

Deoarece există două necunoscute şi o singură ecuaţie, componenta perpendiculară pe direcţia gradientului este necunoscută. Cea mai simplă evidenţiere a acestui lucru este în situaţia în care există gradient spaţial nul şi gradient temporal nenul $I_t \neq 0$, ceea ce înseamnă că există puncte pentru care nu este posibilă verificarea vitezei determinate. Nedeterminarea ce apare în acest caz este evidenţiată în figura 2.3 fiind cunoscută ca problema de apertură [HOR 86].

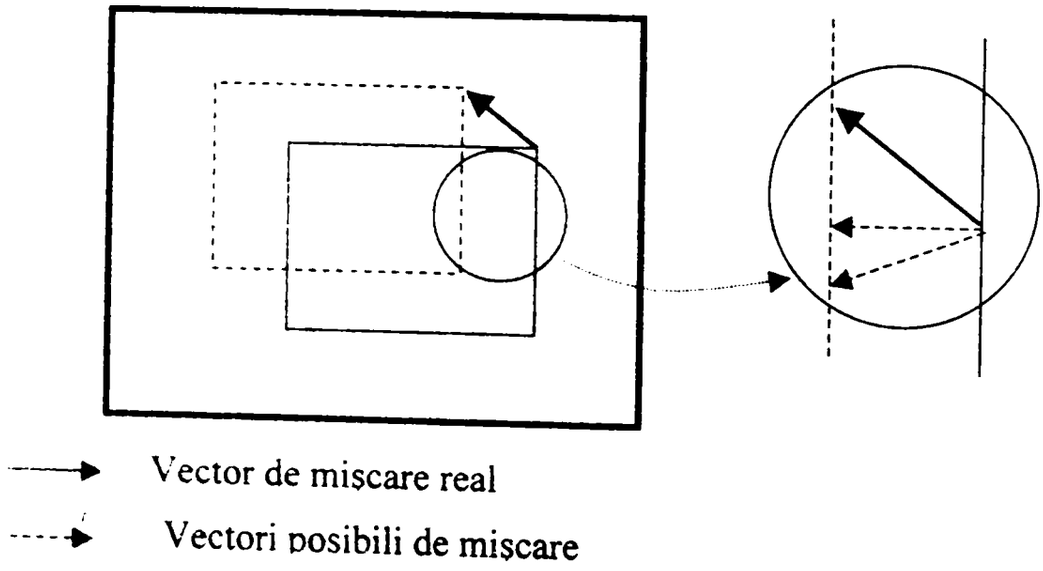


Fig 2.3 *Problema de apertură în estimarea mişcării*

Datorită acestui fapt, estimarea mişcării reprezintă o problemă de nedeterminare, fiind necesară o ecuaţie suplimenară pentru a obţine o soluţie unică pentru viteză (deplasare). Pentru determinarea celei de-a doua ecuaţii există mai multe soluţii prin impunerea unor constrângeri în imagine.

De asemenea, există situaţii în care componenta normală şi modelul global al deplasării curbei (conturului regiunii) nu au soluţii unice. Aceasta, spre deosebire

BUPT

de nedeterminarea locală, detemină o ambiguitate globală a deplasării. Cel mai evident exemplu îl constituie segmentul de linie aflat în translaţie, care, în absenţa informaţiei suplimentare relativă la capetele segmentului, prezintă o infinitate de soluţii privind modelul de aproximare al mişcării.

Această problemă apare ca relevantă în aplicaţiile care au ca scop principal reconstrucţia mişcării şi a structurii obiectelor tridimensionale, fiind determinată de pierderea datelor referitoare la mişcare prin proiecţia obiectelor tridimensionale. Această eroare de percepţie se manifestă şi în cazul percepţiei umane prin iluziile, deja cunoscute (elipsa în rotaţie apare ca o spirală, sau spirala circulară în rotaţie crează impresia unui cilindru), ce apar la interpretarea mişcării. Matematic, această problemă apare atunci când toate punctele curbei au vectorul tangenţial proporţional cu deplasarea acesteia:

$$\frac{dy}{dx} = \frac{v}{u} \tag{2.24}$$

unde u şi v reprezintă vitezele de deplasare ale curbei pentru axele x, respectiv y.

Dacă considerăm un model parametric al deplasării, ecuaţia (2.15), atunci obţinem următoarea ecuaţie diferenţială:

$$\frac{dy}{dx} = \frac{a_4 + a_5 x + a_6 y}{a_1 + a_2 x + a_3 y} \tag{2.25}$$

Considerând matricea A:

$$A = \begin{bmatrix} a_2 & a_3 \\ a_5 & a_6 \end{bmatrix} \tag{2.26}$$

şi presupunând că det$A \neq 0$, o schimbare a coordonatelor prin translaţie $x' = x - x_0$, $y' = y - y_0$ ne conduce la ecuaţia:

$$\frac{dy'}{dx'} = \frac{a_{21}x' + a_{22}y'}{a_{11}x' + a_{12}y'}, \tag{2.27}$$

unde

$$A\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_4 \end{bmatrix} \tag{2.28}$$

soluţia ecuaţiei diferenţiale este cunoscută [SAB 65].

Utilizarea modelului parametric asigură echivalenţa dintre ecuaţiile (2.25) şi (2.27) pentru orice transformare afină a curbei (conturului). În aceste condiţii analiza cazurilor ce rezultă depinde de valorile proprii ale matricii A.

Dacă valorile proprii ale matricei A sunt reale şi distincte $\lambda_1 \neq \lambda_2$, atunci soluţia ecuaţiei diferenţiale (2.27) este:

BUPT

$$y = cx^{\lambda_2/\lambda_1} \tag{2.29}$$

iar curbele ce satisfac această ecuaţie sunt următoarele:

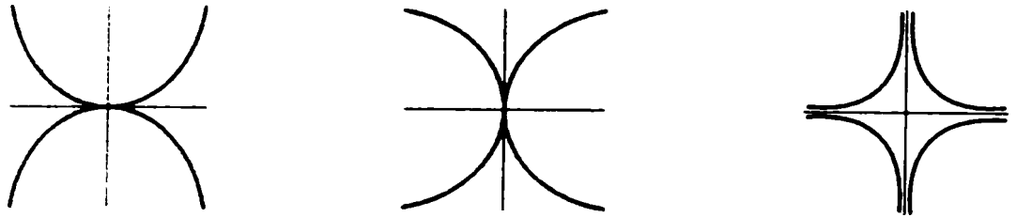


Fig. 2.4 *Familiile de curbe ce satisfac ecuaţia 2.29 şi conduc la ambiguităţi globale ale mişcării [TZI 94]*

Pentru valori proprii reale şi egale, soluţia ecuaţiei (2.27) este:

$$y = (\gamma + \delta \ln|x|)x \tag{2.30}$$

familia curbelor fiind cea din figura 2.5.

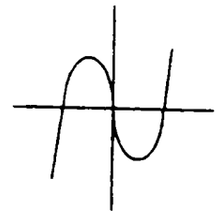


Fig. 2.5 *Familiile de curbe ce satisfac ecuaţia 2.30 şi conduc la ambiguităţi globale ale mişcării [TZI 94]*

În cazul în care matricea A prezintă valori proprii complexe, soluţia ecuaţiei (2.27) este:

$$\rho = ce^{a\theta} \tag{2.31}$$

reprezentând familia spiralelor logaritmice, figura 2.6.

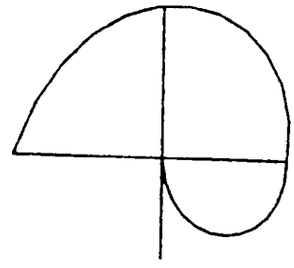


Fig. 2.6 *Familiile de curbe ce satisfac ecuaţia 2.31 şi conduc la ambiguităţi globale ale mişcării [TZI 94]*

BUPT

Deoarece, aceste ambiguităţi sunt inerente în cazul existenţei unor mişcări complexe (de exemplu rotaţie-translaţie), determinarea unei singure soluţii pentru vectorul de deplasare nu este posibilă, cu excepţia reducerii numărului gradelor de libertate ale modelului (impunerea condiţiilor de deplasare).

Se poate concluziona că, este posibilă estimarea deplasării spaţio-temporale a unui contur cu excepţia unor anumite familii de curbe care depind de modelul de mişcare. Această estimare este suficientă pentru a descrie mişcarea în compresia de imagini, deoarece deplasarea în cadrul unei regiuni delimitată de un contur închis este continuă şi liniară [TZI 94]. Astfel, datele referitoare la mişcare sunt compactate iar localizarea conturului şi a deplasării sale se realizează pentru un număr redus de puncte. În plus estimarea deplasării pentru un contur poate constitui condiţia iniţială pentru o estimare densă a deplasării (câmpul vectorilor de mişcare).

Primii ce au utilizat o condiţie suplimentară pentru traseul optic au fost Cafforio şi Rocca [CAF 76]. Ei au încercat descompunerea imaginii într-un fundal fix şi o regiune în mişcare prin utilizarea programării dinamice. În cadrul regiunii în mişcare, Ω, au presupus mişcarea ca fiind constantă şi au căutat minimizarea valorii:

$$\iint_{(x,y)\in\Omega}\left[(\nabla I)^T\,\mathbf{V}+\frac{\partial I}{\partial t}\right]^2 dxdy \qquad (2.32)$$

pentru determinarea valorii lui **V**, care realizează cea mai bună aproximare a regiunii Ω. Această abordare a fost de fapt generalizarea tehnicii bazate pe gradient propuse de Limb şi Murphy [LIM 75].

Horn şi Schunk [HOR 81], impun condiţia netezirii mişcării globale, care se bazează pe presupunerea că punctele vecine ale suprafeţelor în mişcare au o viteză identică. Pentru aceasta ei propun două metode:

1. utilizarea pătratelor amplitudinii gradientului traiectoriei optice

$$\|\nabla v_x\|^2=\left(\frac{\partial v_x}{\partial x}\right)^2+\left(\frac{\partial v_x}{\partial y}\right)^2 \text{ şi } \|\nabla v_y\|^2=\left(\frac{\partial v_y}{\partial x}\right)^2+\left(\frac{\partial v_y}{\partial y}\right)^2; \qquad (2.33)$$

2. pătratele Laplacienilor componentelor traiectoriei

$$\nabla^2 v_x=\frac{\partial^2 v_x}{\partial x^2}+\frac{\partial^2 v_x}{\partial y^2} \text{ şi } \nabla^2 v_y=\frac{\partial^2 v_y}{\partial x^2}+\frac{\partial^2 v_y}{\partial y^2}. \qquad (2.34)$$

Se alege soluţia minimizării deviaţiei lui **V** şi a amplitudinilor derivatelor spaţiale pentru întreaga imagine:

BUPT

$$\iint_{(x,y)\in\Omega}\left\{\left[(\nabla I)^T \mathbf{V}+\frac{\partial I}{\partial t}\right]^2+\lambda^2\left(\|\nabla v_x\|^2+\|\nabla v_y\|^2\right)\right\}dxdy \qquad (2.35)$$

unde parametrul λ determină importanţa termenului de constrângere relativ la traseul optic.

Spre deosebire de metoda Cafforio şi Rocca, care reprezintă o soluţie analitică pentru viteza regiunii care se mişcă, rezolvarea ecuaţiei (2.35) necesită o serie de iteraţii pentru convergenţa minimizării erorii vectorului **V** peste întreaga imagine.

O alternativă la traseul optic propus de Horn şi Schunk, a fost utilizarea aproximărilor de ordinul II. Această metodă se aplică pentru eliminarea limitărilor cauzate de variaţia liniară a intensităţii, prin acceptarea termenilor de ordinul doi în descompunerea în serie Taylor. Problema a fost abordată mai întâi de Nagel şi Enkelman [NAG 82] şi apoi de Nagel [NAG 83], cu scopul de a elimina limitările datorate modelului cu variaţia liniară a nivelurilor de gri.

Dacă se presupune că cele două componente ale vectorului ∇I satisfac ecuaţia traseului optic, se obţin două ecuaţii suplimentare pentru determinarea vectorului de viteză [TRE 84], [URA 88]:

$$\frac{\partial^2 I}{\partial x\partial t}=-\frac{\partial^2 I}{\partial x^2}v_x-\frac{\partial^2 I}{\partial x\partial y}v_y \qquad (2.36)$$

$$\frac{\partial^2 I}{\partial y\partial t}=-\frac{\partial^2 I}{\partial x\partial y}v_x-\frac{\partial^2 I}{\partial y^2}v_y \qquad (2.37)$$

Determinantul matricii ce corespunde acestui sistem de ecuaţii reprezintă curba Gaussiană a intensităţii suprafeţei la un moment dat. Astfel este posibilă estimarea vitezei cu o anumită siguranţă. De asemenea, asigurarea existenţei derivatei a doua reclamă aplicarea de preprocesări a imaginii (filtrarea trece jos). Întrucât aceste condiţii necesită absenţa rotaţiei sau a scalării, metodele sunt limitate atunci când se doreşte aplicarea lor în practică.

Întrucât semnalele video reale sunt versiuni eşantionate ale semnalului continuu $I(x,y;t)$, fiecare cadru constă dintr-un set de pixeli ale căror valori sunt definite pentru o anumită structură spaţială bidimensională. Din acest motiv cantitatea estimată de obicei este vectorul de deplasare $\hat{d}(x,y)=(\Delta x, \Delta y)$ şi nu vectorul viteză **V**. În acest caz se utilizează ecuaţia (2.30), care reprezintă versiunea discretă în timp a ecuaţiei (2.17).

$$I_n(x,y)=I_{n+1}((x,y)+\hat{d}(x,y)) \qquad (2.38)$$

BUPT

Deoarece scopul este determinarea vectorului de deplasare d, este utilă definirea diferenţei cadrelor compensate *DFD* (Displaced Frame Difference), pentru pixelul $\boldsymbol{x}=(x,y)$, aparţinând cadrului *n* şi având deplasarea $\hat{d}$:

$$D_{DFD} = I_{n+1}(\boldsymbol{x}+\hat{d}) - I_n(\boldsymbol{x}) \tag{2.39}$$

Această ecuaţie reprezintă aşa zisa diferenţă înainte utilizată în cazul estimării mişcării înainte. Există şi estimarea mişcării înapoi, ce utilizează diferenţa înapoi definită ca $I_{n-1}(\boldsymbol{x}+\boldsymbol{d}) - I_n(\boldsymbol{x})$, utilizată în schemele de codare cu predicţie pentru predicţia cadrului *n*, folosind cadrul *n*-1 şi vectorul de deplasare $\hat{d}$.

Pentru o deplasare reală $\hat{d}$, D_{DFD} este nulă. Astfel, prin anularea lui D_{DFD} în ecuaţia (2.15) obţinem versiunea discretă a ecuaţiei traseului optic. Obţinerea unei estimări robuste a deplasării $\hat{d}$, se poate realiza presupunând că aceasta este constantă pentru o regiune dată, similar cu metoda Cafforio şi Rocca [CAF 76]. În aceste condiţii cantitatea ce trebuie minimizată este:

$$\sum_{i=1}^{N} w(x_i)^2 D_{DFD_n}^2(x_i,\hat{d}) \tag{2.40}$$

în care *w* reprezintă funcţia de ponderare, care permite ponderarea celei mai mari cantităţi a pixelilor în centrul regiunii $\Omega=(x_1, x_2, \ldots, x_N)$

Dacă utilizăm relaţia (2.40) şi dezvoltăm în serie Taylor cu neglijarea coeficienţilor de ordin superior pentru $I_{n+1}(x+\hat{d})$, vom obţine varianta discretă a criteriului de minimizare Cafforio şi Rocca [CAF 76]:

$$\left(G\hat{d} - \Delta I\right)^T W^2 \left(G\hat{d} - \Delta I\right) \tag{2.41}$$

unde:

$$G = \left[\nabla I_{n+1}(x_1),\ldots,\nabla I_{n+1}(x_N)\right]^T \tag{2.42}$$

W este matricea diagonală având elementele $w(x_i)$,

$$\Delta I = \left[I_n(x_1) - I_{n+1}(x_1),\ldots,I_n(x_N) - I_{n+1}(x_N)\right]^T \tag{2.43}$$

În acest caz, minimizarea se reduce la minimizarea unei funcţii pătratice în $\hat{d}$, a cărei soluţie unică este:

$$d_{\min}(\Omega) = \left[G^T W^2 G\right]^{-1} G^T W^2 \Delta I \tag{2.44}$$

şi care se poate obţine în cazul în care $G^T W^2 G$ este inversabilă.

În concluzie putem spune că algoritmii bazaţi pe gradient sunt foarte bine adaptaţi codării semnalului video datorită obţinerii unui singur vector pentru fiecare regiune şi determinării, în general, rapide a mişcării. Principalul dezavantaj

BUPT

îl constituie condiţiile restrictive ce trebuiesc impuse, care în cele mai multe cazuri sunt greu de obţinut, necesitând o prelucrare prealabilă a imaginilor.

2.3.2 Tehnici de estimare bazate pe algoritmi recursivi pe pixel

Algorimii recursivi pe pixeli au cunoscut o largă utilizare pentru estimarea mişcării în codarea imaginilor. În cazul acestei abordări sunt incluse două faze: o estimare iniţială prin predicţie (aşa numita estimare apriorică) şi o corecţie în concordanţă cu diferenţa cadrelor compensate (aşa numita estimare posterioară). Predicţia este, în general, realizată pentru un punct precedent de pe aceiaşi linie, dar poate fi utilizată orice altă structură de predicţie. Ştiind că ecuaţia este neliniară, poate fi utilă realizarea câtorva iteraţii locale, pentru îmbunătăţirea estimării.

Principalul dezavantaj al metodelor bazate pe gradient îl constituie faptul că pentru deplasări mici, nu se pot ignora termenii de ordin superior din dezvoltarea în serie Taylor, ceea ce înseamnă că soluţia (ecuaţia (2.44)) poate să nu fie validă. Pentru a depăşi această limitare, Netravali şi Robins [NET 79] au introdus primul algoritm de estimare recursivă bazată pe pixel. Ei au presupus existenţa unei deplasări estimate d_k şi au calculat un termen de actualizare u_k:

$$d_{k+1} = d_k + u_k \tag{2.45}$$

Termenul u_k provine din descompunerea în serie Taylor faţă de d_k a valorii I_{n+1}:

$$I_{n+1}(x+\hat{d}_{k+1}) = I_{n+1}(x+\hat{d}_k) + \left(\nabla I_{n+1}(x+\hat{d}_k)\right)^T u_k + e_{n+1}(x+\hat{d}_k) \tag{2.46}$$

în care termenul de eroare e_{n+1} descreşte odată cu reactualizarea lui u_k. Termenul de reactualizare poate fi determinat cu ajutorul unei ecuaţii similare cu ecuaţia 2.44, iar această reactualizare poate fi eficientă prin calcularea cu ajutorul unui algoritm cu convergenţă rapidă pe baza pătratului D_{DFD}.

$$u_k = -\frac{\varepsilon}{2}\nabla_d\left[D_{DFD}(x,d_k)\right]^2 \tag{2.47}$$

Termenul recursiv pe pixel a fost utilizat pentru o variantă particulară a algoritmului iterativ, cu aplicare la codarea video. Această variantă necesită estimarea şi la decodare, ceea ce are avantajul de a nu fi necesară transmiterea câmpului de mişcare. Economia rezultată prin eliminarea câmpului de mişcare din informaţia transmisă, a fost posibilă prin definirea unei regiuni locale constante, care include numai pixelii precedenţi pixelului curent şi utilizează estimarea finală a pixelului precedent ca estimare iniţială pentru pixelul curent. Astfel prin aceste restricţii se obţin erori de predicţie mari, care permit scăderea ratei de transmisie

BUPT

prin omiterea transmiterii informaţiei de mişcare. De aceea algoritmii recursivi pe pixel sunt utilizaţi în sens larg pentru descrierea algoritmilor iterativi bazaţi pe gradient.

Walker şi Rao au propus un algoritm focalizat pe alegerea coeficientului ε, care, în ultimă instanţă, poate permite îmbunătăţirea convergenţei algoritmului (dacă ε are o valoare mare se obţine o convergenţă rapidă dar estimarea este însoţită de zgomot, în timp ce o valoare mică asigură o estimare precisă) [WAL 84].

Pentru eliminarea dependenţei lui u_k de parametrul ε Bergmann [BER 82], introduce în locul acestuia un numărător ce corespunde derivatei a doua a pătratului diferenţei deplasării dintre cadre. Rezultă un termen de îmbunătăţire:

$$u_k = -\frac{\nabla_d \left[D_{DFD}(x,\hat{d})\right]^2}{\frac{\partial^2}{\partial \hat{d}^2}\left[D_{DFD}(x,\hat{d})\right]^2} \tag{2.48}$$

În [MUS 85] se prezintă câţiva algoritmi de estimare ai mişcării recursivi pe pixeli, realizându-se totodată şi o evaluare a acestora. Cea mai apropiată reprezentare de sistemul percepţiei, a fost , probabil, cea propusă de Biemond [BIE 87], care, pe lângă neglijarea termenilor de ordin superior din ecuaţia (2.46), a considerat procesul ca fiind un proces stohastic. Aceasta a condus la utilizarea unei filtrări Wiener [RAB 75], [HOF 96] pentru estimarea lui u_k, care se obţine cu în acest caz cu relaţia:

$$u_k = \left[G_k^T R_e^{-1} G_k + R_d^{-1}\right]^{-1} G_k^T R_e^{-1} \Delta I \tag{2.49}$$

în care G_k şi ΔI sunt termenii definiţi de ecuaţiile (2.42) respectiv (2.43), iar R_d având dimensiunea 2×2 şi R_e de dimensiune $N \times N$, matricile de covarianţă ale deplasării, respectiv erorii.

Ecuaţia termenului de reactualizare (ecuaţia (2.41)) poate fi simplificată dacă presupunem că termenul de eroare este necorelat şi are varianţa identică pentru fiecare pixel, astfel încât $\boldsymbol{R}_e = \sigma_e^2 \boldsymbol{U}$. În acest caz, termenul de reactualizare devine:

$$u_k = \left[G_k^T G_k + \sigma_e^2 R_d^{-1}\right]^{-1} G_k^T \Delta I \tag{2.50}$$

Dacă comparăm ecuaţia (2.44) cu ecuaţia (2.50), se poate observa că singura diferenţă ce apare o reprezintă adăugarea termenului $\sigma_e^2 R_d^{-1}$ matricii $K = G^T G$. Matricea K este neinversabilă în cazul în care gradientul mişcării este constant pentru întreaga regiune şi, deci, nu se poate determina o soluţie unică pentru u_k.

BUPT

Această situaţie corespunde problemei de apertură din figura 2.3. Rezolvarea problemei conduce către două soluţii:

1. încorporarea unei informaţii despre câmpul de deplasare în matricea $\boldsymbol{R}_d$ pentru a transforma matricea K într-o matrice inversabilă, obţinându-se un vector de reactualizare unic;
2. în cazul absenţei unei informaţii despre câmpul de mişcare ($R_d^{-1} = 0$), reactualizarea se defineşte în direcţia valorilor proprii nenule (în direcţie paralelă cu gradientul constant). Aceasta poate fi determinată utilizând relaţia obţinută pentru valoarea singulară a lui K [KOK 96]

$$u_k = \frac{1}{\lambda_{\max}} e_{\max}^T G_k^T \Delta I e_{\max} \tag{2.51}$$

unde $\lambda_{\max}$ reprezintă valoarea proprie diferită de zero iar $e_{\max}$ corespunde vectorului propriu al matricii K.

Această strategie a fost utilizată pentru prima dată de către Biemond [BIE 87], care a presupus pentru matricea R_d o formă diagonală $\sigma_d^2 U$. Pentru îmbunătăţirea performanţelor privind convergenţa şi rezistenţa la zgomot s-a utilizat un termen de stabilizare (amortizare) μU, unde:

$$\mu = \frac{\sigma_e^2}{\sigma_d^2} \tag{2.52}$$

În [BOR 91], s-a trecut la alegerea adaptivă a factorului μ funcţie de K. În acest caz temenul de amortizare utilizat a fost:

$$\mu = \frac{\lambda_{\max}}{\lambda_{\min}} \|\Delta I\| \tag{2.53}$$

Motivul a fost obţinerea unei stabilităţi mai mari, pentru cazurile de nedeterminare în care raportul valorilor proprii este mare şi a unei stabilităţi ce tinde către zero în cazul convergenţei algoritmului şi a scăderii erorii $\|\Delta I\|$.

O alternativă la abordarea recursivă prin utilizarea filtrării Wiener este utilizarea estimării Bayesiene. Aceasta reprezintă o tehnică statistică, care este foarte bine adaptată identificării parametrilor modelului pentru un proces de observare zgomotos. Această abordare a fost realizată de Konrad şi Dubois [KON 92], pornind de la presupunerea că semnalul video corespunde modelului de translaţie dat de ecuaţia (2.38), dar deformată de un termen de eroare:

$$I_{n+1}(x + \hat{d}) = I_n(x) + e_{n+1}(x) \tag{2.54}$$

Termenul de eroare diferenţială e, este o combinaţie între zgomotul produs în procesul formării imaginii şi eroarea modelului de translaţie. Scopul principal al

BUPT

acestei abordări îl constituie determinarea maximului estimării posterioare MAP (Maximum A Posteriori) a modelului de translaţie pentru $\hat{d}$ între două cadre. Legea lui Bayes arată că densitatea de probabilitate este determinată de:

$$p(\hat{d}|\ I_{n+1}(x), I_n(x)) = \frac{p(I_n(x)|\ \hat{d}, I_{n+1}(x))p(\hat{d})}{p(I_n(x)|\ I_{n+1}(x))} \tag{2.55}$$

presupunând că deplasarea este independentă de imagine. Deoarece numitorul este independent de $\hat{d}$, estimarea MAP pentru $\hat{d}$ este dată de:

$$\hat{d} = \max\left\{p(I_n(x)|\ \hat{d}, I_{n+1}(x))p(\hat{d})\right\} \tag{2.56}$$

Dacă se utilizează presupunerea că traiectoria locală este constantă, estimarea MAP pentru întreaga regiune Ω este:

$$\hat{d}(\Omega) = \max\left\{p(I_n(\Omega)|\ \hat{d}, I_{n+1}(\Omega))p(\hat{d})\right\} \tag{2.57}$$

Expresia probabilităţii se poate determina presupunând că eroarea pentru fiecare pixel se poate scrie ca o distribuţie gaussiană, având matricea de covarianţă R_e:

$$p(e_{n+1}) \approx \exp\left(-\frac{1}{2} e_{n+1}^T R_e^{-1} e_{n+1}\right) \tag{2.58}$$

unde:

$$e_{n+1} = \left[e_{n+1}(x_1), \ldots, e_{n+1}(x_N)\right]^T \tag{2.59}$$

Utilizând ecuaţia (2.54) se poate rescrie funcţia de probabilitate în felul următor:

$$p(I_n(\Omega)|\ \hat{d}, I_{n+1}(\Omega)) \approx \exp\left[-\frac{1}{2} D_{DFD}{}^T(\Omega, \hat{d}) R_e^{-1} D_{DFD}(\Omega, \hat{d})\right] \tag{2.60}$$

unde:

$$D_{DFD}(\Omega, \hat{d}) = \left[D_{DFD}(x_1, \hat{d}), \ldots, D_{DFD}(x_N, \hat{d})\right]^T \tag{2.61}$$

În aceste condiţii, pentru determinarea prin estimarea MAP este necesară o informaţie referitoare la $p(\hat{d})$. Simoncelli [SIM 91] alege o distribuţie gaussiană cu medie nulă, având covarianţa R_d, obţinând pentru estimarea MAP o formulare asemănătoare celei date de ecuaţia (2.49):

$$\hat{d}(\Omega) = \min\left\{(G\hat{d} - \Delta I)^T R_e^{-1}(G\hat{d} - \Delta I) + \hat{d}^T R_d^{-1} \hat{d}\right\} \tag{2.62}$$

unde G şi ΔI sunt cei definiţi în relaţiile (2.42) şi (2.43). Soluţia obţinută în acest caz este cea generată de ecuaţia Wiener (relaţia (2.49)), demonstrându-se

BUPT

echivalenţa dintre metodele bazate pe filtrele Wiener şi cele bazate pe decizia Bayesiană. O altă similitudine o reprezintă matricea de covarianţă a estimatei MAP pentru $\hat{d}$ care este dată de $\left[G^T R_e^{-1} G + R_d^{-1}\right]^{-1}$.

Dacă vom considera estimatorul MAP Gaussian neliniar [SIM 91], [BIE 87]

$$\hat{d}(\Omega) = \min\left\{D_{DFD}{}^T(\Omega,\hat{d})R_e^{-1}D_{DFD}(\Omega,\hat{d}) + \hat{d}^T R_d^{-1}\hat{d}\right\} \qquad (2.63)$$

şi presupunând că în modelul translaţional eroarea este independentă între pixelii aceleiaşi regiuni (R_e este matricea diagonală conţinând elementele σ_i^2) se obţine:

$$\hat{d}(\Omega) = \min\left\{\sum_{i=1}^{N}\left[\frac{1}{\sigma_i^2}D_{DFD}(x_i,\hat{d}\right] + \hat{d}^T R_d^{-1}\hat{d}\right\} \qquad (2.64)$$

Rezultatul reprezintă versiunea modificată a criteriului de minimizare dat de ecuaţia (2.60) în care sunt înglobate explicit cunoştinţele anterioare referitoare la $\hat{d}$. Dacă nu există cunoştinţe anterioare ($R_d^{-1} = 0$), estimarea neliniară MAP devine un simplu algoritm de corespondenţă bazat pe regiuni. De asemenea, este posibilă incorporarea unor cunoştinţe mai complexe pentru obţinerea unui estimator MAP mai inteligent. În [GEM 84] au fost introduse modele Markoviene pentru prelucrarea imaginilor, care reprezintă o liniarizare a constrângerilor globale nu foarte diferită de cea propusă în [KON 88]. La fel ca şi în cazul metodelor bazate pe gradient, găsirea câmpului estimat prin metoda MAP utilizând un model Markov, necesită un mare volum de calcule.

2.3.3 Algoritmi de corespondenţă a blocurilor

Ideea de divizare a imaginii în regiuni şi găsirea deplasării $\hat{d}$, care corespunde cel mai bine modelului translaţional descris de ecuaţia (2.30) pentru întreaga regiune, a fost introdusă de Jain [JAI 81]. Astfel, în locul utilizării unei dezvoltări în serie Taylor, pentru a avea acces la termenul necunoscut $\hat{d}$ din derivata intensităţii se utilizează o procedură directă de căutare a lui $\hat{d}$. Astfel punctul de plecare îl reprezintă D_{DFD}, definită de ecuaţia (2.39) dar extinsă la întreaga regiune Ω:

$$D_{DFD}(\Omega,\hat{d}) = \sum_{x\in\Omega} g\left(I_{n+1}(x+\hat{d}) - I_n(x)\right) \qquad (2.65)$$

în care funcţia g reprezintă funcţia de distorsiune, pozitivă şi crescătoare, de cele mai multe ori reprezentând un operator pătratic, care realizează transformarea lui D_{DFD} în diferenţă medie pătratică. De asemenea, poate fi utilizat un operator modul ce conduce la o medie a diferenţelor absolută. În aceste condiţii translaţia locală a regiunii Ω este dată de:

BUPT

$$\hat{d}(\Omega) = \arg \min_{\hat{d} \in R} \left\{ D_{DFD}(\Omega, \hat{d}) \right\} \tag{2.66}$$

unde R reprezintă setul de candidaţi predefiniţi. În aceste condiţii, zona de căutare o reprezintă setul de pixeli $\left\{ x + \hat{d}, x \in \Omega, d \in R \right\}$, iar cea mai simplă implementare se obţine dacă regiunea Ω este un bloc pătrat de pixeli. Mulţimea blocurilor este aleasă astfel încât să acopere întregul cadru. Schema de căutare cunoscută sub denumirea de "căutare completă" (Full Search), este reprezentată în figura 2.7.

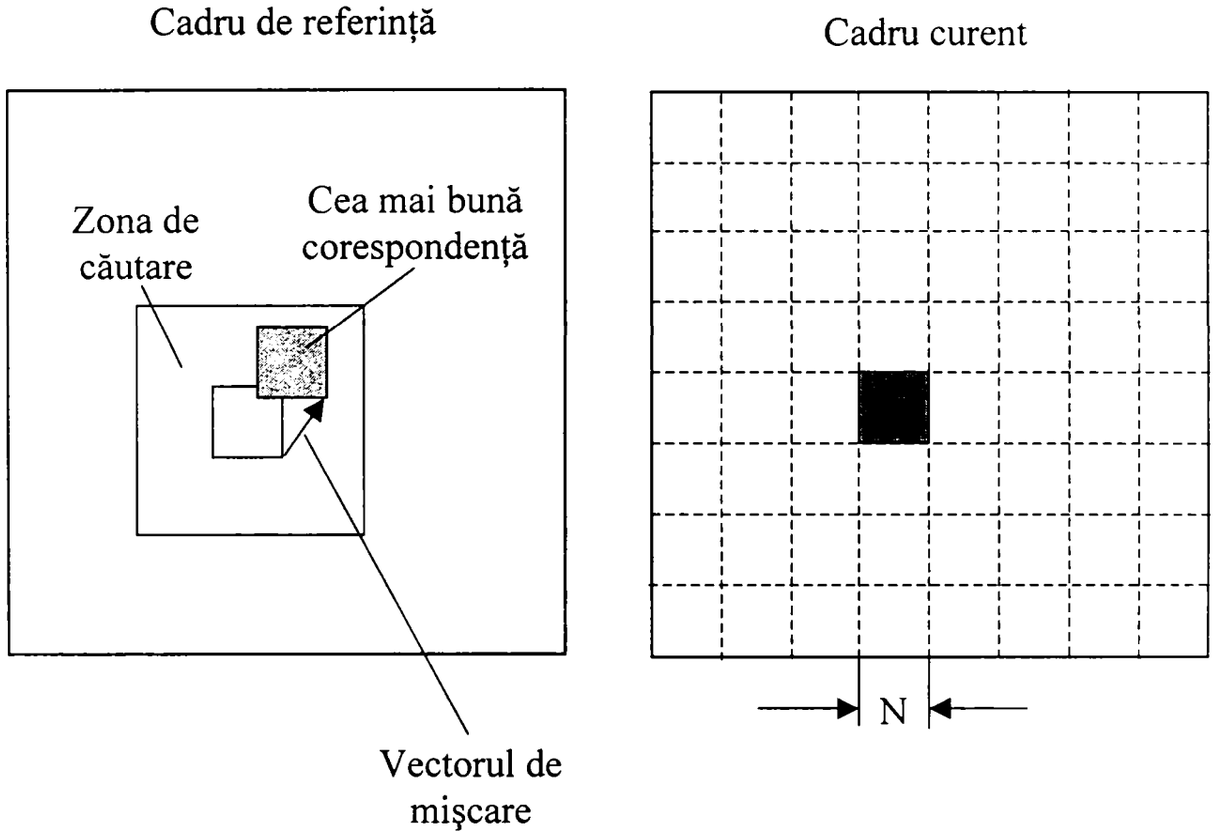


Fig 2.7 *Algoritmul de corespondenţă a blocurilor pentru căutare completă*

Uzual vectorul de deplasare este un multiplu întreg al perioadei de eşantionare sau al jumătăţii acestei perioade. Dacă blocul are N linii şi N coloane, mărimea suprafeţei de căutare, pentru o deplasare egală cu multiplu întreg al perioadei de eşantionare, este $(N + 2\Delta_v)(N + 2\Delta_h)$, iar numărul de vectori posibili, este $(2\Delta_h + 1)(2\Delta_v + 1)$.

Avînd definită suprafaţa de căutare, criteriul de optimizare cel mai des utilizat este media valorilor absolute a estimării, în cazul translaţiei, ecuaţia (2.66) putând fi scrisă sub forma:

BUPT

$$\hat{d} = \min_{\substack{(x,y)\in R \\ |x|\le \Delta_h,\ |y|\le \Delta_v}} \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \left| D_{DFD}(i,j;x,y) \right| \tag{2.67}$$

unde x şi y definesc cei doi parametri ai vectorului de mişcare $\hat{d}$.

Dacă se utilizează operatorul pătratic, ecuaţia (2.66) are următoarea formă.

$$\hat{d} = \arg\min_{\substack{(x,y)\in R \\ |x|\le \Delta_h,\ |y|\le \Delta_v}} \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \left(D_{DFD}(i,j;x,y) \right)^2 , \tag{2.68}$$

iar în cazul în care se utilizează operatorul pătratic normalizat:

$$\hat{d} = \arg\min_{\substack{(x,y)\in R \\ |x|\le \Delta_h,\ |y|\le \Delta_v}} \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \left(D_{DFD}(i,j;u,v) \right)^2}{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} I^2(i,j)} \tag{2.69}$$

Dintre aceste criterii, cel mai des utilizat, este cel ce utilizează valoarea absolută (ecuaţia (2.67)), deoarece este mai economic, chiar dacă are anumite dezavantaje în ceea ce priveşte estimarea vectorului de deplasare. Cel de al doilea criteriu, în care se utilizează funcţia pătratică, a fost şi el utilizat. Algoritmul este similar cu tehnicile bazate pe corelaţie, dezvoltate la începuturile înregistrărilor de imagini [PRA 74]. Acest lucru rezultă din observaţia că, minimizarea diferenţei pătratelor este aproximativ echivalentă cu maximizarea, pentru toţi vectorii aparţinând spaţiului R, a valorii $I_n(x)I_{n+1}(x+\hat{d})$.

Utilizarea algoritmului cu căutare completă permite determinarea oricărei deplasări cu condiţia ca spaţiul de căutare, R, să fie suficient de mare. De asemenea, întrucât se realizează o mediere pentru întregul bloc, această tehnică este suficient de robustă la zgomote. Proprietăţile enumerate, precum şi simplitatea (permite implementarea utilizând arhitecturi paralele [KOM 89], [DOW 94]) au făcut acest algoritm ca fiind unul din cele mai utilizate în codoarele moderne H261 [BOJ 97], MPEG [VRA 98], [ARA 95], [LeG 91].

Analiza algoritmului ne arată că precizia de estimare este dependentă de setul de direcţii posibile R. O implementare simplă utilizează numai deplasări întregi. Dacă se doreşte căutarea, pentru a estima deplasări având o precizie mai mare decât structura de eşantionare, atunci este necesară o interpolare pentru a obţine valori intermediare ale pixelilor din imagine. Se poate folosi o intepolare biliniară

BUPT

[YOU 93] sau se crează o suprafaţă de corespondenţă continuă şi se realizează aproximarea în punctele discrete de interes [ANA 89].

Principalul dezavantaj al algoritmului de estimare prin corespondenţa blocurilor, bazat pe căutare completă, este acela al cantităţii încă foarte mari de calcule. De exemplu, pentru un bloc Ω având $N \times N$ pixeli, un algoritm cu căutare completă având r candidaţi necesită un volum de $3N^2r$ operaţii de adunare şi multiplicare. Aşa cum am arătat, uneori este necesară o precizie de estimare mai bună de un pixel, iar Girod [GIR 93] a demonstrat că, din punctul de vedere al erorii de predicţie, precizia de ¼ pixeli este cea mai bună.

Pentru o deplasare maximă Δ_x în fiecare direcţie, R conţine $(8\Delta_x + 1)^2$ vectori posibili (se consideră o precizie de ¼ pixeli), ceea ce ne conduce la un număr total de operaţii egal cu $3N^2(8\Delta_x + 1)^2$ pentru fiecare bloc. Datorită acestei complexităţi de calcul au fost elaboraţi algoritmi pentru căutare rapidă care reduc numărul de evaluări pentru obţinerea minimului.

2.3.3.1 Căutarea logaritmică bidimensională [JAI 81]

Acest logaritm se bazează pe ipoteza că criteriul utilizat pentru minimizare este convex şi din acest motiv valoarea estimată creşte monoton odată cu distanţa, în relaţie cu vectorul ce ne conduce la valoarea minimă. Pentru fiecare pas, căutarea se realizează numai pentru 5 puncte, în două direcţii principale: orizontală şi verticală. Pentru primul pas, sunt calculate patru puncte între centrul suprafeţei şi margine. Dacă minimul este la mijloc, între cele 5 puncte, atunci suprafaţa de căutare este scăzută prin înjumătăţire. Această procedură continuă până cînd se obţine o suprafaţă de 3×3. În ultimul pas sunt testate toate cele 9 poziţii, alegându-se pentru vectorul de deplasare poziţia minimă (figura 2.8 a).

2.3.3.2 Căutare prin creşterea preciziei [KOG 81]

Algoritmul mai este cunoscut şi sub denumirea de algoritm în trei paşi (Three step algorithm) şi se bazează, de asemenea, pe ipoteza de convexitate. În primul pas se testează vectorul de deplasare peste suprafaţa de căutare cu precizie mică. După determinarea direcţiei vectorului de deplasare ce minimalizează criteriul între cele 9 puncte posibile, se continuă procesul prin creşterea preciziei până la obţinerea preciziei de un pixel. Întrucât algoritmul în trei paşi devine ineficient pentru estimarea deplasărilor reduse, în [LI 94] a fost propus un algoritm îmbunătăţit care utilizează calculul deplasării pentru opt puncte suplimentare în primul pas al algoritmului, conducând în acest fel la scăderea erorii în cazul compensării mişcării (figura 2.8 b).

BUPT

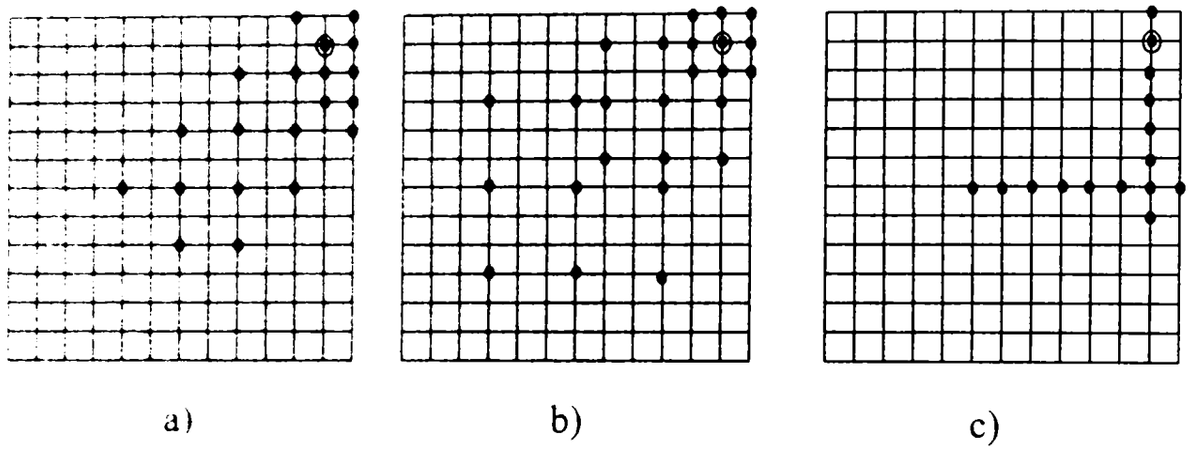


Fig 2.8 *Principalii algoritmi de corespondenţă cu căutare rapidă*

2.3.3.3 Căutarea în direcţii conjugate [SRI 85]

Pornind de la o deplasare nulă (primul punct este plasat în centrul ferestrei de căutare), se estimează mai întâi minimul pentru o direcţie. Odată determinată poziţia minimului, se realizează aceeaşi căutare pentru direcţia perpendiculară pornind de la ultima poziţie a minimului (figura 2.8 c). O variantă a acestui algoritm constă în omiterea ultimului pas, cu restricţii de căutare în două direcţii principale separate. Cele trei metode rapide de căutare a minimului prezentate mai sus sunt ilustrate în figura de mai jos.

2.3.3.4 Un nou algoritm de căutare rapidă în patru paşi [ALE 98a]

Pornind de la algoritmii cu căutare rapidă prezentaţi mai sus, autorul tezei de doctorat a elaborat un algoritm de estimare a mişcării în patru paşi (figura 2.9) [ALE 98a], cu performanţe mai bune decât algoritmul original în trei paşi [KOG 81] şi asemănătoare cu algoritmul în trei paşi îmbunătăţit [LI 94]. Algoritmul propus este mai robust din punct de vedere al acurateţei determinării vectorilor de mişcare şi are performanţe asemănătoare pentru secvenţe de imagini ce conţin schimbări de plan (zoom) sau mişcări rapide. De asemenea rezultatele experimentale au arătat că algoritmul propus este foarte apropiat ca performanţe de algoritmul cu căutare completă, în termeni ai erorii medii pătratice.

Pentru îmbunătăţirea probabilităţii de obţinere a minimului global se utilizează în primul pas mai multe puncte, la fel ca şi în cazul algoritmului de căutare în trei paşi îmbunătăţit. Algoritmul în patru paşi poate fi descris astfel:

Pasul I. Se calculează deplasarea pentru nouă puncte ale unei reţele de 5×5 pixeli. Dacă minimul este găsit în centrul ferestrei de căutare se face un salt la pasul IV, în caz contrar se trece la pasul următor.

Pasul II. Se menţine fereastra de căutare la dimensiunea 5×5, centrul acesteia fiind amplasat în punctul în care s-a găsit minimul pasului anterior.

BUPT

Numărul suplimentar de puncte de calcul sunt: 5 dacă minimul pasului anterior se află într-unul dintre colţurile ferestrei de căutare sau 3 dacă minimul se află situat la mijlocul direcţiei orizontale sau verticale. Dacă minimul este în centrul ferestrei se trece la pasul IV, iar dacă nu se continuă cu pasul III.

Pasul III. Este identic cu pasul II cu observaţia că, indiferent de rezultat se trece la pasul IV.

Pasul IV. Se reduce fereastra la dimensiunea 3×3 obţinându-se ca punct final al estimării minimul celor nouă puncte.

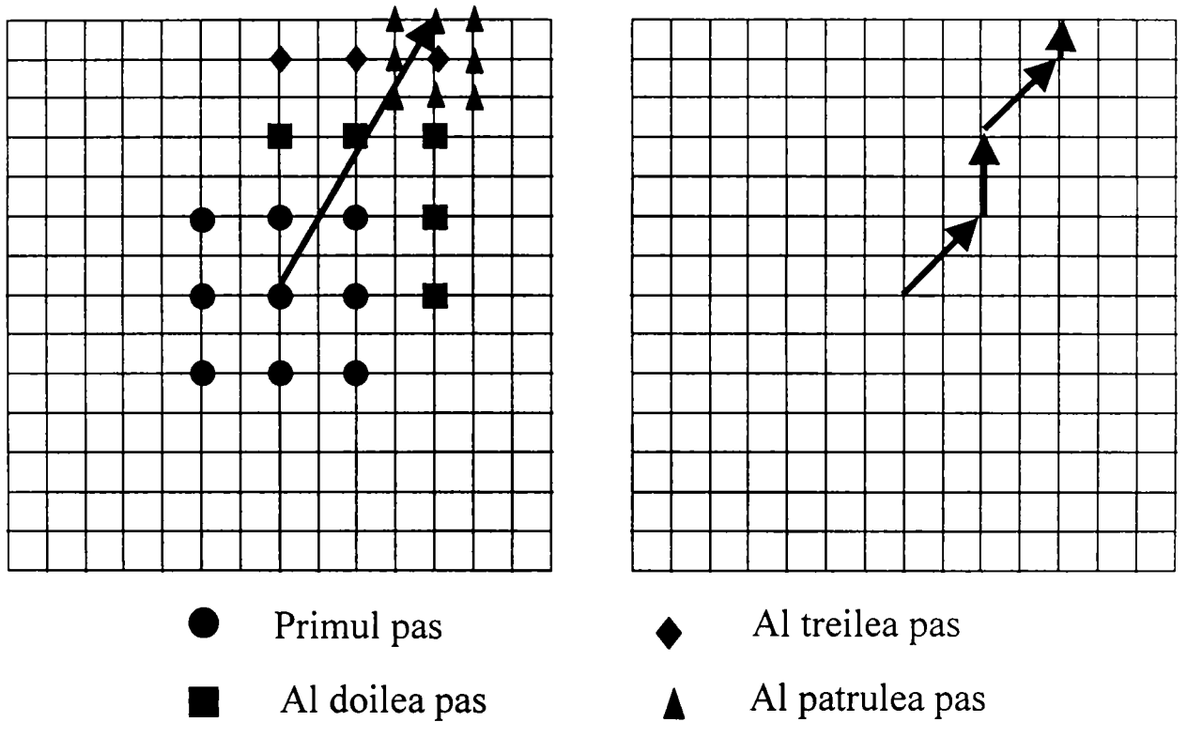


Fig 2.9 *Algoritmul de căutare în patru paşi*

Avantajul acestui algoritm constă în faptul că, se acoperă o fereastră de căutare de dimensiune 15×15 utilizând numai ferestre având dimensiunile de 5×5 şi 3×3 pentru ultimul pas în care este estimat vectorul de mişcare. Prin implementarea celor patru algoritmi am putut realiza compararea lor în termeni ai: complexităţii de calcul, erorii medii pătratice, vitezei de estimare, preciziei de determinare a vectorului de mişcare.

BUPT

Tabelul 2.2 Numărul de puncte şi numărul de paşi de calcul pentru diferiţi algoritmi rapizi de estimare a mişcării, în cazul unei ferestre de căutare de 15×15

Algoritm de estimare a mişcării	Nr. de puncte maxim	Nr. de puncte minim	Nr. de paşi maxim
Cu căutare completă	225	225	225
Căutare logaritmică bidimensională [JAI 81]	21	5	7
Căutare în trei paşi (creşterea preciziei) [KOG 81]	25	9	3
Căutarea în trei paşi îmbunătăţită [LI 94]	33	17	3
Căutare în direcţii conjugate [SRI 85]	17	3	14
Căutare în patru paşi propus de autorul tezei de doctorat [ALE 98a]	27	9	4

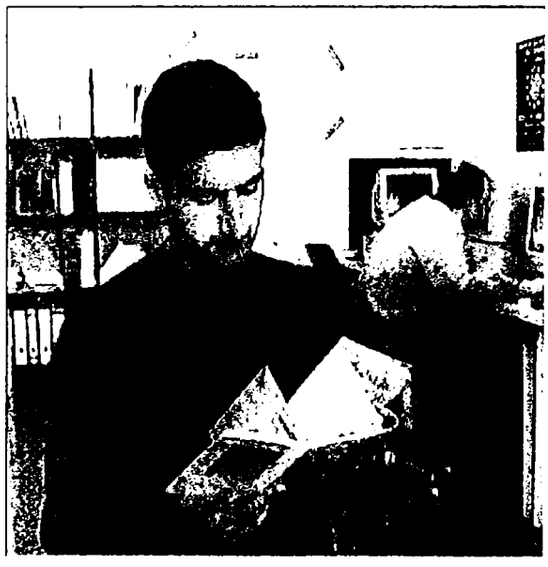
Tabelul 2.3 Eroarea medie pătratică obţinută prin medierea rezultatelor pentru 9 cadre, pentru algoritmii rapizi de estimare a mişcării prezentaţi mai sus

Algoritm de estimare a mişcării	MISSA	Tenis
Căutare logaritmică bidimensională [JAI 81]	450,6	346,57
Căutare în trei paşi îmbunătăţită [LI 94]	330,78	203,94
Căutare în direcţii conjugate [SRI 85]	465,47	356,76
Căutare în patru paşi propus de autorul tezei de doctorat [ALE 98a]	329,8	197,20

BUPT

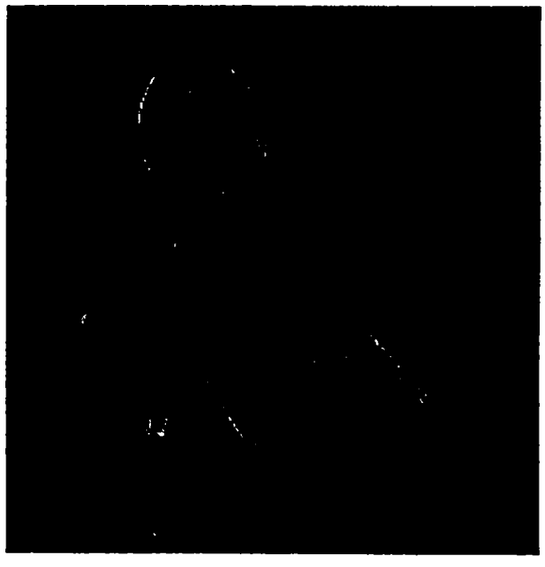
Tabelul 2.4 Factor de rapiditate determinat prin medierea pentru 9 cadre a raportului între numărul de puncte necesar pentru algoritmul cu căutare completă şi cele pentru algoritmii rapizi de estimare a mişcării prezentaţi

Algoritm de estimare a mişcării	MISSA	Tenis
Căutare logaritmică bidimensională [JAI 81]	9	9
Căutare în trei paşi îmbunătăţită [LI 94]	10,86	12,45
Căutare în direcţii conjugate [SRI 85]	9,26	9,85
Căutare în patru paşi propus de autorul tezei de doctorat [ALE 98a]	11,65	12,98

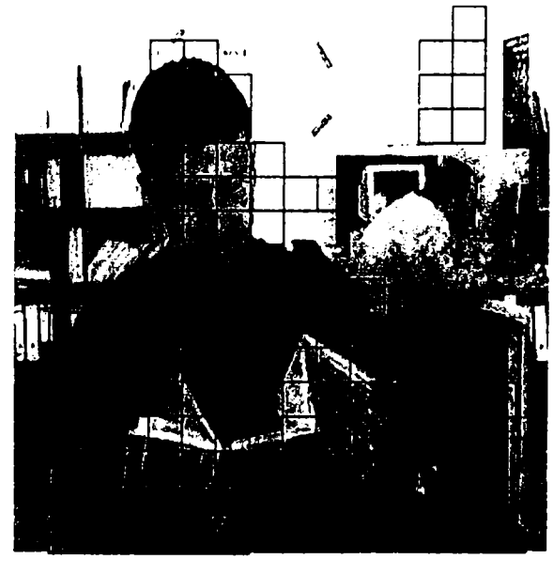


Cadru precedent

Cadru curent



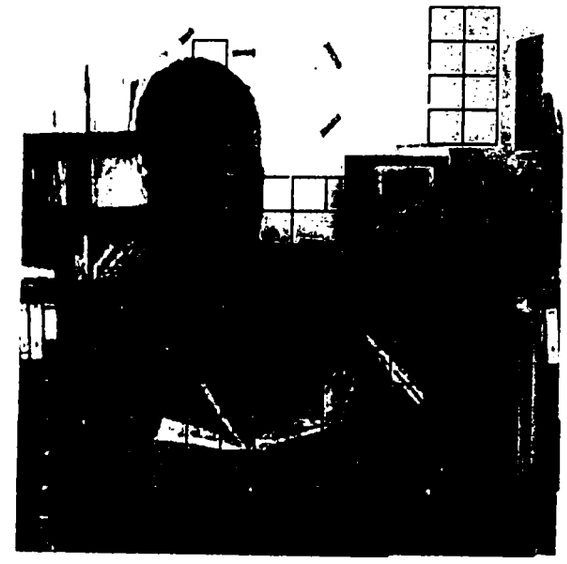
Diferenţa dintre cadre



Estimarea mişcării prin metoda de căutare logaritmică

BUPT

Estimarea mişcării prin metoda de căutare în trei paşi



Estimarea mişcării prin metoda de căutare în direcţii conjugate

Estimarea mişcării prin metoda de căutare în patru paşi

Fig. 2.10 *Estimarea mişcării cu ajutorul diferitelor metode pentru secvenţa MISSA*

Tabelul 2.2 prezintă comparativ, algoritmii rapizi de estimare a mişcării, din punctul de vedere al numărului de paşi minimi şi maximi necesari pentru determinarea vectorului de mişcare a unui bloc cu o deplasare maximă în intervalul (−7, +7) pixeli şi o fereastră de căutare având dimensiunea maximă de 15×15 pixeli.

Tabelul 2.3 prezintă, pentru secvenţele MISSA şi TENIS, eroarea medie pătratică obţinută prin medierea rezultatelor pentru 9 cadre. Se observă că noul algoritm propus de autorul tezei de doctorat [ALE 98a], obţine performanţe mai

BUPT

bune în ceea ce priveşte eroarea medie pătratică, comparativ cu algoritmii rapizi implementaţi, cu o uşoară creştere a complexităţii calculelor faţă de algoritmul în trei paşi îmbunătăţit.

Tabelul 2.4 prezintă comparativ rezultatele obţinute în ceea ce priveşte rapiditatea algoritmilor. Factorul de rapiditate se obţine prin raportarea numărului de puncte de calcul necesare pentru algoritmul cu căutare completă şi numărul de puncte de căutare pentru algoritmii comparaţi. Rezultatele obţinute confirmă rapiditatea algoritmului propus, întrucât factorul de rapiditate cel mai mare este obţinut pentru acesta. De asemenea, se observă faptul că rezultatele sunt asemănătoare cu cele ale algoritmului în trei paşi îmbunătăţit.

În figura 2.10 sunt prezentate rezultatele estimării mişcării pentru cei patru algoritmi analizaţi în tabelul 2.2.

2.3.4 Tehnici de estimare bazate pe reprezentarea în frecvenţă

Tehnicile bazate pe reprezentarea în frecvenţă se referă la metodele care realizează transformarea imaginii într-un alt domeniu şi estimează mişcarea prin coeficienţii transformatei. De cele mai multe ori domeniul transformat este cel Fourier, dar pentru a putea păstra generalitatea, vom folosi denumirea de metode în domeniul transformatelor.

Trăsătura fundamentală a transformatei Fourier tridimensionale exploatată la estimarea mişcării este, în domeniul spaţio-temporal, concentrarea energiei într-un singur plan în domeniu frecvenţă [NAF 95]. În cazul frecvenţelor spaţiale (ω_x, ω_y) şi frecvenţei temporale ω_t, ecuaţia ce se obţine pentru traseul optic este:

$$u\omega_x + v\omega_y + \omega_t = 0 \qquad (2.70)$$

unde u, v reprezintă componentelele vectorului de mişcare.

Analiza pentru întregul domeniu permite estimarea traseului optic (u,v). Deoarece semnalul tridimensional nu este staţionar, este necesară o analiză simultană spaţio-temporală şi în frecvenţă. Câteva alternative pentru această analiză sunt:

- analiza locală Fourier,
- transformarea Wigner [JAC 87],
- transformarea Gabor [HEE 88].

2.3.4.1 Intercorelaţia în domeniul transformat

Aşa cum s-a menţionat în paragraful 2.3.3, metodele de corespondenţă bazate pe găsirea deplasării blocurilor sunt echivalente cu metoda maximizării corelaţiei. În acest scop este necesară localizarea intercorelaţiei maxime pentru regiunea Ω:

$$C_n(\Omega, \hat{d}) = \sum_{x \in \Omega} I_n(x) I_{n+1}(x + \hat{d}) \qquad (2.71)$$

BUPT

Aceasta reprezintă o bună estimare a deplasarii $\hat{d}$ pentru pixelul din centrul regiunii Ω. Intercorelaţia poate fi estimată eficient prin utilizarea transformatei Fourier discrete (DFT – Discrete Fourier Transform), astfel [NAF 95]:

$$C_n(\Omega, \hat{d}) \approx IDFT(\hat{\Omega}_n(m)\hat{\Omega}_{n+1}(m)) \tag{2.72}$$

în care $\hat{\Omega}_n(m)$ reprezintă transformata Fourier bidimensională a blocului Ω din cadrul I_n, iar IDFT reprezintă inversa transformatei Fourier discrete. Rezultatul este valabil doar în cazul în care $\|\hat{d}\|$ este mult mai mică decât dimensiunea blocului

În [YOU 93] este implementată o metodă de estimare bazată pe un model translaţional. Această metodă realizează o corelare a fazei îmbunătăţind robusteţea tehnicii bazate pe corelaţie.

2.3.4.2 Corespondenţa fazei

O altă metodă în domeniul transformatelor o reprezintă corespondenţa fazei. Aceasta rezultă din teorema deplasării în domeniul timp a transformatei Fourier [NAF 95], care stabileşte relaţia dintre translaţia în domeniul timp şi modificarea fazei în domeniul frecvenţă:

$$\begin{aligned} f(t) &\leftrightarrow F(\omega) \\ f(t+\tau) &\leftrightarrow e^{j\omega\tau} F(\omega) \end{aligned} \tag{2.73}$$

Din această teoremă se poate lesne observa că putem estima deplasarea globală prin simpla împărţire a diferenţei de fază cu ω. Cu toate acestea, deoarece transformata Fourier descrie semnalul pe durata întregului interval de timp, deplasările ce sunt localizate într-un anumit interval de timp nu pot fi determinate prin această metodă. Pentru a putea realiza acest lucru este necesară formularea unei descrieri care să reprezinte schimbările locale ale semnalului. În acest scop a fost propusă aşa numita transformată Fourier pe durată scurtă (STFT – Short-Time Fourier Transform) [POR 80], a cărei fereastră bazată pe funcţia Fourier determină o descriere timp-frecvenţă a semnalului *x(t)*:

$$STFT(\tau, \omega) = \int_{\tau-T}^{\tau+T} x(t)w(t-\tau)e^{-j\omega t}\,dt \tag{2.74}$$

unde: *w(t)* reprezintă o funcţie simetrică reală definită pe intervalul [-*T*,*T*]. Această transformare poate fi privită ca o analiză a semnalului prin filtrarea unui domeniu finit în timp şi frecvenţă. Rezoluţia produsului timp frecvenţă depinde de forma ferestrei.

Transformarea Gabor, utilizează o fereastră Gaussiană care permite obţinerea celei mai bune combinaţii între rezoluţia în frecvenţă şi cea în timp. Această

BUPT

analiză cu ajutorul filtrelor Gabor a fost aplicată mai întâi pentru găsirea corespondenţei în stereoviziune [SAN 88], [JEP 89], prin aplicarea filtrării Gabor unidimensionale semnalelor corespunzătoare scenei privite din dreapta *D(x)* şi stânga *S(x)*. Filtrul Gabor folosit, având varianţa σ^2 şi frecvenţa centrală ω_0, este definit de [SAN 88]:

$$G(x;\sigma,\omega_0) = e^{-\frac{x^2}{2\sigma^2}} e^{j\omega_0 x} \tag{2.75}$$

Prin analogie cu Teorema deplasării Fourier, estimarea deplasării $\hat{d}$ în punctul x_0 este dată de:

$$\hat{d}(x_0) = \frac{\Delta\Phi(x_0)}{\omega_0} \tag{2.76}$$

unde $\Delta\Phi(x_0)$ reprezintă diferenţele dintre fazele celor două semnale: drept şi stâng. Această formulare presupune că translaţia unui semnal unidimensional produce o schimbare liniară a fazei funcţiei Gabor, presupunere ce limitază deplasările la valori foarte mici, mult mai mici decât extensia spaţială a ferestrei Gaussiene.

Metoda corespondenţei fazei a fost extinsă şi la domenii bidimensionale [WEN 93], prin utilizarea filtrelor Gabor orientate pentru determinarea celor două componente de deplasare. De asemenea, Fleet şi Jepson [FLE 90], [FLE 91] au construit o familie de filtre Gabor spaţio-temporale (tridimensionale) pentru descompunerea semnalului de intrare în canale acordate pe diferite viteze şi orientări.

Spre deosebire de metodele discutate până acum, care utilizau modelul translaţional pentru estimarea mişcării, a fost introdus modelul afin care ia în consideraţie şi rotaţia şi scalarea regiunilor.

2.3.5 Estimarea mişcării în prezenţa discontinuităţilor

La începutul acestui capitol s-a presupus, implicit sau explicit, că în câmpul vectorilor de mişcare nu există discontinuităţi. Aceasta presupune că vectorii de mişcare diferă foarte puţin faţă de cei vecini, fiind aşa numitul model de "deformare neliniară elastică" (rubber sheet [WAN 94]) şi având forma explicită dată de ecuaţia (2.35). Modelul este adecvat pentru estimarea mişcărilor globale ce rezultă din mişcarea relativă a camerei faţă de o scenă staţionară (fără modificări semnificative ale perspectivei). Acest model nu poate fi utilizat în cazul general al secvenţelor de imagine reale, în care apar discontinuităţi. Cele mai des întâlnite cazuri de discontinuităţi apar la graniţele obiectelor ce se mişcă unele faţă de altele sau atunci când există un fundal având o anumită textură. Estimarea bazată pe

BUPT

modelul "rubber sheet" realizează o netezire a câmpului de mişcare de-a lungul unor astfel de muchii.

Pentru a îmbunătăţi performanţele algoritmilor de estimare ai mişcării au fost propuse soluţii care iau în consideraţie muchiile. Astfel Nagel introduce condiţii orientate de uniformitate a câmpurilor de mişcare, în care nu se impun condiţii restrictive privitoare la uniformitate în vecinătatea muchiilor obţinându-se îmbunătăţirea performanţelor în cazul discontinuităţilor [NAG 83]. Konrad şi Dubois realizează extinderea modelului Markovian, prin eliminarea condiţiilor de netezire în vecinătatea muchiilor în mişcare [KON 92].

În aceste condiţii presupunerea unor traiectorii constante pentru zone limitate, care este luată în consideraţie de mai mulţi dintre algoritmii menţionaţi, prezintă resurse potenţiale pentru rezolvarea problemelor ce apar datorită discontinuităţilor. Pe de altă parte, mişcările muchiilor a două regiuni adiacente, pot fi total necorelate, în ceea ce priveşte vectorii de mişcare reali, după cum se arată în figura 2.11.

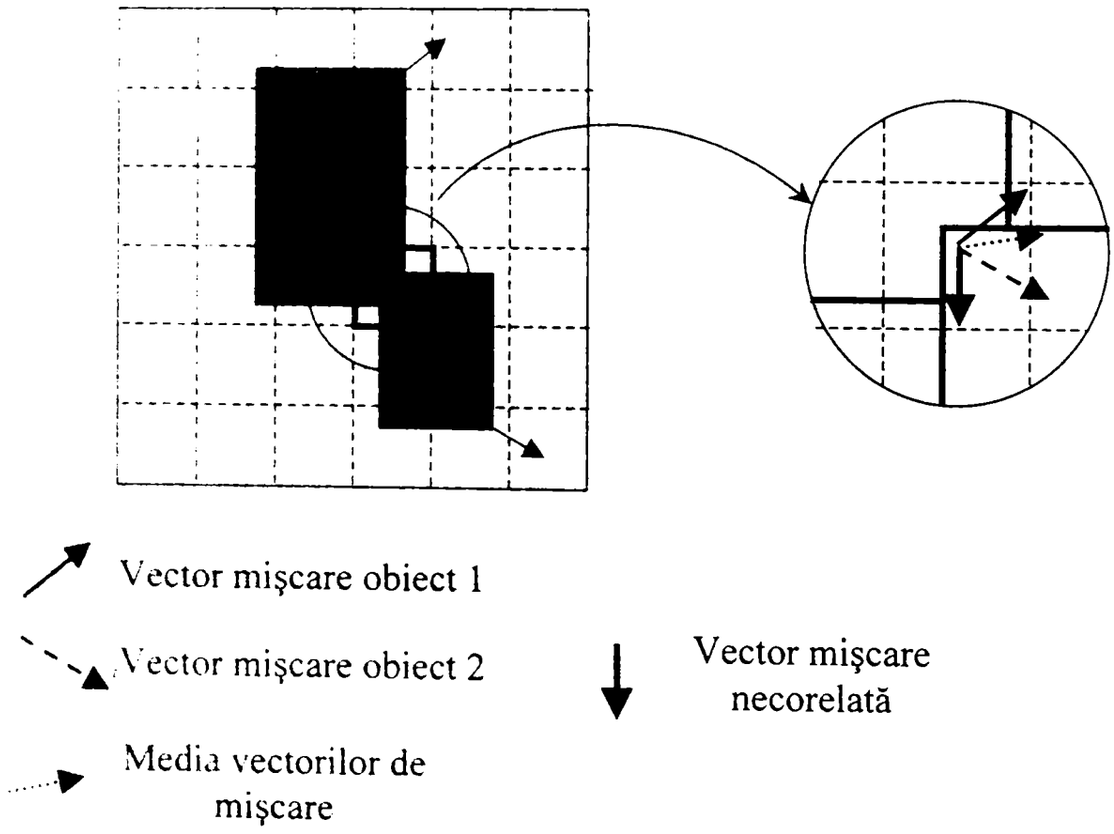


Fig 2.11 *Evidenţierea efectului ce apare în cazul mişcării necorelate a două obiecte învecinate*

Metodele de estimare bazate pe blocuri sunt cele mai susceptibile la apariţia discontinuităţilor, deoarece marginile blocurilor nu au nici o relaţie cu muchiile care sunt în mişcare. Reducerea dimensiunii blocurilor conduce la reducerea proporţiei de blocuri afectate, cu preţul creşterii vulnerabilităţii la zgomote şi,

BUPT

desigur, al scăderii ratei de compresie. În aceste condiţii au fost elaboraţi algoritmi de estimare utilizând blocuri de mărime variabilă [CHA 90], [DUF 95]. În cazul acestor algoritmi blocurile având dimensiuni mari sunt utilizate în cazul mişcărilor foarte apropiate de translaţie, în timp ce blocurile de dimensiuni mici sunt utilizate pentru descrierea mişcărilor complexe.

O metodă mult mai naturală, pentru luarea în considerare a discontinuităţilor, o reprezintă estimarea independentă a mişcării pentru obiectele scenei. Această metodă a fost propusă pentru prima dată de Kunt [KUN 85] prin introducerea aşa ziselor metode de codare din generaţia a doua. Orchad propune un algoritm bazat pe estimarea mişcării blocurilor, în care blocurile ce prezintă discontinuităţi de mişcare sunt resegmentate de-a lungul unei traiectorii arbitrare, aleasă în aşa fel încât, prin segmentare, să se ajungă la reducerea erorilor de predicţie [ORC 93].

O astfel de abordare segmentare-estimare se bazează pe metode iterative, astfel încât muchiile şi parametrii regiunilor sunt estimaţi pe rând, rezultând o descriere semantică a scenei analizate. Aceste iteraţii necesită o creştere substanţială a complexităţii de calcul. La ora actuală există o tendinţă clară spre algoritmi de segmentare-estimare [CHE 98], [MAR 96], [SEZ 93].

2.3.6 Metode de estimare ierarhice

Principalul dezavantaj al metodelor recursive bazate pe gradient este imposibilitatea utilizării lor în estimarea mişcării, pentru cele mai multe din situaţiile practice, datorită în special complexităţii calculelor. Neglijarea termenilor de ordin superior din dezvoltarea Taylor ne ajută numai la reactualizarea structurii semnalului de luminanţă cu neglijarea derivatelor de ordinul 2. Cu cât structura este mai plană cu atât este mai mic nivelul reactualizării reale. În plus, pentru a porni procesul, este necesară o valoare iniţială, în absenţa acesteia valoarea agreată fiind zero. Datorită domeniului de estimare al mişcării limitat, cadrele succesive vor trebui să nu difere foarte mult între ele, sau, cu alte cuvinte, este necesară o supra eşantionare în domeniul timp.

Cele mai uzuale moduri de implementare a metodelor recursive, care nu depind de eşantionarea în domeniul timp, sunt cele ierarhice sau multirezoluţie. Abordarea multirezoluţie porneşte cu descompunerea imaginii într-o structură multirezoluţie, în care fiecare nivel conţine reprezentarea imaginii de intrare la o anumită rezoluţie. Estimarea începe la nivelul superior al piramidei cu valoarea iniţială zero. În continuare rezultatul estimării este folosit ca valoare iniţială a nivelului următor, procesul continuând până când se incorporează toate detaliile imaginii originale. Algoritmul ierarhic bazat pe corespondenţa blocurilor este prezentat în figura 1.12.

Avantajele abordării ierarhice sunt:

BUPT

- creşterea ponderii primului termen din dezvoltarea în serie Taylor a traseului optic în cazul utilizării unei filtrări trece jos (descompunerea piramidală Gauss [ENK 88], [WAN 92]);
- creşterea domeniului efectiv de căutare. O mică deplasare estimată la un nivel inferior devine de două ori mai mare pentru nivelul următor (presupunând o subeşantionare 2:1 între niveluri);

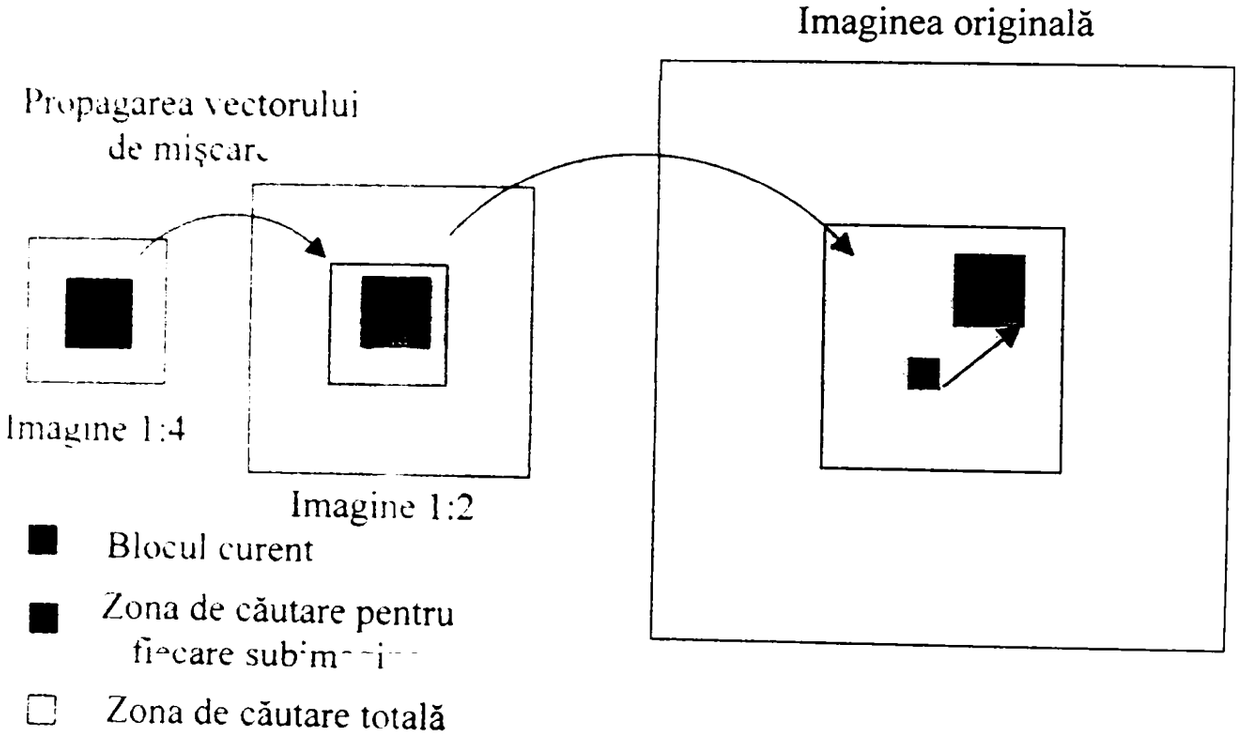
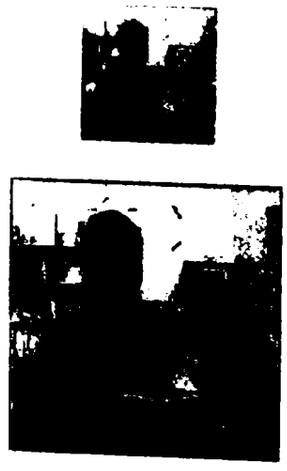


Fig 2.12 *Algoritm ierarhic de corespondenţă a blocurilor*

Cele mai uzuale abordări multirezoluţie sunt cele: piramidale [ROS 84], subbandă [VAI 90], [WOO 91], undişoară [MAL 89] şi spaţiu-frecvenţă [KRU 96]. În cele mai multe cazuri structura multirezoluţie este obţinută printr-o subeşantionare cu 2 în fiecare din cele două direcţii, proces aplicat iterativ până la obţinerea numărului de nivele dorit. În figura 2.13 este reprezentată descompunerea piramidală gausiană a unei imagini.



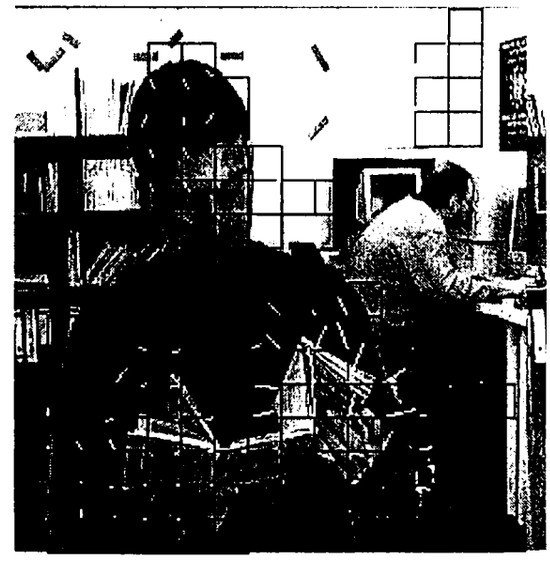
BUPT

Fig 2.13 *Descompunerea piramidală pe trei nivele a cadrului 7 din secvenţa MISSA*

Rezultatele comparative ale vectorilor de mişcare obţinuţi pentru algoritmii de corespondenţă ai blocurilor în trei paşi şi logaritmic, pentru estimare pe imaginea originală şi ierarhică, sunt prezentate în figura 2.14.

Cel mai important dezavantaj al estimării ierarhice a mişcării îl reprezintă determinarea vectorului de mişcare la nivelul de reprezentare cu rezoluţie scăzută a imaginii, ceea ce poate conduce la vectori improprii pentru rezoluţii mari (rezoluţia imaginii originale). Erorile ce apar în estimarea vectorilor de mişcare se pot observa prin compararea rezultatelor estimării din figura 2.14. Acest lucru este posibil a fi rezolvat prin detecţia blocurilor cu vectori neadecvaţi care apoi sunt reanalizaţi şi adaptaţi pentru o eroare de predicţie minimă.

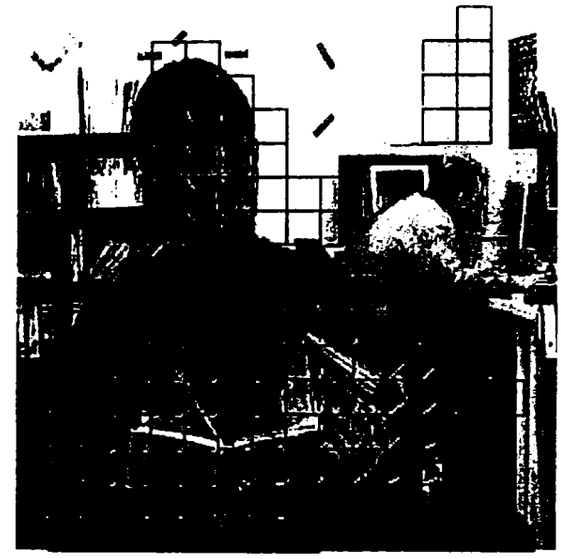
Estimarea mişcării prin algoritmul cu căutare în trei paşi



Estimarea ierarhică a mişcării prin algoritmul cu căutare în trei paşi

BUPT

Estimarea mişcării prin algoritmul logaritmic



Estimarea ierarhică a mişcării prin algoritmul logaritmic

Fig. 2.14 *Prezentarea comparativă a rezultatelor estimării mişcării prin metode clasice şi ierarhice*

2.3.7 Algoritmi de estimare bazaţi pe regiuni

Până acum s-a pus problema estimării mişcării în secvenţele de imagini, luând în consideraţie reprezentarea imaginii în pixeli sau blocuri de pixeli. Aceste tehnici sunt specifice aşa numitelor "tehnici de codare din prima generaţie" [KUN 85] şi se bazează în principal pe codarea formei de undă. În condiţiile unei transmisii la rate foarte scăzute tehnicile primei generaţii prezintă probleme foarte mari la implementare.

Limitările tehnicilor din prima generaţie sunt generate de următoarele:

- imposibilitatea de obţinere a funcţiei de distribuţie reale pentru sursă, motiv pentru care se utilizează modele ale acesteia (Markov, Gibbs);
- dificultatea încorporării proprietăţilor sistemului vizual uman în modelele utilizate pentru reprezentarea imaginii şi a mişcării;
- utilizarea reprezentărilor pe pixeli sau blocuri de pixeli nu permit obţinerea simultană a unei rate foarte scăzute pentru transmitere şi a reprezentării semantice a imaginii (obiecte componente ale scenei tridimensionale).

Aşa cum s-a menţionat, aceste tehnici bazate în principal pe reprezentarea prin blocuri de pixeli a imaginii sunt utilizate în mod curent în principalele standarde H261, H263, MPEG-1, MPEG-2 şi au servit ca punct de plecare în abordarea reprezentării bazate pe regiuni de formă oarecare, utilizate în standardele cele mai moderne propuse, MPEG-4 şi MPEG-7. Principalele dezavantaje ale acestor tehnici sunt:

BUPT

- divizarea oarbă, fără a lua în consideraţie conţinutul semantic al imaginii, care determină apariţia efectelor de bloc la compresii ridicate a imaginii;
- modelele de mişcare sunt aplicate unor blocuri de pixeli având formă pătrată, ceea ce poate conduce la diferenţe faţă de mişcarea reală a obiectelor ce formează imaginea;
- dificultatea încorporării proprietăţilor sistemului vizual.

Soluţia acestor probleme o reprezintă tehnicile de codare din generaţia a doua introduse de Kunt [KUN 85], care realizează încorporarea proprietăţilor sistemului vizual uman prin considerarea obiectelor din imagine.

Principalele proprietăţi ale sistemului vizual uman ce sunt de interes în vederea tehnicilor de codare a secvenţelor de imagini sunt:

- informaţia referitoare la contur şi muchii este apreciată foarte mult de sistemul vizual uman, fiind responsabila principală în ceea ce priveşte percepţia;
- textura are o importanţă diminuată, fiind asociată cu informaţia suplimentară, influenţând percepţia numai atunci când este asociată cu informaţia de contur:
- cele mai multe dintre obiectele naturale pot fi aproximate ca fiind membri ai unor grupări (clase) deterministice, pe baza unor seturi de caracteristici.

Din acest motiv abordarea estimarii bazate pe regiuni ce corespund obiectelor scenei reale, a cunoscut în ultimii ani o dezvoltare exponenţială. Astfel au fost făcute propuneri de: codare orientată pe obiecte [HOT 90], tehnici bazate pe morfologia matematică [SAL 95], [CHA 98], tehnci bazate pe model (model based) [AIZ 93], [AIZ 95]. [PEA 95], [POP 98].

Estimarea mişcării bazată pe regiuni conduce la rezultate mult mai bune în ceea ce priveşte impresia subiectivă, chiar dacă în unele cazuri măsura obiectivă a rezultatelor (vezi Anexa 2) ne arată că imaginea refăcută are raportul semnal/zgomot mai mic. Acesta este unul din motivele pentru care în capitolele 4 şi 5 s-a dezvoltat o metodă de segmentare în obiecte a imaginii, respectiv o metodă de estimare a mişcării bazată pe regiuni.

2.4 Concluzii

În acest capitol au fost prezentate cele mai utilizate strategii în estimarea mişcării precum şi legătura dintre ele.

În paragraful 2.1 a fost făcută o scurtă introducere în domeniul analizei mişcării. Analogia efectuată de autor între nivelurile de cunoaştere şi complexitate rezultate din acestă analiză a secvenţelor de imagini şi nivelurile utilizate în procesul de codare al acestora, conduce spre ideea utilizării informaţiilor semantice din imagine pentru obţinerea de rezultate superioare în procesul de codare.

Paragraful 2.2 introduce conceptul de percepţie a imaginilor, prezentând totodată şi principalele caracteristici ale sistemului vizual uman, care mai apoi vor fi luate în consideraţie la analiza secvenţelor de imagini. Sunt introduse din

BUPT

bibliografie conceptele de modelare a imaginii şi a mişcării prin modelele neparametrice, parametrice şi pseudoparametrice. Aceste modele vor fi utilizate pentru reprezentarea mişcării.

Scopul final al analizei mişcării fiind reprezentarea mişcării şi a obiectelor din imagine pentru codarea acesteia, conduce, la prezentarea în paragraful 2.3, a principalelor metode utilizate în analiza mişcării. După clasificarea tehnicilor de estimare şi prezentarea modelului neparametric aflat în strânsă legătură cu estimarea densă a mişcării ("optical flow"), se trec în revistă strategiile de estimare a mişcării.

Sunt evidenţiate pe rând caracteristicile fiecărei metode. Astfel metodele bazate pe gradient conduc la obţinerea câmpului vectorilor de mişcare apropiat de mişcarea reală, dar prezintă dezavantajele: complexitate de calcul, susceptibilitate la zgomot datorită utilizării derivatelor spaţiale şi temporale, domeniu de măsurare al mişcării limitat.

Metodele bazate pe blocuri (cele mai utilizate în standardele actuale) prezintă o bună imunitate la zgomote, sunt relativ simple, dar necesită complexităţi de calcul mari, prezentând totodată dificultăţi la încorporarea proprietăţilor sistemului vizual uman. S-au elaborat algoritmi rapizi de estimare a mişcării pe baza corespondenţei blocurilor, printre care şi cel propus de autorul tezei de doctorat [ALE 98a], cu rezultate foarte apropiate de cel cu căutare completă (full search) şi rezultate mai bune decât algoritmii rapizi propuşi în literatură în ceea ce priveşte complexitatea, rapiditatea şi erorile (eroarea medie pătratică).

Utilizarea domeniului transformatelor (metodele Fourier) conduce la o precizie mai bună, cu robusteţe în cazul simplificării mişcării la una de translaţie în secvenţele reale, dar prezintă o redusă rezistenţă la zgomote, necesitând o prelucrare suplimentară, ceea ce conduce la complexităţi de calcul foarte mari.

Metodele ierahice prezintă prin estimarea de la nivelurile de rezoluţie joasă la cele înalte o eficienţă în ceea ce priveşte calculul, cu creşterea domeniului de măsurare al mişcării şi îmbunătăţirea imunităţii la zgomote. Principalul dezavantaj constă în dificultăţile de refacere a vectorilor de mişcare improprii mişcării, ce se obţin prin trecerea de la rezoluţie joasă la rezoluţie înaltă.

Dezvoltarea metodelor bazate pe regiuni, care combină caracteristicile sistemului vizual uman cu metodele utilizate pentru corespondenţa regiunilor, conduc la eliminarea multor dezavantaje, cu preţul apariţiei dependenţei de etapa de segmentare. Punctul de start îl reprezintă în acest caz o segmentare spaţială a imaginilor cu ajutorul căreia să se realizeze o estimare cât mai precisă a mişcării regiunilor care reprezintă proiecţia obiectelor din scena tridimensională. În capitolele 4 şi 5 se vor prezenta metodele de segmentare şi cele de estimare bazate pe regiuni implementate de autor.

BUPT

Capitolul 3

CODAREA SECVENȚELOR DE IMAGINI

3.1 Codarea imaginilor

3.1.1 Fundamente

Transmisia semnalului de televiziune numeric necesită o capacitate a canalului foarte mare (pentru un semnal color în standardul 4:2:2 este necesară o capacitate de 216 Mbit/s), iar capacitatea acestora este limitată. Din acest motiv este necesară codarea semnalului. Astfel un sistem de transmisie digitală a semnalului video (ca de altfel orice sistem de comunicație numerică), poate fi definit ca în figura 3.1, fiind compus dintr-un sistem de captare al imaginii, un sistem de codare al sursei şi un sistem de codare al canalului.

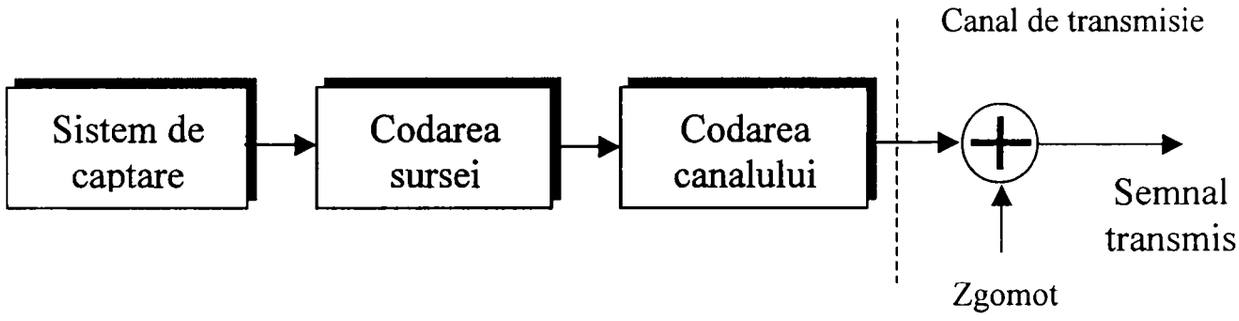


Fig. 3.1 Sistem de transmisie numerică

Blocul de codare al sursei formează partea principală a sistemului de compresie, aşa cum se menționează în [JAI 81], [NET 88], [VAS 98], o posibilă structură fiind prezentată în figura 3.2. Astfel algoritmii utilizați pentru codarea secvențelor de imagini pot fi descompuşi în două părți şi anume:

- analiza imaginilor
- extragerea datelor necesare transmisiei.

Etapa de analiză a imaginii poate utiliza tehnici de prelucrare a imaginii, care uneori nu țin cont de contextul strict al codării imaginii şi are ca scop obținerea de informații suplimentare (care sunt disponibile în scena captată), dar care nu sunt explicite în datele obținute prin noua reprezentare.

Au fost create diferite tipuri de instrumente metodologice pentru analiza imaginilor. O posibilă structură a acestora este următoarea:

BUPT

- elemente de analiză: pixel, contururi, regiuni, obiecte, grupuri de trăsături;
- prelucrări: transformări, filtrări sau dacă este vorba despre analiza mişcării: detecţie, estimare, segmentare, clasificare (etichetare)...;
- metode de abordare: statistice, geometrice, sintactice....

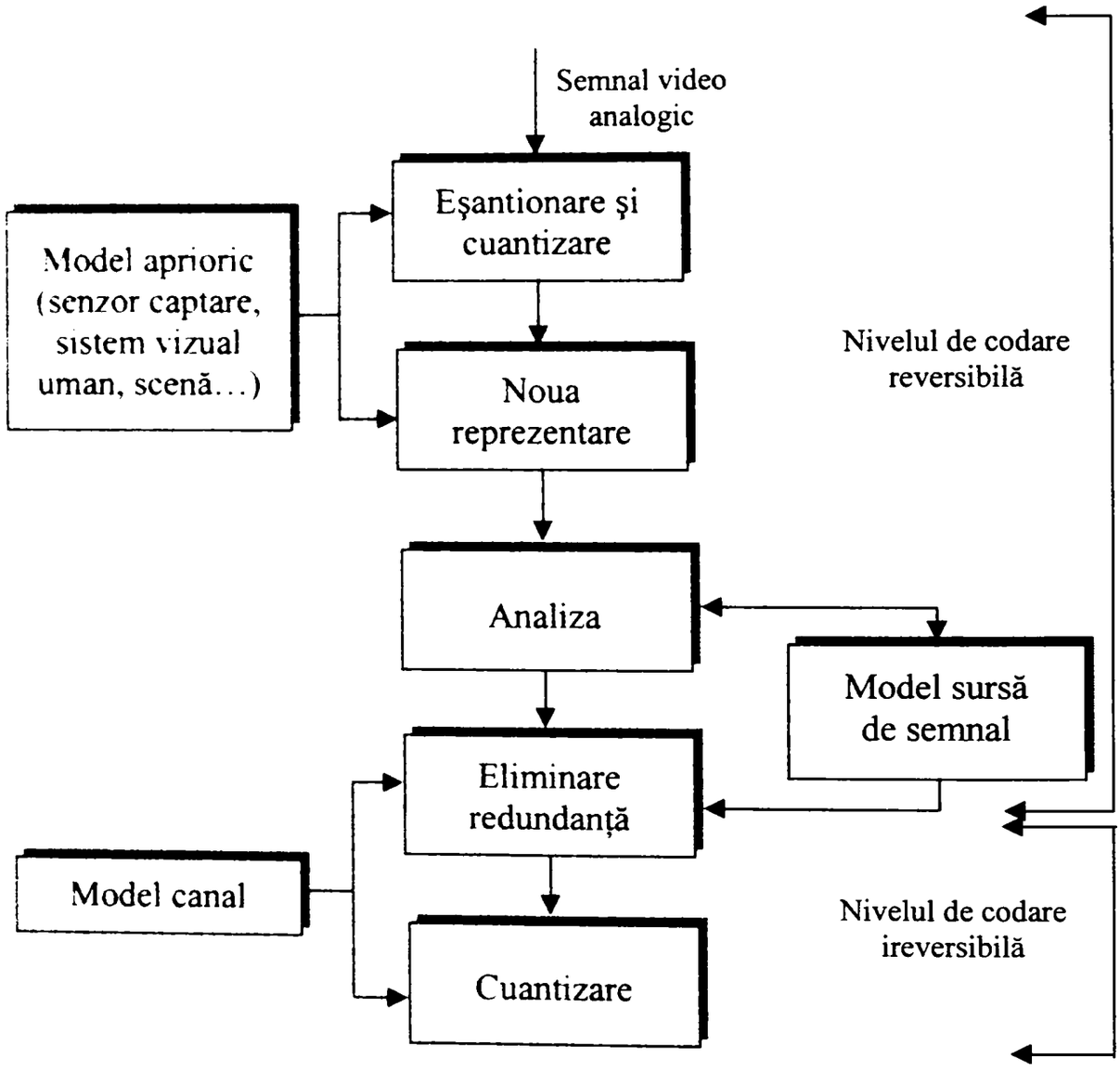


Fig. 3.2 Schema generală a unui sistem de compresie

Blocul de codare al sursei are ca scop principal compresia (reducerea) debitului binar prin reducerea sau suprimarea totală a redundanţei prezente în semnalul de imagine. Schema prezentată conţine două niveluri şi anume:

- nivelul codării reversibile (codare fără pierderi) care analizează corelaţiile prezente în semnalul video şi realizează extragerea redundanţelor;

BUPT

- nivelul codării ireversibile (codare cu pierderi) care completează codarea printr-o cuantizare scalară sau vectorială a semnalului obţinut prin extragerea caracteristicilor.

Prin analiză, se generează pentru oricare pixel având coordonatele $(x,y;t)$ o pereche de date $I(x,y;t)$ şi $C(x,y;t)$ în care $I(x,y;t)$ reprezintă informaţiile despre imaginea originală, iar $C(x,y;t)$ reprezintă caracterizarea imaginii originale de către una din metodele de prelucrare anterioare.

Cea de-a doua fază utilizează aceste date pentru a extrage informaţia utilă necesară redării precise (se vor defini criterii de calitate) a imaginii reconstituite. De exemplu, pentru o codare diferenţială, se poate defini următoarea relaţie :

$$i(x,y;t) = \boldsymbol{I}(x,y;t) - f_p(M(I), \boldsymbol{C}(x,y;t)) \qquad (3.1)$$

unde

- $\boldsymbol{I}$(x,y;t) – semnalul de imagine original
- i(x,y;t) - semnalul extras
- f_p - funcţia de predicţie
- $M(I)$ - reconstrucţia cauzală şi semnalul de imagine memorat
- $\boldsymbol{C}(x,y;t)$ - caracteristicile spaţio-temporale extrase.

Compresia obţinută la acest nivel constă dintr-o codare reversibilă prin analiza corelaţiei prezente în semnal şi eliminarea redundanţelor.

Acest proces este urmat de o cuantizare scalară sau vectorială a semnalului extras i(x,y;t) prin eliminarea unor informaţii care sunt ireversibil pierdute (această cuantizare se aplică şi semnalului caracteristicilor extrase $\boldsymbol{C}(x,y;t)$). Astfel se realizează reducerea sau prevenirea apariţiei informaţiilor subiective, în sensul vizual sau semantic. Rezultă, deci, că performanţele modulului de codare al sursei, în special în privinţa ratei de bit depinde de doi factori:

- compresia pe care semnalul $\boldsymbol{C}(x,y;t)$ o realizează pentru informaţia $\boldsymbol{I}(x,y;t)$ şi
- reprezentarea cât mai compactă pentru $\boldsymbol{C}(x,y;t)$, dacă aceasta se transmite.

În cazul compresiei cu compensarea mişcării (reducerea redundanţei spaţiale şi temporale), scopul principal devine:

- obţinerea unei reprezentări cât mai compacte a graniţelor regiunilor rezultate prin segmentarea spaţio-temporală;
- reprezentarea cât mai compactă a câmpului vectorilor de mişcare rezultaţi în urma estimării mişcării.

Codarea de canal produce un flux de date transformat în aşa fel încât acesta să fie adaptat canalului de transmisie. Procesul de codare a canalului poate implica, în cazul cel mai general, următoarele procese: codarea pentru controlul erorilor, criptarea, translatarea codului şi modularea purtătoarei numerice.

Cea mai simplă protecţie contra erorilor constă într-o disimulare prin înlocuirea unui eşantion eronat cu ultimul eşantion corect. Atunci când, însă, se

BUPT

realizează o protecţie veritabilă sunt utilizate coduri detectoare/corectoare de erori. Acestea sunt coduri corectoare de erori simple: codul Hamming [SPA 71], codul Wyner-Ash [MAT 84], sau în cazul existenţei unor erori aleatoare multiple, coduri mai complexe: Bose-Chaudhuri-Hocquenghem (BCH) (în cazul sistemelor de videoconferinţă) şi Reed-Solomon [SKL 88].

Deoarece se dispune de un mare număr de coduri de protecţie contra erorilor, alegerea se face în funcţie de aplicaţie, urmărindu-se obţinerea de performanţe globale satisfăcătoare, ceea ce pentru sistemele video înseamnă atingerea unei rate a erorilor binare de ordinul 10^{-4} în canalul de transmisie [SAN 90], [ISA 87]. Tehnicile utilizate sunt bazate pe teoria informaţiei (codări statistice, capacitatea şi codarea canalului) şi algebra binară (coduri corectoare de erori, criptografie, ...). Contribuţia acestui domeniu, specific codării secvenţelor de imagini, constă în studierea statistică a lui $C(x,y,t)$, considerat ca un vector de mişcare şi obţinerea unui cod special de protecţie la erori în relaţie cu aceste informaţii particulare.

3.1.2 Entropia semnalului video

Teoria informaţiei furnizează modele matematice pentru caracterizarea eficienţei codării. La baza acestei caracterizări este ideea că generarea informaţiei poate fi modelată printr-un proces probabilistic.

Codarea reversibilă a sursei mai poartă denumirea şi de codare statistică, codare cu lungime variabilă sau codare a entropiei. Din ultima denumire, se poate deduce faptul că această tehnică de codare face apel la conceptul de entropie. În teoria informaţiei aceasta este definită cu ecuaţia (3.2) şi reprezintă lungimea medie a simbolului binar ce permite reprezentarea probabilităţii de apariţie a unui simbol al sursei.

$$H(A) = -\sum_{j=1}^{M} p(a_j)\log_2[p(a_j)] \quad [biţi/simbol] \tag{3.2}$$

unde M reprezintă numărul de simboluri ale sursei, $p(a_j)$ probabilitatea de apariţie a simbolurilor. Această probabilitate mai este denumită şi probabilitate a priori, pentru a o deosebi de probabilitatea a posteriori, care corespunde probabilităţii ca mesajul sosit să fie corect. Deci:

$$p(a_j) \geq 0, \quad j = 1,2,\ldots,M \tag{3.3}$$

şi

$$\sum_{j=1}^{M} p(a_j) = 1 \tag{3.4}$$

BUPT

Se poate aprecia, deci, că entropia este o măsură a incertitudinii, a caracterului aleator al sursei, luând valori între [SPA 71]:

$$0 \le H(A) \le \log_2 M \tag{3.5}$$

În cazul semnalului video, în care sursa nu este staţionară (probabilitatea de apariţie a unui nivel nu este constantă), entropia *H(A)* are doar o mică semnificaţie. De asemenea, entropia semnalului video, calculată cu ecuaţia (3.2), nu reprezintă cu precizie conţinutul mediu de informaţie decât dacă simbolurile sunt statistic independente. Astfel în cazul unei imagini se poate ajunge la rezultate eronate, datorate redundanţei mari existente. Ecuaţia 3.2 reprezintă entropia de ordinul zero în care nu sunt luate în considerare dependenţele statistice faţă de vecini. Întrucât intr-o imagine există o corelare puternică între elementele vecine este de dorit utilizarea dependenţelor statistice faţă de vecini, entropia de ordinul 1 (entropia condiţionată) luând în considerare elementul anterior şi având definiţia dată de ecuaţia:

$$H(X/Y) = -\sum_{j=1}^{M}\sum_{k=1}^{M} p(a_j, a_k)\log_2\left(\frac{p(a_j, a_k}{p(a_k)}\right) \tag{3.6}$$

unde $p(a_j, a_k)$ reprezintă probabilitatea ca $X=r_j$ şi $Y=r_k$, unde Y reprezintă un pixel anterior pixelului X, iar $p(a_j)$ şi $p(a_k)$ reprezintă probabilităţile pentru pixelul X respectiv Y. Aceasta reprezintă, în termenii teoriei informaţiei, media informaţiei conţinute de pixelul X dacă se cunoaşte starea pixelului Y. Similar se pot defini entropii de ordinul 2 sau mai mare [SCH 56], [GRA 90].

Dacă avem o imagine cuantizată uniform cu 256 niveluri de gri (8 biţi pe eşantion), atunci vom observa că există o redundanţă statistică dată de relaţia:

$$R = \log_2 256 - H(A) \tag{3.7}$$

ceea ce indică posibilitatea utilizării unor forme optimale de codare în scopul obţinerii unui debit binar apropiat de H(A) biţi pe eşantion.

În realitate, imaginile digitale au un număr de niveluri de gri egal cu 235 (nivelul de alb) – 16 (nivelul de negru)=219. Deci entropia maximă a unei imagini dintr-o secvenţă reală este în mod real:

$$H = \log_2 219 = 7{,}77 \; biţi / simbol \tag{3.8}$$

Pentru secvenţa MISSA se observă din tabelul 3.1 că există o variaţie temporală (*H(A/X)* reprezintă entropia condiţionată faţă de acelaşi pixel al cadrului anterior), precum şi una spaţială, tabelul 3.2 (*H(A/Y)* reprezintă entropia condiţionată faţă de pixelul anterior al aceleiaşi imagini) a entropiei. De asemenea,

BUPT

prin analiza mai multor secvenţe de imagini se observă şi variaţia entropiei funcţie de imagine.

Tabelul 3.1 Evaluarea stabilităţii entropiei intracadru şi intercadre

Cadru / Entropie	2	3	4	5	6	7	8	9	10
H(A)	7,436	7,437	7,442	7,430	7,435	7,441	7,422	7,428	7,442
H(A/Y)	5,983	5,872	5,700	5,836	5,978	5,887	5,969	5,897	5,772
H(A/X)	4,803	4,672	4,535	4,656	4,788	4,652	4,703	4,683	4,506

Tabelul 3.2 Evaluarea entropiei între diferite secvenţe de imagini

Secvenţă / Entropie	MISSA	TENIS	TREVOR
H(A)	7,434	7,039	6,730
H(A/Y)	5,874	5,770	4,646
H(A/X)	4,660	4,849	3,410

Se ştie că o imagine codată MIC prezintă o corelaţie importantă între eşantioane, aşa încât, pentru realizarea unei transmisiuni eficace, este necesară decorelarea acestora, obţinându-se, astfel, o sursă discretă cu o entropie reală apropiată de entropia zero.

În figura 3.3 este prezentat un decorelator foarte simplu, bazat pe predicţia realizată pornind de la eşantionul precedent.

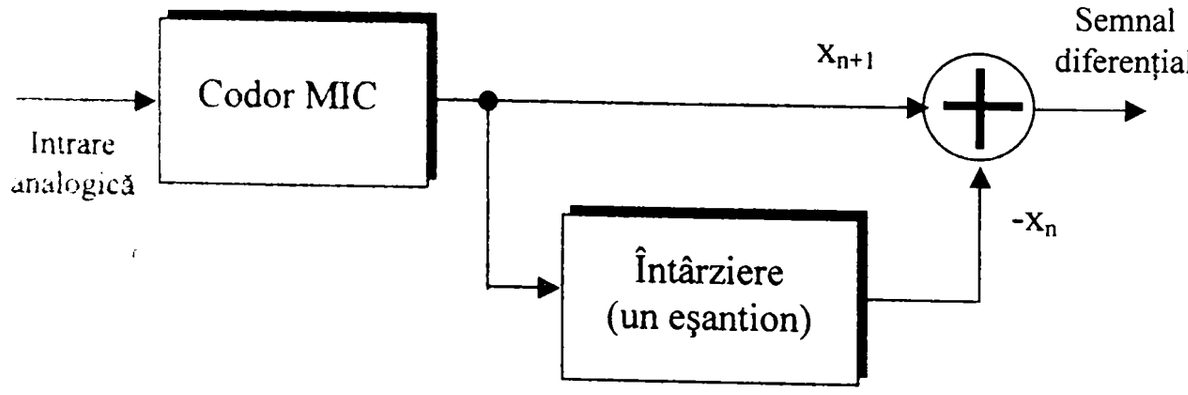


Fig. 3.3 Decorelator

Într-un astfel de sistem se poate considera că probabilitatea condiţionată ca două eşantioane învecinate să fie egale este aproximativ unitară. De aceea este posibil să considerăm eşantioanele diferenţiale ca fiind statistic independente şi densitatea de putere discretă a semnalului diferenţial ca având o distribuţie Gaussiană. De asemenea, din analiza rezultatelor prezentate în tabelele 3.1 şi 3.2 se

BUPT

observă că entropiile condiţionate sunt mai mici decât entropia de ordinul zero şi redundanţa temporală este mai mare decât cea spaţială. Această diferenţă creşte pe măsură ce în secvenţele de imagini mişcarea este mai redusă.

Din acest motiv, este foarte clar că corelaţia temporală dintre două imagini succesive nu este atât de puternică de-a lungul axei temporale, ci în direcţia exactă de deplasare a pixelilor. Mai mult, trebuie să presupunem cunoscută mişcarea, ceea ce ne va ajuta la codarea între-cadre cu compensarea mişcării.

3.2 Tehnici de codare a semnalelor de imagine

În literatura de specialitate sunt propuse diferite clasificări ale metodelor utilizate [JAI 81], [NET 88], [KUN 85], [MUS 85]. O clasificare a acestor metode este prezentată în figura 3.4. Clasificarea ierarhică propusă ţine cont de schema generală a sistemului de codare descrisă în paragraful 3.1.1 şi poate fi interpretată astfel:

- nivelul de codare I – realizează noua reprezentare a spaţiului imaginii, ce permite extragerea redundanţei existente în semnalul ce urmează a fi prelucrat. Astfel putem exemplifica:
  1. redundanţa spaţială (spaţio temporală în cazul secvenţelor de imagine) – rezultată din corelaţia existentă între pixelii adiacenţi, prin utilizarea metodelor PCM, DPCM [JAI 81], transformări unitare şi metode interpolative [JAI 89];
  2. redundanţa spectrală – prezentă între benzile de frecvenţă corelate, prin utilizarea transformărilor multirezoluţie [VET 92];
  3. reprezentarea semantică – care încearcă caracterizarea datelor locale prin atribute generale care realizează identificarea diferitelor obiecte (contururi, regiuni, texturi, mişcare) prezente în imagine, prin utilizarea codării orientată pe obiecte.

  Este evident că aceste metode nu au fost introduse simultan, fiecare dintre ele având avantajele şi dezavantajele proprii. În ceea ce priveşte extragerea redundanţei prezente în imagine (spaţio-temporală şi spectrală), este evidentă superioritatea metodelor hibride, care combină reprezentarea în frecvenţă cu predicţia erorilor.
- nivelul de codare II – a cărui utilizare este opţională şi care creează ireversibilitatea procesului de codare, implică cuantizarea semnalului obţinut prin extragerea informaţiei utile (2.1 paragraful 2.1.3). Convenţional această informaţie este cuantizată scalar sau vectorial [GER 92]. Cuantizoarele pot fi clasificate în:

BUPT

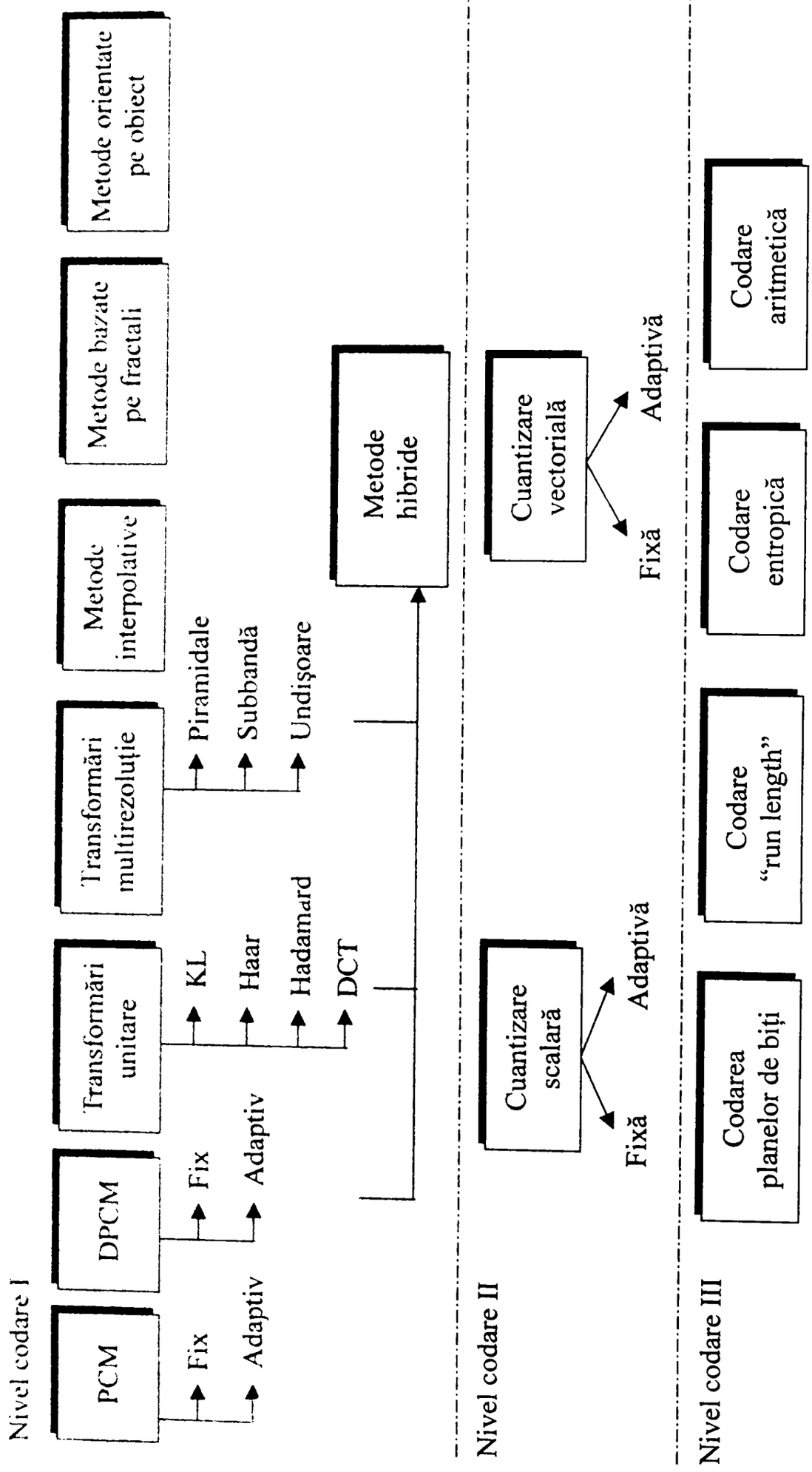


Fig. 3.4 *Clasificarea metodelor de compresie a imaginilor*

BUPT

1. cuantizare bazată pe modele statistice ale datelor de intrare, care presupune cunoaşterea apriorică a modelelor, şi
2. cuantizare bazată pe proprietăţile sistemului vizual uman, care obţin un câştig important pe baza descompunerii adaptive în frecvenţă.

- nivelul de codare III – realizează codarea binară (codarea de canal sau adaptarea la canalul de transmisie), fiind optimală în sensul teoriei informaţiei. Pentru această fază este necesară selecţia metodei celei mai bune din punct de vedere al distribuţiei elementelor ce urmează a fi codate. Cele mai utilizate tehnici de codare binare sunt:
  1. codarea planelor de biţi [HAL 79], [VAS 98];
  2. codarea "run length" [JAI 89], [PRA 78], [BOJ 97] - este o tehnică de codare relativ simplă în care se compară amplitudinea pixelilor de-a lungul unei linii;
  3. codarea entropică [RAB 91], [VIR 74] reprezintă o tehnică de codare a sursei care atribuie cuvinte de cod scurte celor mai probabile simboluri şi cuvinte de cod lungi celor mai puţin probabile simboluri. Dintre cele mai utilizate metode putem aminti: Shannon-Fano [GON 87], Huffman [RAB 91], codarea cu virgulă [GON 87], codare suboptimală [GON 87]
  4. codarea aritmetică [WIT 87], [BLA 87], [NEL 91] este recomandată întrucât este o metodă de codare ce nu se bazează pe blocuri, se acomodează uşor modelului sursei şi este eficientă din punct de vedere al calculului.

  Unul din algoritmii cei mai utilizaţi pentru compresia secvenţială a datelor şi care nu necesită o cunoaştere iniţială a probabilităţilor sursei este algoritmul LZW (Lempel-Ziv-Welch) [ZIV 77], [WEL 84], [YOK 92], [SHE 94].

Întrucât scopul principal al tezei îl reprezintă estimarea mişcării în vederea codării secvenţelor de imagini, se vor prezenta în cele ce urmează principalele tehnici de codare ale secvenţelor de imagini ce utilizează informaţiile referitoare la mişcare. Aceste tehnici corespund în principiu tehnicilor de codare cu compensarea mişcării.

3.2.1 Codarea predictivă cu compensarea mişcării

Aşa după cum s-a arătat, există o corelaţie puternică între elementele de imagine adiacente în spaţiu şi timp, ceea ce face posibilă compresia secvenţelor de imagini prin transmiterea unei informaţii parţiale despre cadrele ce compun secvenţa. Acest lucru poate fi realizat printr-o predicţie a cadrelor pe baza analizei imaginii, informaţia ce urmează a fi transmisă sau memorată fiind redusă la eroarea de predicţie.

Predicţia ce prezintă interes în acest caz este cea bazată pe estimarea mişcării, întrucât principala sursă a variaţiei intensităţii luminoase o reprezintă mişcarea obiectelor sau a senzorului (camerei). Acest tip de predicţie poartă denumirea de compensarea mişcării.

BUPT

Cea mai utilizată tehnică, ce încorporează estimarea mişcării în compresia imaginilor, este predicţia între cadre. Aceasta permite extragerea redundanţei temporale ce există între componentele unei secvenţe de imagini, informaţia ce urmează a fi transmisă sau memorată fiind redusă doar la eroarea de predicţie. Predicţia care interesează este bazată pe estimarea mişcării, iar schema bloc a unui astfel de codor este prezentată în figura 3.5.

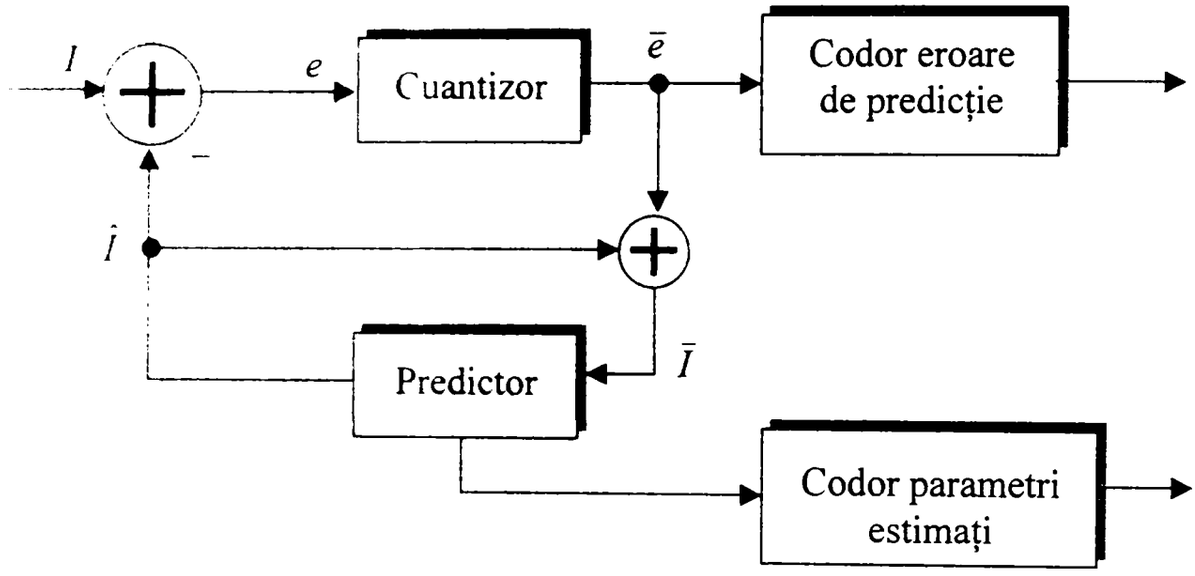


Fig. 3.5 Codarea cu predicţie cu estimare înainte

Funcţionarea codorului este simplă; semnalul de eroare este obţinut prin diferenţa între semnalul real şi semnalul predictat pe baza informaţiilor obţinute din analiza mişcării. Pe lângă acest semnal mai este transmis şi semnalul ce conţine parametrii estimaţi pentru mişcare.

Pentru compensarea mişcării pot fi utilizate următoarele metode de analiză a imaginii (care au fost descrise în capitolul 2):

- algoritmul recursiv pe pixel;
- algoritmul recursiv pe blocuri;
- algoritmul de corespondenţă al blocurilor;

Eficienţa codării depinde de analiza mişcării şi de modelul ales pentru mişcare deoarece cu cât eroarea este mai mare şi modelul vectorilor de mişcare mai complex, cu atât este mai mare cantitatea de informaţie ce trebuie transmisă.

Pentru algoritmii recursivi, predicţia este una înapoi deoarece se bazează pe informaţia deja codată şi care este de asemenea cunoscută în decodor. Spre deosebire de acestea, algoritmii ce se bazează pe corespondenţa blocurilor utilizează o predicţie înainte necesitând cunoaşterea vectorului de deplasare pentru fiecare bloc. În acest caz este aproape întotdeauna necesară cunoaşterea discontinuităţilor ce apar în mişcare, motiv pentru care este necesară, în cazurile practice, transmiterea datelor referitoare la detecţia mişcării şi segmentarea

BUPT

imaginii. Această metodă poate fi aplicată şi în cazul estimării mişcării pe baza regiunilor, cu necesitatea transmiterii informaţiei referitoare la segmentare şi detecţia mişcării.

Principalul scop al compresiei imaginilor îl constituie transmiterea informaţiei codate cu păstrarea degradării calităţii semnalului reconstruit la un anumit nivel. Teoria rată distorsiune este disciplina care se ocupă cu această problemă şi anume: determinarea celui mai mic debit cu care trebuie transmisă informaţia despre sursă către utilizator pentru obţinerea unei fidelităţi prescrise, estimate cu o funcţie de distorsiune dată.

Din acest punct de vedere codorul cu predicţie poate fi analizat din următoarele puncte de vedere:

- codarea intra-cadru, în care se realizează numai eliminarea redundanţei spaţiale din imagine şi pentru care funcţia rată distorsiune, în condiţiile unei codări fără pierderi, este dată de [SCH 97], [VAS 98]:

$$R(D_{DFD}) = \begin{cases} \frac{1}{2}\log_2 \dfrac{\sigma^2}{D_{DFD}} & D_{DFD} \le \sigma^2 \\ 0 & D_{DFD} > \sigma^2 \end{cases} \tag{3.9}$$

  unde σ^2 reprezintă dispersia sursei de semnal şi D_{DFD} reprezintă diferenţa între cadre definită de ecuaţia (2.39).

- codarea între-cadre fără compensarea mişcării, în care se ia în consideraţie numai diferenţa în timp între cadre, pentru care funcţia rată distorsiune este dată de relaţia:

$$R(D_{DFD}) = \frac{1}{2}\log_2\left(\frac{2\sigma^2(1-\rho\exp(-2\pi f_0\sqrt{u^2+v^2}}{D_{DFD}}+1\right) \tag{3.10}$$

  în care: ρ reprezintă coeficientul de corelaţie temporal, u şi v reprezintă deplasările pe direcţiile x respectiv y.

- codarea între cadre cu compensarea mişcării, în care estimarea se realizează luând în consideraţie vectorii de mişcare determinaţi între cele două cadre, pentru care funcţia rată distorsiune este dată de relaţia:

$$R(D_{DFD}) = \frac{1}{2}\log_2\left(\frac{2\sigma^2}{D_{DFD}}\left(1-\frac{\rho}{(1+\pi\sigma_d f_0)^2}\right)+1\right) \tag{3.11}$$

în care σ_d reprezintă dispersia semnalului D_{DFD}.

Pentru cuantizarea semnalului transmis se utilizează cuantizarea scalară, dacă pentru estimarea mişcării sunt utilizate metodele recursive pe pixel, respectiv una

BUPT

vectorială în cazul metodelor bazate pe blocuri [TZI 94]. Cuantizarea vectorială este utilizată pentru cazul în care reprezentarea imaginii este realizată pe blocuri, întrucât se obţin distorsiuni mai reduse faţă de cea scalară pentru o aceeaşi rată dată .

La transmisia pe canal sau memorare, imaginile codate pot fi distorsionate de zgomote aditive, multiplicative sau convolutive. Depinzând de tipul de distorsiune al canalului, fluxul de date liniare este modulat pentru a utiliza cât mai bine banda disponibilă. În toate cazurile există un risc de apariţie al erorilor. Pentru datele binare distingem trei tipuri: inversiunea unui bit, pierderea unui bit sau inserţia unui bit.

În general, cazul considerat este cel al inversiunii unui bit, iar codurile cele mai folosite pentru detecţia sau corecţia acestor tipuri de erori sunt codurile BCH (Base-Chandhuri-Hocqeughem) sau Reed Solomon, care sunt coduri bloc [SKL 88]. Pot fi utilizate şi alte tehnici cum ar fi cele care îmbunătăţesc protecţia la erorile de canal prin folosirea de cuvinte de sincronizare, dar cu preţul creşterii ratei de transmisie.

Această metodă de codare nu este utilizată sub această formă ci, de cele mai multe ori, se combină cu metodele de transformare a blocurilor care asigură o compresie mai bună cu reducerea redundanţei spaţiale din imagine. Metoda rezultată este aşa-zisa tehnică de codare hibridă care face obiectul următorului paragraf.

3.2.2 Codarea hibridă : compensarea mişcării şi transformările ortogonale

Una din metodele de extragere a redundanţei spaţiale a semnalului, cu cele mai bune rezultate în ceea ce priveşte compresia secvenţelor de imagini, este transformarea blocurilor de predicţie a erorii, metodă utilizată în principalele standarde de codare a secvenţelor de imagini MPEG [LeG 92], H261 [LeG 91].

Metodele de codare ce vor fi prezentate în continuare se numesc hibride, deoarece utilizează pentru obţinerea compresiei secvenţelor de imagini, predicţia şi transformările ortogonale. Structurile generale ale codorului şi decodorului bazate pe acest principiu, sunt prezentate în figura 3.6, respectiv figura 3.7. Transformările utilizate în acest caz sunt cele unitare, una din cele mai importante proprietăţi fiind faptul că inversele lor sunt egale cu conjugatele. Cele mai utilizate transformări sunt cosinus discretă (DCT – Discrete Cosine Transform), Walsh-Hadamard şi Haar discretă [GON 93], [MIT 86], [POP 86], [JAI 89], [ALE 95], [CLA 95].

În mod uzual coeficienţii transformărilor sunt cuantizaţi scalar, cu realizarea prealabilă a unei selecţii. Există trei metode de selecţie a coeficienţilor:

BUPT

- metoda zonală – în care se selectează coeficienţii spectrali ai unei zone prestabilite, cum ar fi cei din spectrul de joasă frecvenţă, numărul nivelurilor de cuantizare fiind ales funcţie de dispersia estimată a coeficienţilor selectaţi;
- metoda cu prag – în care sunt selectaţi coeficienţii ce depăşesc un anumit prag:
- metoda cu mască – în care se utilizează o mască ce ţine cont de proprietăţile sistemului vizual uman pentru alegerea coeficienţilor ce vor fi transmişi. Astfel, probabilitatea distribuţiei ce depinde pentru primele două metode de poziţie include particularităţile sistemului vizual uman.

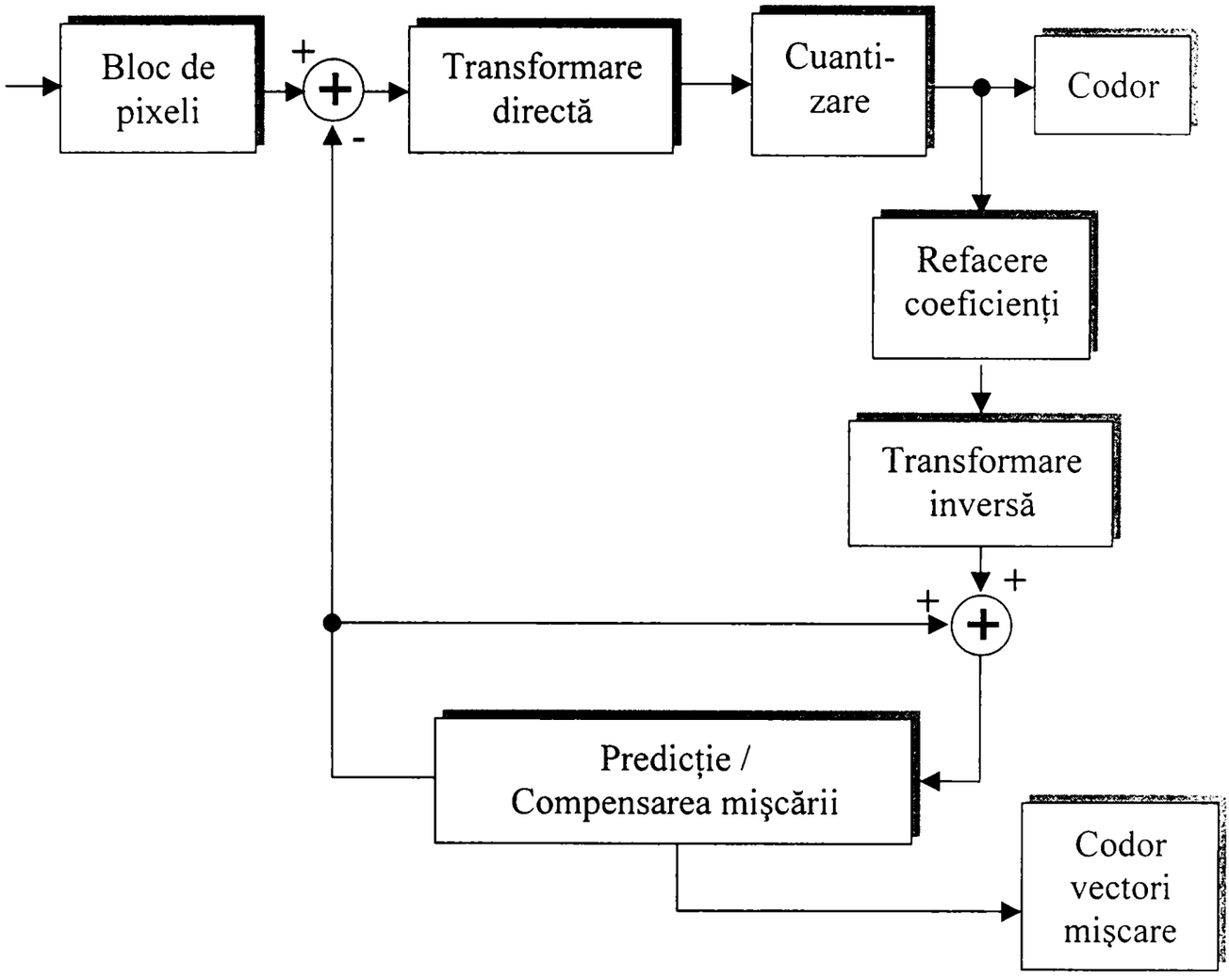


Fig. 3.6 Schema generală de codare hibridă cu compensarea mişcării

Este de remarcat faptul că reducerea ratei de transmisie a semnalului video se realizează, în cazul codării cu transformare, tocmai prin eliminarea anumitor eşantioane din spaţiul transformat, pe baza proprietăţilor statistice ale imaginii. În cazul metodelor hibride cu compensarea mişcării se realizează predicţia şi compensarea mişcării în imaginea originală, urmând ca transmisia semnalului de eroare să se realizeze într-un alt spaţiu, cel transformat.

BUPT

Eliminarea coeficienţilor transformatei se realizează în concordanţă cu sistemul vizual uman, pentru care au existat diferite modelări [NIL 85], [NGA 89], [CHI 90].

La recepţie se obţine din nou semnalul de eroare şi împreună cu parametrii mişcării se refac blocurile imaginii.

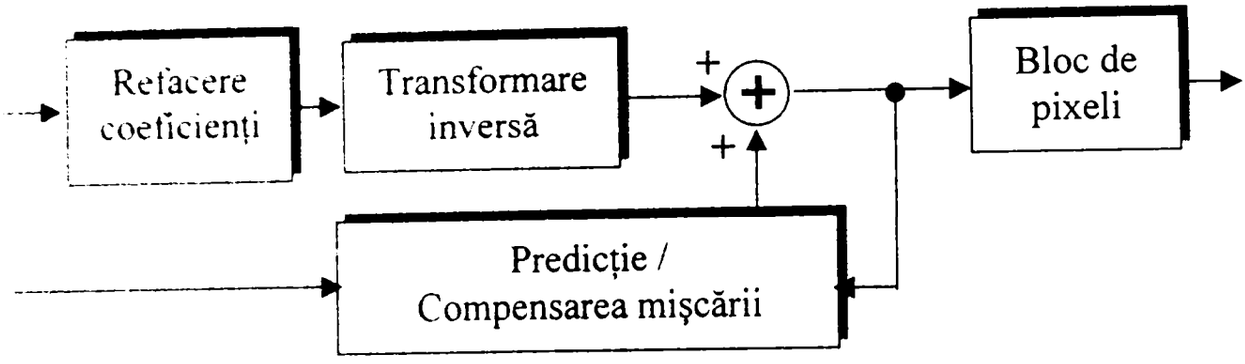


Figura 3.7 Schema de decodare pentru codarea hibridă cu compensarea mişcării

Principiile codării hibride cu compensarea mişcării sunt utilizate în standardele de codare H261, H263, MPEG-1, MPEG-2 [LeG 91], [LeG 92], [ARA 95], [BOJ 97]. Spre deosebire de schema generală prezentată în figura 3.6, standardele prevăd pe lângă transmiterea semnalelor erorii de predicţie şi a parametrilor mişcării şi a unui semnal care precizează modul în care a fost realizată predicţia intra-cadru (pentru blocurile care nu au activitate – mişcare) sau între-cadre (pentru blocurile pentru care a fost detectată mişcare). Diferenţele între diferitele scheme sunt minime şi se datorează, în special, aplicaţiilor pentru care au fost elaborate.

3.2.3 Codarea multirezoluţie cu compensarea mişcării

Vizualizarea obiectelor şi corpurilor ce sunt percepute într-un context oarecare nu reprezintă un concept absolut, el depinzând de mulţi factori. Unul dintre aceşti factori îl reprezintă nivelul de percepţie, pentru care se pot considera două niveluri externe şi anume: nivelul macroscopic şi cel microscopic. Astfel, pentru secvenţele de imagini se poate introduce conceptul de rezoluţie compatibilă, prin utilizarea de rezoluţii diferite la reprezentarea nivelurilor spaţiale şi temporale. În acest scop, se poate introduce conceptul de analiză multirezoluţie cu aplicaţii în domeniul codării secvenţelor de imagini. În capitolul 2 s-au arătat avantajele şi dezavantajele aplicării metodelor ierarhice la estimarea mişcării. Practic, prin analiza multirezoluţie sunt definite funcţiile pătratice integrabile $\left(f \in L^2\left(R^d\right)\right)$ (funcţiile pătratice însumabile $\left(f \in l^2\left(R^d\right)\right)$) ca o limită de aproximări succesive

BUPT

cu funcţii având rezoluţii diferite. Semnalul diferenţă dintre două aproximări succesive corespunde unui semnal de detaliu (rezoluţie mai mare).

Pentru imaginea originală există o limită de rezoluţie maximă care corespunde rezoluţiei iniţiale, astfel încât prin analiza multirezoluţie este relativ uşor de creat o reprezentare de rezoluţie scăzută şi un semnal de aproximare pentru nivelul de rezoluţie următor.

Ţinând cont de acest fapt se va putea descompune semnalul în benzi de frecvenţă spaţială, ce pot fi prelucrate, codate, transmise, decodate şi reconstituite separat, constituindu-se într-o schemă hibridă de compresie cu compensarea mişcării de forma celei prezentate în figura 3.8.

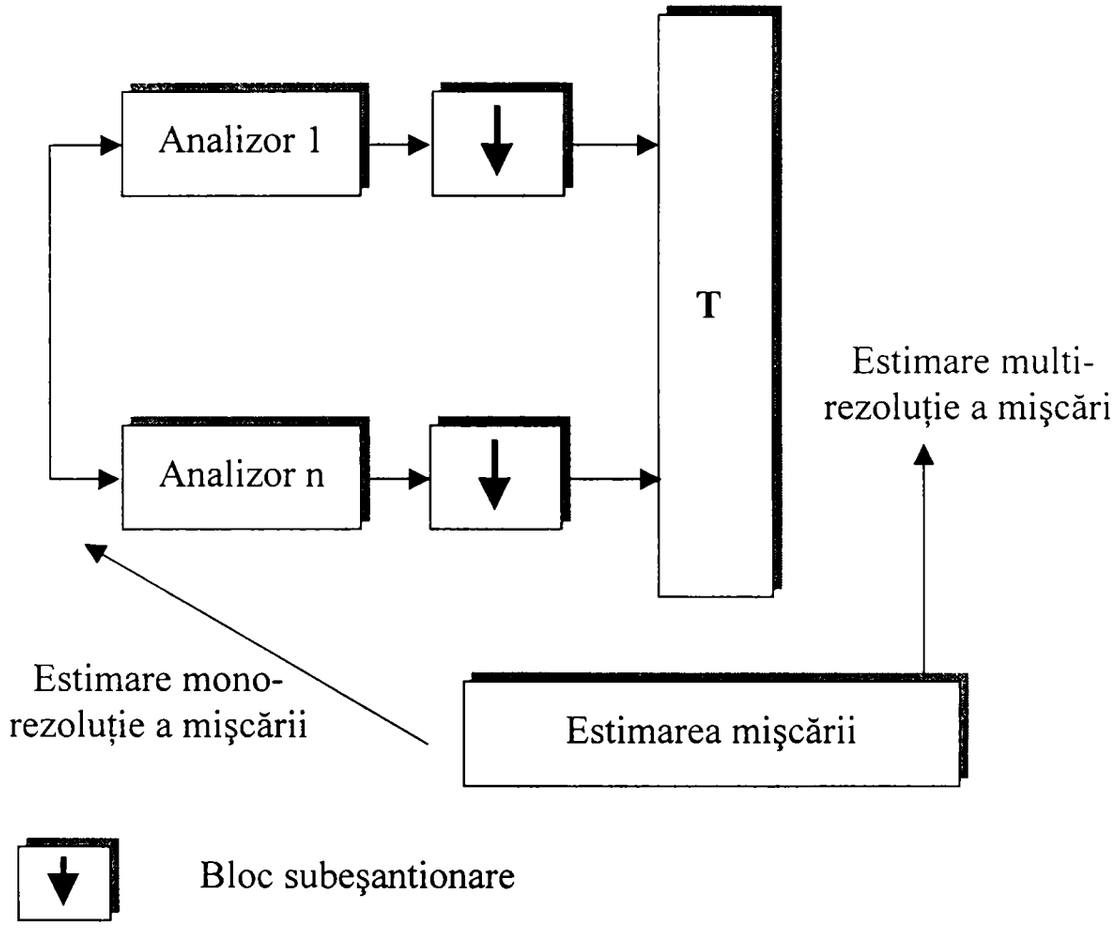


Fig. 3.8 Descompunerea multirezoluţie a imaginii şi estimarea mişcării

Schemele de codare multirezoluţie cu compensarea mişcării reprezintă de fapt scheme hibride, ce asociază descompunerea multirezoluţie (subbandă, undişoară) şi predicţia între cadre cu compensarea mişcării. Acestea realizează în fapt extinderea metodelor de codare hibridă la cazul utilizării descompunerii reconstrucţiei cu utilizarea familiilor de filtre pentru analiză-sinteză. Schema de codare prezentată în figura 3.9 poate fi implementată în următoarele moduri:

- codare predictivă cu estimarea mişcării în imaginea originală;
- codare predictivă cu estimarea mişcării în fiecare subbandă;

BUPT

- codare hibridă subbandă cu determinări multiple.

3.2.3.1 Codare predictivă cu estimarea mişcării în imaginea originală

Schema din figura 3.9 reprezintă cea mai simplă implementare a schemei din figura 3.8, prelucrarea T reducîndu-se la o simplă cuantizare a erorii de predicţie, care are spectrul transformat în $N= n^2$ spectre E_{ij}. La fel ca şi în cazul metodei de codare hibridă cu transformare, predicţia de compensare a mişcării este afectată numai de eroarea de cuantizare, deoarece estimarea mişcării este realizată în cadrul imaginii originale. Pentru reconstrucţie este necesar şi semnalul de predicţie. Îmbunătăţirea rezultatelor codării necesită realizarea unei cuantizări adaptive care să ţină cont de proprietăţile sistemului vizual uman.

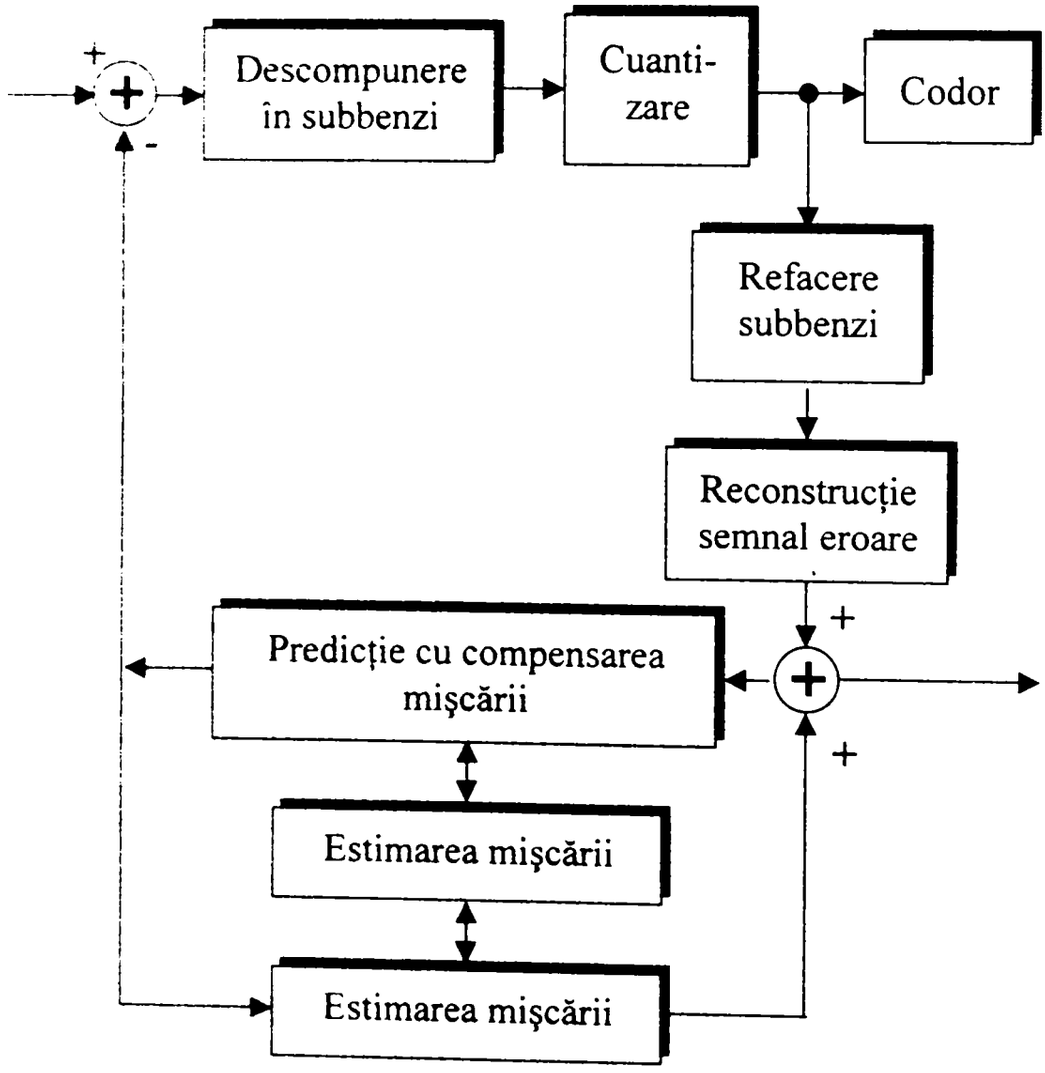


Fig. 3.9 Schema de codare hibridă subbandă cu estimarea mişcării în imaginea originală

BUPT

3.2.3.2 Codare predictivă cu estimarea mişcării în fiecare subbandă

Principalele caracteristici ale schemei din figura 3.10 sunt:

- este paralelă - codarea predictivă este realizată pentru fiecare subbandă obţinută din descompunere. La nivelul de estimare al mişcării, se pot utiliza $N = n^2$ estimatoare de mişcare, fiind acceptată limitarea estimării mişcării numai la benzile având frecvenţele cele mai joase sau la banda având energia cea mai mare [WOO 91].
- este compatibilă - semnalul de rezoluţie joasă poate fi considerat ca o sursă originală la care este aplicată o compresie între cadre standard.

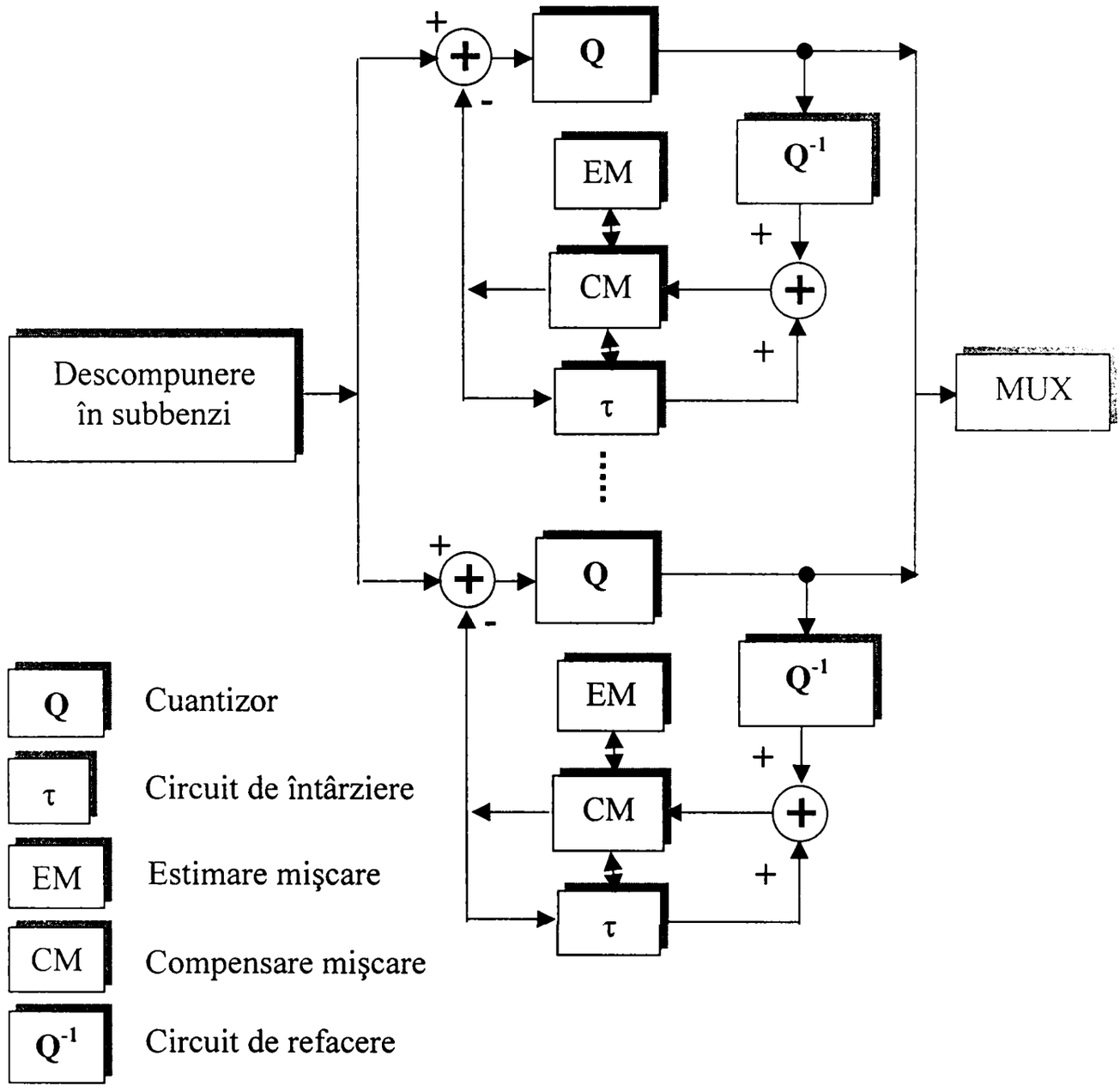


Fig. 3.10 Schema de codare hibridă subbandă cu estimarea mişcării în fiecare bandă

BUPT

3.2.3.3 Codarea hibridă subbandă cu determinări multiple

Cea de-a treia schemă lucrează cu o schemă de estimare a mişcării cu determinări multiple, în cadrul acesteia presupunându-se condiţii de echivalenţă apriorice (figura 3.11).

La fel ca şi în cazul schemei din figura 3.10 se realizează compensarea mişcării în fiecare subbandă, vectorul estimat depinzând numai de localizarea sa spaţială.

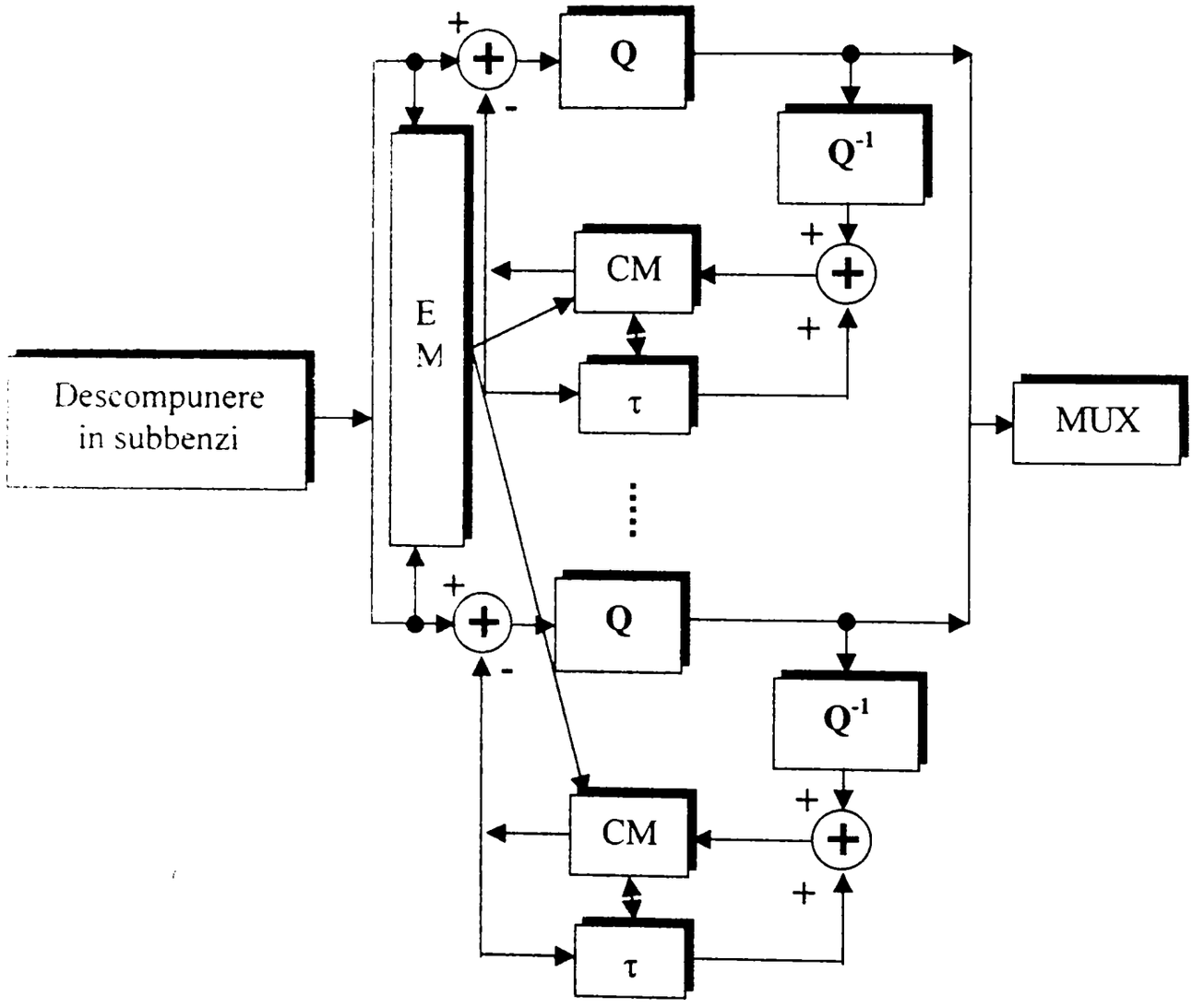


Fig. 3.11 Schema de codare hibridă cu determinări multiple

Există multe variante de implementare a algoritmilor bazaţi pe scheme de codare hibride ce utilizează descompunerea în subbenzi, aceştia diferind prin: descompunerea în subbenzi, algoritmii utilizaţi pentru estimarea mişcării şi modul în care se introduce compensarea mişcării. O altă metodă propusă este cea de codare tridimensională în domeniul frecvenţă, care realizează compensarea

BUPT

mişcării între-cadre (de-a lungul axei temporale), schemă ce poate fi extinsă la metodele undişoare biortogonale [OHM 94]. Pe baza schemelor prezentate mai sus se pot concluziona următoarele:

- Cuantizarea informaţiei transmise (eroarea de predicţie şi vectorii de mişcare) şi regularitatea filtrelor de reconstrucţie joacă un rol important în distorsiunile obţinute prin utilizarea acestor scheme hibride de compresie. Din acest motiv trebuie realizat un compromis între regularitatea filtrelor de reconstrucţie şi complexitatea decodorului (filtrele de sinteză). Această problemă poate fi rezolvată prin utilizarea undişoarelor biortogonale [VET 90].
- Există o analogie între schemele hibride obişnuite, dezvoltate pe baza transformărilor unitare (de exemplu DCT) şi cele dezvoltate în contextul descompunerii în subbenzi prin utilizarea filtrelor.
- Tehnicile de codare multirezoluţie devin foarte atractive, datorită posibilităţii de a defini simplu reconstrucţia compatibilă a structurilor ierarhice. În contextul serviciilor de comunicaţii, ce utilizează procese de codare flexibile ca de exemplu reţelele ATM, astfel de algoritmi şi scheme se adaptează foarte bine.

3.2.4 Codarea bazată pe model

Una din cele mai importante aplicaţii ale estimării mişcării o constituie codarea bazată pe model (MBC - Model Based Coding). Această codare este utilizată în mod special în sistemele de comunicaţii (videotelefon, videoconferinţe) pentru care este necesară pe lângă o rată mică de transmisie şi o estimare foarte exactă a mişcării, mai ales în ceea ce priveşte capul şi umerii. În cele mai multe dintre metode se estimează o mişcare tridimensională după o estimare sau analiză a mişcării bidimensională.

Primul model utilizat pentru detecţia trăsăturilor este cel prezentat în figura 3.12 [FUK 81].

Crinon şi Kolodziej [CRI 94], propun estimarea adaptivă multidimensională a mişcării, ţinând cont de un număr finit de trăsături şi de faptul că sistemul vizual uman se bazează pe un mecanism paralel de identificare al trăsăturilor.

Imaginea poate fi descrisă de următoarele primitive: pixelul ca formă canonică, blocuri, coeficienţii transformărilor, valori predictate sau derivate, dar pe lângă acestea pot fi folosite pentru descriere: textura, contururile, etc.

În contrast cu metodele bazate pe primitive locale, tehnicile orientate pe contururi şi textură încearcă o segmentare a imaginilor în regiuni cu textură uniformă, limitate de contururi ce corespund obiectelor din imagine. Informaţia referitoare la contur şi textură este transmisă separat.

Cele două tehnici de codare ce utilizează o abordare bazată pe contururi şi textură sunt:

- codarea bazată pe creşterea regiunilor, care cuprinde următoarele etape: segmentarea imaginii, codarea contururilor şi codarea texturii.

BUPT

- codare bazată pe descompunere direcţională, având ca scop păstrarea informaţiei referitoare la muchii. Etapele principale ale acestei tehnici de codare sunt: filtrarea direcţională bazată pe relaţia dintre prezenţa muchiilor şi contribuţia lor la spectrul imaginii şi codarea rezultatelor obţinute după filtrare.

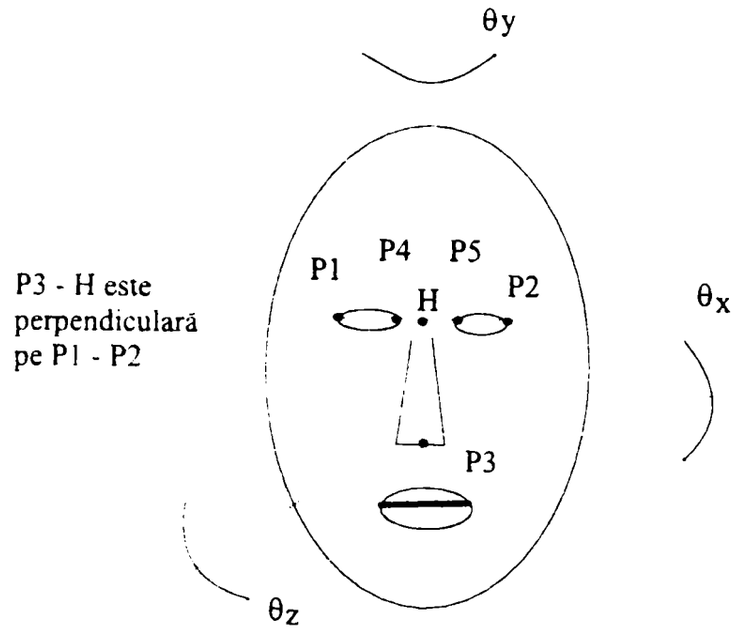


Fig. 3.12 Model pentru detecţia trăsăturilor faciale [FUK 81]

Aizawa prezintă un sistem de codare tridimensională bazată pe model pentru faţa unei persoane, figura 3.13 [AIZ 95].

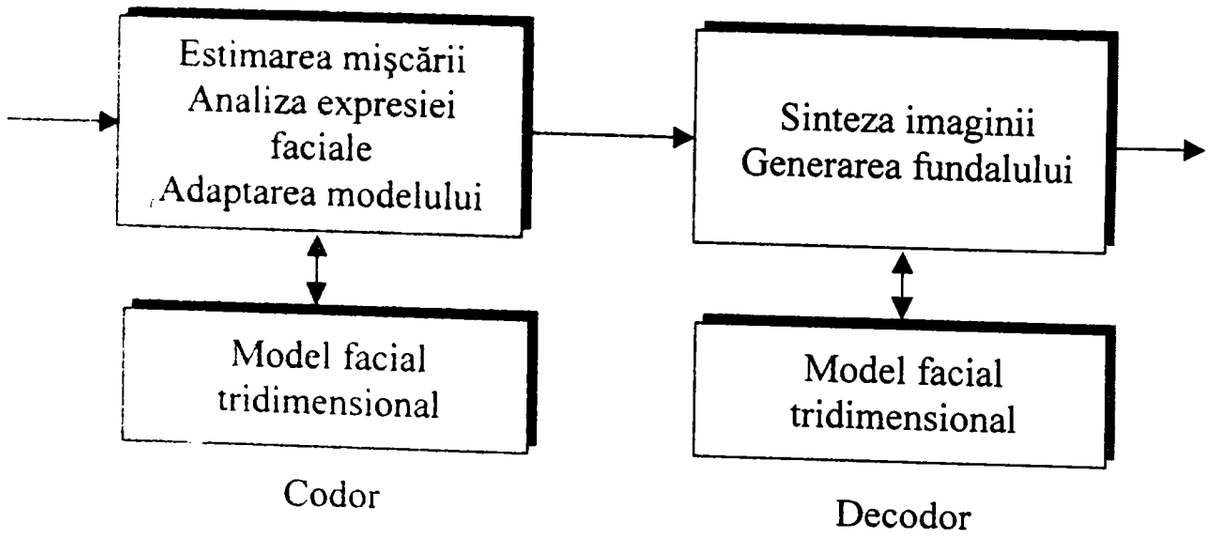


Fig. 3.13 Codare facială bazată pe model [AIZ 95]

BUPT

Cele mai utilizate modele tridimensionale ale feţei sunt realizate sub forma unor reţele de structuri triunghiulare şi au fost propuse de: Aizawa, Parke şi Candide [GER 94], [POP98].

Una dintre aplicaţiile practice ale codărilor la rate foarte scăzute şi care se adaptează foarte bine la modele, este videotelefonul. La aceste codări trebuie să se ţină cont de:

- conţinutul relativ fix al scenei;
- mişcarea limitată în cadrul scenei;
- particularităţi ale semnalului transmis (utilizarea formatului QCIF (144x176) şi frecvenţa cadrelor redusă până la 6-10 Hz).

Modelele de mişcare propuse pentru MBC sunt prezentate în tabelul 3.3 şi se bazează în general pe modele de proiecţie: proiecţie ortogonală şi proiecţie de perspectivă.

Pentru obţinerea unei codări eficiente se realizează încorporarea compensării mişcării precum şi combinarea diferitelor scheme de codare.

Tabelul 3.3 Modele de mişcare propuse în vederea codării bazate pe model

	Rotaţie cameră	Deformare (zoom)	Vibraţii
Modele de mişcare globală	$u = a_0$ $v = b_0$ Rotaţia camerei după o axă paralelă cu imaginea plană.	$u = a_1x$ $v = b_1y$ Translaţia camerei pe axa Z cu distanţa focală fixă sau modificarea distanţei focale a camerei.	$u = a_0 + a_1x + a_2y$ $v = b_0 + b_1x + b_2y$ Mişcări tridimensionale rapide ale camerei.
Modele de mişcare locală	Zonă plană		Zonă parabolică
	$u = a_0 + a_1x + a_2y$ $v = b_0 + b_1x + b_2y$ câmpul de mişcare al zonei plane în proiecţie ortogonală	$x' = \frac{a_0 + a_1x + a_2y}{1 + a_3x + a_4y}$ $y' = \frac{b_0 + b_1x + b_2y}{1 + b_3x + b_4y}$ câmpul de mişcare al zonei plane în proiecţie perspectivă	$u = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy$ $v = b_0 + b_1x + b_2y + b_3x^2 + b_4y^2 + b_5xy$ câmpul de mişcare al zonei parabolice în proiecţie ortogonală

3.2.5 Codarea orientată pe regiuni

Odată cu necesitatea unor rate de transmisie tot mai mici, a fost clar că tehnicile de codare bazate pe reprezentarea statistică a imaginilor (tehnici de

BUPT

codare din prima generaţie [KUN 85]) nu reuşesc să satisfacă aceste cerinţe, chiar şi în condiţiile dezvoltării tehnicilor adaptive hibride utilizate în standardele actuale [ITU 95], [RAM 94]. În aceste condiţii, soluţia pentru realizarea performanţelor impuse a reprezentat-o încorporarea proprietăţilor sistemului vizual uman într-o măsură mai mare. Aceasta presupune în principal reprezentarea imaginilor pe modele bazate pe contur, textură, deoarece, aşa cum am văzut, acestea sunt cele mai importante caracteristici din imagine ce influenţează sistemul vizual uman.

Abordarea compresiei pe baza blocurilor de pixeli conduce la obţinerea de rezultate bune în cazul reconstrucţiei secvenţelor de imagini transmise cu rate mari de biţi (compresii scăzute), dar obţine calităţi scăzute în cazul reconstrucţiei secvenţelor de imagini transmise cu rate scăzute (compresii mari). Aceasta se datorează principiului folosit şi anume: este ignorată corelaţia existentă între blocurile vecine şi este greu de realizat descrierea acestor blocuri ca un proces staţionar, întrucât aceste blocuri nu sunt obţinute funcţie de compoziţia imaginii, ci în mod arbitrar. De asemenea, datorită blocurilor, la reconstrucţie apare efectul de bloc (muchii zimţate) care este foarte deranjant pentru sistemul vizual uman.

Codarea bazată pe regiuni încearcă utilizarea unui model natural pentru reprezentarea imaginii. Astfel algoritmii pentru codarea secvenţelor de imagini, ce utilizează modele bazate pe regiuni, presupun că imaginea este formată din regiuni omogene din punct de vedere al unor proprietăţi şi acestea realizează o descriere completă a imaginii.

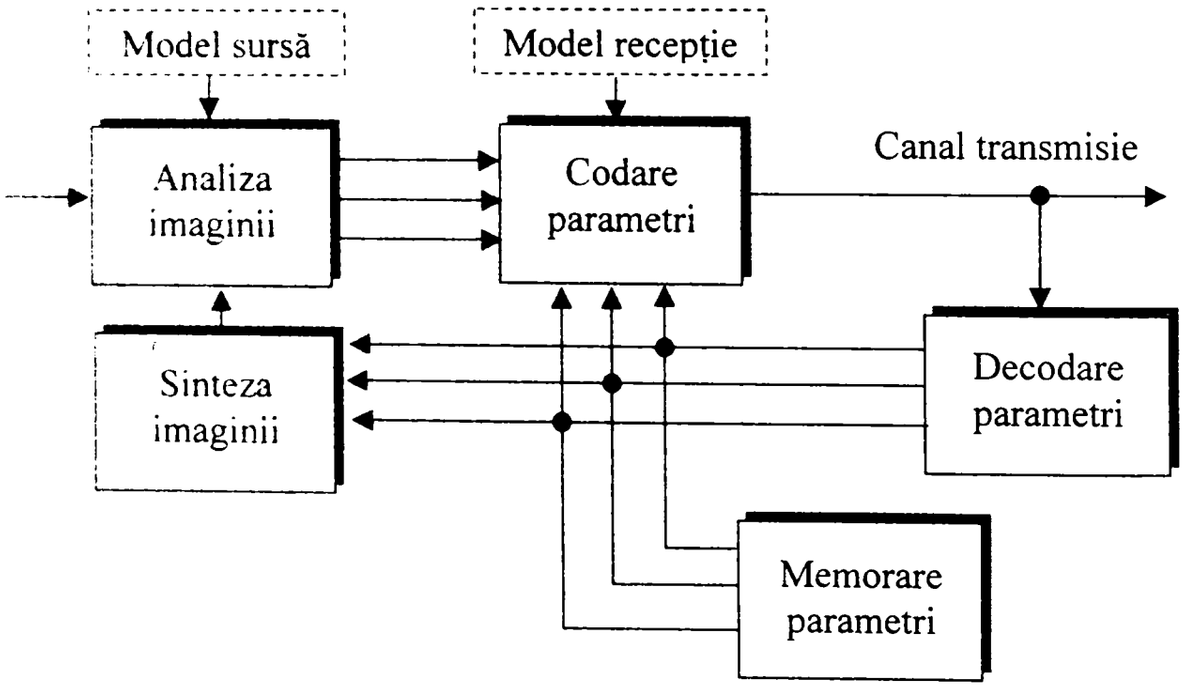


Fig. 3.14 Schema bloc a unui codor orientat pe obiecte [GER 94]

BUPT

O schemă generală de abordare a codării prin analiză-sinteză bazată pe obiecte (Object-Based Analysis-Synthesis Coding) este prezentată în figura 3.14 [HOT 94], [GER 94].

Blocul de analiză a imaginii descompune fiecare imagine a secvenţei în diferite obiecte aflate în mişcare, pentru care se extrag parametrii: forma, vectorii de mişcare şi culoarea (textura). Totodată se face şi o clasificare a obiectelor în obiecte statice şi obiecte în mişcare. Această clasificare este apoi utilizată pentru decizia referitoare la parametrii transmişi. Astfel, pentru obiectele statice, se transmit forma şi textura, iar pentru obiectele în mişcare se transmit forma şi vectorii de mişcare. Pentru fiecare din aceşti parametri ce se transmit se utilizează proceduri de codare specifice. Pentru formă se utilizează tehnici specifice codării contururilor, cu excepţia modificărilor temporale, caz în care se utilizează codarea predictivă. Pentru mişcare este utilizată codarea predictivă, în raport cu parametrii de mişcare estimaţi pentru acelaşi obiect în imaginea anterioară. Pentru informaţia de textură se utilizează tehnici bazate pe forma de undă (codare hibridă transformare-predicţie, cu compensarea mişcării).

Schema generală de abordare a codării orientate pe regiuni este prezentată în figura 3.15. Aceasta implică realizarea unei segmentări pe baza omogenităţii pixelilor, estimarea mişcării bazată pe regiuni şi codarea cu compensarea mişcării a contururilor şi texturii.

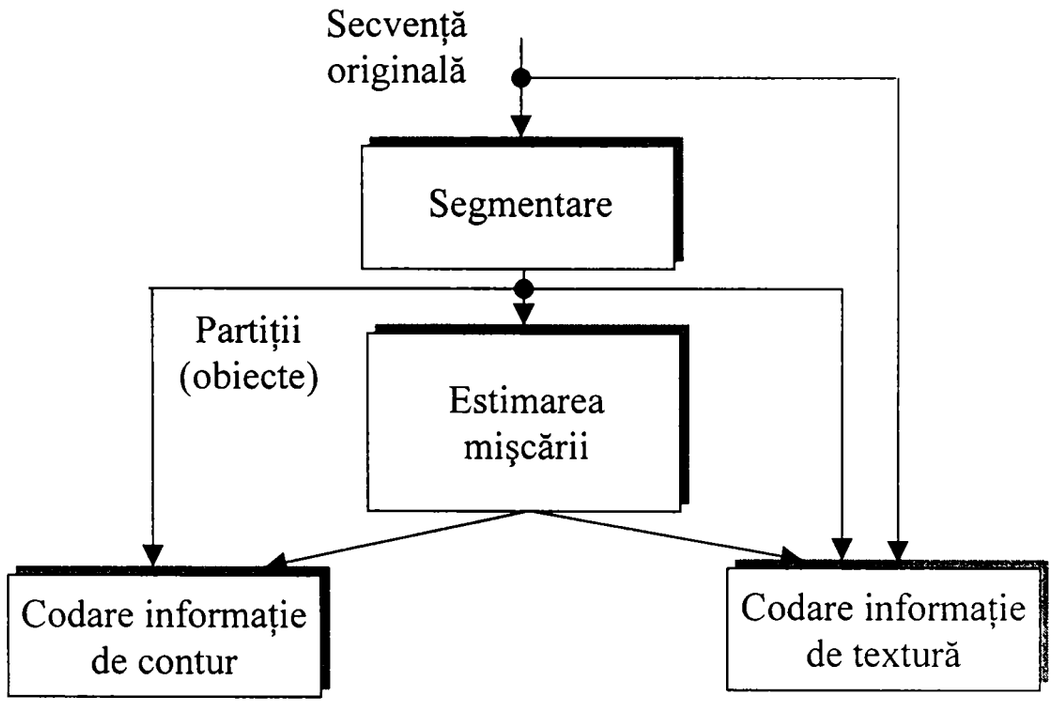


Fig. 3.15 Structura generală a unui codor orientat pe regiuni

BUPT

Cea mai simplă implementare a codării bazate pe regiuni o reprezintă codarea erorii, care utilizează redundanţa temporală existentă în secvenţele de imagini. Astfel, fiecare cadru este obţinut prin predicţia celui anterior, semnalul ce este codat fiind eroarea (diferenţa) dintre cele două cadre: cel curent şi cel predictat.

Schema propusă de autor pentru codare în acest caz este cea din figura 3.16, în care, suplimentar, se realizează o clasificare a regiunilor. Această clasificare este necesară întrucât există posibilitatea ca regiunile obţinute în urma segmentării să nu fie caracterizate de acelaşi model. În această etapă, se poate presupune că avem o segmentare a imaginii în regiuni omogene din punct de vedere al texturii.

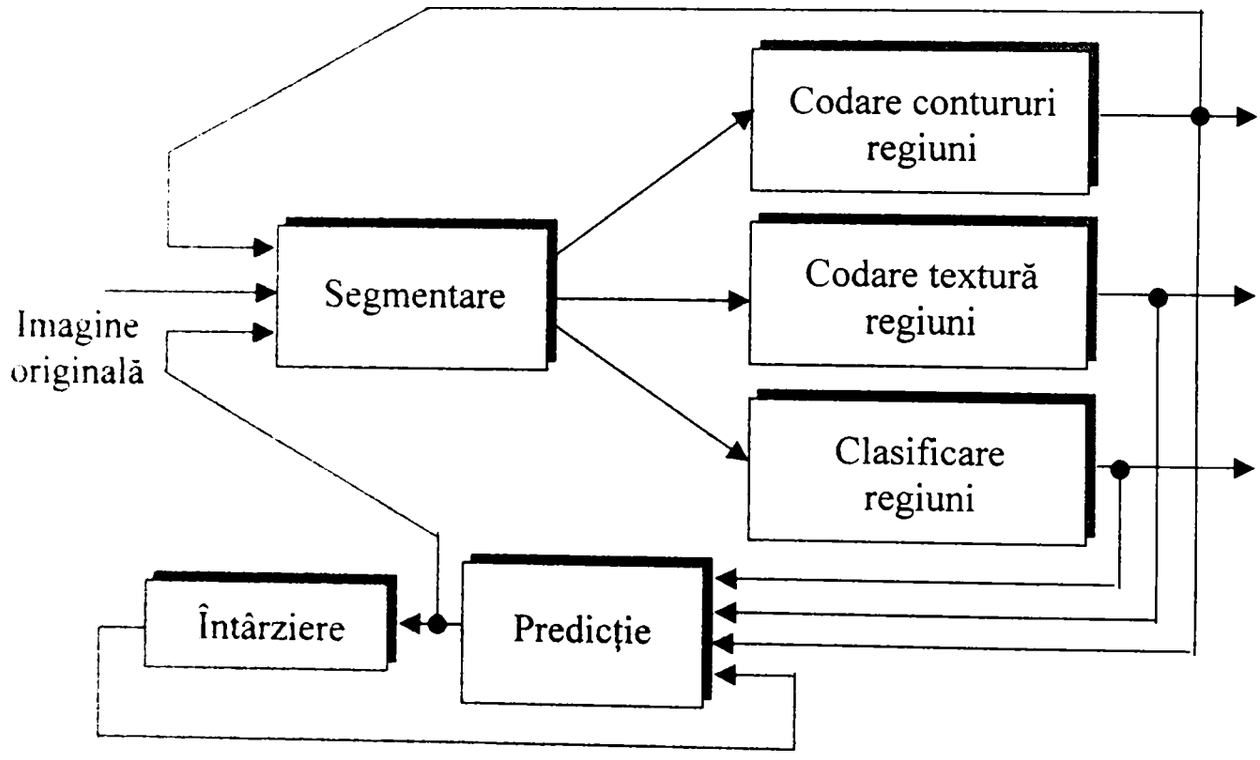


Fig. 3.16 Schema bloc propusă de autorul tezei de doctorat pentru codarea bazată pe regiuni

3.3 Concluzii

În acest capitol au fost analizate principalele metode de codare a secvenţelor de imagini în condiţiile utilizării compensării mişcării. Motivată de aplicaţiile imediate (video-telefonie, video-conferinţe, televiziune digitală, multimedia, etc.), codarea la rate foarte scăzute este principalul ţel al celor mai multe din cercetările actuale referitoare la compresia imaginilor. Întrucât tehnicile de codare bazate pe reducerea statistică a redundanţei din imagine conduc la anumite limitări în cazul ratelor scăzute de biţi, s-a trecut la tehnicile generaţiei a doua care utilizează reprezentarea semantică a imaginilor.

BUPT

Principalul argument în favoarea tehnicilor din prima generaţie îl reprezintă schemele de codare hibridă, care combină predicţia cu compensarea mişcării în domeniul timp cu tehnicile de decorelaţie în domeniul spaţial. Aceste scheme sunt utilizate în mod curent în standardele actuale şi pot fi utilizate ca puncte de start pentru tehnicile de codare ale celei de-a doua generaţii.

Principalele dezavantaje ale acestor scheme de codare sunt:

- divizarea "oarbă" fără a lua în consideraţie conţinutul semantic al imaginii, care determină apariţia efectelor de bloc atunci când sunt cerute compresii mari;
- modelele de mişcare sunt aplicate unor blocuri ce nu corespund în totalitate cu mişcarea obiectelor constituente ale imaginii;
- este dificilă încorporarea proprietăţilor sistemului vizual uman în aceste scheme de codare.

Cu toate acestea, aceste tehnici au condus la înţelegerea mai bună a reprezentării mişcării şi la apariţia metodelor de codare din generaţia a doua, metode ce încorporează particularităţile sistemului vizual uman, care este capabil să recunoască o imagine sau o secvenţă de imagini utilizând informaţia a numai câtorva puncte. În aceste condiţii sarcina este determinarea celor câteva puncte ce conţin majoritatea informaţiei şi contribuie la recunoaşterea acesteia, completată cu detaliile necesare aprecierii informaţiei.

Întrucât rezultatele se adresează sistemului vizual uman, se pune întrebarea care este măsura utilizată pentru evaluarea performanţelor sistemului. Dacă în cazul tehnicilor primei generaţii măsura era reprezentată de teoria rată-distorsiune, prin utilizarea erorii pătratice medii [SCH 97], în cazul tehnicilor de codare din generaţia a doua această măsură corespunde metricii Kantorovich [LI 96].

Una din marile probleme ale sistemelor de codare pentru secvenţele de imagini o reprezintă modelul utilizat pentru estimarea mişcării şi compensarea acesteia. În cazul schemelor de codare din generaţia a doua, întrucât codarea se bazează pe reprezentarea semantică a imaginii, modelul utilizat pentru mişcare a fost constituit pe contur sau regiune omogenă. O posibilă clasificare a schemelor de codare şi a modelelor utilizate este prezentată în tabelul 3.4.

În preambulul primului capitol a fost prezentată necesitatea codării la rate foarte scăzute, realizându-se calculul factorilor de compresie pentru diferite formate şi frecvenţe de transmisie a cadrelor.

Primul paragraf se referă la codarea imaginilor în care a fost propusă de către autorul tezei de doctorat o schemă generală pentru un sistem de compresie, realizându-se totodată şi o analiză a entropiei în secvenţele de imagini. Din analiză a rezultat faptul că entropia unei imagini depinde de conţinutul acesteia, fiind variabilă de la o imagine la alta a secvenţei. De asemenea, prin analiza entropiilor condiţionate spaţial şi temporal se observă că acestea sunt mai mici decât entropia unei imagini. Acest rezultat ne sugerează faptul că se poate realiza, prin codarea secvenţelor de imagini, reducerea redundanţei atât spaţiale cât şi temporale.

BUPT

Tabelul 3.4 Schemele de codare utilizate şi modelele asociate

Schema de codare	Model sursă	
	Model segmentare	Model mişcare
PCM	Pixel	-
MC DCT, etc.	**Dependent statistic de pixel** Bloc	• translaţie
• bazată pe contur • bazată pe regiuni • bazată pe obiect • bazată pe triunghiuri deformabile	**Model bidimensional** • trăsături: muchii, contururi • regiuni rigide bidimensionale • regiuni flexibile bidimensionale • triunghiuri deformabile • pătrate deformabile	• translaţie • transformare biliniară • transformare afină
• bazată pe obiect • bazată pe model tridimensional	**Model tridimensional** • model tridimensional global de suprafaţă • model tridimensional parametrizat	• mişcare globală • mişcare locală
• Hibridă 2D sau 3D • 3D şi compensarea mişcării cu DCT	**Model hibrid** Combinaţii ale celor de mai sus	Combinaţii ale metodelor de mai sus

În continuare s-a prezentat o clasificare a metodelor de codare pe trei niveluri, funcţie de schema de codare propusă, urmărindu-se dezvoltarea principalelor tehnici ce încorporează estimarea mişcării (tehnici de codare cu compensarea mişcării). Sunt dezvoltate tehnicile de codare predictivă cu compensarea mişcării, tehnicile hibride: predicţie cu transformare şi predicţie cu descompunere multirezoluţie, arătându-se faptul ca predicţia cu dezvoltarea multirezoluţie reprezintă o generalizare a codării cu predicţie şi transformare.

Pe baza dezavantajelor metodelor bazate pe codarea statistică sunt prezentate metodele de codare bazate pe reprezentarea semantică a imaginii: codarea bazată pe model şi codarea bazată pe regiuni, autorul tezei de doctorat propunând în finalul capitolului o schemă de codare bazată pe regiuni.

BUPT

Capitolul 4

SEGMENTAREA IMAGINILOR

4.1 Introducere

Imaginile reprezintă un volum de date foarte important ce conțin informații care trebuiesc extrase şi interpretate. Segmentarea are rolul de a extrage indicii vizuali sintetici care să conducă la reprezentarea imaginii sub formă de primitive semantice lipsite de ambiguitate. Această etapă, necesară reducerii volumului informației percepute, depinde de conținutul imaginii, deoarece în analiza diferitelor scene (peisaje, clădiri sau fețe umane) nu sunt utilizate aceleaşi percepte. Într-adevăr, modelarea informației ce aparține unei imagini, având un conținut total necunoscut, necesită utilizarea unui mare număr de informații (textură, contur). Cu toate acestea, uneori, se presupune o cunoaştere prealabilă a scenei pentru a se putea ghida procesul de segmentare.

Pentru scenele de interior care sunt caracterizate de omogenitatea intensității sau a culorii, cea mai mare parte a obiectelor nu au ca şi caracteristică principală textura. Spre deosebire de acestea, pentru obiectele ce evoluează în mediile naturale nu se poate renunța la informația de textură care reprezintă în acest caz o mare parte a informației disponibile.

Cronologic, primitivele extrase din imagine au evoluat spre o complexitate mare: puncte, drepte, curbe, contururi şi în final regiuni, care regrupează punctele imaginii având aceeaşi caracteristică [GON 92]. Există în principal două tipuri de primitive pentru descrierea imaginii: contururi şi regiuni.

Primul pas în analiza imaginilor este, în general, cel de segmentare a acestora. Aceasta înseamnă divizarea imaginilor în părțile lor constituente (obiecte). Operația conduce la identificarea obiectelor de interes din imagine şi obținerea caracteristicilor care permit distingerea acestora față de alte obiecte şi fundal. Nivelul la care se realizează divizarea depinde de problema ce trebuie rezolvată. Din acest motiv segmentarea este operația care se află la granița dintre prelucrarea şi analiza imaginilor.

Segmentarea realizează gruparea pixelilor în vederea structurării imaginilor la înalt nivel, care creează posibilitatea asocierii acestora cu entități ale modelului. Prin segmentare se identifică părți sau regiuni din imagine, pentru a se realiza asocierea acestora cu elementele scenei tridimensionale. Modul în care este realizată segmentarea determină succesul sau insuccesul întregului algoritm de analiză a imaginii.

Algoritmii de segmentare a imaginilor alb-negru se bazează în general pe două proprietăți de bază a nivelelor de gri:

BUPT

- discontinuităţile, reprezentate prin contururi, care permit partiţionarea imaginii pe baza modificării locale a nivelurilor de gri şi
- similaritătile, reprezentate prin zonele de omogenitate relativă (regiuni), care se bazează în principal pe praguri, creşterea regiunilor şi divizarea-contopirea regiunilor.

Idealul descrierii unei imagini pleacă de la aceste două reprezentări semantice care trebuie să fie identice:

- conturul închis ce defineşte o regiune şi
- regiunea ce defineşte un contur.

În practică, rareori se obţin descrieri echivalente plecând de la aceste reprezentări semantice.

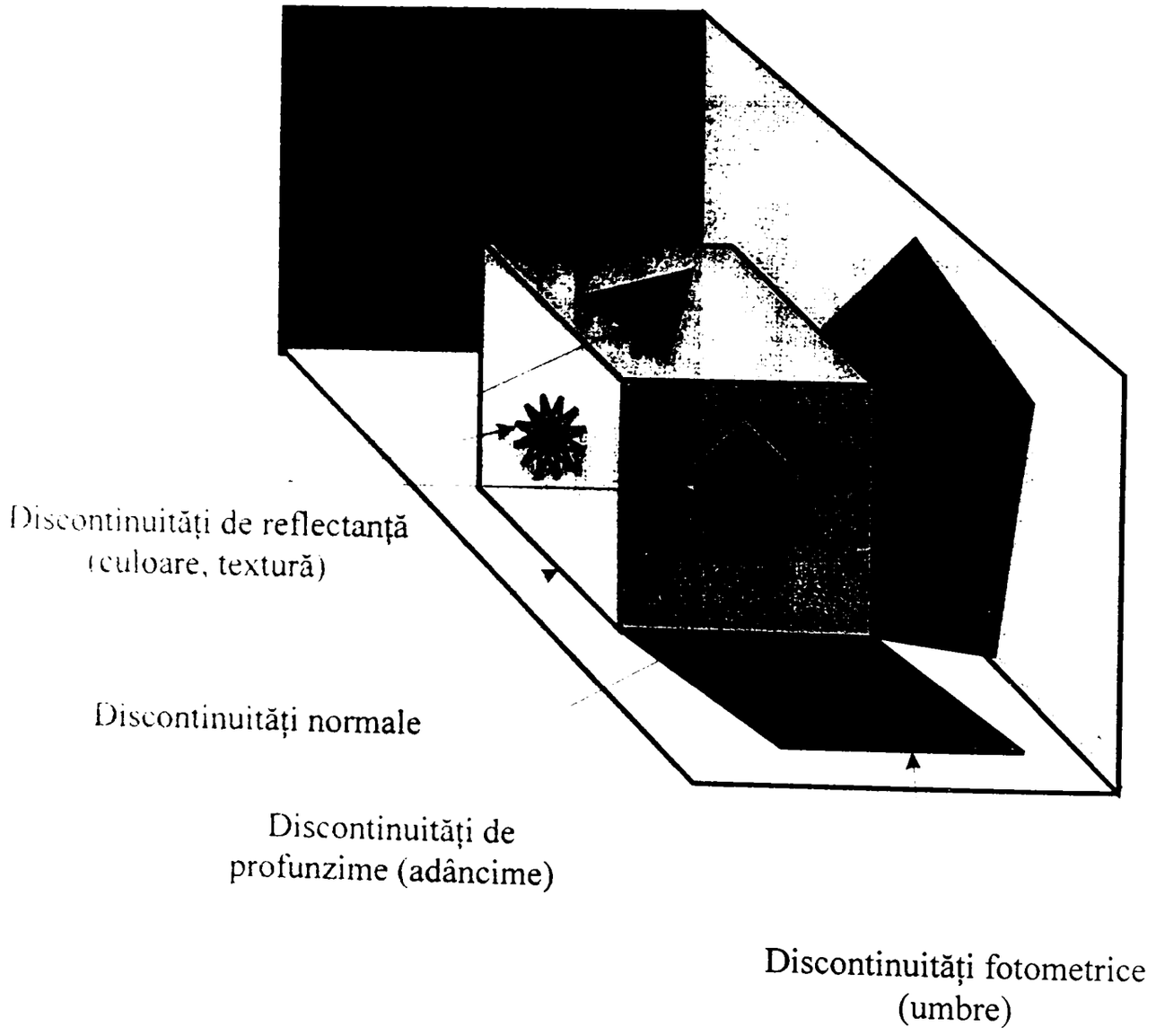


Fig. 4.1 Tipuri de discontinuităţi ce apar în cadrul scenei reale

BUPT

Aşa cum s-a menţionat, contururile reprezintă discontinuităţile imaginii. Să definim, mai întâi, noţiunea de discontinuitate.

Imaginea reprezintă funcţia intensităţii punctelor scenei şi depinde de foarte mulţi factori:

- caracteristicile obiectelor: formă, reflectivitatea suprafeţelor,
- geometria percepţiei: poziţia surselor luminoase, poziţia şi caracteristicile sistemului de captare.

În consecinţă, discontinuităţile care se observă nu sunt toate de aceeaşi natură (figura 4.1). Se pot distinge:

- discontinuităţi geometrice determinate de schimbarea sursei de iluminare (umbre),
- discontinuităţi normale (separarea între două feţe),
- discontinuităţi de profunzime (frontiere determinate de adâncimea imaginii),
- discontinuităţi determinate de reflectivitatea diferită a suprafeţelor (culoare, textură).

Fiecare din aceste tipuri de discontinuităţi are o semantică proprie, a cărei cunoaştere este necesară pentru a înţelege mai bine modul de obţinere a imaginii cu ajutorul sistemului de captare.

O clasificare a algoritmilor de segmentare poate fi:

- metode bazate pe pixel, în care se utilizează numai valorile individuale ale nivelurilor de gri ale pixelilor,
- metode bazate pe muchii, în care se determină muchiile după care se încearcă urmărirea acestora,
- metode bazate pe regiuni, care analizează valorile nivelurilor de gri pentru suprafeţe mai mari.

Metodele bazate pe pixeli caută discontinuităţile în imagini. Acestea pot fi puncte, linii sau muchii. În practică, cea mai obişnuită metodă de determinare a discontinuităţilor este aceea de a utiliza un operator liniar. Utilizarea acestuia are ca scop obţinerea accentuării contururilor şi a detaliilor fine. Prin utilizarea unui astfel de operator se realizează o filtrare trece sus a imaginii. Schema de prelucrare cu ajutorul operatorilor liniari poate fi reprezentată ca în figura 4.2.

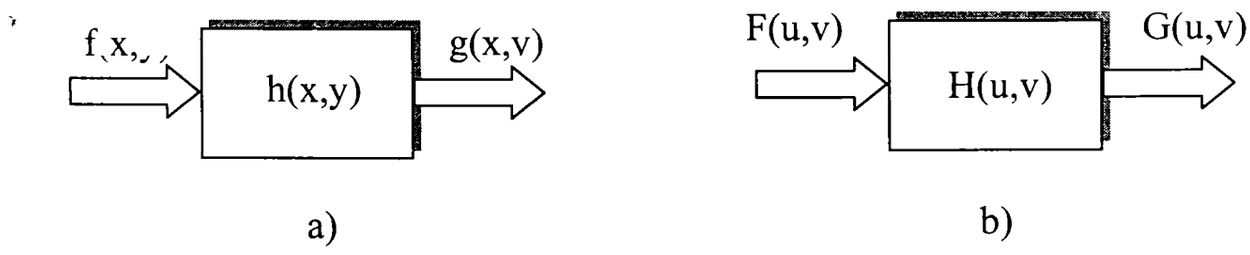


Figura 4.2 *Acţiunea unui sistem liniar a) în domeniul timp b) în domeniul frecvenţă*

BUPT

În modul cel mai general acţiunea unui astfel de operator poate fi descrisă de ecuaţia:

$$g(x,y) = h(x,y) * f(x,y) = \int_{-\infty}^{+\infty}\int h(s,t) f(x-s, y-t)\,ds\,dt \qquad (4.1)$$

Având în vedere că este vorba de o imagine numerică (discretă în timp), răspunsul unui operator într-un punct (i,j) al imaginii este dat de:

$$g(i,j) = \sum_{m=1}^{M}\sum_{k=1}^{K} h(m,k) f(i-m+1, j-k+1) \qquad (4.2)$$

Dacă considerăm operatorul de dimensiune 3×3 prin aplicarea acestuia unui punct (j,k) al imaginii (figura 4.3) vom obţine răspunsul de forma:

$$R = \sum_{i=1}^{9} c_i l_i \qquad (4.3)$$

unde l_i reprezintă nivelul de gri al pixelului asociat coeficientului c_i al operatorului. Uzual răspunsul sistemului este definit în raport cu centrul operatorului iar atunci când acesta este centrat pe un punct al marginii imaginii răspunsul se calculează utilizând periodicitatea, oglindirea sau repetarea rândurilor faţă de margine pentru eliminarea erorilor de margine. Aceşti operatori mai poartă şi denumirea de măşti.

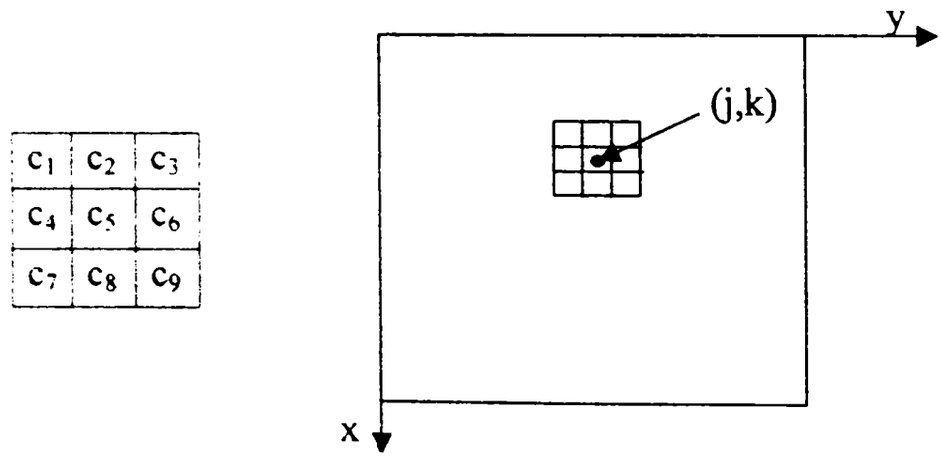


Fig. 4.3 Operatorul 3×3 generalizat şi aplicarea lui asupra imaginii

BUPT

4.2 Extragerea contururilor

4.2.1 Detecţia punctelor

Detecţia punctelor izolate din imagine este o primă aplicaţie a acestei filtrări. Astfel, utilizarea măştii din figura 4.4, bazându-se pe faptul că punctul prezintă nivel de gri diferit de vecinii săi, conduce la evidenţierea unui punct izolat dacă

$$|R| > T \tag{4.4}$$

unde T reprezintă un prag pozitiv ales iar R se obţine din ecuaţia (4.3). Această formulare reprezintă de fapt evidenţierea diferenţei dintre nivelurile de gri ale punctului central şi vecinii săi.

-1	-1	-1
-1	8	-1
-1	-1	-1

Fig. 4.4 Operator liniar pentru detecţia punctelor izolate din imagine

4.2.2 Detecţia liniilor

Creşterea nivelului de complexitate al detectorului presupune detecţia liniilor din imagine. În acest scop se pot utiliza măştile din figura 4.5. Aceste măşti aplicate unei imagini vor avea cel mai bun răspuns în cazul detecţiei liniilor, orizontale – masca din figura 4.5.a, înclinate la 45° – masca din figura 4.5.b, verticale – masca din figura 4.5.c şi respectiv înclinate la –45° – masca din figura 4.5.d.

-1	-1	-1
2	2	2
-1	-1	-1

a)

-1	-1	2
-1	2	-1
2	-1	-1

b)

-1	2	-1
-1	2	-1
-1	2	-1

c)

2	-1	-1
-1	2	-1
-1	-1	2

d)

Fig. 4.5 Operatori liniari pentru detecţia liniilor

În cazul aplicării acestor operatori asupra imaginii, rezultatul ne arată direcţia cea mai probabilă a liniei de care aparţine punctul pentru care a fost realizată prelucrarea. Dacă în urma aplicării operatorilor unui punct al imaginii va rezulta $|R_i| > |R_j|$, pentru $\forall\ i \neq j$ ($i,j \in \{1, 2, 3, 4\}$), atunci putem spune că punctul are cea mai probabilă asociere la o linie având direcţia operatorului i.

BUPT

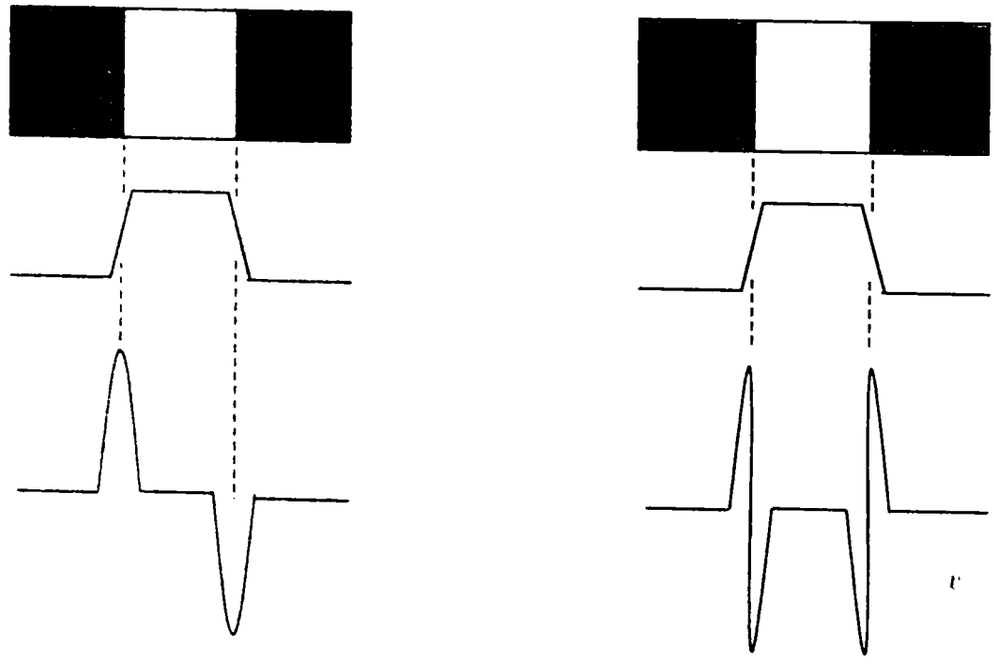
4.2.3 Detecţia muchiilor

Fără discuţie, punctele şi liniile reprezintă elemente necesare ale segmentării, respectiv îmbunătăţirii calităţii imaginilor, dar detecţia muchiilor reprezintă de departe scopul final al detecţiei discontinuităţilor în imagine. Aceasta, deoarece punctele izolate şi liniile nu sunt caracteristice celor mai multe aplicaţii, iar contururile au un rol important în modelarea, extragerea şi reprezentarea principalelor componente ale imaginilor.

În aceste condiţii muchia reprezintă graniţa dintre două regiuni având niveluri de gri relativ distincte. În cele ce urmează se presupune că regiunile au o anumită omogenitate care să permită determinarea tranziţiei dintre ele pe baza discontinuităţilor nivelurilor de gri [QIA 96], [TAG 90].

Ideea de bază ce stă la determinarea muchiilor o reprezintă determinarea maximului derivatei (gradientul funcţiei) de-a lungul acesteia. Astfel se pot distinge două metode de detecţie a muchiilor:

- măsurarea gradientului funcţiei de luminanţă $l(x,y)$ de-a lungul unei direcţii θ în cazul în care există tranziţii bruşte ale nivelelor de gri, figura 4.6.a, şi
- măsurarea derivatelor de ordinul II ale funcţiei de luminanţă în cazul în care regiunea de tranziţie este mai largă, figura 4.6.b.



a) b)

Fig. 4.6 Localizarea muchiilor cu ajutorul operatorilor de derivare

După cum se observă din figură, în cazul utilizării primei derivate se obţine un maxim, pozitiv la trecerea de la un nivel mai scăzut spre unul mai ridicat sau negativ la trecerea de la un nivel ridicat la unul mai scăzut, în raport cu mijlocul tranziţiei. În cazul utilizării derivatei a doua se observă că se obţine o trecere prin

BUPT

zero chiar la mijlocul tranziţiei între nivelurile de gri, ceea ce determină o localizare mai bună a muchiilor.

Pe baza detecţiei conturului utilizând gradientul s-au dezvoltat tehnici bazate pe operatori liniari. Aceştia pot fi operatori gradient sau operatori unghiulari. Gradientul unei imagini într-un punct (x,y) este un vector dat de relaţia:

$$\overline{\nabla f} = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \\ \dfrac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} G_x \\ \\ G_y \end{bmatrix} \tag{4.5}$$

În detecţia muchiilor o mărime importantă o constituie modulul vectorului gradient, denumit simplu gradient ∇f, şi care are valoarea:

$$\nabla f = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} = \sqrt{G_x^2 + G_y^2} \tag{4.6}$$

Direcţia acestui gradient, notată prin θ_g, reprezintă de asemenea o mărime importantă fiind unghiul vectorului faţă de axa x în punctul (x,y).

$$\theta_g = artctg\left(\frac{\dfrac{\partial f}{\partial y}}{\dfrac{\partial f}{\partial x}}\right) = arctg\left(\frac{G_y}{G_x}\right) \tag{4.7}$$

reprezentând direcţia relativă a muchiei faţă de imaginea înconjurătoare.

Aşa cum s-a amintit mai sus, calculul gradientului într-un punct pentru o imagine discretă se realizează prin determinarea gradienţilor după cele două direcţii. Un element important în utilizarea operatorilor gradient îl constituie alegerea coeficienţilor. Acest obiectiv este important mai ales din perspectiva direcţiei gradientului, deoarece orice eroare poate determina obţinerea de contururi eronate.

Cei mai simpli operatori utilizaţi pentru calculul gradientului sunt operatorii având dimensiunea 2×2 (figura 4.7.a) şi 3×3 (figurile 4.7.b, 4.7.c, 4.7.d). Deoarece aceştia sunt foarte sensibili la zgomot au fost înlocuiţi de operatorii 5×5 (figura 4.8).

Operatorii Roberts
[CAN 86]

1	0
0	-1

0	1
-1	0

a)

BUPT

Operatorii Prewitt [PRE 70]

-1	-1	-1
0	0	0
1	1	1

-1	0	1
-1	0	1
-1	0	1

b)

Operatorii Sobel [GON 92]

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

c)

Operatorii izotropi [FRE 77]

1	$\sqrt{2}$	1
0	0	0
-1	$-\sqrt{2}$	-1

-1	0	1
$-\sqrt{2}$	0	$\sqrt{2}$
-1	0	1

d)

Fig. 4.7 Operatorii gradient utilizaţi la detecţia muchiilor

Dacă în imagine există obiecte relativ mari şi dacă aceasta este zgomotoasă atunci este recomandată utilizarea unui operator de dimensiuni 9×9. Creşterea dimensiunii operatorilor face ca rezultatul să devină sensibil la prezenţa muchiilor învecinate şi conduc de asemenea la o localizare mai puţin exactă a muchiilor.

-10	-10	0	10	10
-17	-17	0	17	17
-20	-20	0	20	20
-17	-17	0	17	17
-10	-10	0	10	10

a)

10	17	20	17	10
10	17	20	17	10
0	0	0	0	0
-10	-17	-20	-17	-10
-10	-17	-20	-17	-10

b)

Figura 4.8 Operatorii Sobel 5×5

Una din metodele de implementare a metodei de extragere a muchiilor cu ajutorul operatorilor gradient este prezentată în figura 4.9:

În figura 4.10 se prezintă rezultatele prelucrării imaginii Lena cu ajutorul algoritmului de detecţie a muchiilor prezentat în figura 4.9 şi prin utilizarea operatorilor Sobel prezentaţi în figura 4.8.

BUPT

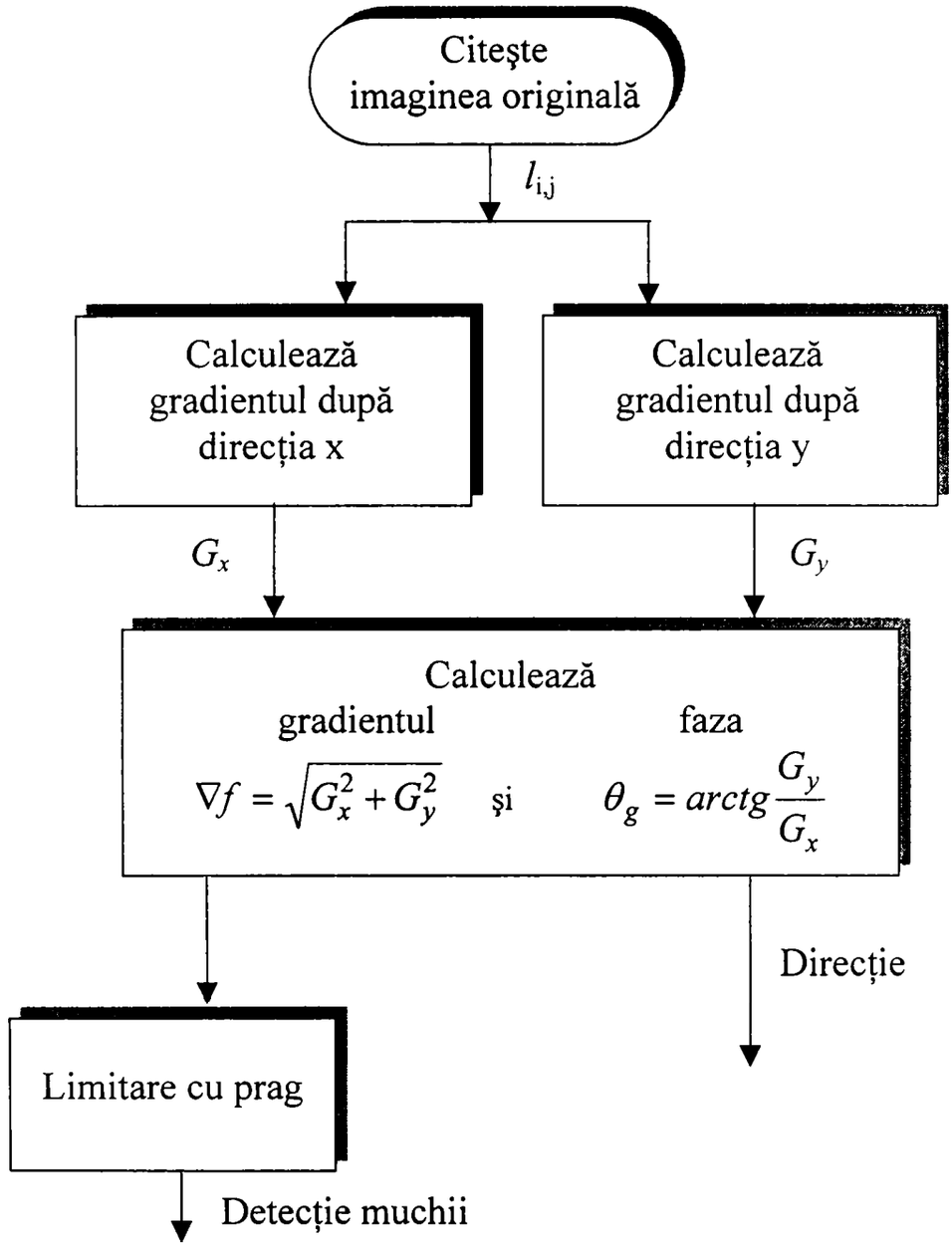
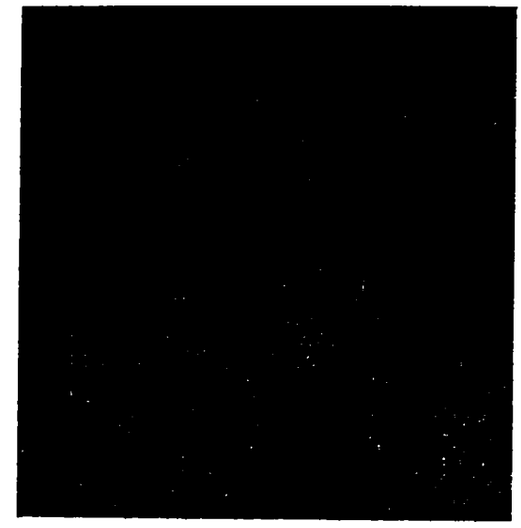


Fig. 4.9 Algoritmul de detecţie a muchiilor cu ajutorul operatorilor gradient

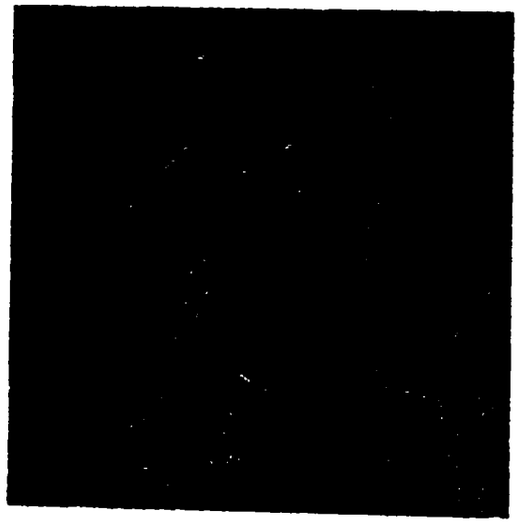
Pe lângă operatorii gradient mai pot fi utilizaţi pentru detecţia muchiilor şi aşa numiţii operatori cardinali cu ajutorul cărora se calculează gradientul într-un număr de direcţii stabilit (de obicei numărul acestora este 8). În figura 4.11 se prezintă patru tipuri de operatori cardinali, care realizează detecţia muchiilor orientate spre nord. Prin rotirea cu un unghi de 45° a elementelor măştii în sens trigonometric se obţin operatori pentru detecţia gradientului pe direcţii rotite cu 45°.

BUPT

Imaginea originală



Imaginea gradientului după axa x



Imaginea gradient după axa y

Imaginea gradient

Imagine gradient obținută
după limitare

Figura 4.10 *Detecția contururilor cu ajutorul operatorilor gradient*

BUPT

1	1	1
1	-2	1
-1	-1	-1

a)

5	5	5
-3	0	-3
-3	-3	-3

b)

1	1	1
0	0	0
-1	-1	-1

c)

1	2	1
0	0	0
-1	-2	-1

d)

Fig. 4.11 Operatori cardinali orientaţi pentru detecţia liniilor pe direcţia Nord

În aceste condiţii gradientul pentru locaţia (x,y) este definit:

$$\nabla f = \max_{k \in O_c} \{ |\nabla f_k| \} \tag{4.8}$$

unde $|\nabla f_k|$ reprezintă modulul gradientului după direcţia k obţinut prin aplicarea operatorului pentru această direcţie iar O_c reprezintă mulţimea operatorilor cardinali.

Întrucât în cazul sistemului vizual uman există o tendinţă de filtrare a frecvenţelor spaţiale şi temporale funcţie de nivelul de iluminare, este mai avantajoasă utilizarea derivatelor de ordinul doi, în locul gradienţilor, la detecţia contururilor. Un operator de ordinul doi des întâlnit este operatorul Laplace:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{4.9}$$

Ca şi în cazul gradientului, ecuaţia (4.9) poate fi implementată în formă numerică în mai multe moduri. Cele mai frecvente implementări discrete ale operatorilor Laplacieni sunt prezentate în figura 4.12. Cerinţele de bază ale acestor operatori sunt:

- coeficientul asociat pixelului de interes din imagine este pozitiv,
- coeficienţii asociaţi celorlalţi pixeli sunt de cele mai multe ori negativi,
- suma coeficienţilor operatorului este nulă (acest lucru conduce la obţinerea unui rezultat nul ori de câte ori punctul pentru care se calculează Laplacianul are aceeaşi valoare cu vecinii săi).

0	-1	0
-1	4	-1
0	-1	0

a)

-1	-1	-1
-1	8	-1
-1	-1	-1

b)

1	-2	1
-2	4	-2
1	-2	1

c)

Fig. 4.12 Operatori Laplacieni

BUPT

Deoarece apar derivatele de ordinul II, operatorul Laplacian este mult mai sensibil la zgomot decât cei anteriori. De asemenea, datorită apariţiei muchiilor duble (figura 4.6) şi a incapacităţii de detecţie a direcţiei muchiilor, operatorul nu este utilizat în această formă. O modalitate mai eficientă de detecţie a muchiilor este aceea de a utiliza trecerile prin zero ale derivatei de ordinul doi [CLA 89], [PIE 90]. În acest scop este utilizat Laplacianul funcţiilor Gaussine, care are următoarea formă:

$$h(x,y) = \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{4.10}$$

unde σ reprezintă lăţimea funcţiei gaussiene. Dacă notăm $r^2 = x^2 + y^2$ atunci derivata a II-a în raport cu r este:

$$\nabla^2 h = \left(\frac{r^2-\sigma^2}{\sigma^4}\right)\exp\left(-\frac{r^2}{2\sigma^2}\right) \tag{4.11}$$

Se observă că trecerile prin zero ale acestei funcţii au loc pentru $r = \pm\sigma$, iar forma funcţiei ∇h^2 ne conduce la operatorii din figura 4.12.

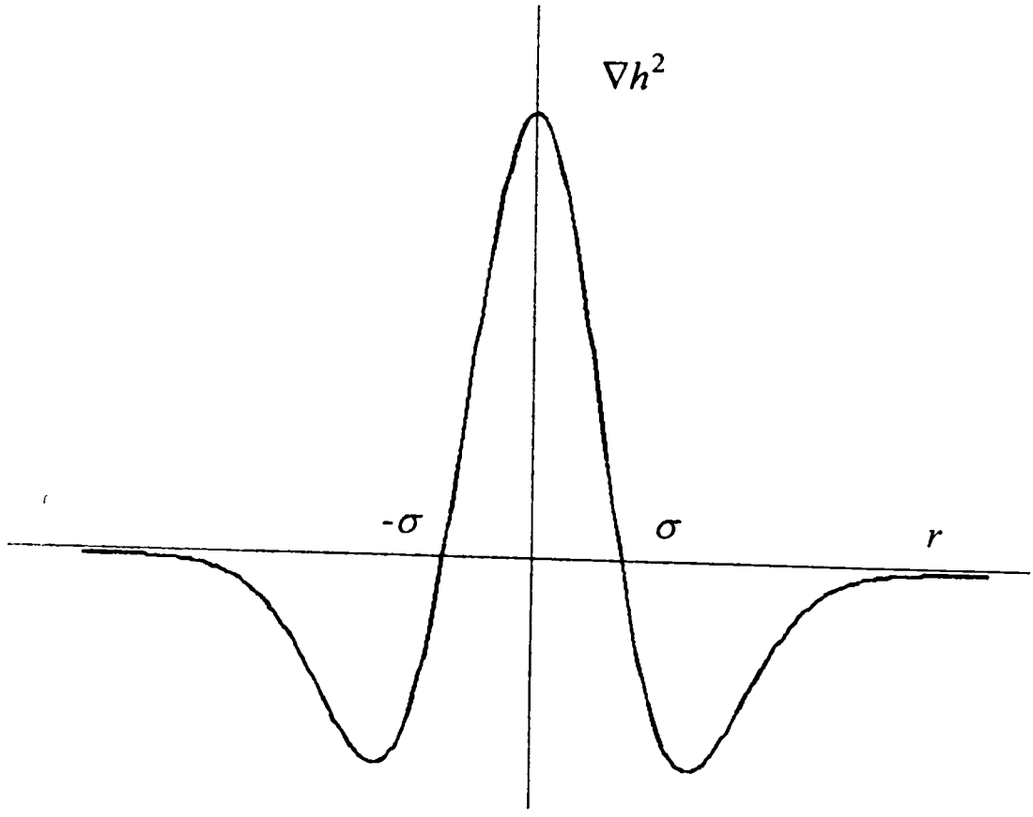


Fig. 4.13 Reprezentarea funcţiei ∇h^2

Dacă se consideră imaginea apple (figura 4.14.a) şi se realizează convoluţia acesteia cu operatorul din figura 4.12.a, atunci se obţine rezultatul din figura

BUPT

4.14.b. Dacă se realizează convoluţia imaginii Lena (figura 4.15.a) se obţine rezultatul din figura 4.15.b

a) b)

Fig. 4.14 Imaginea originală şi rezultatul convoluţiei cu operatorul Laplacian în cazul imaginii apple

a) b)

Fig. 4.15 Imaginea originală şi rezultatul convoluţiei cu operatorul Laplacian în cazul imaginii Lena

Din rezultatele obţinute se poate observa că detecţia muchiilor cu ajutorul gradientului conduce la rezultate destul de bune în cazul imaginilor având tranziţii abrupte ale intensităţii şi zgomot scăzut. Trecerile prin zero ale operatorului Laplacian ne ajută la detecţia cu mai bună precizie a muchiilor, la reducerea efectelor zgomotelor, cu dezavantajul creşterii complexităţii şi a duratei calculului.

BUPT

4.3 Segmentarea bazată pe prag

Segmentarea bazată pe prag reprezintă una din cele mai utilizate metode pentru obţinerea unei partiţionări a imaginii [LEE 90], [CHA 98], [CHA 96]. Tehnicile din această categorie se dovedesc foarte utile atunci când există o separare relativ clară între nivelurile de gri ale obiectelor analizate şi fundal (histograme bimodale). În multe situaţii, pentru a se ajunge la acest caz favorabil, sunt necesare prelucrări iniţiale care să permită obţinerea unor histograme cu o formă avantajoasă.

Acest tip de segmentare face parte din cele bazate pe pixeli, pentru decizia de apartenenţă a acestora la obiect sau fundal luându-se în consideraţie numai valorile nivelurilor de gri. În acest caz obţinerea unei imagini segmentate este condiţionată de determinarea nivelului de gri ce caracterizează obiectul de interes. Acest nivel reprezintă pragul după care se realizează segmentarea. După modul în care se realizează determinarea acestuia putem deosebi mai mulţi algoritmi de segmentare:

- segmentare prin determinarea pragului global
- segmentare prin determinarea pragului optim
- segmentare prin determinarea pragului pe baza caracteristicilor muchiilor
- segmentare prin determinarea pragului pe baza mai multor variabile

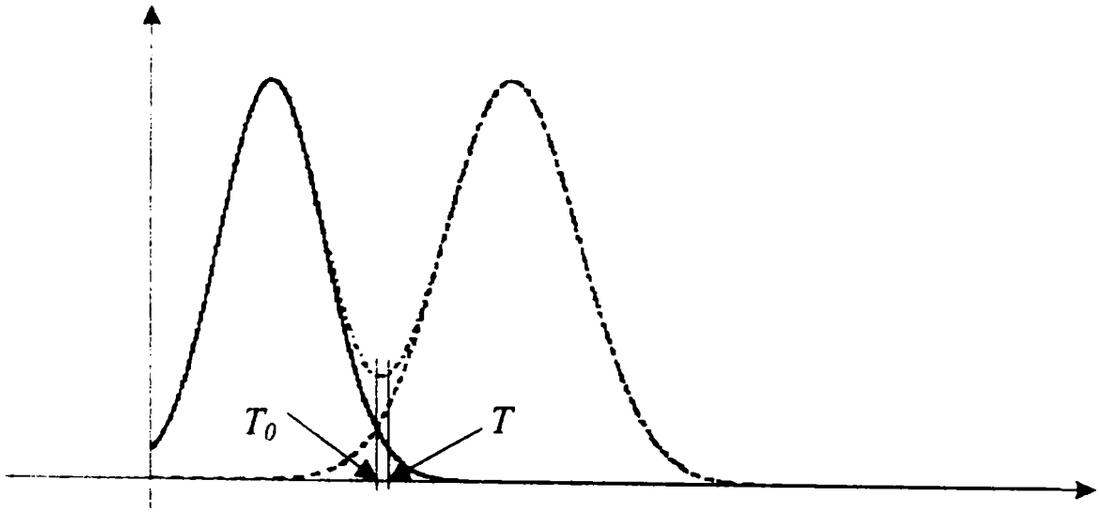


Fig. 4.16 Poziţiile pragurilor în cazul segmentării cu prag

4.3.1 Segmentare prin determinarea pragului global

Segmentarea cu prag unic reprezintă metoda de bază, fiind în acelaşi timp o tehnică simplă, rapidă şi eficientă în multe cazuri. Ea poate fi aplicată în cazul unei imagini formată dintr-un obiect amplasat pe un fundal. Determinarea pragului T se realizează prin analiza histogramei, segmentarea obţinându-se prin explorarea imaginii şi atribuirea unei etichete fiecărui pixel. Prin această etichetare se

BUPT

realizează practic determinarea apartenenţei unui pixel la obiect sau fundal. Succesul acestei metode depinde de alegerea pragului T.

Un exemplu de determinare a pragului este prezentat în figura 4.16, în care pragul este ales ca fiind minimul local al histogramei, în care s-a considerat că histograma intensităţilor este obţinută prin aproximarea unei duble funcţii Gaussiene pentru un zgomot normal distribuit.

4.3.2 Segmentare prin detecţia pragului optim

Poziţia minimului histogramei intensităţilor nu este cea mai bună alegere a pragului datorită efectelor de minime şi maxime locale ce apar în histograma. Din acest motiv pragul optim T_0 (din punct de vedere al erorii) se determină la intersecţia celor două distribuţii (figura 4.15) şi este dat de ecuaţia [Gon 93]:

$$\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}\right)T_0^2+2\left(\frac{\mu_2}{\sigma_2^2}-\frac{\mu_1}{\sigma_1^2}\right)T_0+\frac{\mu_1^2}{\sigma_1^2}-\frac{\mu_2^2}{\sigma_2^2}+2\ln\frac{p_2\sigma_1}{p_1\sigma_2}=0 \qquad (4.12)$$

unde p_1 şi p_2 definesc proporţiile pixelilor în jurul maximelor, μ_1 şi μ_2 ($\mu_1 < \mu_2$) definesc poziţiile maximelor, iar σ_1 şi σ_2 definesc lăţimile celor două distribuţii. În cazul în care cele două distribuţii sunt egale ($\sigma_1 = \sigma_2$) se obţine un prag optim dat de relaţia:

$$T_0=\frac{2\sigma^2\ln(p_2/p_1)+\mu_1^2-\mu_2^2}{2(\mu_1-\mu_2)} \qquad (4.13)$$

Această condiţie restrictivă impusă de algoritmul lui Otsu [OTS 79], şi care impune forma histogramei, a fost evitată în algoritmii elaboraţi de Kittler şi Illingworth [KIT 85], Kurita ş.a. [KUR 92], Grennhill şi Davies [GRE 95], arătând totodată că metoda se comportă mai bine în anumite situaţii datorită numărului mai mic de parametri necesari.

Pentru histograme reale, având distribuţii apropiate, dar nu strict Gaussiene, selecţia pragului în punctul minim al histogramei constituie un rezultat satisfăcător. Prezenţa zgomotului, care generează numeroase neregularităţi ale histogramei, conduce la dificultăţi în determinarea minimului din histogramă, făcând inoperant algoritmul de tipul:

- se caută maximele locale şi se declară ca moduri ale histogramei nivelurile determinate,
- se determină nivelul de gri minim situat între cele două maxime şi se alege ca fiind pragul de segmentare.

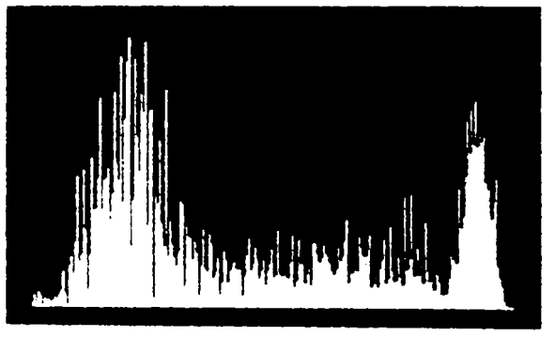
Există astfel posibilitatea de a se introduce minime şi maxime locale ce reprezintă discontinuităţi în forma histogramei (figura 4.17.b) şi care pot conduce la erori foarte mari. Soluţia o reprezintă aplicarea unei operaţii de netezire a histogramei, prin aplicarea unui operator de mediere de tipul:

BUPT

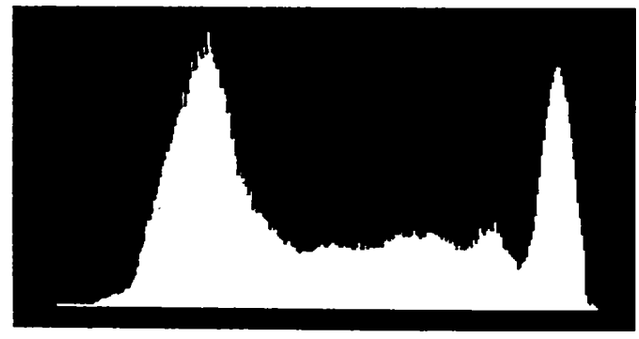
$$w = \frac{1}{10}[1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1] \tag{4.14}$$

care conduce la histograma din figura 4.17.c.

a)



b)



c)

Fig. 4.17 Netezirea histogramei pentru determinarea pragurilor de segmentare

Aplicarea acestei soluţii prezintă dezavantajul depărtării minimului de poziţia optimă în cazul în care unul din cele două moduri ale histogramei este mult mai mic în comparaţie cu celălalt. Însă prin revenirea la histograma iniţială se elimină minimele şi maximele locale, determinându-se un prag optim.

4.3.3 Segmentare bazată pe caracteristicile muchiilor

Unul din cele mai importante aspecte ale selecţiei pragului de segmentare îl reprezintă capacitatea de selecţie a maximelor corespunzătoare fiecărui obiect într-o anumită histogramă. Această capacitate reprezintă una din cele mai importante caracteristici pentru selecţia automată a pragurilor, în special în cazurile în care se pot modifica caracteristicile imaginii într-un anumit domeniu de

BUPT

distribuţie al intensităţilor. Alegerea unui prag de segmentare cât mai bun este determinată de forma histogramei (maxime cât mai înalte, înguste, simetrice şi separate de văi adânci).

În cazul în care avem obiecte mici aflate pe un fundal este clar că histograma depinde de dimensiunile relative ale acestora. Dacă se consideră numai pixelii din zonele de separaţie dintre obiecte, atunci cele două maxime datorate obiectului şi fundalului vor avea aproximativ aceiaşi amplitudine, se va îmbunătăţi simetria histogramei şi se va accentua valea dintre cele două vârfuri.

Principala dificultate în abordarea anterioară o constituie presupunerea cunoaşterii graniţelor dintre obiect şi fundal. Aceasta, însă, nu este o informaţie cunoscută pe durata segmentării. Din acest motiv se poate utiliza informaţia gradientului sau Laplacianului pentru segmentare.

Aplicarea metodei propuse se realizează prin crearea unei imagini având trei niveluri după cum urmează:

$$i_s = \begin{cases} 0 & dacă \nabla f < T \\ + & dacă \nabla f \geq T şi \nabla^2 f \geq 0 \\ - & dacă \nabla f \geq T şi \nabla^2 f < 0 \end{cases} \qquad (4.15)$$

unde 0, +, – reprezintă simbolurile pentru cele trei niveluri ale imaginii. Această metodă este foarte utilă în cazul în care se doreşte generarea unei imagini segmentate binare.

Metoda anterioară are ca scop identificarea nivelurilor de gri având o importanţă esenţială în ceea ce priveşte schimbările de nivel de gri ale suprafeţelor mai mari (determinarea nivelurilor de gri din zona muchiilor). Aşa cum s-a arătat la detecţia contururilor, operatorul Laplacian este unul din cele mai exacte în determinarea poziţiei muchiilor. În acest scop, se va realiza o convoluţie între imagine şi operatorul definit în figura 4.12, obţinând o imagine a Laplacianului.

Întrucât ne interesează numai punctele în care Laplacianul are valorile cele mai mari se va realiza o segmentare a imaginii Laplacianului prin identificarea unui prag, sub care se consideră că imaginea nu mai conţine detalii semnificative. Acest prag se determină punând condiţia ca numărul nivelurilor ce depăşesc pragul să reprezinte aproximativ 10% din numărul total al pixelilor din imagine. În acest fel rezultă o imagine mască ce conţine evidenţiaţi pixelii reprezentând zonele de interes din imagine din punct de vedere al segmentării. Pentru a obţine nivelul real al punctelor de interes din imagine se realizează un ŞI între imaginea originală şi imaginea mască (nivelul pixelului este 0, dacă nivelul corespunzător măştii este 0 şi cel real dacă nivelul corespunzător măştii este 255). Din această imagine se obţine histograma ce generează pragurile pentru segmentare.

Figura 4.18 prezintă un exemplu de implementare a algoritmului pe o imagine dintr-o secvenţă. Se observă că se obţine un număr foarte mare de regiuni (2218).

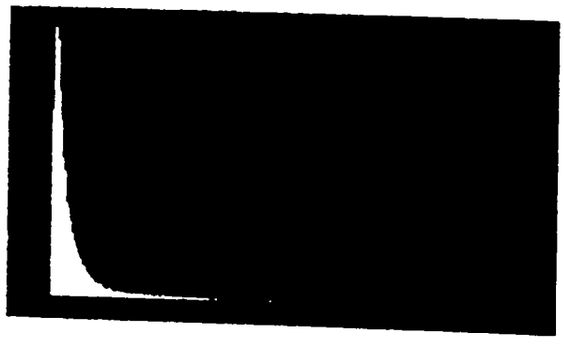
BUPT

Imaginea originală

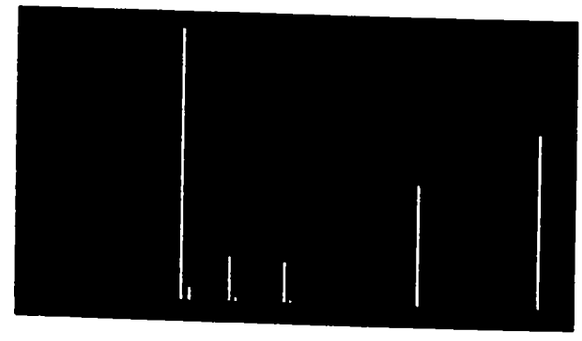
Imaginea Laplacianului

Imaginea mască

Imaginea punctelor de interes



Histograma imaginii Laplacianului

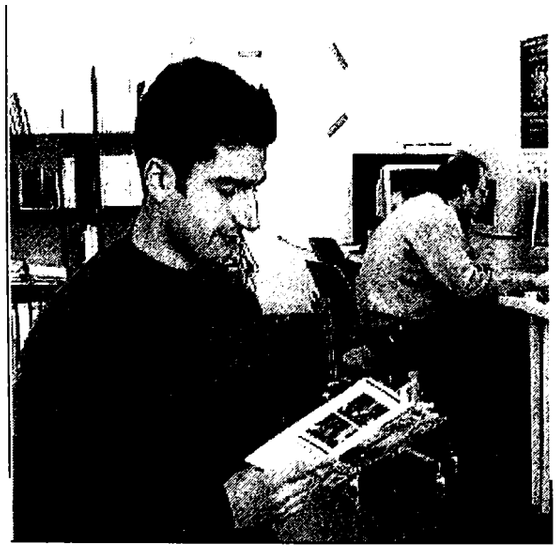


Histograma pragurilor

BUPT

Imaginea segmentată 2218 regiuni

Fig. 4.18 Segmentarea imaginii pe baza caracteristicilor muchiilor



Imaginea originală

Imaginea segmentată 1574 regiuni

Fig. 4.19 Rezultatul segmentării cu ajutorul algoritmului ISODATA

4.3.4 Segmentare cu căutare automată a pragului (ISODATA)

Algoritmul are ca scop determinarea automată a unui prag pentru segmentare, care urmăreşte împărţirea optimă a imaginii pe niveluri de gri. Aceasta se realizează prin analiza histogramei şi identificarea unui punct care îndeplineşte condiţia medie a centrelor de greutate ale zonelor histogramei. Rezultatul segmentării este prezentat în figura 4.19, iar structura algoritmului este prezentată în figura 4.20.

BUPT

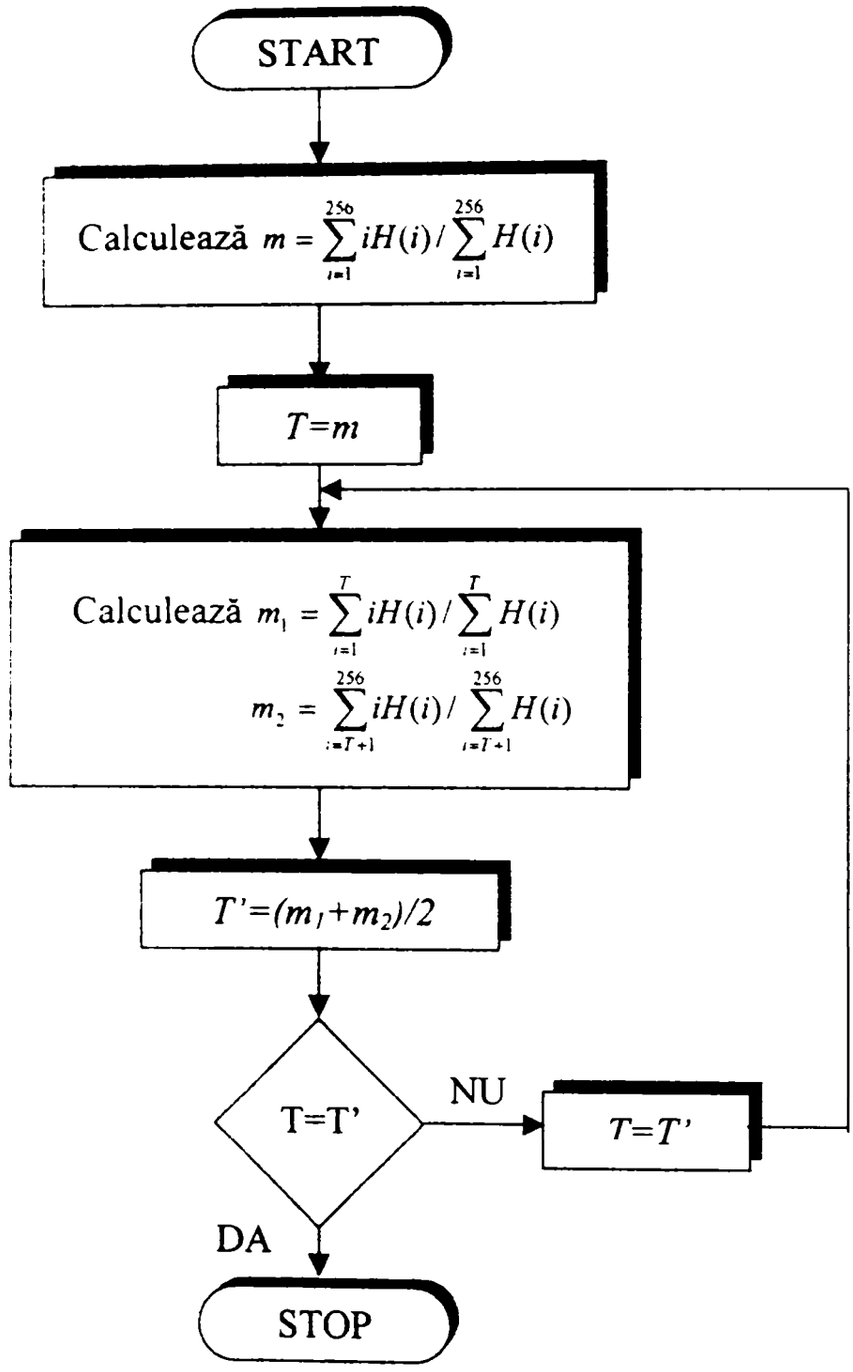


Fig. 4.20 Ordinograma de determinare a pragului prin algoritmul ISODATA

Algoritmul se poate implementa şi pentru mai multe praguri care se pot obţine într-o etapă anterioară sau se pot calcula impunând gradientul maxim pentru care se consideră o regiune ca fiind omogenă. După determinarea pragurilor se aplică o segmentare simplă cu prag, înlocuindu-se valorile pixelilor dintr-o regiune cu media nivelurilor de gri (centrul de greutate dintre cele două praguri). Această metodă are avantajul minimizării erorii medii pătratice dintre imaginea originală şi cea segmentată.

BUPT

4.4 Segmentare orientată pe regiuni

Dacă până acum s-a abordat problema segmentării din punct de vedere al determinării graniţelor dintre regiuni pe baza discontinuităţilor de intensitate (paragraful 4.1.1) sau pe determinarea unui prag pe baza distribuţiei pixelilor (paragraful 4.2), în acest paragraf se abordează problema segmentării bazată pe determinarea directă a regiunilor.

4.4.1 Definirea problemei

Problema segmentării imaginilor reprezintă o problemă larg abordată în literatura privind prelucrarea imaginilor şi vederea artificială [BAL 82], [GON 92], [HAR 92], [LEV 85]. Definiţiile utilizate sunt aproape identice şi din analiza procesului de segmentare rezultă următoarea definiţie [GON 92].

Segmentarea reprezintă divizarea imaginii I în n regiuni $R_1, R_2, \ldots, R_n$ în aşa fel încât să avem îndeplinite condiţiile:

(a) $I = \bigcup_{i=1}^{n} R_i$,

(b) R_i este o regiune conectată, $i = 1,2,\ldots,n$,

(c) $R_i \cap R_j = \Phi$ pentru $\forall\ i \neq j$,

(d) $P(R_i) =$ ADEVĂRATĂ pentru orice pixel din regiunea R_i,

(e) $P(R_i \cup R_j) =$ FALSĂ pentru $i \neq j$,

unde $P(R_i)$ reprezintă criteriul de omogenitate acceptat pentru pixelii regiunii R_i.

Condiţia (a) arată o segmentare completă. Cea de-a doua condiţie (b) reclamă existenţa conectivităţii pentru toate punctele regiunii, aceasta fiind o condiţie care nu apare în cazul definirii segmentării orientată pe regiuni de către Levine [LEV 85]. Prin condiţia (c) se arată faptul că regiunile sunt disjuncte. Condiţia (d) arată că pixelii unei regiuni trebuie să respecte criteriul de omogenitate ales (aceeaşi intensitate), în timp ce condiţia (e) impune ca în regiuni diferite condiţiile de omogenitate să fie diferite. Acest din urmă criteriu poate fi completat cu încă o condiţie, şi anume ca regiunile să nu fie adiacente deoarece poate exista un acelaşi criteriu de omogenitate dacă regiunile nu au graniţe comune [HOO 96].

Pe lângă micile diferenţe ce apar în definirea procesului de segmentare, există serioase dificultăţi în utilizarea acestei definiţii pentru realizarea segmentării. Cu toate acestea, există foarte mulţi algoritmi de segmentare bazată pe regiuni. O posibilă clasificare a lor este:

- creşterea regiunilor,
- algoritmi de tipul divide şi contopeşte (split and merge),
- algoritmi de segmentare morfologică.

BUPT

4.4.2 Creşterea regiunilor

Aşa după cum se observă şi din denumirea algoritmului, această procedură se bazează pe existenţa unui grup iniţial de pixeli (germene – seeds), din care, prin adăugarea de noi pixeli cu proprietăţi similare (nivel de gri, textură, culoare), se formează regiunea finală [BEV 89], [ADA 94], [ZHU 96], [HAJ 98], [LEO 95].

Pentru a ilustra această procedură se consideră că puncte iniţiale sunt punctele evidenţiate în figura 4.21.a. Ca rezultat se obţin cel puţin trei regiuni asociate acestor puncte. Cel mai simplu criteriu de includere a unui pixel într-o regiune este acela de a testa apartenenţa nivelului de gri a acestuia la un interval (diferenţa absolută dintre nivelurile de gri ale pixelului şi pixelului iniţial să nu difere cu mai mult de un anumit prag T). Pentru pixelii ce satisfac simultan condiţia pentru mai multe puncte iniţiale, atribuirea se face în mod arbitrar regiunii pentru care diferenţa este pozitivă. În exemplul prezentat, presupunând T=4, rezultă regiunile ca în figura 4.21.b. Dacă se selectează pragul ca fiind egal cu 7 rezultă regiunile din figura 4.21.c.

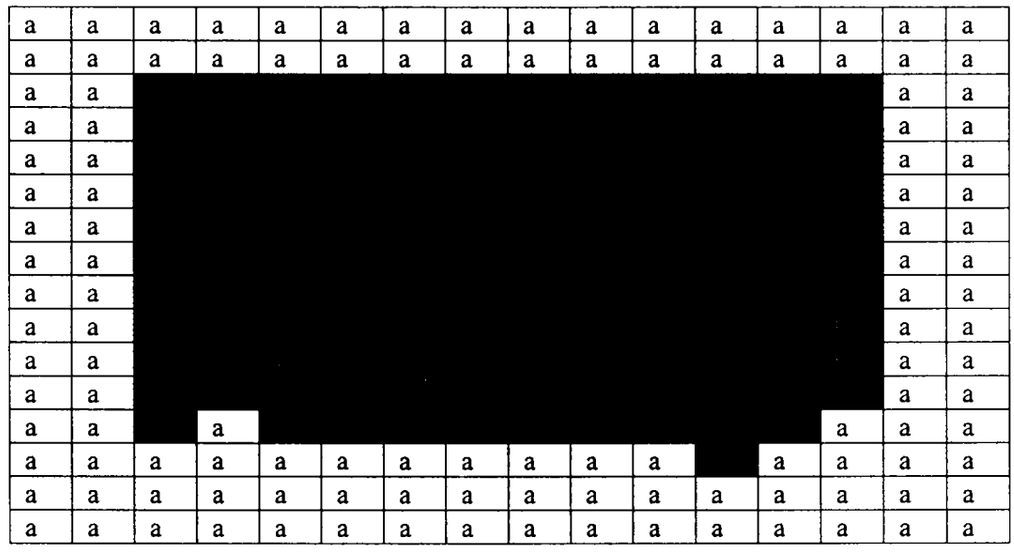
1	1	1	1	1	2	1	2	1	1	1	2	1	1	1	2
2	2	1	2	3	3	2	2	3	1	2	1	1	2	1	1
1	2		11	10	12	11	11	10	10	10	10	10	9	3	2
3	4	10	10	10	12	13	14	13	13	13	12	11	10	1	3
1	3	10	10	9	13	15	17	19	20	16	12	11	10	2	1
1	2	10	11	10	11	16	20	21	20	18	11	12	10	5	1
1	1	9	10	11	11	15	19		18	17	13	10	8	4	1
1	2	10	10	12	13	18	20	22	21	15	14	11	10	1	1
1	1	9	10	10	13	17	19	18	17	14	14	12	11	3	1
1	3	10	11	10	12	11	11	11	12	12	13	10	10	2	1
1	2	10	10	12	10	10	11	12	12	10	11	11	9	1	1
1	2	9	10	11	10	12	11	10	11	11	10	11	10	1	2
1	2	10	8	10	9	11	10	10	9	10	9	10	8	2	1
1	1	2	3	4	3	3	2	3	4	4	10	3	3	2	1
1	1	1	3	2	2	1	2	1	1	2	1	2	1	1	2
1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1

a)

a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
a	a													a	a
a	a													a	a
a	a													a	a
a	a													a	a
a	a													a	a
a	a													a	a
a	a													a	a
a	a													a	a
a	a													a	a
a	a													a	a
a	a													a	a
a	a													a	a
a	a	a	a	a	a	a	a	a	a	a		a	a	a	a
a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a

b)

BUPT



c)

Fig. 4.21 Creşterea regiunilor pornind de la puncte iniţiale cunoscute

Din acest exemplu rezultă două probleme ce apar în acest caz şi anume:

- selecţia punctului iniţial care reprezintă cel mai bine regiunea, care depinde în cea mai mare măsură de natura problemei ce trebuie rezolvată, şi
- selecţia criteriului de omogenitate după care se face includerea unui pixel în regiunea de interes, care nu depinde numai de natura problemei ci şi de imaginea asupra căreia se aplică.

Imagine iniţială

Imagine după filtrare mediană

Fig. 4.22 Acţiunea filtrului median asupra unei imagini înecate în zgomot binar

Pentru a elimina problemele de suprapunere ce apar în cazul selecţiei mai multor puncte iniţiale, algoritmul poate pleca de la un singur punct faţă de care se

BUPT

creează o anumită regiune prin ataşarea tuturor pixelilor vecini ce îndeplinesc condiţia de omogenitate aleasă. Regiunea următoare se construieşte plecând de la primul pixel, din vecinătatea primei regiuni create, care nu îndeplineşte condiţia de omogenitate. Această metodă este foarte sensibilă la zgomote. Pentru a se elimina posibilitatea obţinerii de regiuni ce nu corespund obiectelor reale, este necesară o prelucrare suplimentară pentru reducerea zgomotului fără a afecta contururile. În acest scop poate fi utilizată o filtrare mediană cu ajutorul unui operator [PIT 90], exemplificată în figura 4.22.

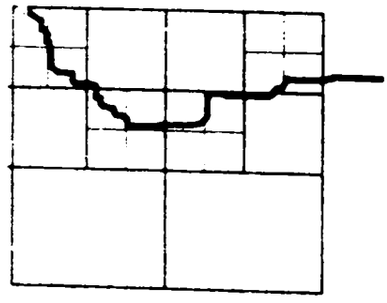
4.4.3 Algoritmi de tip divide şi contopeşte

Aceşti algoritmi de tip divide şi contopeşte s-au dezvoltat în conjucţie cu metodele de codare bazate pe blocuri [LEE 94], [STR 91], [TO 96] şi a reprezentării multirezoluţie a imaginilor [TAB 97], [VIN 97], [ROS 84], [WOO 91]. Pe de altă parte, acest algoritm a apărut ca şi alternativă la algoritmul de creştere a regiunilor. Plecând de la o împărţire arbitrară a imaginii, algoritmul constă în divizarea regiunilor neuniforme, urmând ca apoi să se realizeze o regrupare ţinând cont de criteriul utilizat.

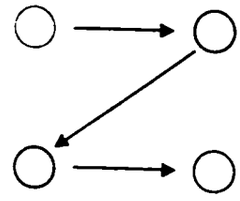
Structurile arborescente (quadtree) reprezintă o modalitate eficientă de descriere a regiunilor bidimensionale [SAM 90]. Utilizarea unei descompuneri regulate (blocuri pătrate) este echivalentă cu divizarea imaginii în blocuri de mărime fixă, care sunt apoi segmentate în 4 subblocuri egale. Imaginea segmentată, modul de segmentare şi structura arborescentă sunt prezentate în figura 4.23.

După fiecare operaţie de divizare mărimea unui subbloc este un sfert din blocul iniţial. Această operaţie este recursivă fiind repetată până când nu mai este necesară divizarea sau s-a ajuns la dimensiunea minimă posibilă. Fiecare operaţie de divizare este condiţionată de un test de omogenitate. În cazul în care testul este pozitiv blocul poate fi reprezentat printr-un set de parametri ai modelului folosit, iar în cazul în care este negativ blocul va fi divizat în patru blocuri egale. Această structură poate fi construită în două moduri, şi anume:

- de sus în jos, aşa numita construcţie top-down (figura 4.24),
- de jos în sus, construcţia bottom-up (figura 4.25).

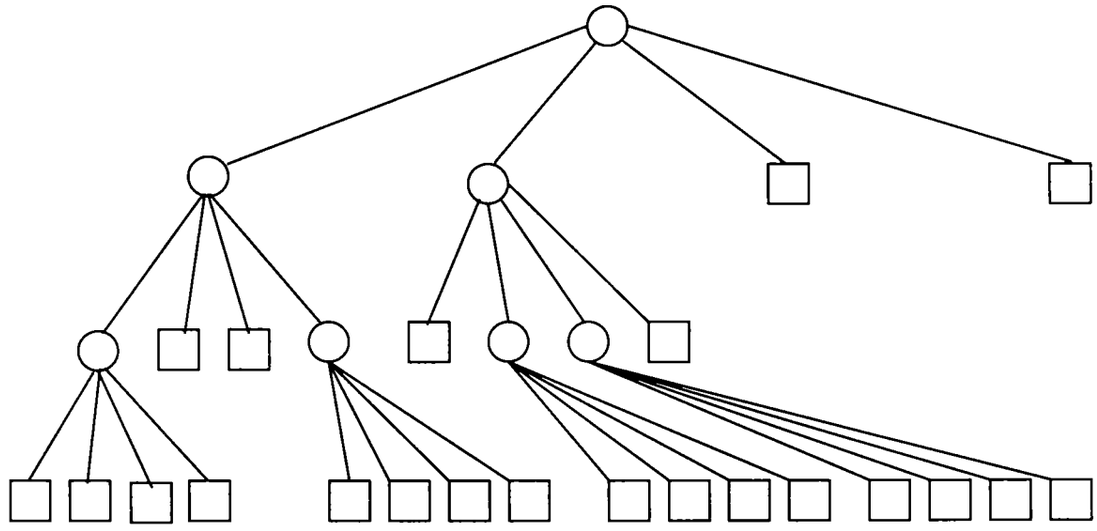


a)



b)

BUPT



c)

Fig. 4.23 Algoritmul quadtree pentru segmentarea unei imagini bidimensionale

Fig. 4.24 Construcţia top-down a structurii quadtree prin divizare

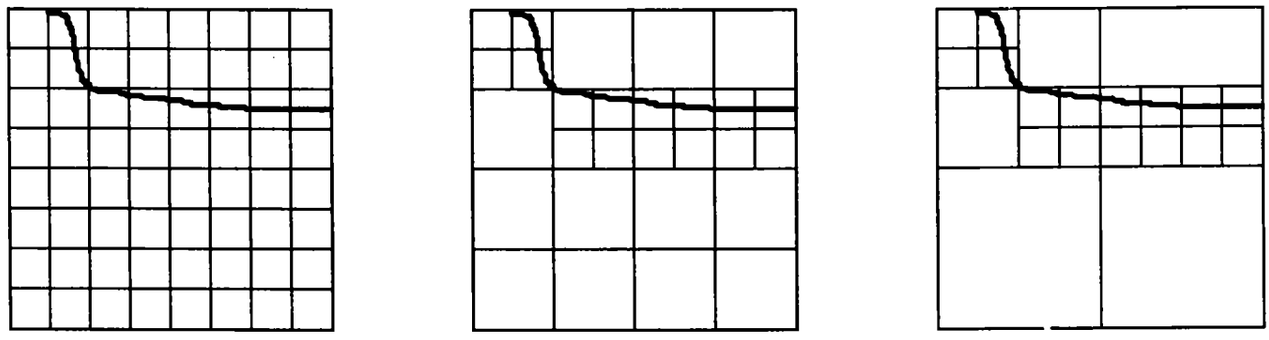
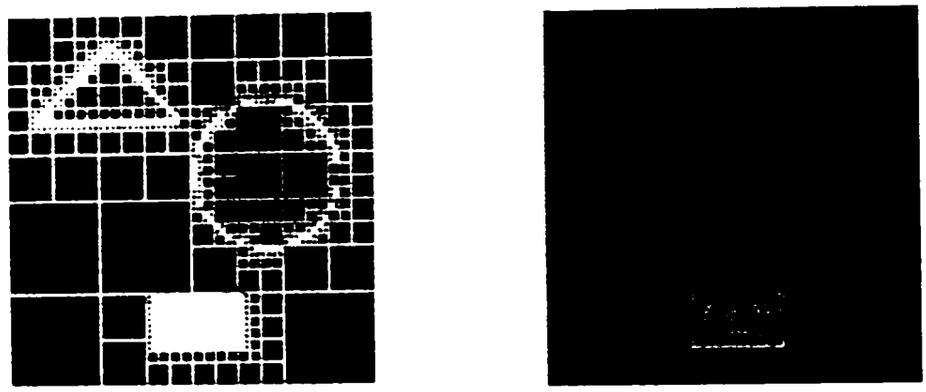


Fig. 4.25 Construcţia bottom-up a structurii quadtree prin contopire

Regiunile omogene astfel obţinute sunt apoi contopite utilizând testul de omogenitate. Algoritmul se încheie atunci când nu mai este posibilă nici o contopire sau nici o divizare.

BUPT

Un exemplu de utilizare al algoritmului pentru o imagine de test 128×128, în care regiunile sunt practic cunoscute, este prezentat în figura 4.25 (imaginea după divizare figura 4.26.a, conţine 980 regiuni iar imaginea după contopire, figura 4.26.b conţine 4 regiuni).



a) b)

Fig. 4.26 Structura quadtree obţinută pentru imaginea Test.

4.4.4 Segmentarea morfologică

Dezvoltată ca alternativă geometrică a prelucrării semnalelor, morfologia matematică a devenit foarte utilă, la început în analiza imaginilor, pentru ca apoi să fie luată în calcul pentru codarea imaginilor [MAR 96], [SAL 97]. Morfologia matematică este un instrument foarte atractiv pentru segmentarea imaginilor în regiuni, deoarece lucrează cu trăsăturile geometrice ale acestora: dimensiuni, formă, contrast şi conectivitate. Din acest motiv poate fi utilizată în continuare pentru codarea orientată pe trăsături a imaginilor.

Una din metodele clasice de segmentare constă în extragerea unor caracteristici prin utilizarea algoritmului watershed asupra imaginii ce trebuie segmentată [SHA 97], [MEY 90]. Deoarece în imaginile ce conţin mişcare este foarte importantă pierderea de informaţie, s-a demonstrat în [SAL 96] că utilizarea gradientului trebuie evitată, iar algoritmul watershed trebuie modificat pentru a se acţiona direct asupra semnalului şi nu asupra gradientului său. Algoritmul rezultat este o variantă a algoritmului de creştere a regiunilor, în sensul că porneşte de la punctul de identificare a interiorului regiunilor şi le extinde până la ocuparea întregului spaţiu disponibil.

O implementare eficientă a algoritmului watershed modificat necesită o explorare inteligentă a imaginii, definită prin cozi ierarhice. O astfel de coadă ierarhică este formată dintr-un set de şiruri FIFO având priorităţi diferite. Elementele prelucrate de aceste cozi sunt poziţiile pixelilor, cozile fiind utilizate pentru definirea explorării. Pixelii sunt extraşi în ordinea intrării mai întâi din coada de prioritate maximă şi apoi, după golirea acesteia, din celelalte cozi în

BUPT

ordinea priorităţilor. Algoritmul lucrează în doi paşi distincţi: iniţializarea cozii şi creşterea regiunilor.

Pasul de iniţializare constă în aşezarea tuturor locaţiilor pixelilor ce aparţin regiunilor în coada de aşteptare de prioritate maximă, deoarece probabilitatea de apartenenţă la o anumită regiune este maximă şi aceasta corespunde priorităţii cozii de aşteptare.

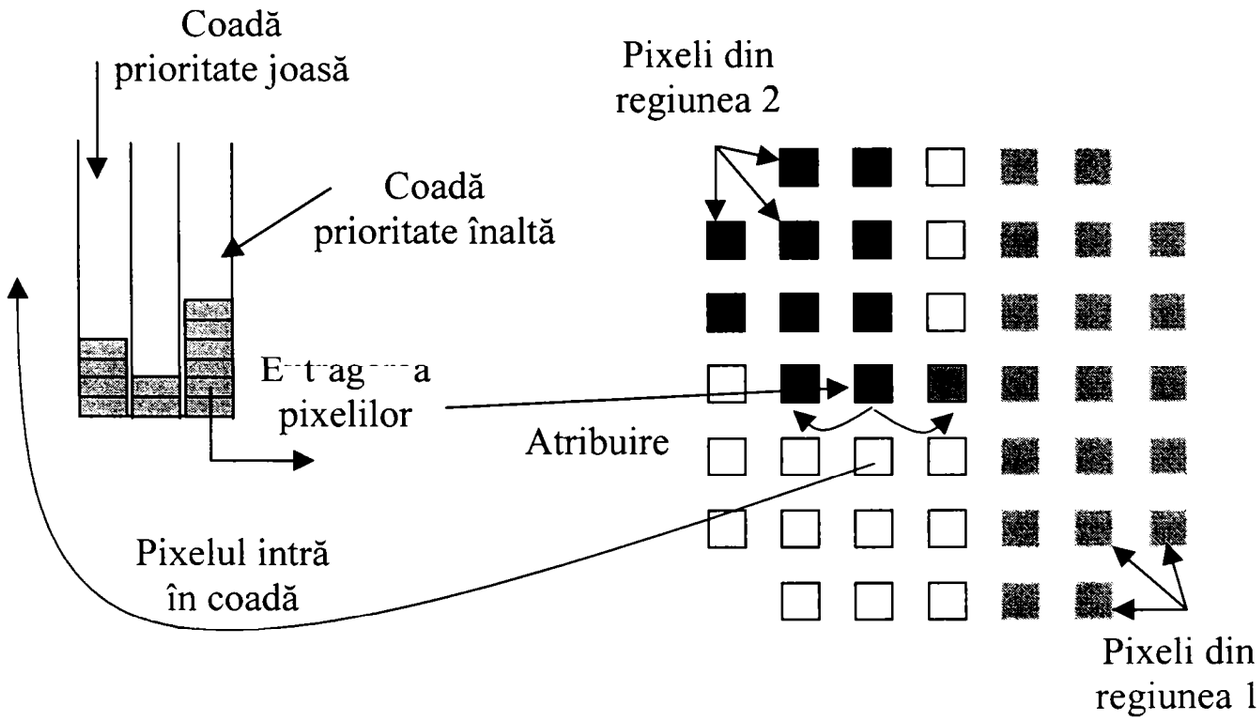


Fig. 4.27 Algoritmul de creştere al regiunilor

Pasul de creştere a regiunilor constă în extragerea unui pixel din coada de aşteptare. Dacă pixelul nu aparţine încă unei regiuni, acest lucru este cunoscut prin procedura de umplere, acesta va fi asignat regiunii pentru care se obţine distanţa minimă. Dacă pixelul curent are vecini care nu aparţin nici unei regiuni aceştia sunt introduşi în coada de aşteptare. Astfel orice pixel care este în coada de aşteptare are cel puţin un vecin, ceea ce face posibilă decizia cu privire la apartenenţă atunci când este extras din coada de aşteptare. Procedura de creştere a regiunilor este ilustrată în figura 4.27.

Această metodă poate fi utilizată în cadrul secvenţelor de imagini realizându-se şi o abordare multirezoluţie [PAD 94]. Metoda propusă este o abordare ierarhică (top-down) şi este prezentată în figura 4.28.

În cadrul fiecărui nivel ierarhic există patru paşi:

- *Simplificarea* - imaginea este simplificată pentru a realiza o cât mai uşoară segmentare. Pentru aceasta se utilizează filtrele morfologice care elimină obiectele mai mici decât o dimensiune dată dar păstrează informaţiile referitoare la conturile obiectelor rămase.

BUPT

- *Extragerea markerului (marcarea)* - pasul realizează marcarea de identificare a regiunii ce va fi segmentată. Regiunile corespunzătoare segmentării anterioare sunt păstrate şi incluse în marker.
- *Decizia* - contururile precise ale obiectelor detectate în etapa de extragere a markerului sunt determinate cu ajutorul algoritmului watershed [MEY 90].
- *Modelarea* - următorul nivel ierarhic are de-a face cu componente ce nu sunt segmentate fizic. Pentru a avea informaţii despre acestea fiecare regiune este modelată, la nivelul următor transmiţându-se o diferenţă între secvenţa modelată şi cea originală.

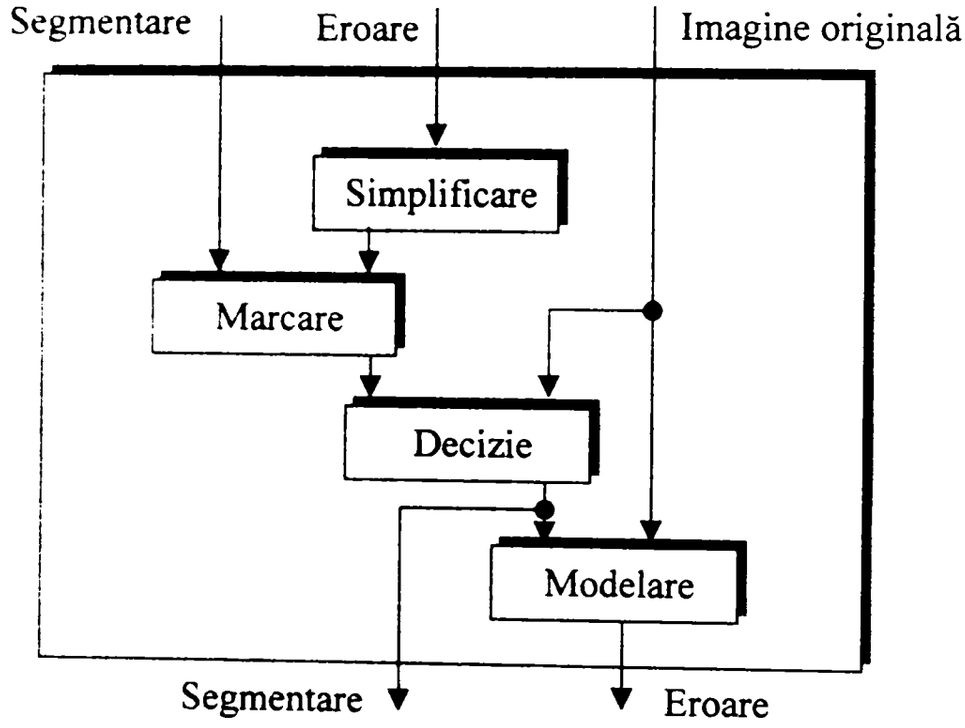


Fig. 4.28 Schema de segmentare ierarhică morfologică

Schema de segmentare este completată cu o abordare recursivă în timp, obţinându-se o segmentare grupată pe un număr de cadre. Rezultatul segmentării împreună cu un model afin sunt apoi utilizate pentru estimarea mişcării.

4.4.5 Algoritm de segmentare de tip divide şi contopeşte adaptiv [ALE 98]

Întrucât, aşa cum s-a evidenţiat în paragraful 4.1, contururile joacă un rol important în păstrarea informaţiilor semantice în imagine, la realizarea unei bune segmentări a imaginii este esenţială conservarea acestora. Mai mult chiar, este de dorit ca graniţele regiunilor obţinute prin segmentare să coincidă cu cele obţinute prin detecţia contururilor.

BUPT

Obţinerea unei segmentări bune este condiţionată de utilizarea unui model matematic sau statistic pentru descrierea modului de variaţie a nivelului de gri în cadrul unei regiuni. Astfel, pentru orice regiune obţinută prin segmentare, nivelul de gri poate fi reprezentat printr-un set de funcţii bidimensionale. Pentru a preveni crearea de distorsiuni prin segmentare este necesară, aşa cum s-a amintit mai sus, conservarea informaţiei privitoare la contur. Acest lucru este necesar atât pentru păstrarea informaţiei reale de contur, cât şi pentru a preveni crearea de contururi false.

Una din problemele ce trebuie rezolvată în cadrul algoritmului este aproximarea fiecărei regiuni obţinute prin segmentare. Problema a fost pe larg dezbătută în literatură [BOR 97], [AHU 96], urmând ca în cele ce urmează să fie abordată din punctul de vedere al segmentării.

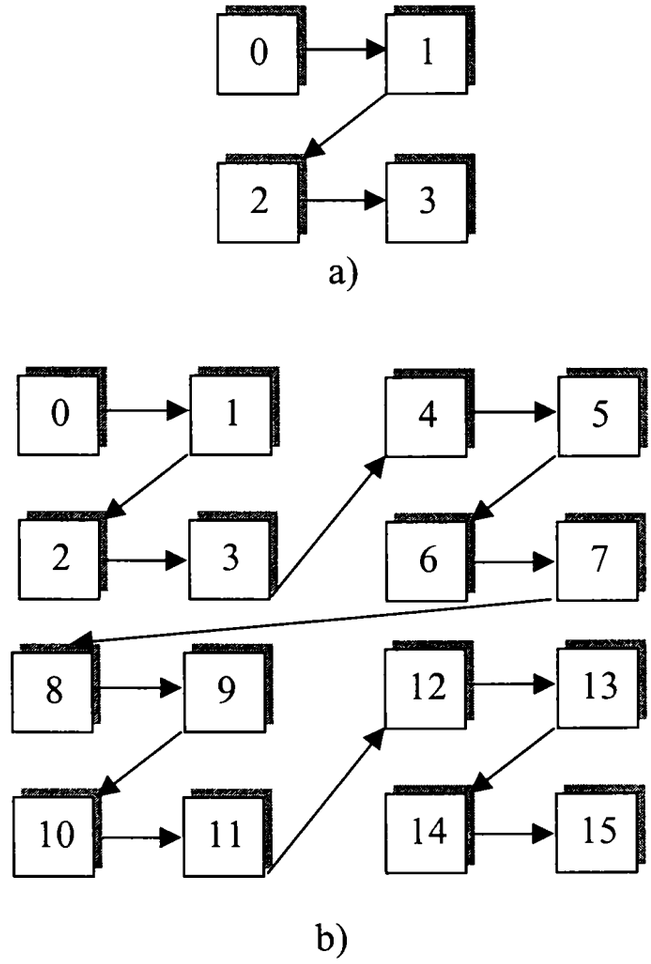


Fig. 4.29 Explorarea Z

Una din cele mai utilizate aproximări este cea liniară, care asigură modelarea aproximativă a regiunii:

$$g(x, y) = a + bx + cy \tag{4.16}$$

BUPT

în care: a reprezintă media nivelurilor de gri a blocului (regiunii), iar b şi c gradienţii după direcţiile x şi respectiv y. Această funcţie realizează practic aproximarea unor regiuni din imagine prin plane oblice. Asocierea acestei aproximări cu o structură arborescentă necesită serializarea datelor, mai ales din punctul de vedere al calculului erorii de aproximare:

$$e(x,y) = |f(x,y) - g(x,y)| \tag{4.17}$$

Din acest motiv prelucrarea blocurilor se realizează după algoritmul Z [STR 91], în aşa fel încât să se exploreze întreaga imagine. Această explorare Z are proprietatea de a fi aplicată recursiv, conducând la o structură simplă a secvenţelor de adrese şi asigură faptul că regiunile curente sunt prelucrate în dimensiuni mai mici decât cele ale regiunii ce urmează a fi analizată. Acest lucru este evidenţiat în figura 4.29.a şi b.

Explorarea Z este complet determinată de matricea de explorare $\boldsymbol{E}_N$. Această matrice pentru un bloc de dimensiune $N = 2$ are următoarea formă:

$$\boldsymbol{E}_2^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \tag{4.18}$$

Întrucât această explorare oferă posibilitatea recursivităţii, prin generalizare se poate obţine matricea de explorare având forma:

$$\boldsymbol{E}_N^T = \begin{bmatrix} \boldsymbol{I}_N^T \\ \boldsymbol{X}_N^T \\ \boldsymbol{Y}_N^T \end{bmatrix} \tag{4.19}$$

unde x_N şi y_N reprezintă vectorii de coordonate definiţi de relaţiile:

$$\boldsymbol{X}_N = [\, x_N(0), x_N(1), \ldots\ldots, x_N(N^2\text{-}1)\,]^T \tag{4.20}$$

$$\boldsymbol{Y}_N = [\, y_N(0), y_N(1), \ldots\ldots, y_N(N^2\text{-}1)\,]^T \tag{4.21}$$

iar $\boldsymbol{I}_N$ reprezintă vectorul unitate de dimensiune N^2 $\boldsymbol{I}_N = [\,1,1,\ldots\ldots,1\,]^T$.

Elementele vectorilor de coordonată $\boldsymbol{X}_N$ şi $\boldsymbol{Y}_N$ pot fi generaţi automat pentru un N arbitrar utilizând următoarele relaţii:

$$x_N(j) = \sum_{n=0}^{\log_2 N} 2^n MOD_2[DIV_k(j)] \quad unde\ k = 2^{2n}, \quad iar\ 0 < j < N^2 - 1 \tag{4.22}$$

BUPT

$$y_N(j) = \sum_{n=0}^{\log_2 N} 2^n MOD_2[DIV_l(j)] \quad unde\ l = 2^{2n}, \quad iar\ 0 < j < N^2 - 1 \qquad (4.23)$$

în care $DIV_k(j)$ reprezintă valoarea întreagă a raportului j/k şi $MOD_k(j)$ reprezintă modulul împărţirii lui j la k.

Dacă se iau în considerare definiţiile vectorilor de coordonate (relaţiile (4.21) şi (4.22)) se poate demonstra uşor că matricea E_{2N}^T poate fi construită din matricea E_N^T în modul următor:

$$E_{2N}^T = \begin{bmatrix} E_N^T & E_N^T + N\begin{bmatrix} O_N^T \\ O_N^T \\ I_N^T \end{bmatrix} & E_N^T + N\begin{bmatrix} O_N^T \\ I_N^T \\ O_N^T \end{bmatrix} & E_N^T + N\begin{bmatrix} O_N^T \\ I_N^T \\ I_N^T \end{bmatrix} \end{bmatrix} \qquad (4.24)$$

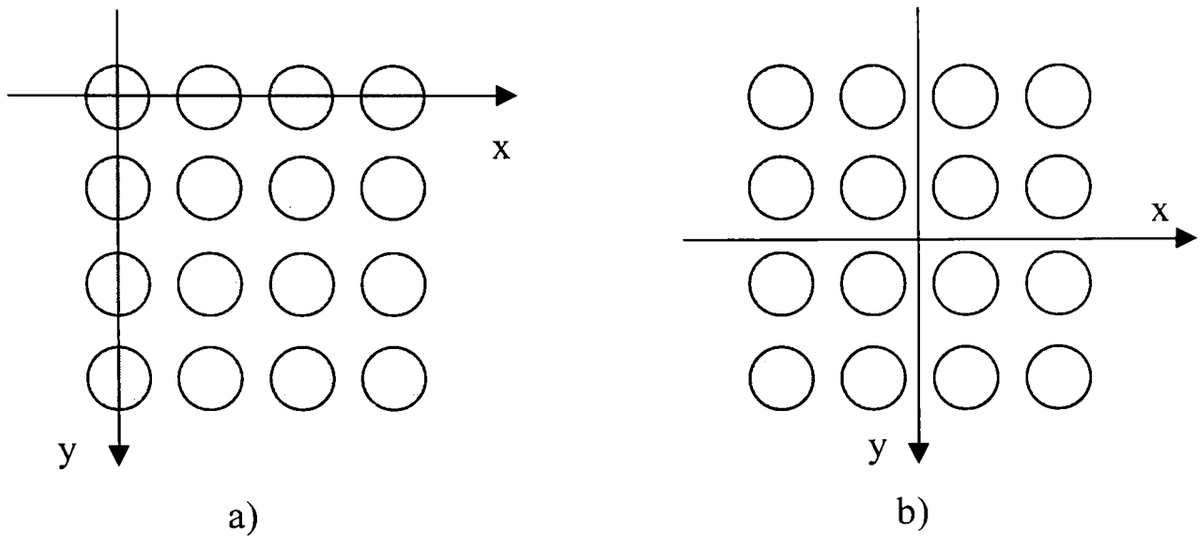


Fig. 4.30 Definirea axelor de coordonate pentru modelarea imaginii

Definiţiile matricii de explorare E_N^T arată că acestea sunt valabile numai pentru cazul modelului liniar, în care originea axelor de coordonate este localizată în colţul din stânga sus a blocului respectiv (figura 4.30.a). În cazul imaginilor plane cu model constant (media pixelilor regiunii) localizarea sistemului nu este importantă din punct de vedere al erorii, însă în cazul în care este utilizat modelul liniar (ecuaţia (4.16)) eroarea de cuantizare fiind funcţie pătratică de distanţa de la origine până la pixel, este important din punct de vedere al erorii ca definirea axelor de coordonate să se facă în centrul regiunii (figura 4.30.b).

În aceste condiţii matricea de definiţie a structurii $E_N^{'}$ respectă aceeaşi proprietate ca şi matricea E_N având structura de recurenţă definită de relaţia:

BUPT

$$E'^{T}_{2N} = \left[E'^{T}_{N} + \frac{N}{2}\begin{bmatrix} O^T_N \\ -I^T_N \\ -I^T_N \end{bmatrix} E'^{T}_{N} + \frac{N}{2}\begin{bmatrix} O^T_N \\ I^T_N \\ -I^T_N \end{bmatrix} E'^{T}_{N} + \frac{N}{2}\begin{bmatrix} O^T_N \\ -I^T_N \\ I^T_N \end{bmatrix} E'^{T}_{N} + \frac{N}{2}\begin{bmatrix} O^T_N \\ I^T_N \\ I^T_N \end{bmatrix} \right] \tag{4.25}$$

Faptul că cele două matrici de explorare sunt în concordanţă una cu cealaltă rezultă din modul în care pot fi definite:

$$\begin{aligned} E'_N &= [I_N, X_N - C_N, Y_N - C_N] \\ E_N &= [I_N, X_N, Y_N] \end{aligned} \tag{4.26}$$

unde C_N reprezintă o constantă (deplasarea de adresă), dată de relaţia:

$$C_N = \left[\frac{N-1}{2}, \frac{N-1}{2}, \ldots, \frac{N-1}{2} \right]^T \tag{4.27}$$

Pentru calculul coeficienţilor modelului blocului se utilizează minimizarea energiei erorii. În acest scop se defineşte semnalul de eroare al unui bloc de dimensiune $N \times N$:

$$\boldsymbol{e}_N = \boldsymbol{f}_N - \bar{\boldsymbol{f}}_N \tag{4.28}$$

în care $\bar{f}_N$ reprezintă estimata funcţiei f_N, iar indicele N se referă la dimensiunea blocului pentru care se calculează eroarea. Dacă se ţine cont de matricea de explorare E_N vom obţine pentru eroare expresia:

$$\boldsymbol{e}_N = \boldsymbol{f}_N - \boldsymbol{E}_N \boldsymbol{p}_N \tag{4.29}$$

în care vectorii e_N şi f_N conţin valorile erorii şi semnalului în concordanţă cu matricea de explorare E_N, iar p_N reprezintă vectorul de aproximare a planului blocului

$$\boldsymbol{p}_N = [a_N, b_N, c_N] \tag{4.30}$$

fiind presupus constant pentru întregul bloc de dimensiune $N \times N$. Elementele vectorului p_N reprezintă coeficienţii polinomului de aproximare al planului (ecuaţia (4.16)).

În aceste condiţii energia erorii pentru blocul de dimensiuni $N \times N$ este reprezentată de ecuaţia:

$$W_N = e_N^T e_N = f_N^T f_N - p_N^T E_N^T f_N - f_N^T E_N p_N + p_N^T E_N^T E_N p_N \tag{4.31}$$

a cărei minimizare poate fi obţinută prin rezolvarea ecuaţiei:

BUPT

$$\frac{\partial W_N}{\partial p_N^T} = O_N = -E_N^T f_N + E_N^T E_N p_N \tag{4.32}$$

Soluţia ecuaţiei (4.20) este dată de:

$$p_N = \left(E_N^T E_N\right)^{-1} E_N^T f_N \tag{4.33}$$

Această ecuaţie poate fi rezolvată atâta timp cât matricea $E_N^T E_N$ nu este singulară.

Algoritmul constă în trei etape, şi anume:

- etapa de divizare, care constă în definirea regiunilor omogene din punct de vedere al nivelului de gri şi care nu conţin puncte de contur determinate printr-o tehnică de detecţie a muchiilor [ALE 98],
- etapa de contopire, care constă în regruparea regiunilor până când contururile acestora coincid cu cele detectate şi
- etapa de eliminare a regiunilor mici, echivalentă cu o etapă de post-prelucrare în care sunt eliminate eventualele regiuni având dimensiuni mai mici decât o valoare aleasă.

4.4.5.1 Divizarea

Pentru partiţionarea imaginii într-un set iniţial de regiuni, se utilizează o procedură iterativă asemănătoare celei de obţinere a structurii arborescente descrise în paragraful 4.4.3 şi se foloseşte schema de explorare Z. Procedeul de divizare este reprezentat în figura 4.31 şi se aplică în condiţiile în care regiunea conţine puncte de contur sau nu respectă condiţiile de omogenitate. În acest caz condiţia de omogenitate înseamnă că eroarea minimă pătratică (relaţia (4.29)) nu depăşeşte o anumită valoare.

Condiţia de terminare a procesului este definită de dimensiunea minimă a elementului rezultat, de nesingularitatea matricii $E_N^T E_N$ sau de îndeplinirea condiţiei de omogenitate în blocurile rezultate. Divizarea regiunilor are loc în subregiuni având formă regulată, în cazul nostru formă pătrată, ceea ce presupune că regiunile de la care se pleacă au dimensiunea $2^q \times 2^q$. Astfel, după fiecare divizare, rezultă patru regiuni având dimensiunile $2^{q-1} \times 2^{q-1}$, proces ce este continuat până când este îndeplinită condiţia de omogenitate. Ultimul nivel de divizare este l, fiind definit de blocul de dimensiune minimă $2^{2(q-l+1)} \times 2^{2(q-l+1)}$. Procesul poate fi realizat şi pentru o descompunere în regiuni de orice dimensiune şi formă dar se obţine o complexitate mare de calcul. Regiunile care conţin elemente de contur sunt divizate fără a se mai testa condiţia de omogenitate.

Procesul de divizare porneşte de la regiuni având dimensiuni de 64×64, deoarece situaţiile în care există regiuni omogene, pentru imagini de dimensiuni mai mari (256×256 sau 512×512), sunt extrem de rare. Algoritmul conduce la

BUPT

obţinerea unei structuri arborescente care conţine coeficienţii a, b, c asociaţi fiecărei regiuni finale. Aceşti coeficienţi sunt obţinuţi prin minimizarea erorii pătratice pentru fiecare regiune în parte. Schema bloc simplificată a algoritmului de divizare este prezentată în figura 4.32.

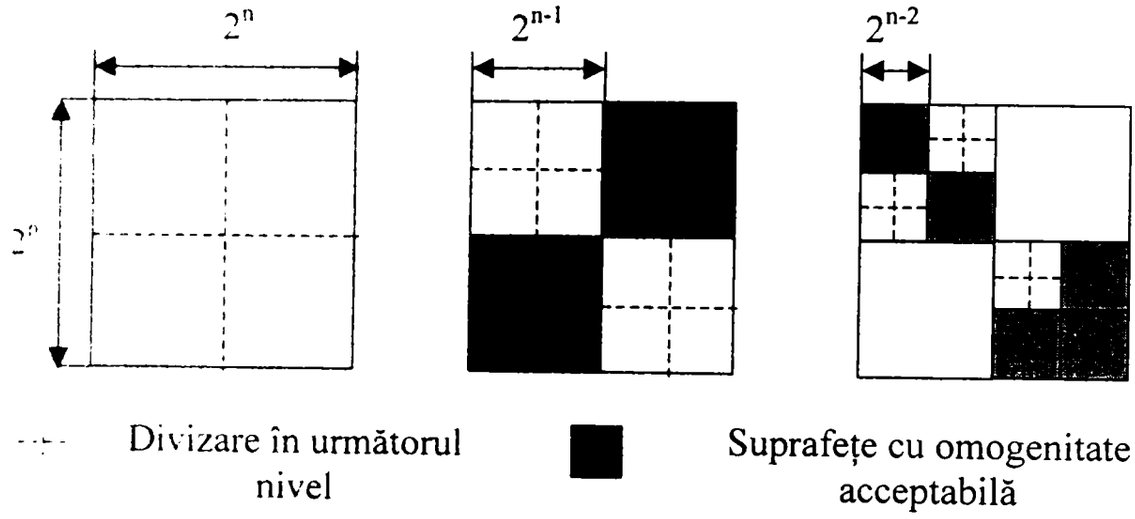


Fig. 4.31 Procedeul de divizare utilizat

În aceste condiţii este avantajos de realizat prelucrarea tuturor blocurilor având aceeaşi mărime, întrucât matricea $E_N^T E_N$ are aceeaşi expresie pentru blocuri având aceeaşi dimensiune. Întrucât s-a presupus o aproximare liniară a planului fiecărui bloc, inversa matricii $E_N^T E_N$ are o dimensiune de 3×3 fiind dată de relaţia:

$$\left(E_N^T E_N\right)^{-1} = \frac{1}{N^2}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \dfrac{12}{N^2-1} & 0 \\ 0 & 0 & \dfrac{12}{N^2-1} \end{bmatrix} \tag{4.34}$$

Această matrice depinde numai de dimensiunea blocului N şi deci este foarte uşoară determinarea parametrilor planului de aproximare, fiind necesară numai calcularea produsului $E_N^T f_N$.

Aproximarea regiunii utilizând eroarea minimă pătratică LSE (Least Square Error), ecuaţia (4.31), ajută la divizarea blocurilor, deoarece aceasta este folosită ca şi criteriu de uniformitate. Astfel, dacă valoarea LSE obţinută este sub un anumit prag T_1 regiunea este considerată omogenă nefiind necesară divizarea, iar în cazul în care aceasta depăşeşte acest prag este necesară modificarea formei regiunii. Se observă că LSE dată de ecuaţia (4.31) reprezintă o aproximare globală a calităţii

BUPT

imaginii, ceea ce nu mai necesită o măsură suplimentară în context cu informaţia prezentă în regiune.

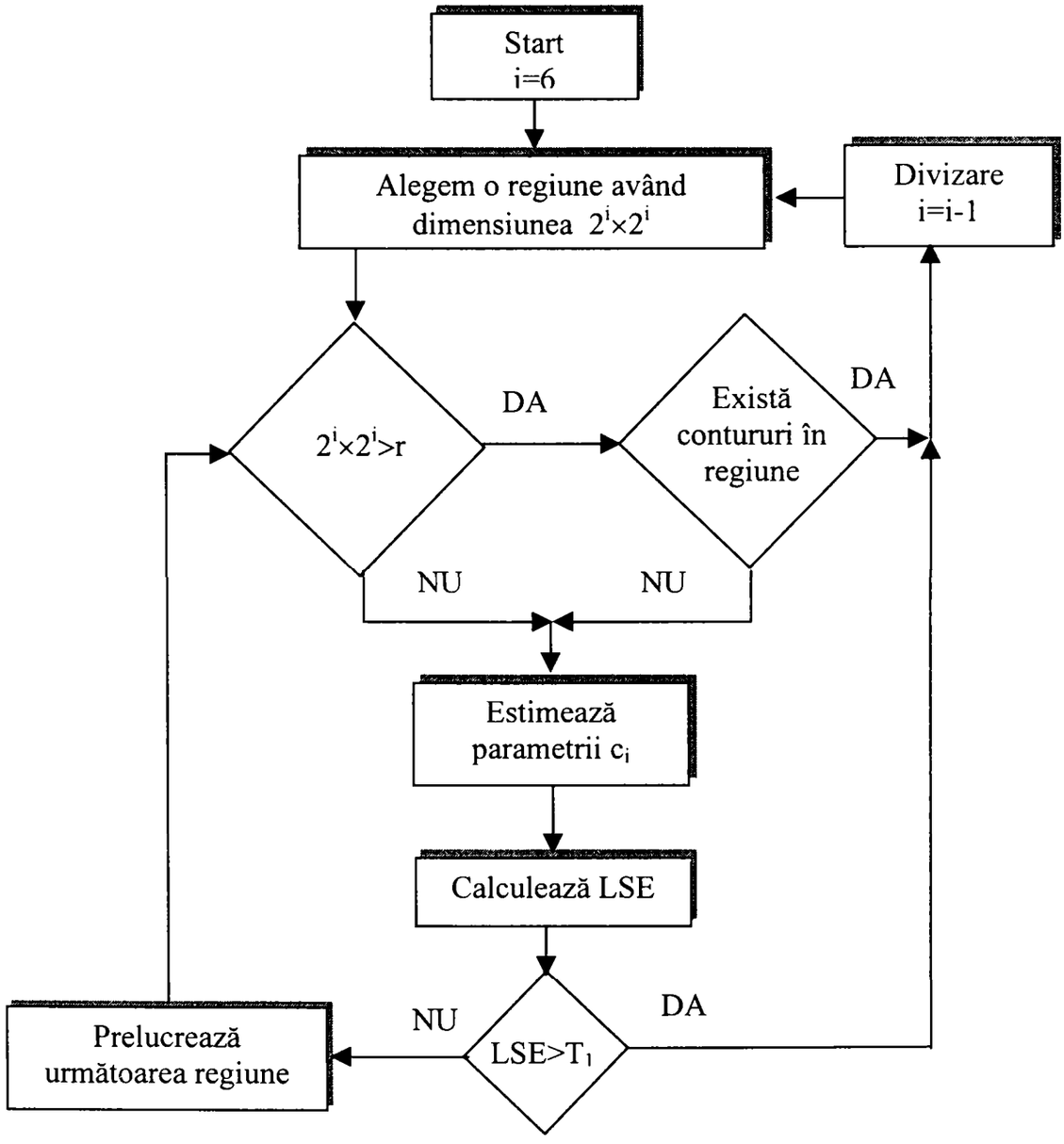


Fig. 4.32 Algoritmul de divizare simplificat

Păstrarea informaţiei de contur necesită însă alegerea unei valori foarte mici pentru LSE. Datorită acestui fapt, prin segmentare se obţin un număr foarte mare de regiuni, ineficient în vederea codării. Dacă s-ar utiliza ca şi criteriu de uniformitate eroarea medie pătratică MSE (Mean Square Error - Anexa 2) , eroarea de aproximare ar fi mai bună (rezultatul fiind independent de dimensiunea regiunii), dar s-ar putea ajunge la pierderea de contururi importante.

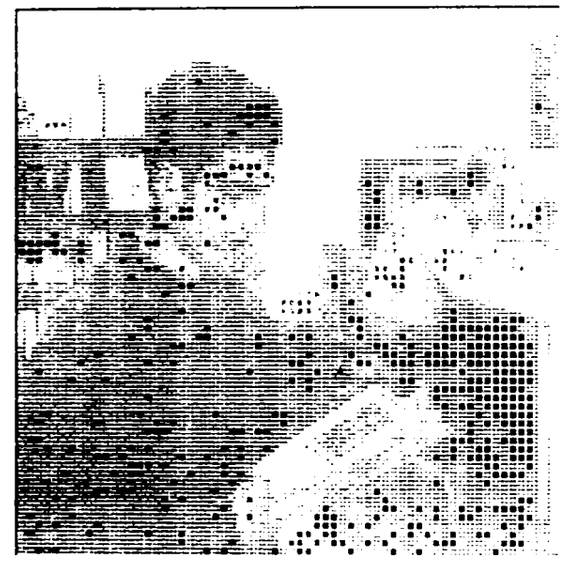
Din acest motiv este necesară utilizarea unui criteriu sensibil la eroarea de contur, care să ia în consideraţie existenţa contururilor ce traversează regiunile de

BUPT

interes. Acest criteriu este îndeplinit de testarea existenţei punctelor de contur în interiorul regiunii.

Rezultatul procesului de divizare al cadrului 1 din secvenţa MISSA, precum şi a imaginii APPLE sunt prezentate în figura 4.33, respectiv, figura 4.34.

Imaginea originală



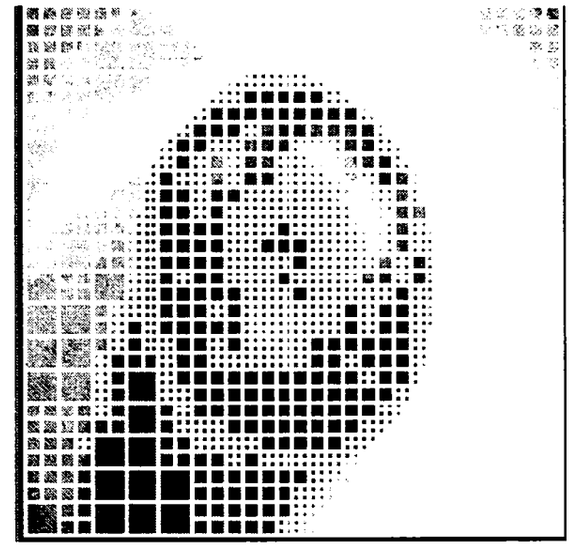
Grila de divizare fără prelucrarea suplimentară a imaginii

Grila de divizare cu medierea prealabilă a imaginii

Fig. 4.33 Rezultatul procesului de divizare pentru cadrul 1 din secvenţa MISSA

BUPT

Imaginea originală



Grila de divizare fără prelucrarea suplimentară a imaginii

Fig. 4.34 Rezultatul procesului de divizare pentru imaginea APPLE

4.4.5.2 Contopirea

Contopirea reprezintă un pas intermediar în procesul de segmentare, cu scopul de a obţine regiuni ale căror contururi să corespundă celor ce delimitează obiecte din imaginea iniţială. Astfel, imaginea obţinută va conţine contururi de orice formă şi dimensiune. În această etapă sunt contopite regiunile adiacente care, prin această operaţie, nu conduc la degradarea conţinutului imaginii în raport cu alte asociaţii ale regiunilor. Acest lucru se poate exprima prin aceea că, două regiuni sunt contopite dacă aparţin aceleiaşi regiuni omogene din imagine.

Rezolvarea eficientă a problemei presupune alegerea unui prag de contopire corespunzător, care să permită acumularea erorilor fără a conduce la distorsionarea imaginii. În acest scop, pe durata procesului de divizare, se realizează graful adiacenţelor [BAL 82], [WU 93], [BOR 98], [LAN 98], în care fiecare nod reprezintă o regiune şi fiecare legătură defineşte adiacenţa. Schema bloc simplificată a algoritmului de contopire bazat pe similitudinea regiunilor este prezentată în figura 4.35.

BUPT

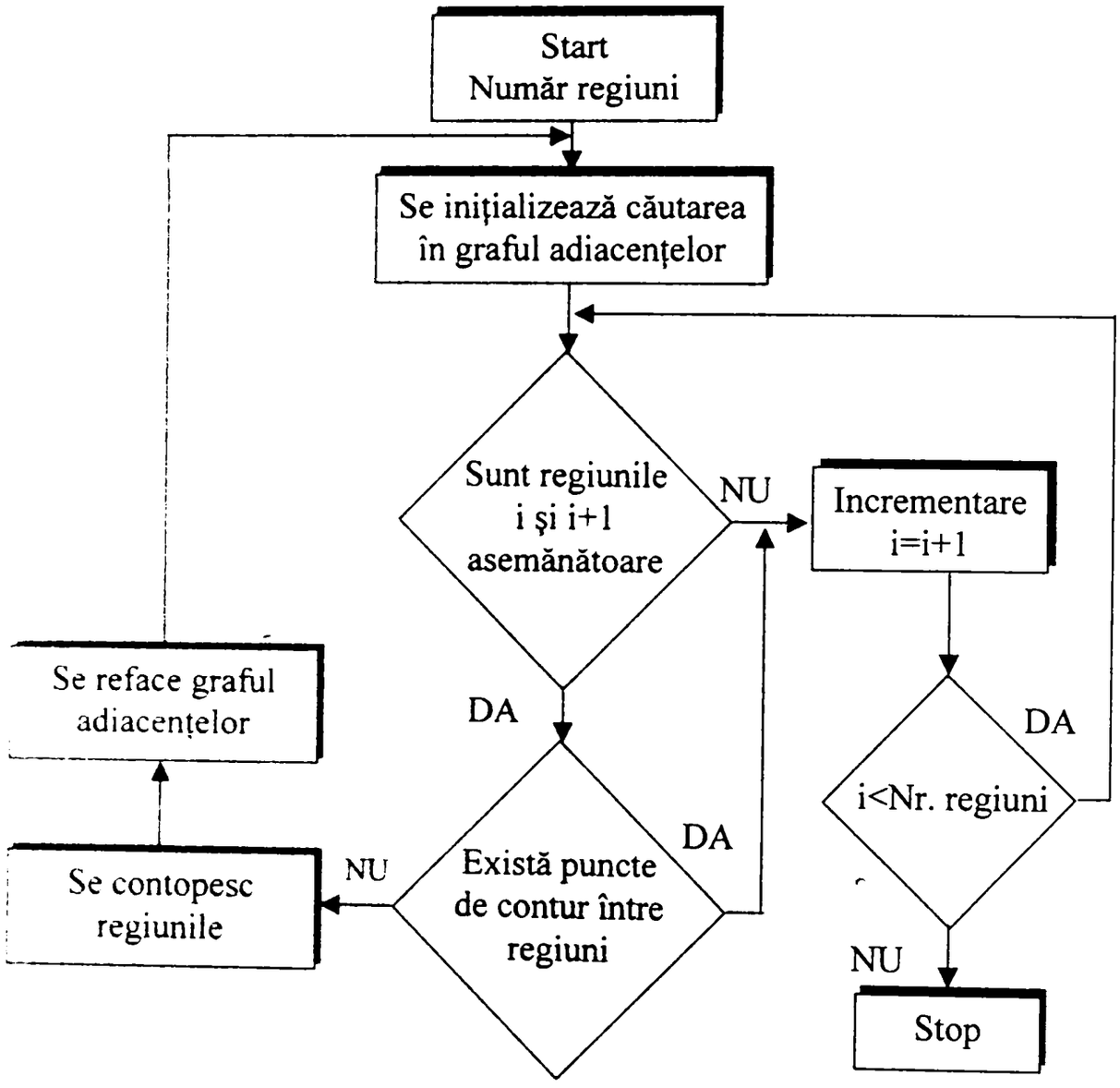
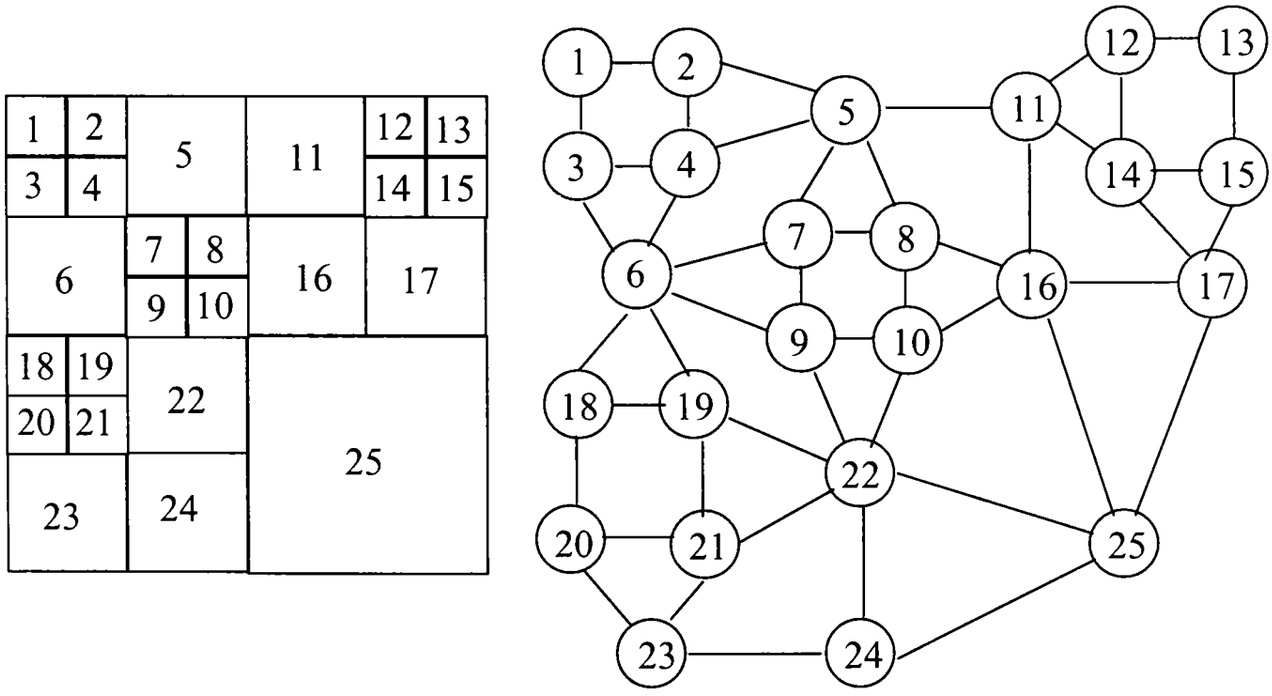


Fig. 4.35 Schema simplificată a algoritmului de contopire bazat pe similitudinea regiunilor

După obţinerea grafului adiacenţelor (figura 4.36), contopirea constă practic în parcurgerea acestui graf, punct cu punct, şi verificarea similitudinii caracteristicilor regiunilor alăturate. Primul test pentru contopirea regiunilor îl reprezintă compararea funcţiei de aproximare pe direcţia x. Dacă acestea au media şi gradientul pe direcţia x egale, şi nu există puncte de contur între regiuni, atunci se realizează contopirea, iar funcţia de aproximare rezultată pentru noua regiune se obţine prin medierea celor două funcţii de aproximare, ecuaţia (4.35). După realizarea contopirii după direcţia x, se trece la contopirea după direcţia y, utilizând aceeaşi regulă. Graful adiacenţelor este refăcut după fiecare contopire.

BUPT

$$g'(x,y) = \frac{g_i(x,y) + g_{i+1})x,y)}{2} = \frac{a_i + a_{i+1}}{2} + \frac{b_i + b_{i+1}}{2}x + \frac{c_i + c_{i+1}}{2}y \tag{4.35}$$



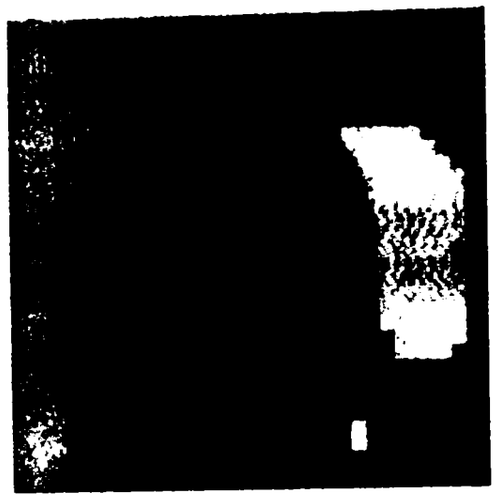
Imagine divizată Graful adiacenţelor

Fig. 4.36 Construcţia grafului adiacenţelor

După această contopire, care se bazează pe similitudinea regiunilor adiacente, se trece la contopirea regiunilor pe baza criteriului erorii medii pătratice (MSE), deoarece utilizarea criteriului LSE depinde în această etapă foarte mult de dimensiunea regiunilor. Deoarece prin contopire se obţin regiuni având suprafeţe mari criteriul LSE devine inoperant. În urma acestei etape de contopire se realizează contopirea tuturor regiunilor care conduce la apariţia de erori sub un anumit prag. Întrucât scopul final îl reprezintă obţinerea a cât mai puţine regiuni, etapa de segmentare poate include un pas suplimentar pentru eliminarea regiunilor mai mici de o anumită valoare. În acest scop contopirea este realizată astfel încât să se obţină o eroare minimă şi să nu existe puncte de contur între regiunile contopite.

În urma procesului de contopire, se obţine imaginea segmentată, care este utilizată pentru estimarea vectorilor de mişcare ai obiectelor obţinute. În figura 4.37.a se prezintă procesul de segmentare pentru imaginea apple, iar în figura 4.37.b pentru cadrul 5 din secvenţa Claire.

BUPT



Imagine segmentată în 9 regiuni

a)

Imagine segmentată în 19 regiuni

b)

Fig. 4.37 Rezultatul segmentării imaginilor apple şi Claire5

4.5 Concluzii

Necesitatea unor rapoarte de compresie tot mai mari a determinat evoluţia tehnicilor de codare spre cele ce realizează incorporarea proprietăţilor sistemului vizual uman. Din acest motiv, este necesară reprezentarea imaginilor în funcţie de conţinutul lor semantic. Segmentarea are, din acest motiv, un rol esenţial deoarece conduce la obţinerea indicilor vizuali ce ne ajută la reprezentarea imaginii sub formă de primitive semantice lipsite de ambiguitate (contururi, regiuni).

În acest context capitolul a fost structurat astfel:

- Paragraful 1 – prezintă o introducere în segmentarea imaginilor cu evidenţierea principalelor caracteristici ale imaginilor (discontinuităţi, similarităţi) şi metodele de evidenţiere ale acestora.
- Paragraful 2 – prezintă metode de extragere a contururilor, pornind de la puncte şi linii şi evidenţiind algoritmii de extragere ai acestora bazaţi pe gradient şi respectiv derivata a doua. De asemenea sunt prezentate rezultate experimentale ale implementării acestor algoritmi.
- Paragraful 3 – prezintă metodele de extragere a regiunilor din imagini prin metodele bazate pe prag, cu prezentarea principalelor rezultate în cazul implementării algoritmilor cu determinarea pragului global şi ISODATA.
- Paragraful 4 – prezintă metodele de segmentare orientate pe regiuni (obiecte). Mai întâi este definită problema segmentării orientată pe regiuni [GON 92], după care sunt definiţi principalii algoritmi existenţi:
    1. Creşterea regiunilor [ADA 94], [BEV 89]
    2. Algoritmi de tip divide şi contopeşte [LEE 94],

BUPT

3. Algoritmi bazaţi pe matematica morfologică [MAL 96], [SAL 97],

În continuare este prezentat algoritmul adaptiv propus de autor [ALE 98b], care realizează segmentarea imaginilor bazându-se pe algoritmii de tip divide şi contopeşte, dar care ia în consideraţie proprietăţile sistemului vizual uman referitoare la informaţia de contur şi textură. Astfel etapa de divizare şi contopire se realizează pe baza minimizării erorii pătratice în funcţie de informaţia de contur din imagine obţinută cu ajutorul unui operator Sobel. Pe baza erorii pătratice se determină coeficienţii funcţiei liniare de modelare a informaţiei referitoare la textură.

Calculul parametrilor funcţiei de aproximare a conturului se realizează prin minimizarea erorii pătratice. În acest scop se utilizează explorarea Z, introducându-se o matrice caracteristică pentru blocurile de dimensiune 2×2, care ajută la serializarea datelor. Matricea caracteristică definită pentru fiecare regiune permite construcţia recursivă a matricii caracteristice regiunii de nivel superior, conducând la reducerea duratei de calcul necesară determinării coeficienţilor funcţiei de aproximare a regiunii.

Dacă în cazul aproximării informaţiei de textură prin valoarea medie a acesteia nu este importantă referinţa pentru calculul erorii, utilizarea funcţiei liniare de aproximare a erorii ne conduce la modificarea referinţei. Astfel cea mai bună referinţă este în centrul regiunii întrucât se obţin erorile cele mai mici.

Etapa de contopire este realizată după un algoritm în trei paşi care ţine cont de: eroare, număr de regiuni şi contururile existente în imagine, conducând la obţinerea de regiuni de formă oarecare. În primul pas al algoritmului se realizează contopirea funcţie de modelul de aproximare ţinând cont de gradienţii după cele două direcţii, cel de-al doilea pas realizează contopirea pe baza unui prag de eroare iar cel de-al treilea pas realizează eliminarea regiunilor sub un anumit număr de pixeli pentru reducerea numărului acestora.

Algoritmul de segmentare propus este utilizat pentru estimarea mişcării în secvenţele de imagini, în vederea codării orientate pe regiuni a secvenţelor de imagini.

BUPT

Capitolul 5

ESTIMAREA MIŞCĂRII BAZATĂ PE REGIUNI ÎN SECVENŢELE DE IMAGINI

5.1 Introducere

În acest capitol este prezentat un algoritm de estimare al mişcării bazat pe regiuni în vederea codării secvenţelor de imagini propus de autorul tezei de doctorat. Problemele şi abordările principiale ale estimării mişcării bazate pe regiuni au fost prezentate în paragraful 2.3.7. Pentru determinarea regiunilor (obiectelor) din imagine se poate utiliza o segmentare: spaţială, temporală sau spaţio-temporală a imaginilor. Segmentarea reprezintă cea mai mare provocare în analiza secvenţelor de imagini şi a fost prezentată în capitolul 4.

Aşa cum reieşe din capitolul 3, tehnicile de codare, din punct de vedere al reprezentării secvenţelor de imagini, se pot clasifica în principal în: tehnici de codare din prima generaţie şi tehnici de codare din cea de a doua generaţie.

Tehnicile de codare din prima generaţie se bazează pe reprezentarea secvenţelor de imagini ca un proces stohastic, iar codarea se rezumă la reducerea redundanţei statistice prin codarea pixelilor sau a blocurilor de pixeli. Cele mai cunoscute sunt tehnicile hibride ce utilizează transformarea cosinus discretă (standardele MPEG-1, MPEG-2, H261, H263). În acest context, tehnicile de estimare a mişcării bazate pe corespondenţa blocurilor (block matching) [MUS 85], [DUF 95] sunt cele mai potrivite, fiind utilizate pentru reducerea redundanţei temporale.

Tehnicile din prima generaţie se bazează pe divizarea imaginii în blocuri de dimensiuni reduse cărora li se atribuie mişcări de translaţie, prin minimizarea unui criteriu de eroare referitor la corespondenţa acestora. Informaţie ce este apoi utilizată pentru codarea între cadre. Eficienţa acestor tehnici constă, în cea mai mare parte, în abilitatea de a determina cu precizie predicţia temporală prin minimizarea directă a diferenţei de deplasare a cadrelor.

Mai mult, datorită faptului că se realizează o reprezentare pe blocuri a imaginii, aceste tehnici se adaptează foarte bine schemelor de codare bazate pe blocuri şi necesită puţine informaţii suplimentare. Un avantaj important al tehnicilor de codare din prima generaţie îl constituie implementarea hardware foarte uşoară.

Cu toate că se obţin performanţe bune din punct de vedere al calităţii obiective (eroare medie pătratică, raport semnal/zgomot), ele suferă din cauză că, datorită reprezentării sub formă de pixeli sau ca blocuri de pixeli, aceste imagini nu au nici o semnificaţie din punct de vedere semantic. Astfel, la rate de compresie

BUPT

ridicate se obţin erori majore din punct de vedere al sistemului vizual uman. Din acest motiv nu pot fi obţinute îmbunătăţiri majore ale performanţelor compresiei.

În plus, lipsa informaţilor referitoare la conţinutul imaginii limitează funcţionalităţile disponibile la recepţie, nefiind posibilă manipularea şi personalizarea informaţiei recepţionate, aşa cum se doreşte în cadrul sistemelor de transmisie multimedia.

În contrast cu acestea, tehnicile de codare din cea de a doua generaţie introduse de Kunt [KUN 85], [KUN 87], se bazează pe o reprezentare semantică a imaginii. Prin implementare s-a încercat imitarea funcţiilor sistemului vizual uman şi în acest scop algoritmii s-au bazat pe descrierea imaginii sub forma obiectelor care o compun. Compresia se obţine luând în considerare proprietăţile sistemului vizual uman şi anume alocarea mai multor biţi în zonele importante şi reducerea numărului acestora în rest.

Ca rezultat al folosirii reprezentării semantice a imaginii, aceste tehnici permit o rată de compresie foarte ridicată, cu obţinerea unor deteriorări vizuale ale imaginii foarte mici, iar informaţiile despre conţinutul imaginii oferă funcţionalităţi suplimentare la recepţie. În particular, utilizatorului final îi este accesibilă interactivitatea.

În aceste condiţii robusteţea estimării mişcării este mult mai importantă decât obţinerea unei predicţii temporale uşoare, în vederea asigurării unei calităţi corespunzătoare a secvenţei de imagini reconstituite. Pentru obţinerea unei rate de compresie ridicată este necesară determinarea unui câmp al mişcării reprezentativ deoarece:

- se doreşte eliminarea erorilor de predicţie care sunt transmise către decodor, motiv pentru care este important să nu se introducă distorsiuni suplimentare deranjante pentru sistemul vizual uman,
- se doreşte subeşantionarea temporală a secvenţei de imagini înaintea codării.

În acest caz, informaţia referitoare la mişcare este utilizată la decodare pentru interpolarea cadrelor eliminate la codare. În condiţiile menţionate, tehnicile clasice de corespondenţă a blocurilor nu sunt adecvate pentru estimarea deplasărilor reale din imagini, în special, datorită faptului că se bazează pe blocuri care introduc distorsiuni.

Pentru creşterea performanţelor metodelor bazate pe blocuri au fost propuse tehnici îmbunătăţite de segmentare şi estimare a mişcării în secvenţele de imagini:

- blocuri cu dimensiune variabilă [CHA 90];
- estimarea multigrid adaptivă local [DUF 95];
- segmentarea corespunzătoare eliminării erorilor determinate de utilizarea modelului de bloc [ORC 93], [DUF 94], [CHA 97].

Cu toate că aceste tehnici îmbunătăţite de corespondenţă a blocurilor au condus la obţinerea de performanţe mai bune prin eliminarea unora din limitările determinate de utilizarea blocurilor, segmentarea obţinută a rămas una de rezoluţie

BUPT

scăzută. Mai mult, nu se realizează o reprezentare a scenei prin descompunerea acesteia în obiecte, ci se încearcă eliminarea dezavantajelor introduse prin utilizarea modelului bazat pe blocuri de pixeli.

Aceste consideraţii, au condus la elaborarea de tehnici de estimare a mişcării bazate pe obiecte [AIZ 95], [HOT 94], [GER 94], [KWA 98], care se bazează pe o reprezentare semantică prin primitive a imaginii, luând în consideraţie câteva elemente de bază.

- Existenţa unui model pentru descrierea mişcării. În algoritmii bazaţi pe segmentare, estimarea mişcării este realizată pe obiecte şi nu pe blocuri de pixeli. Din acest motiv, mişcarea atribuită fiecărui obiect este modelată printr-un set de parametri de mişcare (modelul parametric) [ANA 93].
- Este posibilă determinarea simultană a obiectelor şi a mişcării lor. Segmentarea spaţio-temporală defineşte regiunile care sunt caracterizate de o mişcare coerentă şi care pot fi identificate ca fiind obiecte în mişcare ce aparţin scenei reale [BOU 93], [DUF 95b]. În mod evident există interdependenţă foarte mare între cele două procese, segmentarea spaţio-temporală şi estimarea mişcării, problemă cunoscută sub denumirea de "problema oului şi a găinii" (egg and chiken problem). Pe de o parte estimarea mişcării depinde de regiuni, motiv pentru care este necesară o segmentare foarte bună, iar pe de altă parte dacă regiunea este definită prin existenţa unei coerenţe a mişcării acesteia, atunci obţinerea unei bune segmentări este determinată de precizia estimării mişcării. Prin această metodă se estimează parametrii de mişcare relativ la fiecare obiect obţinut prin procesul de segmentare [WU 90], [PAR 94], [DUF 95b], [CHA 97b].
- Obţinerea unei segmentări coerente în timp, necesită utilizarea unor tehnici de căutare-urmărire foarte bune [AGG 81], [BAR 88], care trebuie să definească traiectoria obiectelor, detectate a fi în mişcare, în timp. Suplimentar se tinde spre realizarea unei stabilizări a segmentării spaţio-temporale şi a estimării mişcării. În contexul codării, cunoştinţele referitoare la traiectoria obiectelor elimină necesitatea transmiterii informaţiei despre segmentare pentru fiecare cadru codat.

Problemele menţionate mai sus şi anume: segmentarea spaţio-temporală şi estimarea mişcării bazată pe segmentare şi urmărire au făcut obiectul cercetărilor în domeniile analizei secvenţelor de imagini şi al vederii artificiale. De altfel, codarea secvenţelor de imagini implică abordări foarte diferite deoarece scopul este acela de a realiza o reprezentare cât mai compactă a informaţiei de mişcare, motiv pentru care este foarte importantă precizia de estimare a mişcării şi cantitatea de informaţii suplimentare ce vor fi transmise spre decodor [GIR 94], [MOS 93]. Aceasta implică următoarele:

- controlul procesului de segmentare;
- posibilitatea alegerii modelului de mişcare;

BUPT

- cuantizarea parametrilor de mişcare.

De asemenea, în scopul utilizării secvenţelor codate în cadrul serviciilor interactive, este necesară considerarea întârzierii care se obţine în procesul de codare.

5.2 Tehnici de estimare bazate pe segmentare

5.2.1 Estimarea mişcării utilizând modelul parametric

Mişcarea poate fi descompusă într-o mişcare globală datorată camerei (rotire sau scalare) şi o mişcare locală datorată deplasării obiectelor în scenă. Deoarece se doreşte eliminarea mişcării datorate camerei, estimarea mişcării se poate realiza în două etape aşa cum este propus în [MOS 95].

Tehnicile de estimare a mişcării luate în discuţie se bazează pe estimarea locală a mişcării, cea globală fiind inclusă în aceasta. O caracteristică importantă în domeniul codării secvenţelor de imagini şi în special în cazul serviciilor interactive, o reprezintă întârzierea în codare. Din acest motiv, în ceea ce urmează se va considera problema estimării mişcării numai între două cadre, cu posibilitatea extinderii la N cadre de imagine. De asemenea, în aplicaţiile uzuale de codare a secvenţelor de imagini se utilizează estimarea mişcării faţă de cadrul anterior, realizîndu-se o predicţie între cadrul curent şi cel anterior, presupus cunoscut (vezi capitolul 3), adică:

$$\hat{I}(\vec{r},t) = I(T(\vec{r}),t-1) \tag{5.1}$$

iar eroarea de predicţie este dată de:

$$e(\vec{r},t) = I(\vec{r},t) - \hat{I}(\vec{r},t) \tag{5.2}$$

unde $I(\vec{r},t)$ reprezintă intensitatea imaginii în punctul $\vec{r}$ şi la momentul t, $\hat{I}(\vec{r},t)$ valoarea estimată a intensităţii cu compensarea mişcării, iar $T(\vec{r})$ reprezintă pixelul corespunzător cadrului anterior. Ecuaţia (5.1) reprezintă de fapt principiul de codare predictivă cu compensarea mişcării.

Aşa cum s-a arătat în capitolul 2, problema estimării mişcării este în cele mai multe cazuri o problemă nedeterminată. Pentru rezolvarea acesteia, toate tehnicile de estimare a mişcării au nevoie de condiţii impuse. Aceste constrângeri pot fi implicite (modelul parametric) sau ele necesită evidenţierea lor explicită (modelul neparametric).

Pentru implementarea estimării mişcării pe baza modelului parametric există două posibilităţi:

- calculul indirect, presupune o estimare bazată pe modelul neparametric urmată de modelarea vectorilor de mişcare prin utilizarea unui model parametric [ADI 85], [WAN 95], [RAJ 97], [MEM 98];

BUPT

- calculul direct, presupune estimarea directă a mişcării pe baza unui model ce cuprinde implicit constrângerile. În acest caz, cele mai utilizate sunt tehnicile diferenţiale bazate pe descompunerea în serie Taylor a semnalului de luminanţă [ANA 93], [WU 90], având ca alternativă tehnicile de corespondenţă [DUF 95b], [MOS 95].

Tehnicile de calcul direct al vectorilor de mişcare conduc la erori în cazul unor imagini zgomotoase sau în cazul unei segmentări imprecise, motiv pentru care sunt utilizaţi algoritmi de estimare robustă [EGG 97]. O alternativă o reprezintă implementarea în scheme multirezoluţie [ANA 93], [STR 98].

Estimarea parametrică a mişcării, prezentată mai sus, este caracterizată de prezenţa unor vectori de mişcare a căror valoare depăşeşte domeniul de estimare şi care conduc la deterioarea estimării. Ele pot rezulta din prezenţa zgomotului sau prin definirea eronată a suportului pentru estimarea mişcării (segmentarea), ce nu corespunde unei suprafeţe caracterizată de o mişcare coerentă.

Din acest motiv, pentru estimare se pot utiliza tehnici robuste, a căror sensibilitate la valori ce depăşesc domeniul este foarte mică [MEE 91]. Performanţele acestor tehnici sunt indicate de punctul de separare, care defineşte cea mai mică fracţiune de eşantioane cu depăşirea domeniului ce se obţine prin estimare. Tehnicile de estimare propuse pot fi grupate în trei categorii [MEE 91]:

- **M** - minimizează $\sum \rho(r_i)$ unde r_i reprezintă erorile (definite de ecuaţia (5.2)) iar ρ o funcţie simetrică pozitivă având o valoare minimă unică pentru $r_i=0$). Cel mai semnificativ exemplu în acest sens este metoda bazată pe LSE în care $\rho(r_i) = r_i^2$. Pentru reducerea efectelor unor erori mari ce pot conduce la valori în afara domeniului au fost propuse tehnici de estimare ce se bazează pe următoarea funcţie:

$$\rho(r_i) = \frac{\frac{r_i^2}{s^2}}{1 + \frac{r_i^2}{s^2}} \qquad (5.3)$$

în care $s = 1{,}4826 \times \frac{1}{n}\sum_{i=1}^{n}|r_i|$. Punctul extrem obţinut în acest caz este mai mic decât $\frac{1}{n+1}$, unde n reprezintă numărul parametrilor ce urmează a fi estimaţi.

- **L** - reprezintă combinaţii liniare ale statisticii de diferite ordine. Cea mai utilizată în acest caz este valoarea mediană pentru cazul în care cel mult jumătate din date sunt cu erori, punctul de separare obţinut fiind maximum 0,5.
- **R** – care se bazează pe testul de ierarhizare.

BUPT

5.2.2 Estimarea mişcării bazată pe segmentarea spaţio-temporală

Tehnicile de estimare a mişcării prezentate în paragraful precedent s-au bazat pe existenţa unei segmentări. Spre deosebire de acestea, tehnicile bazate pe segmentare spaţio-temporală au ca scop final determinarea regiunilor omogene atât din punct de vedere al semnalului de luminanţă, cât şi din punct de vedere al mişcării. În aceste condiţii regiunile pot fi identificate ca fiind obiectele din imagine. Noţiunea de mişcare coerentă depinde şi în acest caz de modelul utilizat, fiind mai apropiată de modelul fundamental al mişcării.

Nedeterminarea rezultată în urma aplicării algoritmilor de segmentare spaţio-temporală poate fi rezolvată prin realizarea unei segmentări tridimensionale bazată pe informaţia de luminanţă [SAL 94]. Metoda rezolvă problema corespondenţei de-a lungul axei timp, dar utilizează ca şi criteriu de segmentare luminanţa uniformă în locul coerenţei mişcării.

O altă propunere se bazează pe estimarea mişcării (traseul optic), urmând ca mai apoi să se definească pixelii a căror mişcare corespunde modelului ales [WAN 94]. Limitările acestui algoritm se datorează preciziei limitate în ceea ce priveşte mişcarea muchiilor, deoarece estimarea traseului optic utilizează un model neparametric [ANA 93].

Eliminarea dezavantajelor metodelor anterioare este posibilă prin realizarea simultană a segmentării spaţio-temporale şi a estimării [BOU 93]. Problema ce trebuie rezolvată în acest caz este reprezentarea unui model de relaxare ce se bazează pe modelarea câmpului Markov MRF (MRF – Markov Random Field) şi a criteriului Bayesian.

Datorită echivalenţei dintre MRF şi distribuţia Gibbs, modelul realizează o reducere a energiei pentru vecinătăţile locale. În acest caz probabilitatea p a sistemului de a se afla într-o stare particulară este dată de :

$$p = \frac{1}{Z} e^{\frac{-U}{T}} \tag{5.4}$$

unde: Z reprezintă constanta de normalizare (sau funcţia de partiţie), T este o constantă care corespunde temperaturii din mecanica statică, iar U reprezintă funcţia de energie.

Această schemă de estimare este aplicată iterativ, iar pentru coerenţa segmentării în timp, câmpul mişcării obţinut pentru momentul t este utilizat pentru predicţia segmentării la momentul t+1.

O altă abordare constă în îmbunătăţirea secvenţială a segmentării şi mişcării. În [PEL 90] se realizează estimarea pe baza unei mişcări dominante, după care pe baza erorilor de predicţie mari (obţinute pentru regiunile aflate în mişcare), se recalculează noii parametri de mişcare. Procedeul se aplică iterativ şi se bazează pe presupunerea că obiectele sunt mici în raport cu fundalul.

BUPT

O procedură asemănătoare este aplicată în [DUF 95b]. Astfel se descompune mişcarea în: mişcare globală datorată camerei (rotire sau zoom) şi locală datorată deplasării obiectelor în scenă. Pentru obţinerea mişcării globale se realizează o estimare a mişcării bazată pe un model afin, printr-o tehnică de corespondenţă ierarhică. Fiind utilizată o segmentare statică, estimarea mişcării se aplică şi regiunilor caracterizate de o mişcare coerentă, obţinându-se în acest mod o estimare precisă şi robustă. Cu toate acestea, pentru a elimina erorile datorate erorilor de segmentare, se utilizează un estimator robust de tip M, definit de ecuaţia (5.3), care este mai puţin sensibil la valori excepţionale (care ies din domeniu).

Deoarece unele din caracteristicile de contrast scăzut ale imaginii se pierd în etapa de segmentare, pentru regiunile a căror compensare a mişcării nu este eficientă (erorile depăşesc un anumit prag), se ia decizia divizării suplimentare a acestora. Această îmbunătăţire a segmentării statice permite reconstrucţia câtorva dintre caracteristicile de contrast, având o mişcare semificativ distinctă, care se presupun a fi importante din punct de vedere vizual.

Avantajul principal al algoritmului de segmentare spaţio-temporală constă în poziţionarea precisă a muchiilor, deoarece acestea sunt determinante în semnalul de luminanţă. În particular, există problemele referitoare la acoperire (dispariţia regiunilor) sau apariţie. Pe lângă aceasta, algoritmul de estimare al mişcării se caracterizează prin robusteţe şi reducerea influenţei zgomotului, datorită celor două etape de estimare: globală şi locală. Din acest motiv, algoritmul este eficient în găsirea entităţilor spaţio-temporale semificative ale scenei.

Prezenţa acoperirilor şi apariţiilor ridică probleme pe durata operaţiilor de segmentare. Datorită acestora mişcarea aparentă a texturii şi a contururilor unor obiecte poate apărea ca fiind diferită. Pentru a evita această problemă, se poate estima şi transmite mişcarea conturului împreună cu mişcarea corespunzatoare texturii. Principalele situaţii posibile în cazul apariţiilor şi acoperirilor sunt prezentate în figura 5.1

O altă posibilitate de a evita problemele datorate apariţiei şi dispariţiei regiunilor o constituie utilizarea tehnicilor de căutare. Cu ajutorul acestora se defineşte traiectoria obiectelor în timp, putându-se recunoaşte apariţia unui obiect prezent în scenă în cadrele anterioare.

Problema fundalului ce apare şi care are o mare importanţă pentru codare, poate fi rezolvată prin utilizarea unei memorii pe termen lung ce stochează toată informaţia referitoare la fundal [IRA 94], [KER 95]. Eficienţa acestor algoritmi se observă cu precădere în aplicaţiile la rate foarte scăzute pentru: video conferinţe sau video telefonie.

BUPT

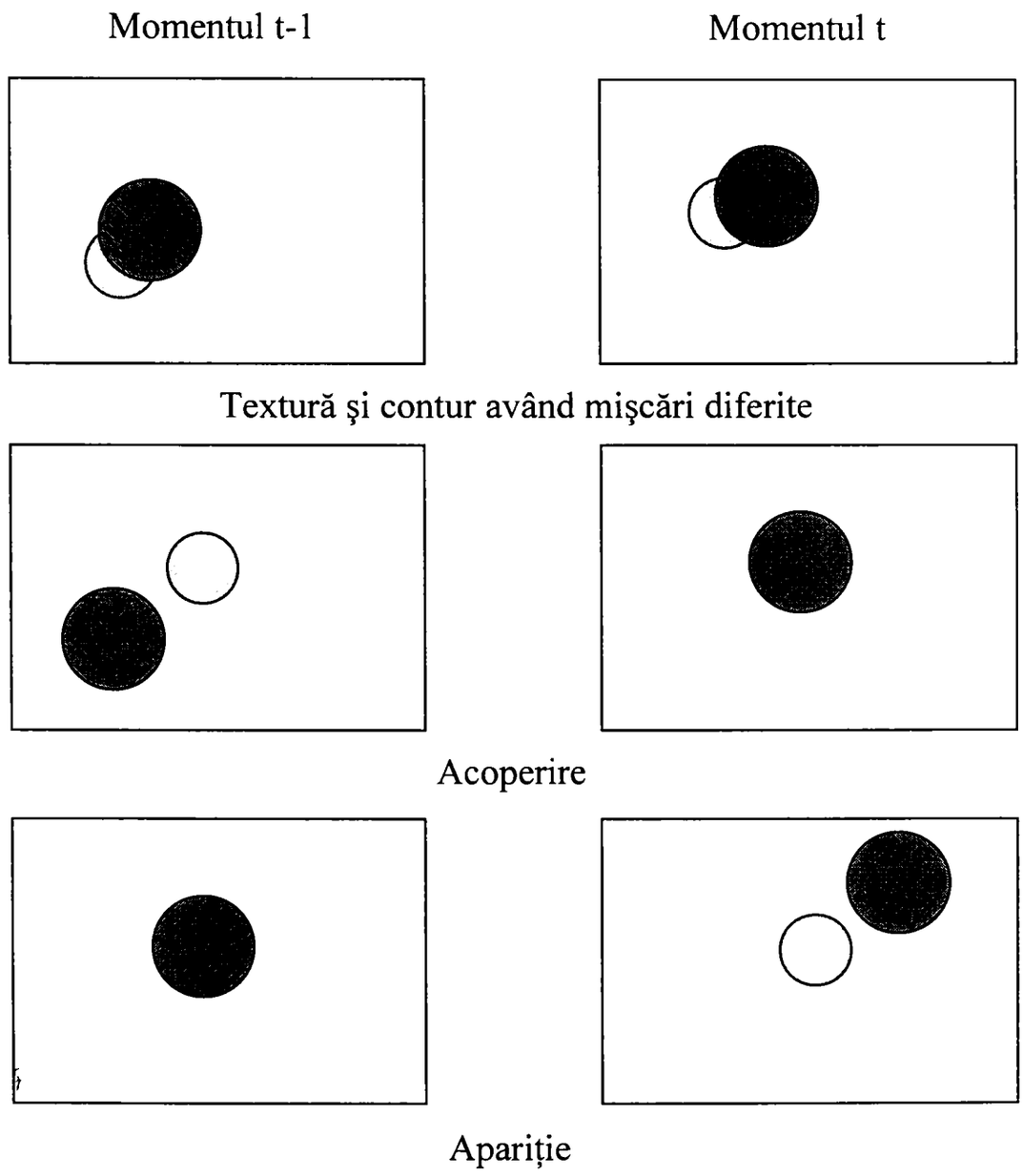


Fig. 5.1 Principalele situaţii posibile referitoare la apariţie şi acoperire

5.2.3 Urmărirea

Tehnicile ale căror scop îl reprezintă obţinerea unei segmentări spaţio-temporale trebuie să se bazeze pe informaţia de mişcare, pentru a obţine o reprezentare completă a scenei şi a permite refacerea imaginii cu distorsiuni vizuale nesemnificative. Oricum, cele mai multe dintre tehnici iau în considerare două cadre consecutive, motiv pentru care prezintă următoarele dezavantaje:

- imprecizia estimării ,
- zgomotul sau absenţa unei informaţii importante pot conduce la alterarea interpretării scenei.

BUPT

Aceasta poate conduce către o segmentare spaţio-temporală fără o coerenţă în timp. Rezolvarea acestei probleme o constituie aplicarea unei proceduri de căutare pe baza informaţiei obţinute în imaginile precedente.

5.2.3.1 Urmărirea obiectelor din scenă

Dacă segmantarea este obţinută prin tehnici spaţio-temporale, atunci este de aşteptat ca regiunile să corespundă obiectelor reale din scenă, iar evoluţia lor să fie predictibilă şi fără existenţa unor discontinuităţi. În aceste condiţii procesul de urmărire poate fi definit ca: procesul de estimare a stării dinamice a obiectului bazat pe măsurarea mişcării.

Astfel, evoluţia obiectelor poate fi aproximată prin traiectoria în timp. Implementarea procedurii necesită, din acest motiv, definirea filtrării temporale precum şi a unui model explicit al traectoriei. Această tehnică permite combinarea informaţiilor de la mai multe imagini, conducând la scăderea incertitudinii pentru estimarea unei stări, prin acumularea mai multor măsurători în timp. Astfel procedura de urmărire este capabilă să ia în considerare în mod explicit imprecizile existente în segmentarea spaţio-temporală la fel ca şi în cazul estimării mişcării bazată pe segmentarea statică.

Impreciziile se datorează mai multor factori:

- distorsiuni introduse la nivelul formării imaginii (distorsiuni geometrice, neclaritate, discretizare ,cuantizare);
- segmentarea spaţio-temporală inexactă;
- textura şi conţinutul ce pot avea mişcări aparent diferite (paragraful 5.2.2).

Pentru a exploata proprietăţile şi caracteristicile procesului de urmărire este necesară găsirea traiectoriei obiectelor, ceea ce presupune rezolvarea problemei de corespondenţă definită de Aggarwall şi alţii [AGG 81]. Implementări ale acestei tehnici au fost prezentate în [CHE 96], [IRA 94], [LEG 82].

5.2.3.2 Algoritmi de filtrare Kalman

Deoarece regiunile obţinute prin segmentare sunt foarte apropiate de obiectele prezente în scenă, mişcările, în perspectiva evoluţiei temporale, pot fi descrise prin legile clasice ale cinematicii. Această observaţie permite definirea modelului traiectoriei utilizând componentele cinematice şi anume: poziţia, viteza şi accelerаţia. Astfel se poate descrie evoluţia prin ecuaţia dinamicii (ecuaţia lui Plant), a cărei relaţie în situaţia unei eşantionări temporale este descrisă de ecuaţia:

$$\vec{x}(t_{k+1}) = F(t_k)\vec{x}(t_k) + G(t_k)\vec{u}(t_k) + \vec{v}(t_k) \qquad (5.5)$$

în care $F(t_k)$ este matricea ce încorporează modelul traiectoriei.

Relaţia dintre sistem şi vectorul de intrare cunoscut $\vec{u}(t_k)$ este definită de matricea $G(t_k)$, vectorul de intrare $\vec{u}(t_k)$ reprezentând influenţele exterioare asupra sistemului.

BUPT

Ecuaţia (5.5) descrie starea sistemului dinamic la momentul t_k, atunci când a fost realizată o măsurare. Relaţia dintre $\vec{x}(t_k)$ şi valoarea măsurată $\vec{z}(t_k)$ este dată de:

$$\vec{z}(t_k) = H(t_k)\vec{x}(t_k) + \vec{w}(t_k) \tag{5.6}$$

Ecuaţiile (5.5) şi (5.6), care definesc relaţia de predicţie şi cea de recursivitate, formează împreună aşa-numitul algoritm de filtrare Kalman ce poate fi utilizat pentru urmărirea obiectelor în cazul în care este definit un model al traectoriei [KIM 98], [ZIL 97].

Întrucât definirea explicită a modelului traiectoriei poate deveni mult prea complexă, se pot utiliza tehnici de integrare temporală, care permit implementarea unei proceduri de urmărire fără necesitatea unui model al traiectoriei. În acest caz informaţia referitoare la mişcarea obiectului se păstrează sub forma unei imagini, care se obţine prin combinarea imaginilor succesive $I(t_k)$ în raport cu obiectul urmărit, obţinându-se o reprezentare dinamică pentru acesta.

Integrarea temporală conduce la o funcţie de filtrare temporală, ce se obţine prin combinarea imaginilor anterioare şi a celor curente, conducând la obţinerea unei anumite imunităţi la impreciziile de măsurare a mişcării. Având un obiect secvenţa corespunzătoare imaginii integrate temporal A_v poate fi definită astfel :

$$A_v(0) = I(t_0) \tag{5.7}$$

$$A_v(t_{k+1}) = mI(t_{k+1}) + (1-m)reg(A_v(t_k), I(t_{k+1})) \tag{5.8}$$

unde $m \in [0,1)$ iar *reg(K,L)* semnifică înregistrarea imaginii K în raport cu imaginea L cu măsurarea mişcării obiectului, fiind calculată între momentele t_K şi t_L [IRA 92].

Prin calculul mişcării obiectului dintre noul cadru şi imaginea integrată corespunzătoare se realizează creşterea preciziei de măsurare a mişcării, cu dezavantajul că, în imaginea înregistrată obiectele care nu sunt supuse urmăririi sunt neclare, faţă de cele supuse urmăririi şi care sunt clare.

5.3 Algoritmul propus de autor pentru estimarea mişcării în vederea codării bazate pe regiuni a secvenţelor de imagini

Având în vedere tendinţa actuală spre codarea la rate foarte scăzute de biţi cu aplicaţii în video-telefonie, video-conferinţă, multimedia, precum şi propunerea de standardizare pentru MPEG-4, este clar că soluţia o reprezintă implementarea tehnicilor de codare bazate pe regiuni [COR 98], [GUN 98], [SIK 97], [PER 97].

În acest scop autorul a propus un algoritm de estimare a mişcării pentru codarea secvenţelor de imagini care se bazează pe o segmentare statică adaptivă a imaginilor [ALE 98c]. Schema bloc a algoritmului propus pentru estimarea şi codarea secvenţelor de imagini este prezentată în figura 5.2 [ALE 98e].

BUPT

Ca date de intrare algoritmul utilizează imaginea curentă originală şi segmentată. De asemenea, există o întârziere de două cadre înaintea generării primei imagini codate, întrucât se utilizează o codare cu predicţie, motiv pentru care este necesară cunoaşterea atât a imaginii transmise cât şi a imaginii curente. Deoarece, aşa cum s-a arătat, tehnicile de estimare a mişcării depind foarte mult de calitatea segmentării s-a utilizat pentru segmentare algoritmul adaptiv propus de autor în capitolul 4.

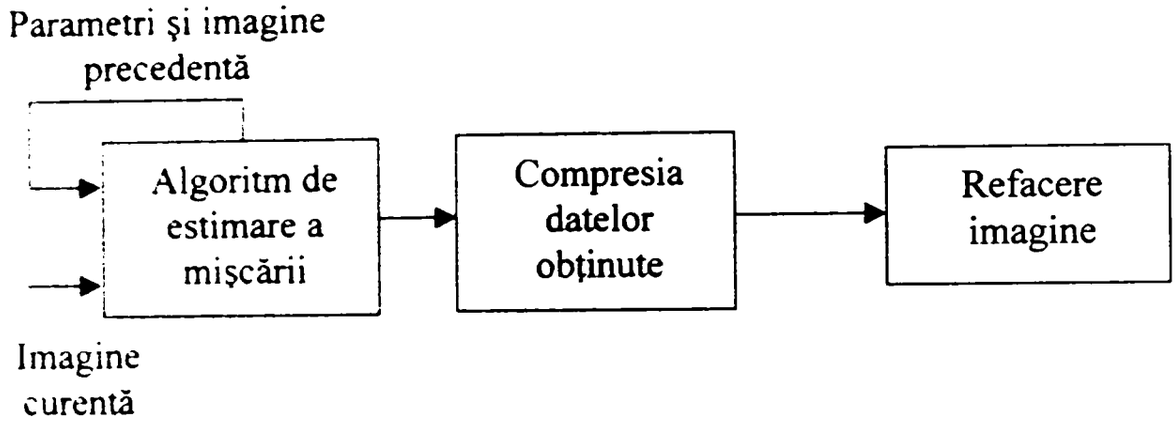


Fig. 5.2 Schema bloc a algoritmului propus pentru estimarea şi codarea secvenţelor de imagini

Organigrama algoritmului propus este prezentată în figura 5.3.

Problema principală care se pune este modul în care poate fi descris un obiect care se mişcă în secvenţă în aşa fel încât să se obţină o compresie cât mai mare a informaţiei. Tehnica propusă este o tehnică de căutare globală pe baza unor parametri de formă şi de poziţie. În continuare se vor prezenta algoritmii utilizaţi pentru determinarea acestor parametri, pentru estimarea mişcării, determinarea vectorilor de mişcare şi o posibilă codare a informaţiilor pe baza cărora se poate reface secvenţa de imagini la recepţie.

După cum se observă din schema bloc prima etapă o constituie segmentarea imaginii. Etapa de filtrare a regiunilor mici corespunde etapei de contopire a regiunilor mici din algoritmul de segmentare propus (paragraful 4.4.5). Etichetarea reprezintă o etapă necesară, în cazul în care imaginea segmentată nu conţine şi imaginea de etichete, deoarece astfel se realizează clasificarea regiunilor şi a parametrilor acestora.

Pe baza imaginilor originale curente şi precedente, a imaginilor segmentate şi a caractristicilor regiunilor se realizează estimarea mişcării prin urmărirea regiunilor şi determinarea corespondenţelor dintre acestea. În etapa de căutare se întocmeşte un tabel cu cele mai bune corespondenţe pe baza grafului adiacenţelor, calculul vectorilor de mişcare realizându-se pe baza erorii medii pătratice minime prin aplicarea modelului afin al mişcării (translaţie, rotaţie, scalare) [ALE 98d].

BUPT

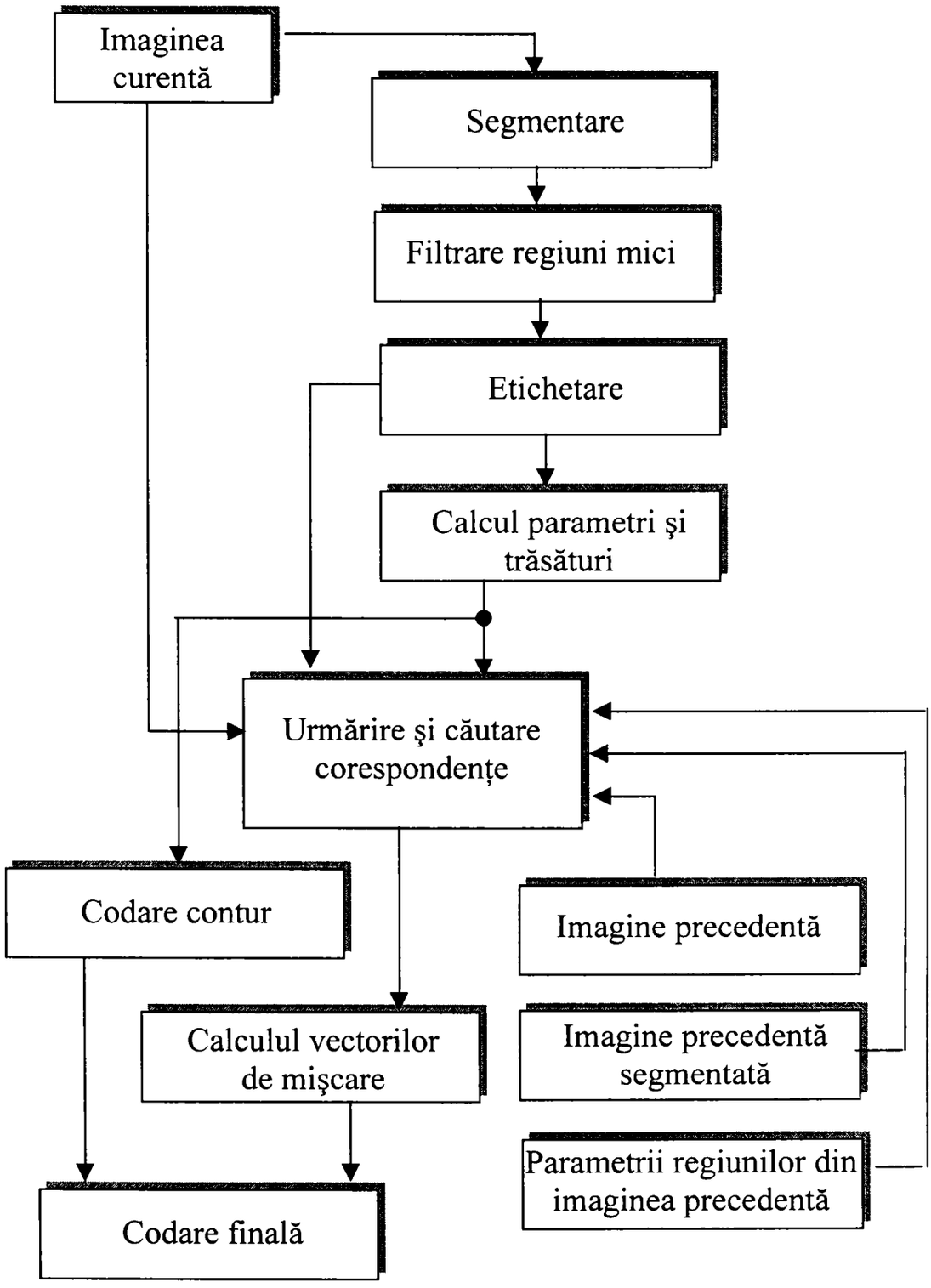


Fig. 5.3 Algoritmul de estimare a mişcării în secvenţele de imagini pe baza segmentării statice

Pe baza vectorilor de mişcare se realizează o compresie cu predicţie a informaţiei, transmiţându-se eroarea, vectorii de mişcare, eticheta regiunii şi corespondenţa precum şi coeficienţii funcţiei de aproximare a luminanţei. În cazul în care regiunile nu au corespondenţă în cadrul anterior se transmite conturul

BUPT

regiunii precum şi informaţia referitoare la coeficienţii funcţiei de aproximare a luminanţei. Algoritmul are posibilitatea unei codări ierarhice în sensul că se poate transmite şi informaţia referitoare la textură dacă se doreşte o codare fără pierderi.

5.3.1 Etichetarea

Întrucât se realizează segmentarea imaginii în regiuni a căror modele sunt aproximate prin funcţii liniare, este foarte important ca în procesul prelucrărilor în vederea extragerii caracteristicilor să nu introducem erori suplimentare. Din acest motiv în procesul de etichetare a regiunilor se atribuie o etichetă fiecărei regiuni în aşa fel încât nu mai există posibilitatea modificării formei sau a pierderii de caracteristici.

Prelucrarea imaginii segmentate se face pixel cu pixel prin parcurgerea imaginii de la stânga la dreapta şi de sus în jos. Algoritmul calculează aria fiecărei regiuni şi atribuie câte o etichetă fiecăreia. Datele obţinute sunt păstrate într-un tabel de descriere a regiunilor. Pentru fiecare pixel analizat se stabileşte regiunea din care face parte, se incrementează aria regiuni şi se depune eticheta corespunzătoare în imaginea de etichete. Determinarea regiunii din care face parte fiecare pixel se realizează prin analiza conectivităţii.

Etichetarea este realizată în cadrul etapei de segmentare, întrucât pentru etapa de filtrare a regiunilor mici este necesară cunoaşterea etichetei fiecărei regiuni şi de asemenea constituie o bază pentru realizarea extragerii contururilor regiunilor.

5.3.2 Calculul parametrilor regiunilor

Determinarea parametrilor regiunilor reprezintă finalizarea etapei de analiză a imaginii, scopul fiind acela de a extrage principalele caracteristici ale regiunilor din imagine. Caracteristicile trebuie alese cu grijă întrucât este necesar ca acestea să fie invariante la principalele transformări ce rezultă din mişcarea regiunilor. În această etapă s-a presupus că obiectele corespunzătoare regiunilor obţinute prin segmentare sunt rigide (nedeformabile).

Aceste caracteristici trebuie de asemenea să caracterizeze regiunile în aşa fel încât interpretarea lor să fie singulară (să nu existe ambiguităţi de clasificare). Studiile în domeniul estimării şi compensării mişcării au utilizat un model predefinit fără a căuta adaptarea acestuia la variaţiile mişcării prezente în imagine. De aceea în cazul algoritmului dezvoltat, modelul de mişcare va fi identificat local prin analiza a două imagini.

Adaptarea locală a modelului de mişcare este avantajoasă din următoarele motive:

- alegerea unui model prea simplu (de exemplu model constant) într-o regiune în care mişcarea observată fizic este complexă (mişcări tridimensionale ale corpurilor rigide), poate conduce la o reconstituire cu erori mari prin utilizarea

BUPT

tehnicilor cu compensarea mişcării sau la o suprasegmentare a imaginii, care în acest caz devine costisitoare în termeni ai volumului de date transmise;

- alegerea unui model complicat (spre exemplu un model pătratic) într-o regiune a cărei mişcare nu este complexă (translaţia bidimensională), conduce la o estimare ce necesită o cantitate mare de informaţii transmise pentru reconstrucţie, incluzând şi parametrii corespunzători unei mişcări ce nu poate fi natural identificată. De fapt criteriul de minimizare în estimarea parametrului de mişcare este adesea global deoarece este dependent simultan de toate componentele vectorului de mişcare identificat. Astfel toate componentele de mişcare ce nu sunt observabile introduc o eroare sistematică în identificarea componentelor reale ale mişcării.

Metodele bazate pe corelare, necesită din acest motiv, determinarea unor trăsături invariante pentru obiecte care să ne conducă la o soluţie de deplasare unică. Acest lucru are avantajul că odată determinată mişcarea trăsăturilor, deplasarea obiectului ca întreg poate fi sugerată de deplasarea trăsăturii. Astfel pentru analiza regiunilor se utilizează trăsături corespunzătoare cazurilor în care ele caracterizează toţi pixeli regiunii (corpuri rigide).

Trăsăturile utilizate în cadrul acestui algoritm sunt:

- funcţia de aproximare a luminanţei regiunii (care în unele cazuri poate fi chiar media luminanţei pixelilor);
- suprafaţa regiunii (aria);
- conturul regiunii (perimetrul);
- centru de greutate (care determină poziţia obiectului şi care mai este cunoscut sub denumirea de moment de ordinul I);
- momentele de ordinul II care ne permit calculul:
    - caracteristicii de formă, şi
    - unghiul dintre axa x şi axa de alungire a obiectului.

Aceste caracteristici vor fi înscrise într-o structură de descriere a regiunilor care reprezintă punctul de plecare pentru următoarea etapă de calcul. Mai jos este prezentat modul de calcul al acestor caracteristici aşa cum a fost implementat în cadrul programului de simulare a algoritmului.

Aria: Este numărul pixelilor ce fac parte dintr-o regiune şi se calculează la etichetarea imaginii, reprezentând numărul total al pixelilor închişi de conturul regiunii.

Funcţia de aproximare a luminanţei regiunii: este descrisă de relaţia

$$f(x, y) = a + bx + cy \tag{5.9}$$

unde: a reprezintă media de gri a regiunii calculată prin însumarea valorilor de luminanţă a pixelilor şi împărţirea sumei la aria regiunii, b reprezintă gradientul mediu pe direcţia x, iar c gradientul mediu pe direcţia y, calculaţi funcţie de centrul de greutate al regiunii.

BUPT

Conturul (Perimetrul): este numărul pixelilor de margine al unei regiuni; calcularea perimetrului se face prin algoritmul descris în paragraful 5.3.3.

Centrul de greutate: Este centru de greutate al fiecărei regiuni şi se utilizează pentru determinarea poziţiei unei regiuni în imagine. Se calculează pentru fiecare regiune conform relaţiei:

$$(\bar{x},\bar{y})^T = \left(\frac{1}{n}\sum_{i=0}^{n} x_i, \frac{1}{n}\sum_{i=0}^{n} y_i\right)^T \tag{5.10}$$

unde: n este numărul pixelilor din regiune; $\bar{x}$ şi $\bar{y}$ ordonata respectiv abscisa centrului regiunii, iar x_i şi y_i ordonata respectiv abscisa pixelului curent al regiunii.

Momentele unei regiuni: Sunt necesare pentru calculul unui parametru de formă a regiunilor utilizat la estimare şi a unghiului dintre axa x şi axa de alungire maximă a regiunii. În general momentul de ordinul II este utilizat pentru determinarea unghiului de rotaţie al regiunii.

Momentele regiunii sunt definite asfel [JAH 95]:

$$m_{p,q} = \int (x - x_i)^p (y - y_i)^q g(x) dxdy \tag{5.11}$$

unde:

$$x_i = \frac{\int x_i g(x) d^2x}{\int g(x) d^2x}, \quad y_i = \frac{\int y_i g(x) d^2x}{\int g(x) d^2x} \tag{5.12}$$

în care se definesc momentele de inerţie pentru un obiect prin funcţia $g(x)$, care poate fi privită ca densitatea obiectului, integrala este pe suprafaţa obiectului iar vectorul $(\bar{x},\bar{y})^T = (x_i, y_i)$ reprezintă centrul de greutate al regiunii.

Pentru o imagine binară calculul momentelor se reduce la ecuaţia:

$$m_{p,q} = \sum_{j=0}^{n} (x_j - x_i)^p (y_j - y_i)^q \tag{5.13}$$

unde suma include toţi pixelii aparţinând obiectului, x_j şi y_j reprezentă coordonatele pixelului curent.

5.3.3 Reprezentarea şi codarea conturului

Conturul unei regiuni constă din acele puncte ale imaginii care au cel puţin 4 vecini în regiunea care se constituie ca fundal al obiectului sau este vecina acestuia. Ca punct iniţial al conturului poate fi ales orice punct al acestuia. Deşi au fost calculate un număr relativ ridicat de caracteristici ale regiunilor, acestea sunt insuficiente pentru reconstrucţia imaginii numai pe baza lor. Din acest motiv este

BUPT

necesară transmiterea conturului regiunii cu ajutorul căruia este posibilă reconstrucţia obiectului.

Un lanţ de contur porneşte dintr-un punct iniţial al conturului şi se sfârşeşte în acelaşi punct, realizând conectarea tuturor punctelor de contur. Acesta se obţine printr-o succesiune de paşi făcuţi în 4 sau 8 direcţii funcţie de metoda de codare aleasă. În general codarea se realizează după 8 direcţii întrucât lungimea codului obţinut este mai scurtă.

Codarea Freeman pe opt direcţii se bazează pe asocierea unei cifre fiecărei direcţii, ca în figura 5.4, realizându-se în acest fel descrierea fără nici o nedeterminare a conturului imaginii. Specificarea suplimentară a coordonatelor absolute a punctului iniţial ales pentru contur permite o poziţionare exactă a conturului pe rastrul imaginii.

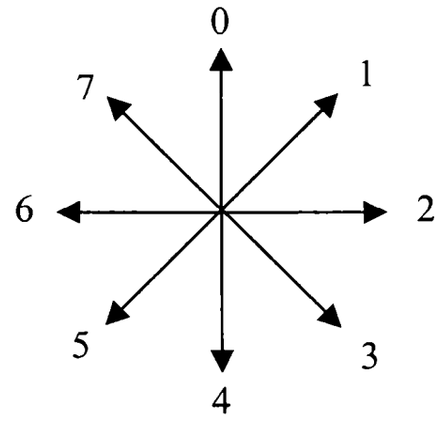


Fig. 5.4 Notarea direcţilor pentru codarea pe 8 direcţii a conturului.

Algoritmul de codare al contururilor este o procedură fundamentală în domeniul prelucrării imaginilor. El este utilizat foarte des în contextul analizei formelor, a extragerii caracteristicilor geometrice ale obiectului sau al compresiei imaginilor. Codurile pentru contur au avantajul că permit calcularea caracteristicilor geometrice ale obiectului prin manipulări numerice ale codului lanţ al conturului.

În general algoritmii pentru codarea contururilor sunt foarte eficienţi în ceea ce priveşte timpul de calcul, codul lanţ al conturului permiţând reconstrucţia exactă a obiectului. Procesul de determinare al conturului constă în două faze, una de localizare a obiectului şi cealaltă de trasare a conturului. Cele două faze ale algoritmului alternează pe durata întregului proces până la codarea contururilor tuturor obiectelor din imagine. Organigrama algoritmului de trasare este prezentată în figura 5.5.

Faza de trasare a conturului conduce la determinarea secvenţei binare a codului conturului şi la memorarea acestei secvenţe. În prima fază se calculează

BUPT

perimetrul fiecărei regiuni pentru a şti lungimea aproximativă a cuvântului de cod ce rezultă în urma codării Freeman, în vederea alocării dinamice a memoriei.

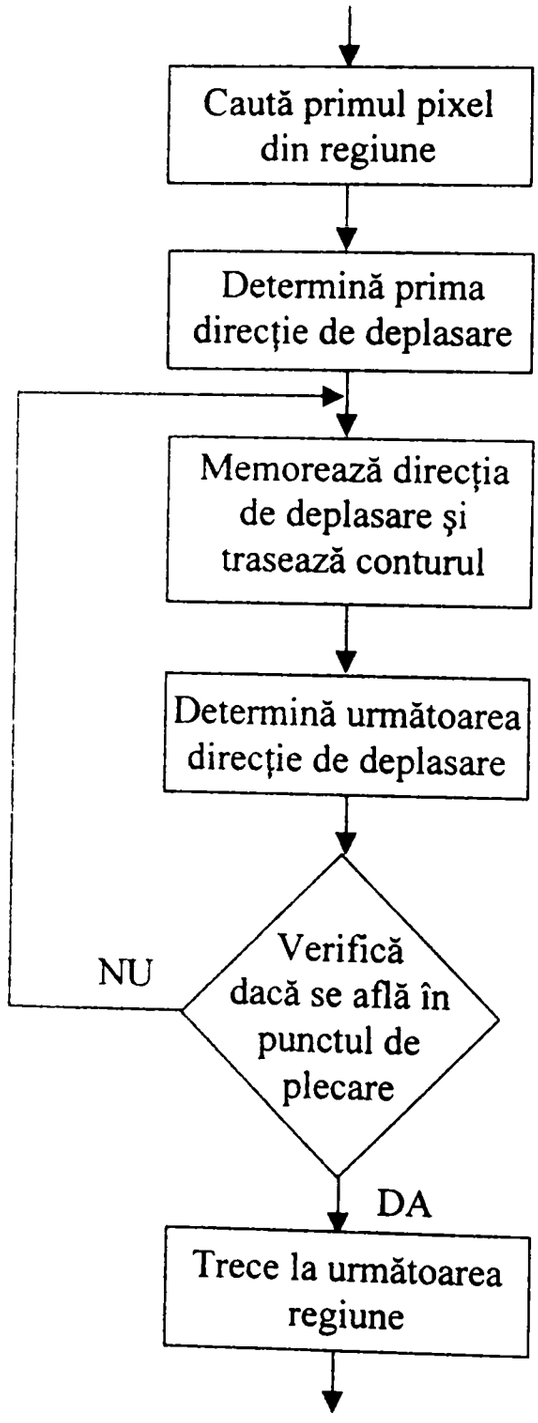


Fig. 5.5 Organigrama algoritmului de trasare

Deplasarea pe contur este asigurată după algoritmul "ţinînd mâna stângă pe perete". Direcţiile de deplasare ale fiecărui pas vor fi înscrise în pointerul aferent fiecărei regiuni, constituindu-se codul Freeman aferent conturului regiunii respective. Concomitent cu codarea se construieşte şi o imagine în care cu puncte

BUPT

albe pe fond negru este ilustrată urmărirea conturului. Funcţia de urmărire construieşte o imagine a contururilor fără a trasa contururi duble.

După terminarea codării rezultatele vor fi utilizate de algoritmul de compresie, după care memoria utilizată este disponibilizată.

5.3.4 Calculul vectorilor de mişcare

Aşa cum s-a arătat este esenţial să se determine cu precizie caracteristicile regiunilor pentru a putea identifica cu exactitate regiunile şi mişcările acestora între două cadre succesive. Acest lucru este esenţial în alegerea modelului de mişcare al regiunii.

Pentru un obiect rigid, care reprezintă obiectul solid din cinematică, se poate defini de asemenrea o mişcare rigidă, care înseamnă că distanţa între oricare două puncte ale obiectului este invariantă în timp. Mişcarea rigidă poate fi descrisă ca sumă a unor mişcări de rotaţie, translaţie şi scalare faţă de o axă fixă. În aceste condiţii dacă se consideră un spaţiu Euclidian bidimensional, în care poziţia unui punct este definită de vectorul:

$$\bar{x} = (x, y)^T \tag{5.14}$$

se poate descrie mişcarea unui anumit punct ce aparţine unei regiuni cu ajutorul relaţiei:

$$\bar{x}' = R\bar{x} + \bar{d} \tag{5.15}$$

în care vectorul $\bar{d}$ reprezintă translaţia, iar R reprezintă matricea de rotaţie care are proprietatea:

$$R^T R = I, \quad |R| = 1 \tag{5.16}$$

Matricea bidimensională este definită de:

$$R = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{5.17}$$

În aceste condiţii putem defini CR', ca rezultat al aplicării transformării rigide pentru CR, astfel:

$$CR' = \{\bar{x}' \mid \bar{x}' = R\bar{x} + \bar{d}, x \in CR\} \tag{5.18}$$

Astfel este posibil calculul parametrilor de mişcare a două regiuni, considerând numai formele acestora măsurate la momentele t_1 şi t_2. Având în vedere faptul că în condiţiile unei mişcări rigide avem o ecuaţie liniară, autorul a dezvoltat un algoritm pentru estimarea mişcării fără necesitatea cunoaşterii apriori a punctelor de corespondenţă a regiunilor.

Analizând proprietăţile corpului rigid putem enunţa următoarele:

BUPT

Propoziţia 1. Dacă considerăm centrele de greutate g şi g' ale regiunilor CR şi respectiv CR' atunci este îndeplinită relaţia:

$$g' = g + a \tag{5.19}$$

unde $g = (\bar{x}, \bar{y})^T$ şi $g' = (\bar{x}', \bar{y}')^T$.

Propoziţia 2. Pentru orice triplet de vectori, ***x, y, z*** se poate admite că $\|x - y\|$ (norma euclidiană) şi $(x - z)^T (y - z)$ sunt invariante în cazul unei mişcări euclidiene.

În aceste condiţii se poate enunţa următoarea teoremă:

TEOREMA: Dacă CR reprezintă o formă plană (CR reprezintă un domeniu finit al spaţiului real bidimensional) şi ∂CR reprezintă conturul regiunii CR, se poate defini pentru orice punct g, aparţinând lui CR, un set de puncte

$$K(r) = \partial CR \cap \{x \mid |x - g| = r, r > 0\} \tag{5.20}$$

În aceste condiţii numărul de elemente ale mulţimii K(r), $|K(r)|$ este invariant în cazul unei mişcări euclidiene (translaţie, rotaţie).

DEMONSTRAŢIE:

Pentru $K(r) \subset CR$ şi ***x, y, z*** $\in$CR şi $x', y', z' \in CR'$, ţinând cont de faptul că $g' = g + a$, se obţine:

$$K'(r) = \partial CR' \cap \{x' \mid |x' - g'| = r, r > 0\} \tag{5.21}$$

aceasta ne conduce în mod evident către concluzia că $|K(r)| = |K'(r)|$, dacă cele două regiuni sunt identice.

Teorema de mai sus ne arată că într-un spaţiu continuu, se poate rezolva problema punctelor de corespondenţă prin utilizarea lui $|K(r)|$.

Pentru estimarea centrului de greutate se poate utiliza ecuaţia (5.10) sau ecuaţia (5.22) care reprezintă aproximarea recursivă a ecuaţiei (5.10).

$$g_{n+1} = \frac{n}{n+1} g_n + \frac{1}{n+1} x \tag{5.22}$$

pentru $x \in CR$ şi $g_0 = 0$. Centrul de greutare calculat cu ajutorul ecuaţiei (5.22) converge către valoarea reală a centrului de greutate după un număr de iteraţii maxim egal cu numărul de pixeli ai regiunii.

Propoziţia 1 ne arată că translaţia regiunii este translaţia centrului de greutate, fiind uşor de determinat prin cunoaşterea poziţiei acestora în cadrul curent şi cel precedent.

BUPT

Pentru determinarea rotaţiei regiunii se poate utiliza fie momentul de ordinul doi fie definirea unor puncte aleatoare pe conturul regiunii, cu ajutorul corespondenţei acestora putându-se determina rotaţia regiunii.

Prin utilizarea momentelor de ordinul doi se poate defini un tensor de inerţie al obiectului:

$$J = \begin{bmatrix} m_{0,2} & -m_{1,1} \\ -m_{1,1} & m_{2,0} \end{bmatrix} \tag{5.23}$$

cu ajutorul căruia se poate calcula unghiul dintre axa x şi axa de alungire a obiectului:

$$\phi = \frac{1}{2}\arctan\frac{2m_{1,1}}{m_{2,0} - m_{0,2}} \tag{5.24}$$

Unghiul de rotaţie se determină prin diferenţa unghiurilor obţinute pentru regiune în cele două imagini succesive.

O a doua metodă pentru determinarea unghiului de rotaţie a regiunii, propusă de autorul tezei de doctorat, se bazează pe teorema de mai sus şi utilizează pentru aceasta un set de puncte obţinute prin intersecţia conturului regiunii cu un cerc având raza r cu proprietatea că $r_1 \le r \le r_2$, unde r_1 şi r_2 reprezintă distanţa minimă şi respectiv maximă a punctelor de contur faţă de centrul de greutate.

Astfel dacă considerăm două mulţimi de puncte $\overline{\mathbf{X}} = \{x_i\}_{i=1}^n$ şi $\overline{\mathbf{X}}' = \{x_i'\}_{i=1}^n$ în care $x_i \in K(r) \cap \partial CR$ şi respectiv $x_i' \in K'(r) \cap \partial CR'$, obţinem două mulţimi $\overline{\mathbf{X}} = \{\overline{x}_i\}_{i=1}^n$ şi $\overline{\mathbf{X}}' = \{\overline{x}_i'\}_{i=1}^n$, ale căror elemente sunt definite astfel: $\overline{x}_i = x_i - g$ şi $\overline{x}_i' = x_i' - g'$. În aceste condiţii, $\overline{\mathbf{X}}$ şi $\overline{\mathbf{X}}'$ reprezintă vârfurile poligonului de pe cercul de rază *r din* figura 5.6.

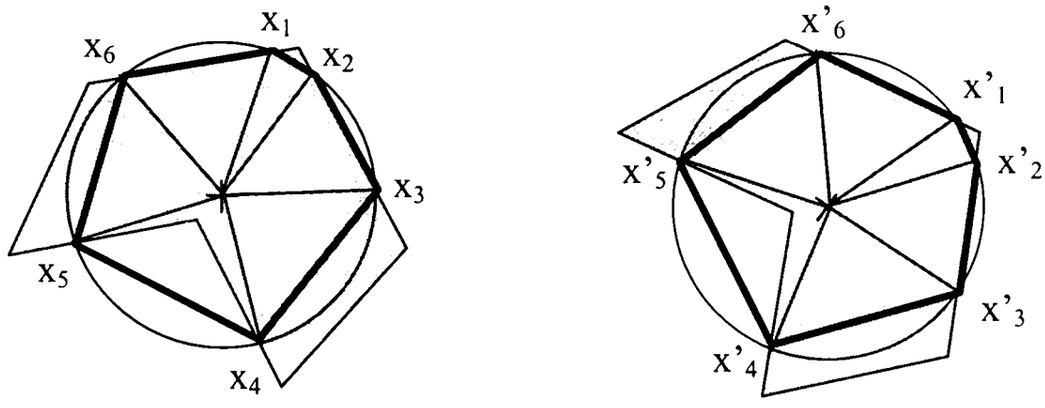


Fig. 5.6 Poligonul obţinut prin unirea punctelor de intersecţie a cercului cu conturul regiunii

BUPT

Dacă x_i şi x'_{i+k} corespund atunci este îndeplinită relaţia:

$$\bar{x}'_{i+k} = R\bar{x}'_i, \text{ pentru } i=1, 2, \ldots, n \tag{5.25}$$

Definind unghiul dintre două două vârfuri ale poligonului:

$$\theta_i = \arccos\left(\frac{\bar{x}_i^T \bar{x}_{i+1}}{\|\bar{x}_i\|\|\bar{x}_{i+1}\|}\right) \text{ şi,} \tag{5.26}$$

$$\Phi_i = \arccos\left(\frac{\bar{x}_i'^T \bar{x}'_{i+1}}{\|\bar{x}'_i\|\|\bar{x}'_{i+1}\|}\right) \tag{5.27}$$

se pot determina listele de unghiuri:

$$L(\bar{X}) = \langle \theta_1, \theta_2, \ldots, \theta_n \rangle \tag{5.28}$$

$$L(\bar{X}') = \langle \Phi_1, \Phi_2, \ldots, \Phi_n \rangle \tag{5.29}$$

Dacă se aplică ecuaţia (5.25) pentru toate rotaţiile posibile, se poate defini un criteriu de eroare astfel:

$$E(j) = \sum_{k=1}^{n} \left|\theta_k - \Phi_{j+k}\right| \tag{5.30}$$

Minimizarea acestui criteriu de eroare conduce la determinarea punctelor candidate pentru corespondenţă. În acest caz determinarea unghiului de rotaţie este echivalentă cu determinarea unghiului:

$$\alpha = \frac{1}{n}\sum_{k=1}^{n} \alpha_k \tag{5.31}$$

unde

$$\alpha_k = \arccos\left(\frac{\bar{x}_i^T \bar{x}'_{i+k}}{\|\bar{x}_i\|\|\bar{x}'_{i+k}\|}\right) \tag{5.32}$$

Pentru determinarea măririi respectiv micşorării regiunilor se calculează o caracteristică de formă a obiectului conform ecuaţiei (5.33):

$$cf = \frac{perimetru^2}{4\pi \times aria} \tag{5.33}$$

Un alt mod utilizat pentru determinarea caracteristicii de formă este cel bazat pe ecuaţia (5.34), în care se utilizează momentele de ordinul 1 şi 2. În algoritmul implementat este utilizată caracteristica determinată cu ajutorul ecuaţiei 5.34),

BUPT

deoarece este mai puţin sensibilă la variaţiile de contur datorate segmentării sau zgomotului.

$$cf_1 = \frac{(m_{2,0} - m_{0,2})^2 + 4m^2{}_{1,1}}{(m_{2,0} + m_{0,2})^2} \tag{5.34}$$

Estimarea coeficientului de mărire/micşorare se realizează prin calculul raportului caracteristicilor de formă pentru regiunile corespondente din cele două imagini succesive. Utilizînd aceşti parametri se caracterizează fiecare regiune a imagini.

Pentru a putea estima mişcarea se utilizează o căutare pe baza parametrilor de descriere a regiunilor, realizându-se determinarea corespondenţei între reginile din două imagini succesive. Pe baza corespondenţei şi prin utilizarea parametrilor ce descriu mişcarea (translaţie, rotaţie şi scalare) se poate estima noua poziţie a unei regiuni din imaginea precedentă în cea curentă.

Modelul de descriere a mişcării care se alege în cazul acestei aplicaţii este un model parametric definit prin comparaţia parametrilor de descriere a regiunilor (poziţia centrelor de greutate pentru translaţie, unghiul dintre axa de alungire şi ordonată sau unghiul de rotaţie determinat prin metoda cercului pentru rotaţie, raportul caracteristicilor de formă pentru scalare).

Pentru modelul de mişcare ales se impun o serie de limitări. Acestea sunt: valorile maxime ale translaţiei (16 pixeli), rotaţiei (20 grade) şi măririi/micşorării (50%), şi se aleg pe baza considerentului vitezei maxime de deplasare a obiectelor şi modificării maxime a distanţei focale pentru obţinerea a două imagini succesive. Aceste limitări sunt necesare pentru a se evita corespondenţa între regiuni asemănătoare din punctul de vedere al formei dar care nu corespund din punctul de vedere al mişcării reale. Apariţia de corespondenţe false, poate fi explicată prin faptul că pot apare regiuni noi în imagine, care sunt asemănătoare cu regiuni din imaginile anterioare.

Varianta de căutare a corespondenţelor aleasă pentru acest algoritm este una globală în cadrul imaginilor. Ca date de intrare algoritmul de căutare utilizează imaginea precedentă şi cea curentă precum şi structurile cu parametrii de descriere a regiunilor din imagini. Se parcurg regiunile din imaginea curentă. Pe baza parametrilor de poziţionare a regiunilor în imagine şi pe baza parametrului de formă definit anterior se determină regiunile din imaginea precedentă care asigură cea mai bună corespondenţă cu regiunea curentă. Decizia finală de corespondenţă se ia pe baza calculului erorii medii pătratice.

Algoritmul de căutare este prezentat în figura 5.7 şi se realizează pentru toate regiunile din imaginea curentă prin parcurgerea lor o singură dată.

BUPT

În scopul reducerii timpului necesar pentru calcule, regiunile pentru care a fost stabilită corespondenţa nu mai sunt luate în calcul pentru corespondenţele regiunilor următoare.

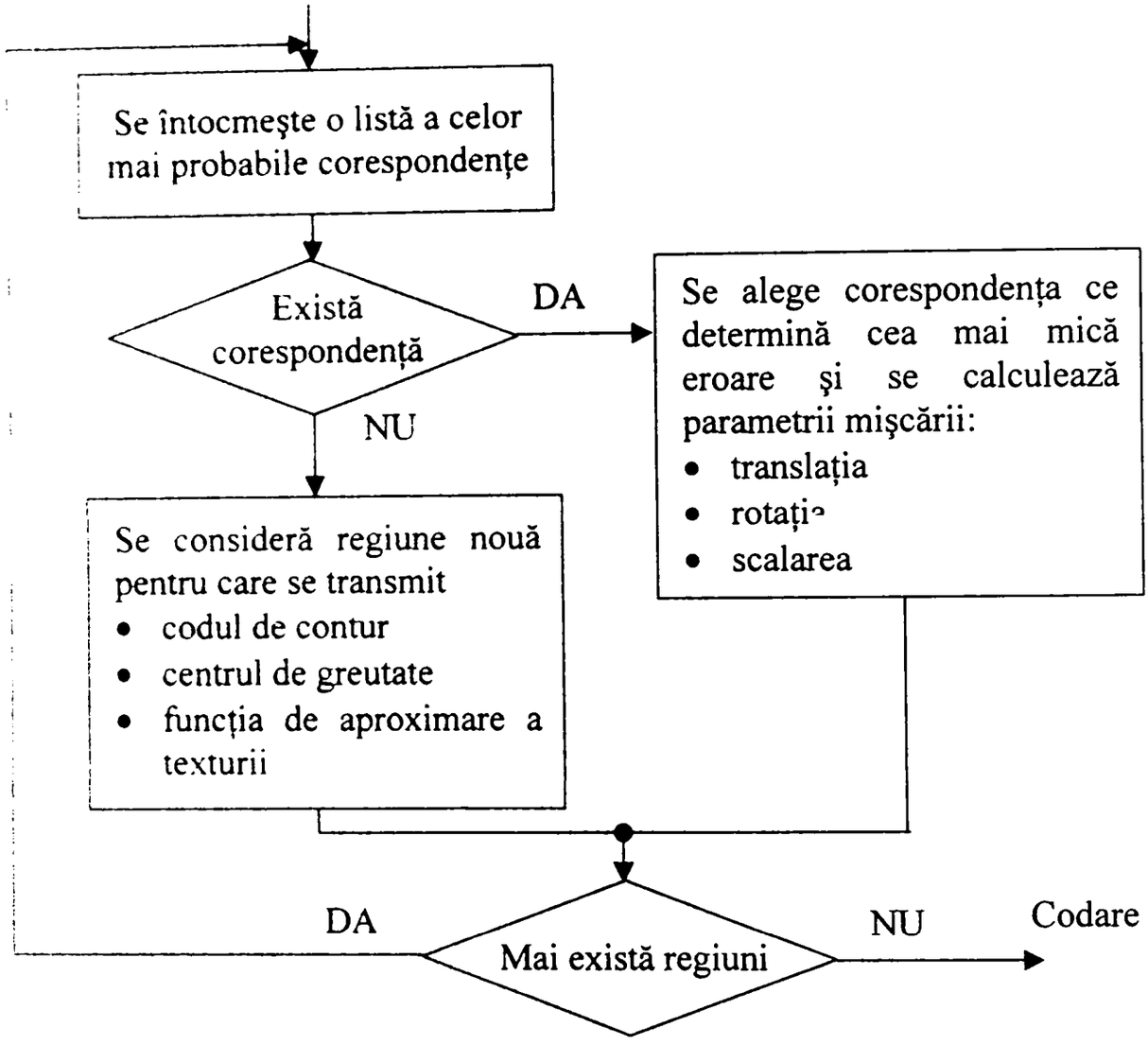


Fig. 5.7 Organigrama algoritmului de corespondenţă a regiunilor

Datele obţinute la căutarea corespondenţelor sunt utilizate pentru determinarea vectorilor de mişcare. Parametrii determinaţi sunt:

- translaţia, prin diferenţa dintre coordonatele centrelor de greutate;
- rotaţia, prin estimarea unghiului de rotaţie;
- scalarea (mărirea/ micşorarea), prin determinarea raportului dintre ariile celor două regiuni.

După determinarea vectorilor de mişcare se generează o imagine a acestora în care se specifică şi tipul de mişcare a regiuni prin următoarele simboluri:

- linie, pentru translaţie (realizează legătura între centrele de greutate a regiunilor corespondente);

BUPT

- patru pixeli poziţionaţi pe axele x, y în jurul centrului de greutate a regiunii din imaginea precedentă, pentru rotaţie;
- patru pixeli poziţionaţi pe axe oblice în jurul centrului de greutate a regiunii din imaginea precedentă, pentru mărire/ micşorare.

5.3.5 Codarea datelor obţinute prin estimarea mişcării

Acest algoritm realizează o compresie prin: transmiterea vectorilor de mişcare pentru imaginea curentă, în cazul regiunilor având o corespondenţă în imaginea precedentă şi transmiterea conturului, centrului de greutate şi modelului texturii pentru regiunile ce nu au o corespondenţă în imaginea precedentă. Compresia se realizează în ipoteza că la recepţie există deja imaginea precedentă şi imaginea de etichete a acesteia.

Pe baza acestei ipoteze s-au determinat parametrii de caracterizare a regiunilor care sunt necesari la recepţie pentru refacerea imagini curente. Parametrii sunt diferiţi funcţie de tipul regiunii (dacă este regiune nouă sau nu) şi de modelul de mişcare adoptat pentru aceasta.

Datele obţinute sunt depuse într-un fişier care este apoi codat prin utilizarea unui cod aritmetic. Cu excepţia datelor referitoare la contur, care sunt codate prin utilizarea unui cod run length, fişierul conţine date primare privind vectorii de mişcare, modelul utilizat pentru textură, eticheta regiunii, precum şi tipul modelului utilizat pentru estimarea mişcării (translaţie, rotaţie-translaţie, scalare-translaţie, rotaţie scalare, translaţie-rotaţie-scalare).

Secvenţa binară obţinută conţine următoarele date:

$D(1)$	numărul regiunilor imaginii curente (1 octet)
$D_1(2)$	numărul regiunilor de tipul 0 (regiuni fixe având corespondenţă în imaginea anterioară) (1 octet)
$R_{11}(1)$	tipul regiunii (3 biţi)
$R_{11}(2)$	eticheta regiunii (1 octet)
$R_{11}(3)$	eticheta regiunii corespondente (1 octet)
$R_{11}(4)$	modelul de aproximare al texturii (3 octeţi)
...	
$R_{i1}(1)$	tipul regiunii (3 biţi)
$R_{i1}(2)$	eticheta regiunii (1 octet)
$R_{i1}(3)$	eticheta regiunii corespondente (1 octet)
$R_{i1}(4)$	modelul de aproximare al texturii (3 octeţi)
$D_2(3)$	numărul regiunilor de tipul 1 (regiuni având corespondenţă aproximate prin translaţie) (1 octet)
$R_{12}(1)$	tipul regiunii (3 biţi)
$R_{12}(2)$	eticheta regiunii (1 octet)
$R_{12}(3)$	eticheta regiunii corespondente (1 octet)
$R_{12}(4)$	modelul de aproximare al texturii (3 octeţi)

BUPT

$R_{12}(5)$	deplasarea pe axa x (4 biţi)
$R_{12}(6)$	deplasarea pe axa y (4 biţi)
...	
$D_3(4)$	numărul regiunilor de tipul 2 (regiuni având corespondenţă aproximate prin rotaţie) (1 octet)
$R_{13}(1)$	tipul regiunii (3 biţi)
$R_{13}(2)$	eticheta regiunii (1 octet)
$R_{13}(3)$	eticheta regiunii corespondente (1 octet)
$R_{13}(4)$	modelul de aproximare al texturii (3 octeţi)
$R_{13}(5)$	centrul de greutate al regiunii corespondente (2 octeţi)
$R_{13}(6)$	unghiul de rotaţie (1 octet)
...	
$D_4(5)$	numărul regiunilor de tipul 3 (regiuni având corespondenţă aproximate prin translaţie-rotaţie) (1 octet)
$R_{14}(1)$	tipul regiunii (3 biţi)
$R_{14}(2)$	eticheta regiunii (1 octet)
$R_{14}(3)$	eticheta regiunii corespondente (1 octet)
$R_{14}(4)$	modelul de aproximare al texturii (3 octeţi)
$R_{14}(5)$	centrul de greutate al regiunii corespondente (2 octeţi)
$R_{14}(6)$	deplasare pe axa x (4 biţi)
$R_{14}(7)$	deplasare pe axa y (4 biţi)
$R_{14}(8)$	unghiul de rotaţie (1 octet)
...	
$D_5(6)$	numărul regiunilor de tipul 4 (regiuni având corespondenţă aproximate prin mărire/micşorare) (1 octet)
$R_{15}(1)$	tipul regiunii (3 biţi)
$R_{15}(2)$	eticheta regiunii (1 octet)
$R_{15}(3)$	eticheta regiunii corespondente (1 octet)
$R_{15}(4)$	modelul de aproximare al texturii (3 octeţi)
$R_{15}(5)$	raportul factorilor de formă (1 octet)
$R_{15}(6)$	semnul scalării (1 bit)
...	
$D_6(7)$	numărul regiunilor de tipul 5 (regiuni având corespondenţă aproximate prin translaţie-mărire micşorare) (1 octet)
$R_{16}(1)$	tipul regiunii (3 biţi)
$R_{16}(2)$	eticheta regiunii (1 octet)
$R_{16}(3)$	eticheta regiunii corespondente (1 octet)
$R_{16}(4)$	modelul de aproximare al texturii (3 octeţi)
$R_{16}(5)$	deplasare pe axa x (4 biţi)
$R_{16}(6)$	deplasare pe axa y (4 biţi)
$R_{16}(7)$	raportul factorilor de formă (1 octet)

BUPT

$R_{16}(8)$	semnul scalării (1 bit)
...	
$D_7(8)$	numărul regiunilor de tipul 6 (regiuni având corespondenţă aproximate prin rotaţie-mărire/micşorare) (1 octet)
$R_{17}(1)$	tipul regiunii (3 biţi)
$R_{17}(2)$	eticheta regiunii (1 octet)
$R_{17}(3)$	eticheta regiunii corespondente (1 octet)
$R_{17}(4)$	modelul de aproximare al texturii (3 octeţi)
$R_{17}(5)$	centrul de greutate al regiunii corespondente (2 octeţi)
$R_{17}(6)$	raportul factorilor de formă (1 octet)
$R_{17}(7)$	semnul scalării (1 bit)
$R_{17}(8)$	unghiul de rotaţie (1 octet)
...	
$D_8(9)$	numărul regiunilor de tipul 7 (regiuni având corespondenţă aproximate prin modelul afin complet (rotaţie-translaţie-mărire/micşorare)) (1 octet)
$R_{18}(1)$	tipul regiunii (3 biţi)
$R_{18}(2)$	eticheta regiunii (1 octet)
$R_{18}(3)$	eticheta regiunii corespondente (1 octet)
$R_{18}(4)$	modelul de aproximare al texturii (3 octeţi)
$R_{18}(5)$	centrul de greutate al regiunii corespondente (2 octeţi)
$R_{18}(6)$	raportul factorilor de formă (1 octet)
$R_{18}(7)$	deplasare pe axa x (4 biţi)
$R_{18}(8)$	deplasare pe axa y (4 biţi)
$R_{18}(9)$	semnul scalării (1 bit)
$R_{18}(10)$	unghiul de rotaţie (1 octet)
...	
$D_9(10)$	numărul regiunilor noi (1 octet)
$R_{19}(1)$	numărul de paşi al conturului z (2 octeţi)
$R_{19}(2)$	coordonatele punctului iniţial (2 octeţi)
$R_{19}(3)$	eticheta regiunii (1 octet)
$R_{19}(4)$	modelul de aproximare al texturii (3 octeţi)
$R_{19}(5)$	centrul de greutate al regiunii (2 octeţi)
$R_{19}(6)$	direcţia primului pas de contur (3 biţi)
$R_{19}(z+5)$	direcţia de deplasare a ultimului pas de contur (3 biţi)
...	

În fişierul rezultat aceste date sunt stocate succesiv. La refacerea imaginii se recunoaşte structura datelor putându-se reface structura de caracterizare a regiunilor. Algoritmul presupune transmiterea primei imagini complete (parametrii necesari pentru regiuni fiind cei corespunzători cazului în care toate regiunile sunt

BUPT

noi). Se poate observa că alegerea adaptivă a modelului de estimare al mişcării conduce la creşterea factorului de compresie.

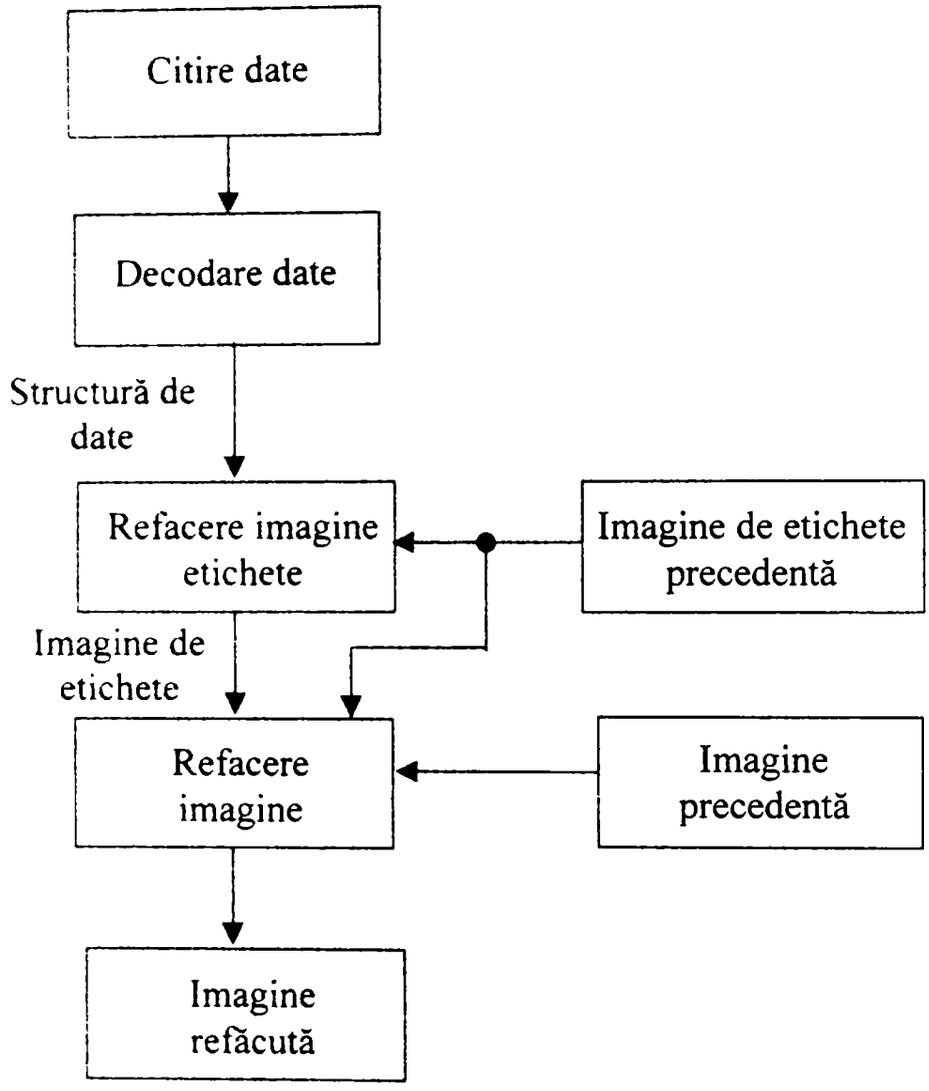


Fig. 5.8 Organigrama algoritmului de refacere.

5.3.6 Refacerea imaginii

Algoritmul de refacere a imaginii, prezentat în cele ce urmează, propune o metodă de refacere a unei imagini dintr-o secvenţă de imagini pe baza parametrilor transmişi şi ai imaginii precedente. Pentru a se putea reface imaginea se consideră că există vectorii de mişcare, imaginea precedentă şi imaginea sa de etichete din parcurgerea anterioară a algoritmului de refacere.

Figura 5.8 prezintă organigrama algoritmului de refacere. Întrucât există două tipuri principale de regiuni şi anume, regiuni noi (nu au corespondenţă în imaginea precedentă) şi regiuni cu corespondenţă, au fost implementate două funcţii separate pentru refacerea regiunilor.

BUPT

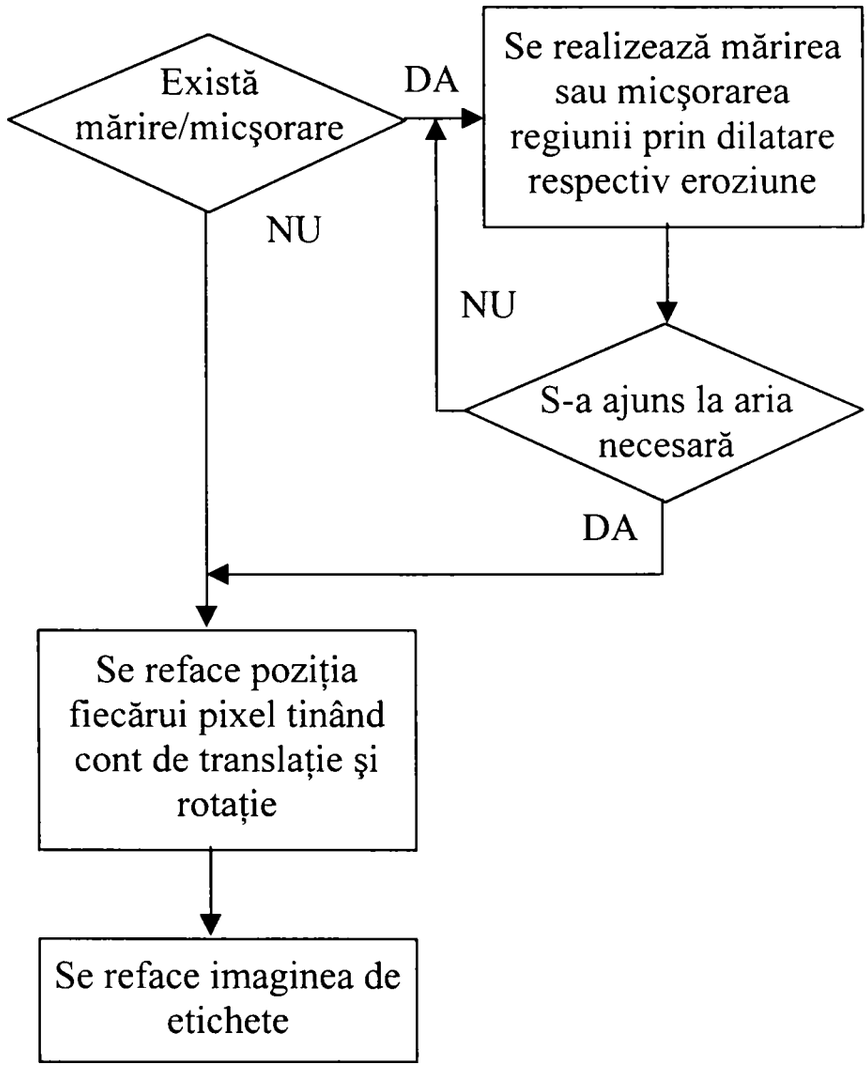


Figura 5.9 Organigrama algoritmului de refacere pentru regiunile având corespondenţă

Refacerea regiunilor cu corespondenţă se realizează prin deplasarea imaginii de etichete ale imaginii precedente conform vectorilor de deplasare definiţi pentru regiuni, care în cazul general conţin următorii parametri: translaţia, rotaţia şi respectiv mărirea/ micşorarea. Algoritmul de refacere în acest caz este prezentat în figura 5.9.

Refacerea imaginii regiunilor noi se bazează pe codul şir de contur şi pe modelul de aproximare al texturii. Algoritmul urmează următorii paşi:

- se trasează conturul regiuni pe baza codului şir prin plecarea din punctul iniţial ale cărui coordonate sunt cunoscute;
- concomitent se memorează coordonatele tuturor pixelilor de pe contur;
- se memorează coordonatele a patru pixeli care definesc dreptunghiul circumscris regiunii definită de contur;
- se parcurge zona din imagine definită de cei patru pixeli şi se etichetează toţi pixelii înscrişi în contur;

BUPT

- se verifică dacă s-au etichetat toţi pixeli prin compararea ariei cu numărul pixelilor găsiţi în interiorul conturului;
- se încheie algoritmul prin returnarea ariei regiuni refăcute.

Algoritmul de refacere finală a imaginii realizează de fapt refacerea regiunilor ţinând cont de vectorii de mişcare prin folosirea directă a regiunilor din imaginea precedentă. Algoritmul introduce unele erori la refacere prin înlocuirea regiunilor noi cu modelul de aproximare al texturii. Aceste erori nu pot fi eliminate din cauza faptului că modelul de aproximare al texturii reprezintă singurul indiciu care se transmite referitor la conţinutul regiunilor noi.

Rezultatele obţinute cu ajutorul algoritmului propus sunt prezentate în capitolul următor.

5.4 Concluzii

Estimarea mişcării are un rol important în obţinerea unor performanţe înalte la codarea video. În domeniul metodelor de codare din generaţia a II-a este clară necesitatea dezvoltării de noi tehnici de estimare a mişcării care să se bazeze pe o reprezentare semantică a scenei.

Acest capitol face o sinteză a tehnicilor de estimare a mişcării în contexul schemelor de codare din generaţia a doua şi prezintă principalele direcţii dezvoltate în literatura de specialitate în abordarea semantică a estimării mişcarii:

- estimarea pe baza modelului mişcării,
- estimare mişcarii bazată pe segmentare spaţio-temporală,
- estimarea pe baza căutării regiunilor (bazată pe segmentarea spaţio-temporală sau a algoritmilor de filtrare Kalman).

De asemenea sunt prezentate caracteristicile principale ale algoritmilor propuşi. În continuare este propus, de către autorul tezei de doctorat, un algoritm de estimare al mişcării bazat pe segmentare şi căutare. Algoritmul se bazează pe segmentarea statică obţinută prin aplicarea algoritmului propus de autor în capitolul 4 şi care realizează partiţionarea imaginii în regiuni de formă oarecare. Astfel estimarea mişcării se realizează pentru regiuni ce corespund obiectelor din scena reală, ceea ce constituie un avantaj din punct de vedere al funcţionalităţii şi al erorilor subiective.

Algoritmul realizează o determinare a corepondenţelor între regiunile din cadre succesive. Pe baza caracteristicilor alese pentru caracterizarea regiunii şi anume: media luminanţei pixelilor, aria, coordonatele centrului de greutate, momentele de inerţie de ordinul II şi caracteristica de formă, sunt determinaţi adaptiv vectorii de mişcare. Adaptivitatea este echivalentă în acest caz cu atribuirea fiecărei regiuni a vectorului de mişcare ce descrie cu cele mai mici erori evoluţia regiunii. Altfel spus se atribuie cel mai simplu model de mişcare, adică conţine numai acei parametri ai mişcării care sunt suficienţi pentru descrierea

BUPT

acesteia. Acest lucru este posibil dacă se utilizează presupunerea că regiunile din imagine reprezintă corpuri rigide (nedeformabile).

Se prezintă modul de calcul al parametrilor ce caracterizează regiunile şi se explică adaptarea locală a modelului vectorilor de mişcre cu scopul reducerii cantităţii de informaţie necesară a fi transmisă.

Întrucât nu este posibilă o corespondenţă a tuturor regiunilor este necesară determinarea codului de contur pentru regiunile noi, motiv pentru care este prezentat algoritmul utilizat pentru codarea conturului.

Corespondenţa între regiunile din carele succesive se determină prin crearea unei liste de posibile corespondenţe din care este aleasă cea care determină eroarea minimă şi vectorii de mişcare estimaţi se încadrează în anumiţi parametri (translaţie de maximum 16 pixeli, rotaţie de maximum 30° şi mărire/micşorare de maximum 50%).

Pentru determinarea mişcării de rotaţie autorul propune o metodă nouă, care se bazează pe determinarea corespondenţei dintre punctele caracteristice ale regiunii obţinute prin intersecţia conturului cu un cerc. Pe baza acestor puncte se determină unghiul de rotaţie al regiunii prin minimizarea unui criteriu de eroare determinat de corespondenţa unghiurilor caracteristice obţinute. Precizia de determinare a rotaţiei este de 1° şi depinde de performanţele algoritmului de segmentare. Avantajul acestui algoritm îl constituie faptul că nu este necesară cunoaşterea unor puncte caracteristice a regiunii pentru determinarea mişcării plane.

Codarea se bazează în principal pe caracteristicile regiunilor din imaginea anterioară. Pentru regiunile pentru care a fost stabilită corespondenţă cu regiuni din cadrul precedent se transmit: corespondenţa şi vectorul de mişcare, iar pentru regiunile considerate ca fiind noi se transmit: codul de contur şi coeficienţii funcţiei de aproximare a informaţiei de textură.

În final este prezentat algoritmul de refacere al imaginii care se bazează pe existenţa imaginii precedente.

BUPT

Capitolul 6

REZULTATE EXPERIMENTALE

6.1 Introducere

În acest capitol se prezintă rezultatele obținute prin aplicarea algoritmului propus pentru estimarea mişcării şi codarea secvențelor de imagini pe baza segmentării statice în regiuni. Imaginile utilizate pentru evaluarea algoritmului sunt imagini standard: "tenis", "claire", "salesman" sau imagini de test sintetizate. De asemenea este utilizată imaginea "missa", care prezintă şi un fundal cu regiuni statice.

Algoritmul propus realizează estimarea mişcării bazându-se pe proprietățile sistemului vizual uman:

- informația referitoare la contur şi muchii este apreciată foarte mult de sistemul vizual uman, fiind responsabila principală în ceea ce priveşte percepția;
- textura are o importanță diminuată, fiind asociată cu informația suplimentară, influențând percepția numai atunci când este asociată cu informația de contur, putând fi transmisă, de cele mai multe ori, doar cu ajutorul unui bit.

Algoritmul propus pentru estimarea mişcării se bazează pe utilizarea reprezentării semantice a imaginii. Pentru determinarea regiunilor (obiectelor) din imagine se poate utiliza o segmentare: spațială, temporală sau spatio-temporală a imaginilor. Această problemă reprezintă cea mai mare provocare în analiza secvențelor de imagini şi a fost prezentată în capitolul 4.

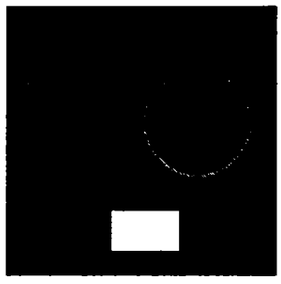
Primul pas al algoritmului corespunde unei segmentări adaptive pe baza informației de contur, cu aproximarea informației de textură printr-o funcție liniară. Algoritmul de segmentare conduce la partiționarea imaginii în regiuni de formă oarecare, cu ajutorul cărora, pe baza corespondențelor dintre parametri caracteristici, se estimează vectorii de mişcare.

Problemele şi abordările principiale ale estimării bazate pe regiuni au fost prezentate în paragraful 2.3.7 şi capitolul 5.

6.2 Imaginile utilizate

Secvențele de imagini utilizate au fost alese în aşa fel încât să prezinte o mare varietate a conținutului, de la imagini cu figuri geometrice regulate (triunghi, pătrat, cerc), la fețe umane (claire, salesaman), imagini cu mişcări semnificative (tenis) şi imagini complexe (fețe umane şi fundal – missa).

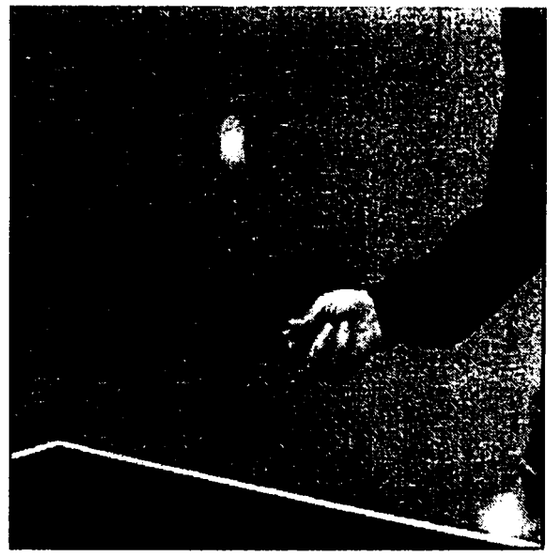
BUPT



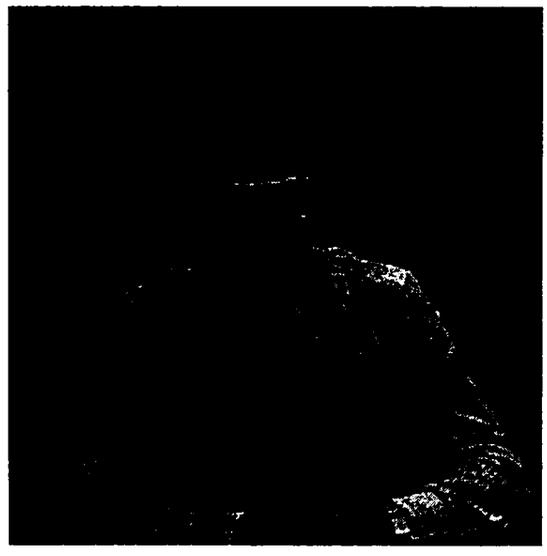
TEST (128×128)

CLAIRE (176×144)

MISSA (256×256)



TENIS (256×256)



SALESMAN (256×256)

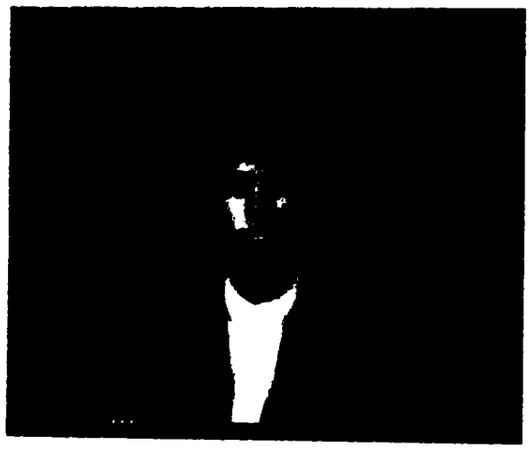
BUPT

6.3 Segmentarea imaginilor

Acest paragraf prezintă rezultatul segmentării pentru diferite cadre aparţinând secvenţelor de imagini din figura 6.1. Rezultatele sunt prezentate comparativ între doi algoritmi de segmentare (un algoritm cu prag şi cel propus de autorul tezei de doctorat) doar pentru o imagine din fiecare secvenţă din motive de spaţiu. Se observă că în cazul algoritmului propus de autor numărul de regiuni rezultate în urma segmentării este mult mai mic, iar erorile obiective obţinute au aproximativ acelaşi nivel.

Pentru imaginea Claire, am ilustrat segmentarea pentru cadrul 5 din secvenţă.

Imagine cadru 5



Segmentare

Algoritm cu prag

Număr regiuni 42
SNR= 21,23 dB
PSNR=27,38

Algoritm propus

Număr regiuni 21
SNR= 19,38 dB
PSNR=24,75

BUPT

Pentru imaginea Missa, am ilustrat segmentarea cadrului 15 din secvență.

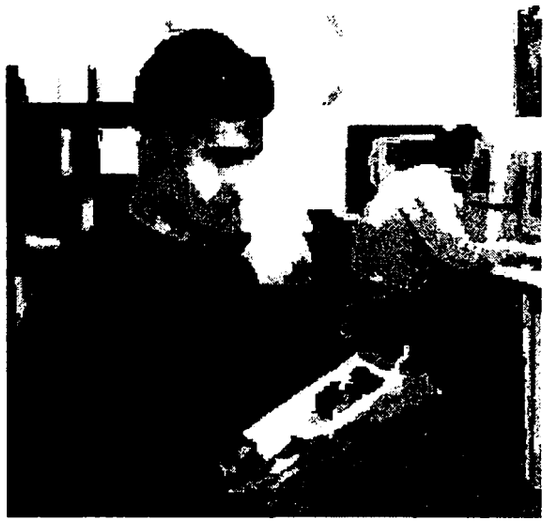
Imagine cadru 15

Segmentare

Algoritm cu prag

Număr regiuni 243
SNR= 19,43 dB
PSNR=24,38

Algoritm propus

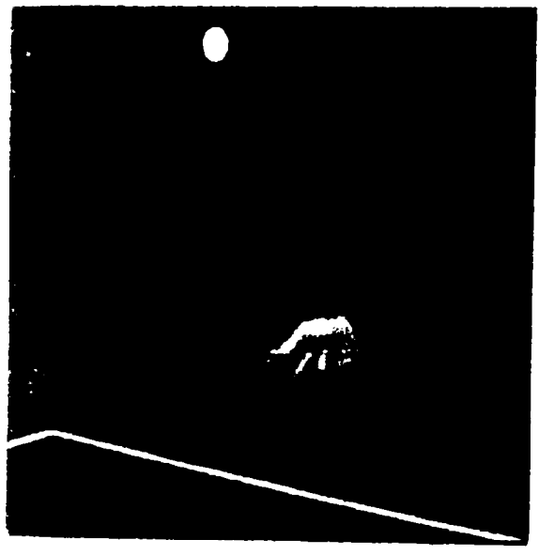


Număr regiuni 58
SNR= 17,63 dB
PSNR=21,18

BUPT

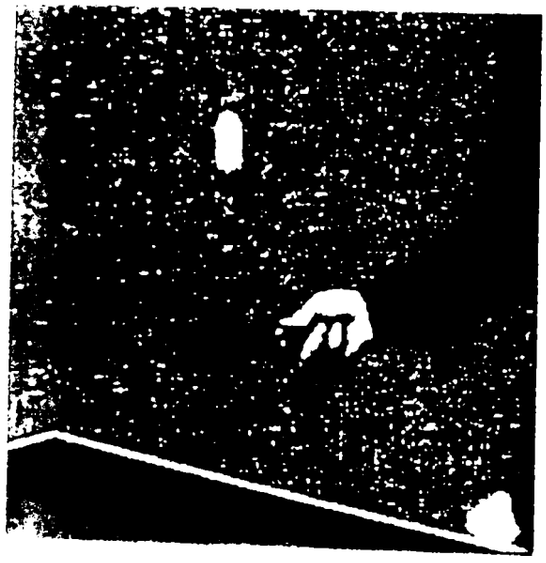
Pentru imaginea Tenis am ilustrat segmentarea cadrului 25 din secvenţă.

Imagine cadru 20



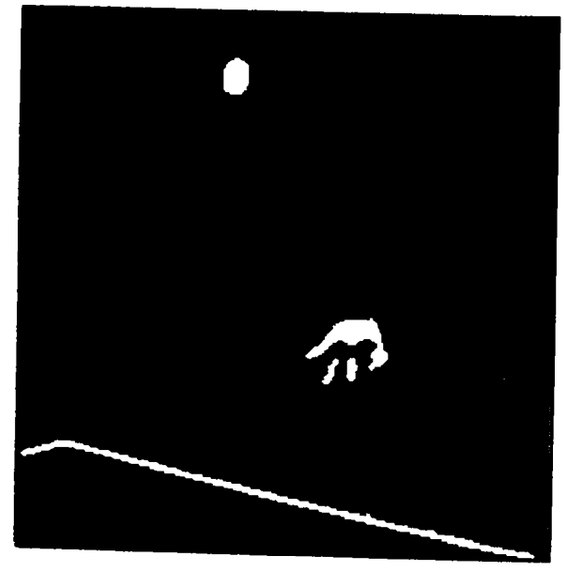
Segmentare

Algoritm cu prag



Număr regiuni 245
SNR= 19,63 dB
PSNR=26,18

Algoritm propus



Număr regiuni 9
SNR= 16,83 dB
PSNR=20,16

BUPT

Pentru imaginea Salesman am ilustrat segmentarea cadrului 5 din secvenţă.

Imagine cadru 5

Segmentare

Algoritm cu prag

Număr regiuni 185
SNR= 21,63 dB
PSNR=28,68

Algoritm propus

Număr regiuni 67
SNR= 18,63 dB
PSNR=23,18

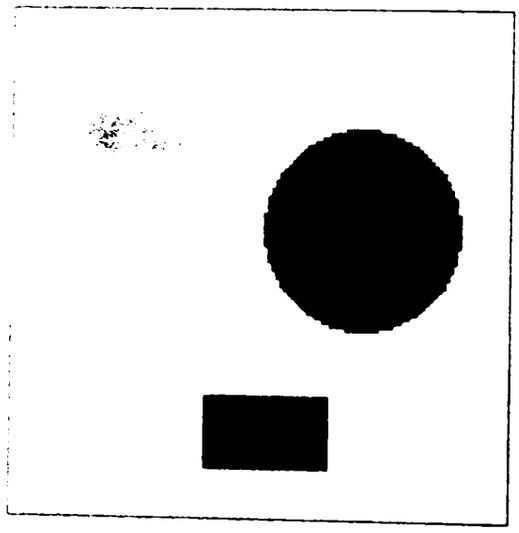
BUPT

6.4 Estimarea mișcării

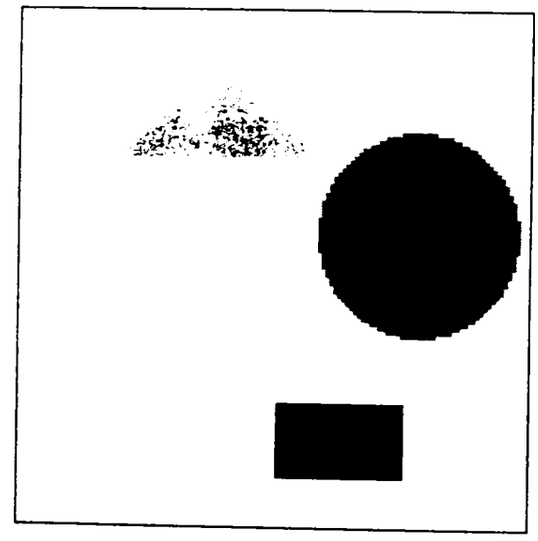
În continuare se prezintă rezultatul estimării vectorilor de mișcare pentru secvențele de imagini utilizate.

6.4.1 Rezultate ale estimării mișcării pentru imagini sintetice

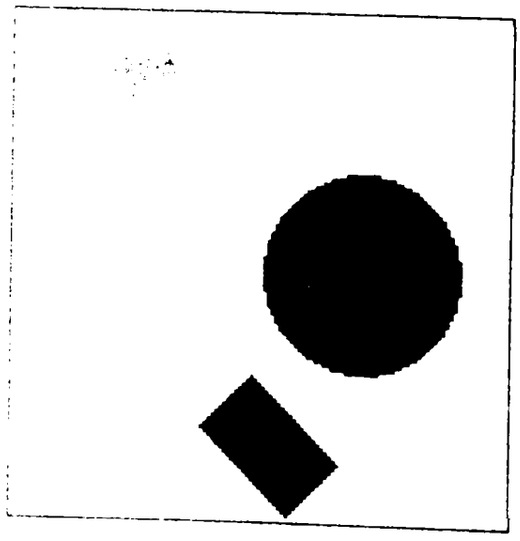
Secvența Test a fost creată pentru calculul vectorilor estimați în diferite cazuri cunoscute (rotație, translație, scalare, mișcare complexă). Astfel este posibilă compararea rezultatelor estimării cu vectorii de mișcare reali, putându-se evidenția acuratețea de determinare a vectorilor în cazul unei segmentări de precizie, precum și raportul de compresie obținut prin utilizarea algoritmului.



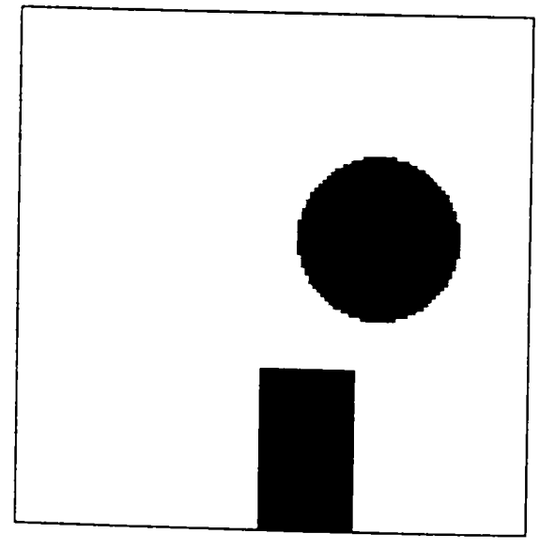
Imagine originală Test1



Imagine originală Test2

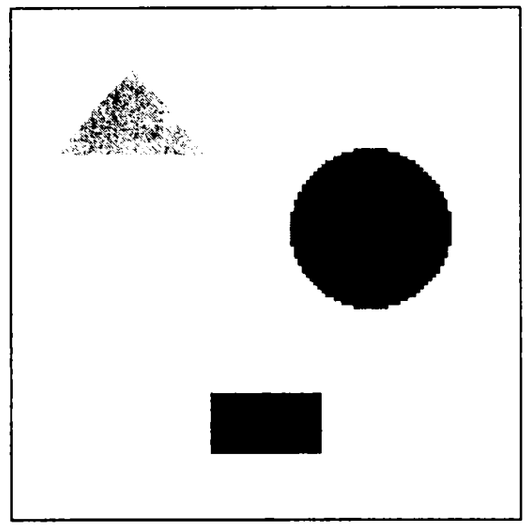


Imagine originală Test3



Imagine originală Test4

BUPT



Imagine originală Test5

a) translaţie

Parametri imagine Test1

reg	Medie	Pixeli contur	arie	Cg	Unghi	Factor forma
0	64	101	703	50, 27	0	0,2072
1	112	139	2025	100, 55	0	0
2	0	834	13048	57, 64	2,14	0,0012
3	224	97	608	80, 106	0	0,2309

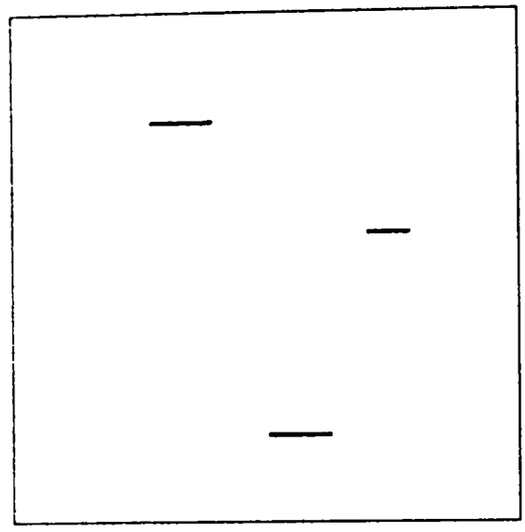
Parametri imagine Test2

reg	Medie	Pixeli contur	arie	Cg	Unghi	Factor forma
0	64	101	703	35, 27	0	0,2072
1	112	139	2025	90, 55	0	0
2	0	834	13048	57, 64	2,14	0,0002
3	224	97	608	65, 106	0	0,2309

Parametri estimare

Reg	etic	medie	Arie	Dx	dy	Zoom	Unghi	Erori estimare	
								dx	dy
0	0	64	703	15	0	1	0	0	0
1	1	112	2025	10	0	1	0	0	0
2	2	0	13048	0	0	1	0	0	0
3	3	224	608	15	0	1	0	0	0

BUPT

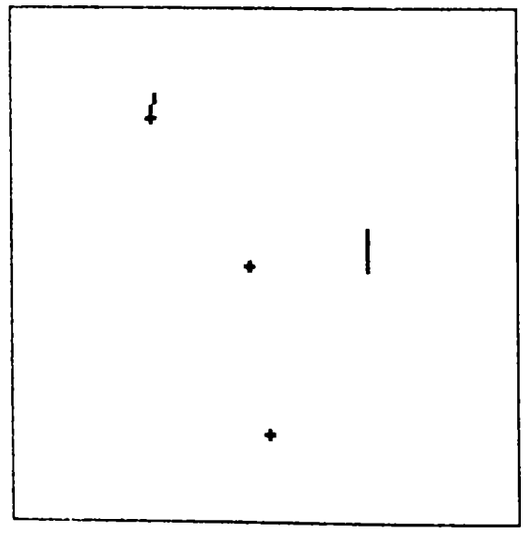


Vectori de mişcare estimaţi

b) rotaţie

Parametri imagine Test3

reg	Medie	Pixeli contur	arie	cg	Unghi	Factor forma
0	64	101	703	37, 20	44,99	0,2500
1	112	139	2025	90, 65	0	0
2	0	97	608	66, 108	44,99	0,2239
3	224	834	13048	60, 63	2,48	0,0028



Vectori de mişcare estimaţi

BUPT

Parametri estimare

Reg	etic	Medie	Arie	dx	dy	Zoom	Unghi	Erori	
								dx, dy	unghi
0	0	64	703	2	-7	1	45	0	0,01
1	1	112	2025	0	10	1	0	0	0
2	2	224	608	1	2	1	45	0	0,01
3	3	0	13048	0	-1	1	0,34	0	0,02

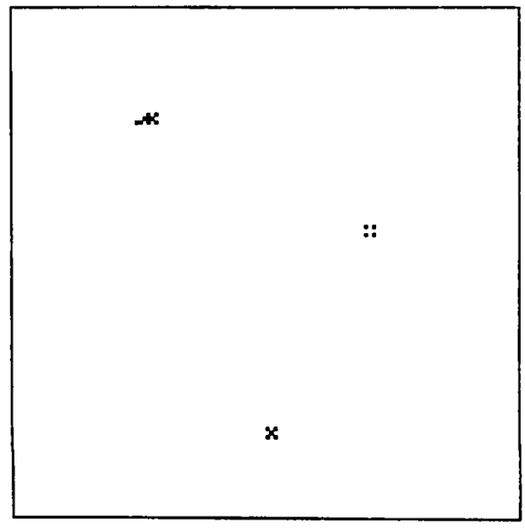
c) scalare

Parametri imagine Test5

Reg	Medie	Pixeli contur	arie	cg	Unghi	Factor forma
0	64	81	462	30, 29	0	0,1984
1	112	111	1313	90, 55	0	0
2	0	770	14189	62, 64	34,64	0,0002
3	224	81	420	63, 104	0	0,3096

Parametri estimare

Reg	Etic	medie	Arie	dx	Dy	Zoom	unghi	Erori	
								dx, dy	zoom
0	0	64	462	-5	2	0,672	0	0	0,02
1	1	112	1313	0	0	0,648	0	0	0
2	2	0	14189	0	0	0,92	0	0	0,08
3	3	224	420	-2	-2	0,69	0	0	0



Vectori de mişcare estimaţi

BUPT

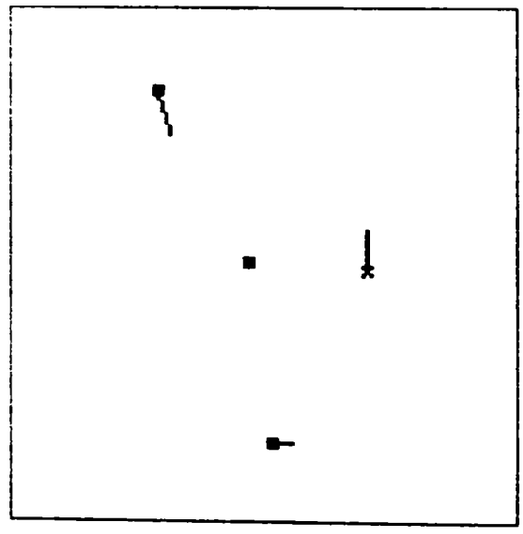
d) transformare complexă (rotaţie, scalare, translaţie) (Test3 – Test4)

Parametri imagine Test4

reg	Medie	Pixeli contur	arie	cg	Unghi	Factor forma
0	64	107	703	37, 20	44,99	0,2500
1	112	139	2025	90, 65	0	0
2	0	69	608	66, 108	44,99	0,2239
3	224	822	13048	60, 63	-37,48	0,0028

Parametri estimare

Reg	medie	Arie	dx	dy	Zoom	unghi	Erori		
							dx, dy	unghi	zoom
0	64	441	4	11	0,627	-45	0	0,01	0,007
1	112	1313	0	10	0,648	0	0	0	0
2	0	13670	1	2	1,04	-37,84	0	0,36	0,008
3	224	960	0	-1	1,58	-45,17	0	0,17	0,004



Vectori de mişcare estimaţi

Din analiza rezultatelor se poate concluziona că se realizează o estimare corectă (erorile relative sunt sub 1% pentru rotaţii şi modificări ale obiectelor) şi întrucât imaginile originale sunt simple, factorii de compresie obţinuţi sunt cuprinşi între 745,18 şi 546,46. Variaţia factorului de compresie se datorează modelului de mişcare ales pentru fiecare regiune, corespunzătoare estimării.

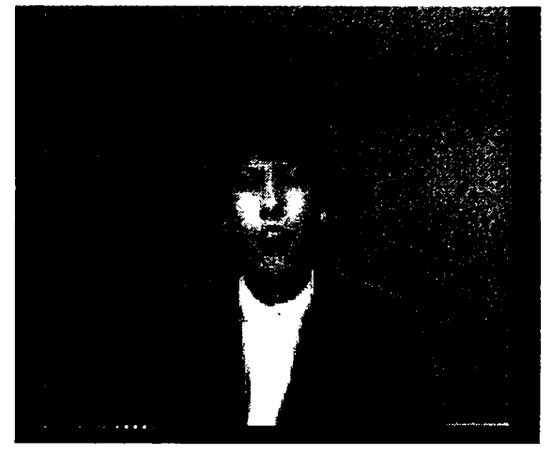
BUPT

6.4.2 Rezultate ale estimării mişcării în secvențe de imagini reale de referință

Pentru secvențele de imagini utilizate (Claire, Tenis, Missa, Salesman) se prezintă în cele ce urmează rezultatele algoritmului de estimare a mişcării propus. Se evidențiază informațiile utilizate la codarea, respectiv refacerea imaginilor. De asemenea se prezintă performanțele obținute de algoritmul propus comparativ cu performanțele algoritmului propus de Dufaux şi Mosheini [DUF 96].

Cadru 5

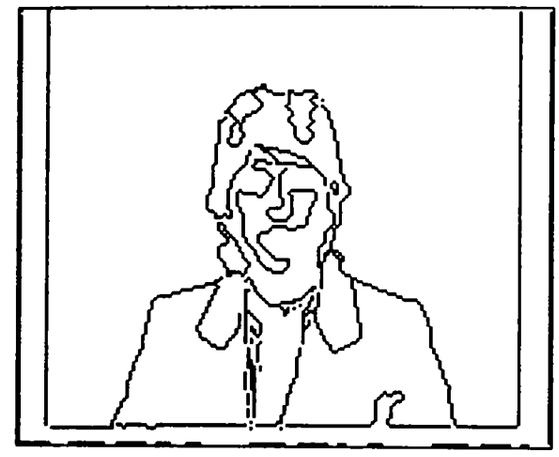
Cadru6



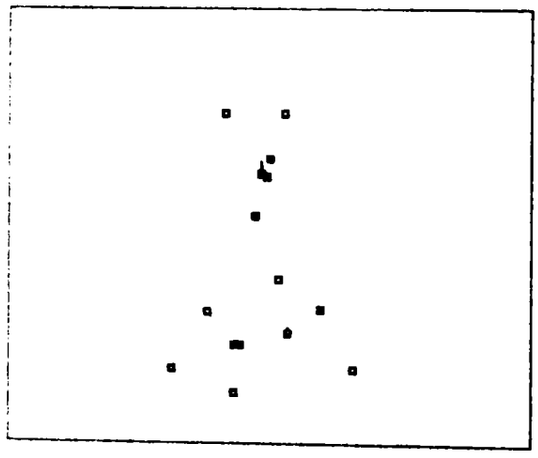
Imagini originale

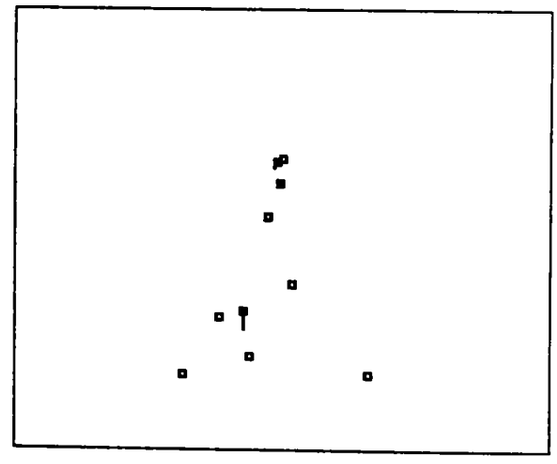
Imaginile de etichete

BUPT



Imagini ce conţin contururile regiunilor





Imaginea vectorilor de mişcare faţă de cadrele anterioare

Parametrii estimaţi între cadrele 5 şi 6.

Reg	Etic	Medie	Arie	Centru de greutate		dx	dy	zoom	Unghi	Reg_cor
0	0	39	137	9	69	0	0	1	0	0
1	1	77	134	93	34	0	0	0,95	3,83	1
2	2	80	121	73	34	0	0	0,96	3,99	2
3	3	80	67	88	48	0	-1	1,06	1,45	3
4	4	17	11	104	59	0	0	1	0	5
5	5	175	436	83	67	0	-1	1,01	1,27	6

BUPT

6	6	35	104	70	80	0	0	1	0	2000
7	7	48	863	86	49	1	-5	0,81	33,14	11
8	8	126	15395	87	56	0	1	1,06	0	7
9	9	167	25	75	98	0	0	1	0	2000
10	10	218	14	89	98	0	0	1	0	2000
11	11	10	331	105	98	0	-1	1	-0,14	10
12	12	192	66	77	13	1	2	1,1	3,93	9
13	13	221	24	96	97	0	0	1	0	2000
14	14	169	18	95	104	1	-3	0,69	5,2	14
15	15	12	262	67	100	0	0	0,99	0,005	12
16	16	220	34	78	112	0	1	0,94	2,83	13
17	17	117	24	76	127	0	0	0,96	-0,22	15
18	18	248	525	85	115	0	0	1	0	2000
19	19	36	1374	55	119	0	0	0,99	-0,17	17
20	20	36	1674	116	119	0	0	1	0,38	18
21	21	10	3705	91	89	0	0	0,99	-0,01	19

Parametrii estimaţi între cadrele 6 şi 7 sunt

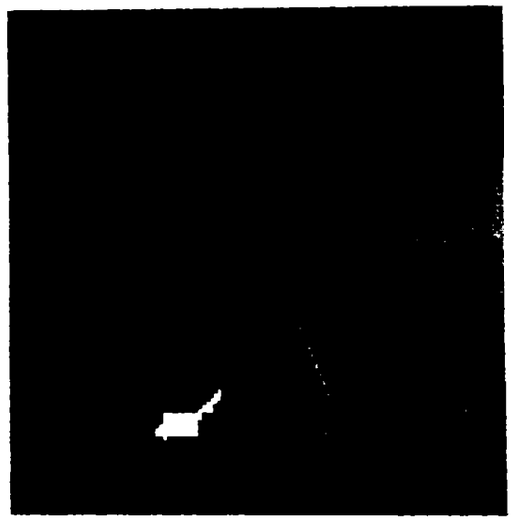
Reg	Etic	Medie	Arie	Centru de greutate		dx	dy	zoom	Unghi	Reg_cor
0	0	39	137	9	69	0	0	1	0	0
1	1	82	98	72	34	0	0	1	0	2000
2	2	78	50	88	48	0	0	0,74	-0,38	3
3	3	174	460	83	67	0	0	1,05	-5,72	5
4	4	49	1071	84	52	-2	3	1,24	-21,67	7
5	5	125	14624	87	55	0	-1	0,95	-0,02	8
6	6	130	811	87	73	0	0	1	0	2000
7	7	182	11	98	92	0	0	1	0	2000
8	8	121	12	75	104	0	6	0,48	-1,83	9
9	9	190	63	77	113	0	0	0,95	0,18	12
10	10	17	341	104	99	0	0	1	0	2000
11	11	11	264	67	100	0	-1	1	-0,12	15
12	12	117	24	76	127	0	0	1	0	17
13	13	244	614	85	113	0	0	1	0	2000
14	14	36	1379	55	119	0	0	1	-0,92	19
15	15	35	16772	116	119	0	0	1	-0,03	20
16	16	10	3708	91	89	0	0	1	0,023	21

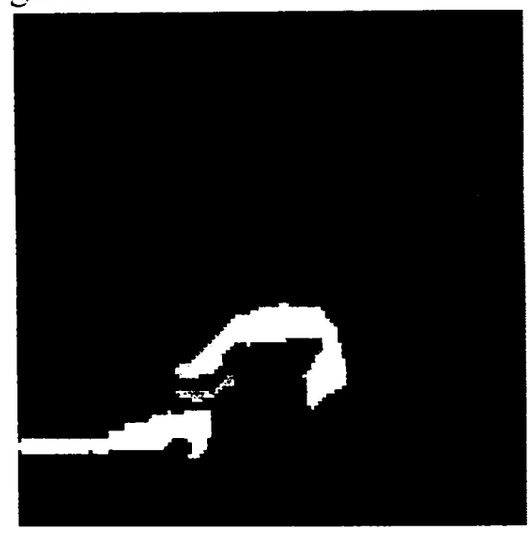
BUPT

Detaliu Tenis 1

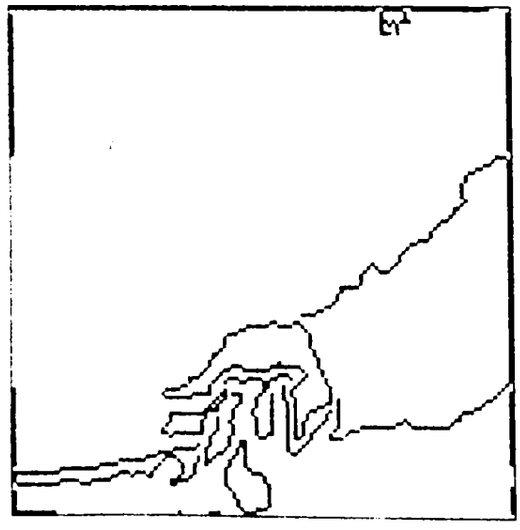
Detaliu Tenis2

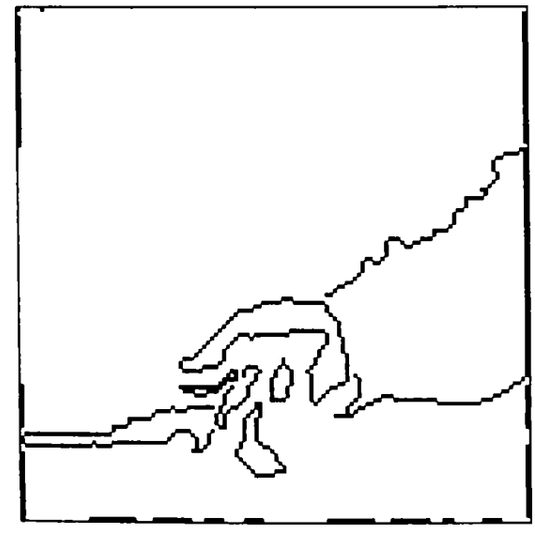
Imagini originale





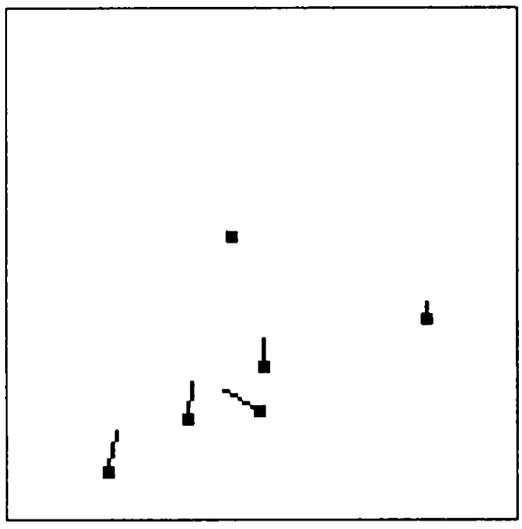
Imaginile de etichete

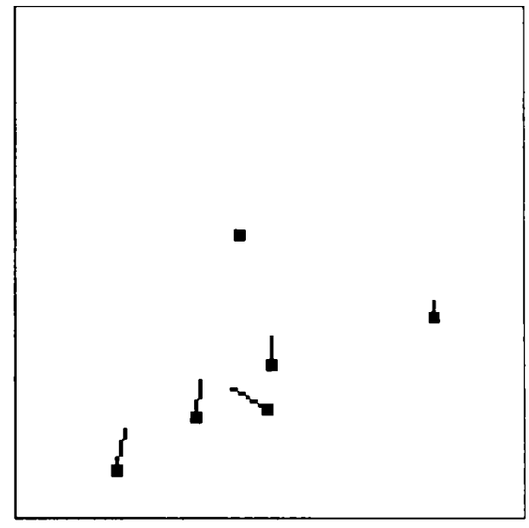




Imagini ce conțin contururile regiunilor

BUPT



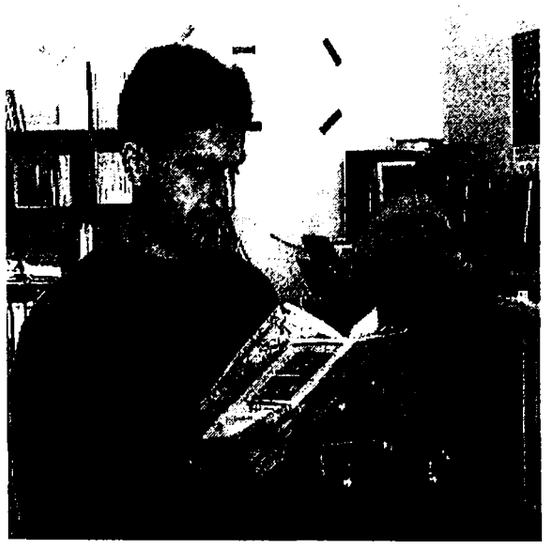


Imaginile ce conţin vectorii de mişcare faţă de cadrele anterioare

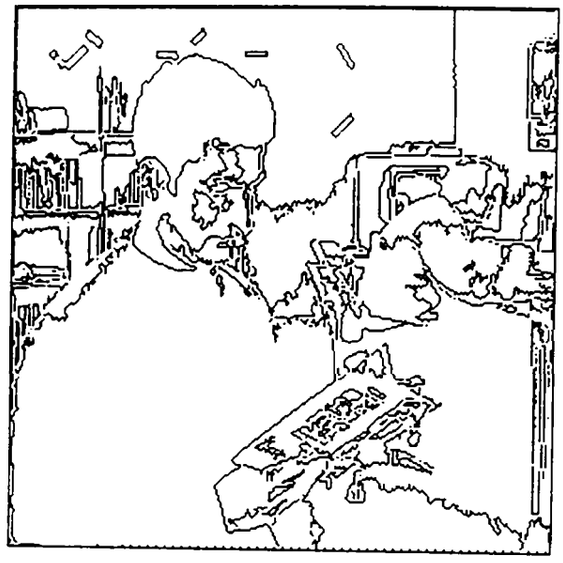
Parametri estimaţi între cadrele 2 şi 3 în cazul secvenţei Tenis sunt:

Reg	Etic	Medie	Arie	Centru de greutate		dx	dy	zoom	Unghi	Reg_cor
0	0	190	419	65	82	1	-8	0,92	0,95	1
1	1	196	46	65	92	0	0	1	0	2000
2	2	73	34	47	93	2	-10	0,43	11,92	4
3	3	187	70	54	95	-9	-6	1,14	36,29	3
4	4	57	2109	106	73	1	-5	0,91	1,85	5
5	5	62	288	28	105	3	-11	1,29	-0,35	8
6	6	63	114	59	109	0	0	1	0	2000
7	7	131	13304	57	59	1	2	1,03	-14,19	9

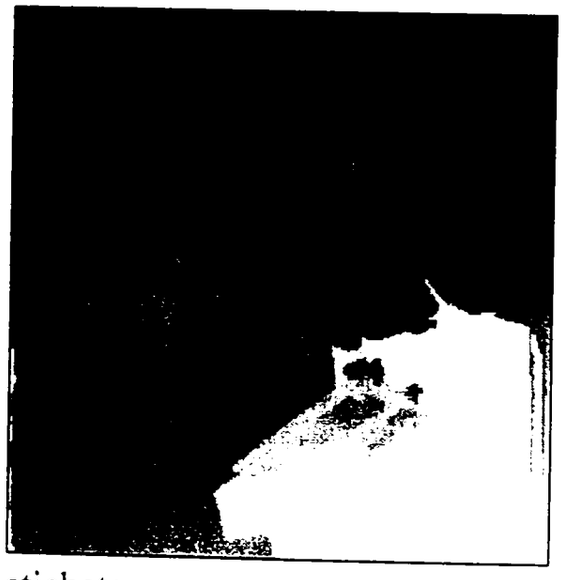
Cadru 20 Imagini originale Cadru 21



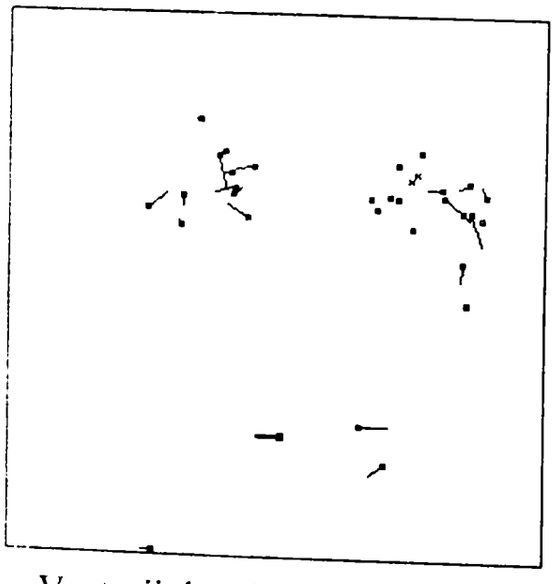
BUPT



Imagini ce conțin contururile regiunilor

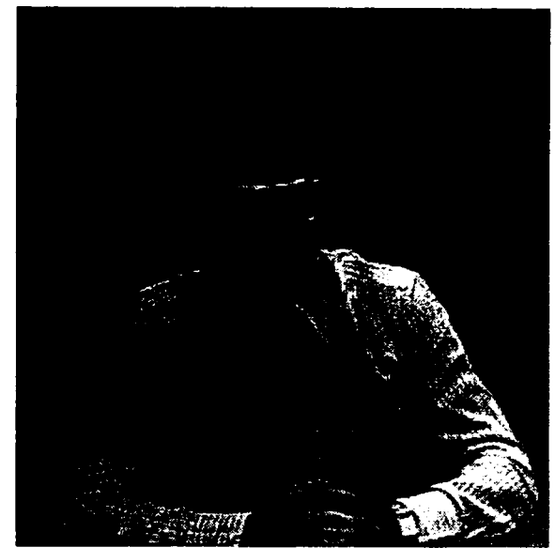


Imaginile de etichete



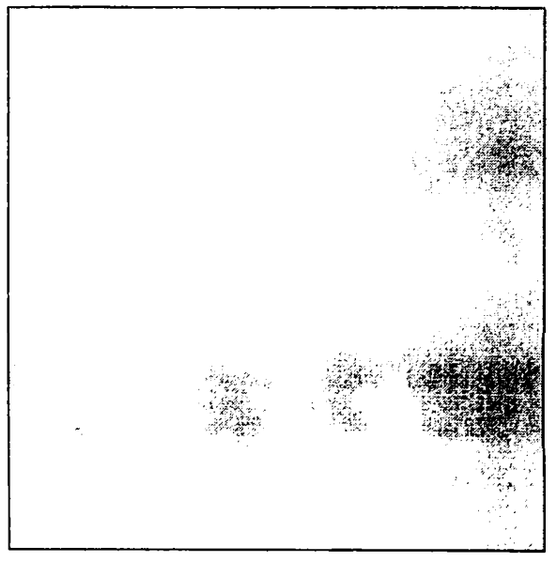
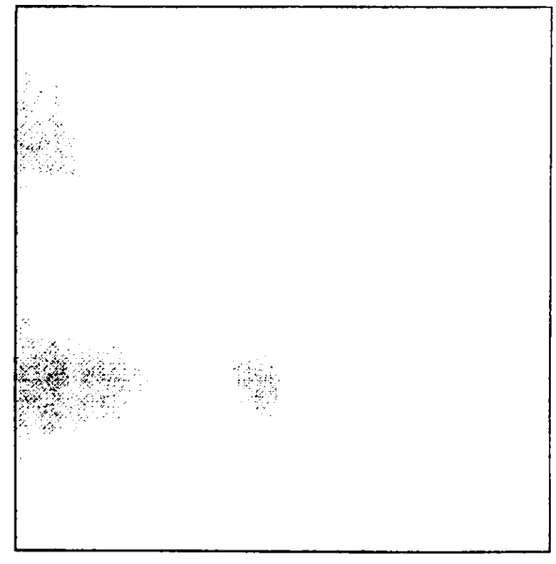
Vectorii de mișcare estimați

BUPT



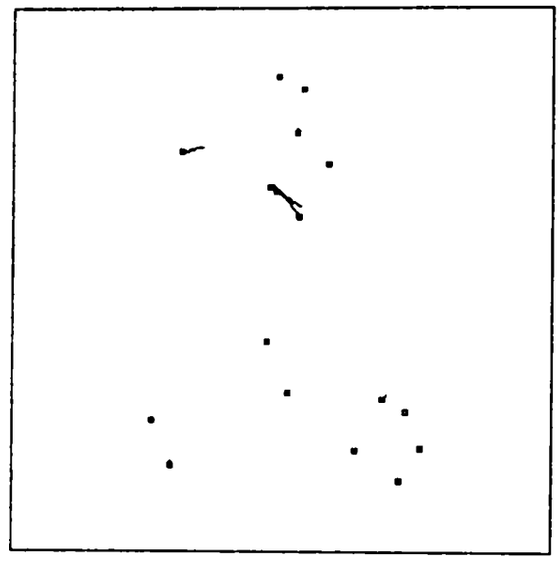
Imagini originale

Imagini ce conțin contururile regiunilor

Imagini de etichete ale regiunilor

BUPT



Vectorii de mişcare estimaţi

Întrucât numărul de regiuni pentru imaginile din secvenţele MISSA şi Salesman este destul de mare, 58 respectiv 67, nu se mai prezintă tabelele cu rezultatele obţinute pentru estimare, dar sunt prezentate graficele cu raportul semnal/zgomot, respectiv compresia obţinută în urma codării imaginilor segmentate.

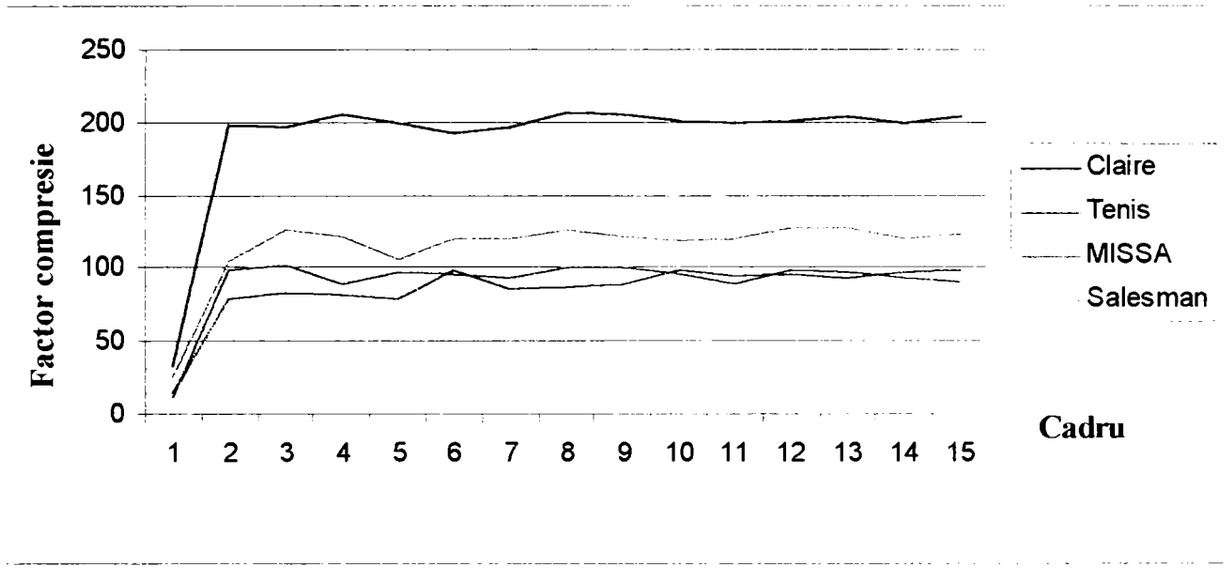
Tabelul 6.1 Factorul de compresie obţinut pentru primele 7 cadre din secvenţă

Secvenţa/ Cadru	1	2	3	4	5	6	7
Claire	32,46	129,45	126,35	145,32	132,76	121,56	127,45
Tenis	13,92	79,46	83,89	82,35	78,43	98,56	86,25
MISSA	10,82	99,37	102,66	88,45	97,35	96,78	93,26
Salesman	26,46	104,83	126,53	121,79	105,67	120,48	120,32

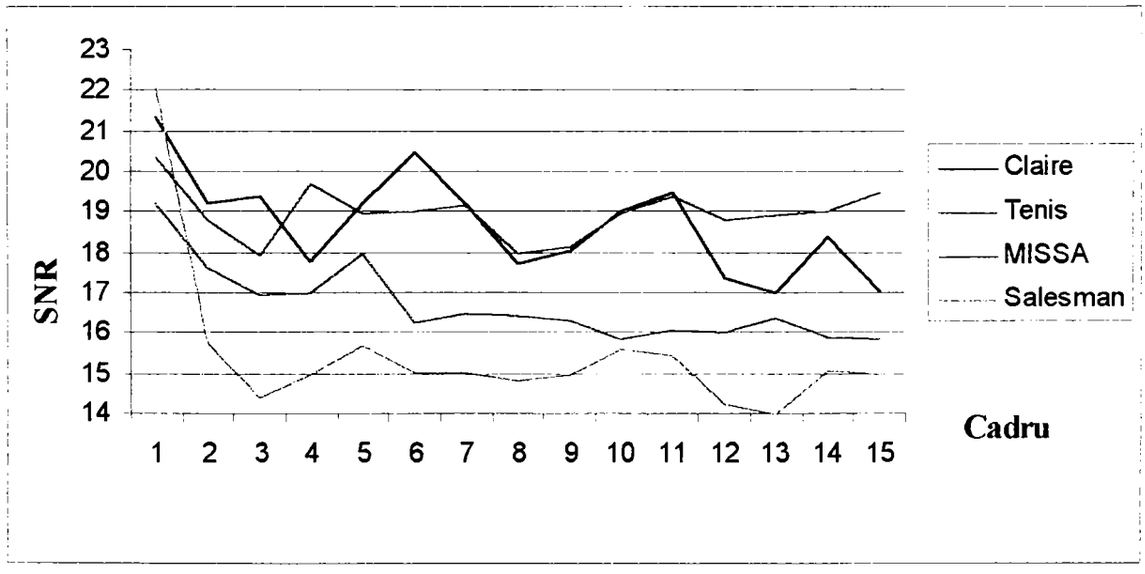
Tabelul 6.2 Raportul semnal/zgomot obţinut la reconstrucţia imaginilor din secvenţă

Secvenţa/ Cadru	1	2	3	4	5	6	7	8
Claire	19,25	19,38	19,03	19,78	19,21	19,47	19,81	19,01
Tenis	17,63	17,96	17,67	16,73	17,08	17,93	17,62	17,28
MISSA	16,83	17,02	16,54	16,36	16,49	16,98	16,88	16,74
Salesman	18,63	18,71	18,36	18,54	18,36	18,44	18,21	18,62

BUPT



Reprezentarea grafică a tabelului 6.1



Reprezentarea grafică a tabelului 6.2

6.5 Imaginile de test reconstruite

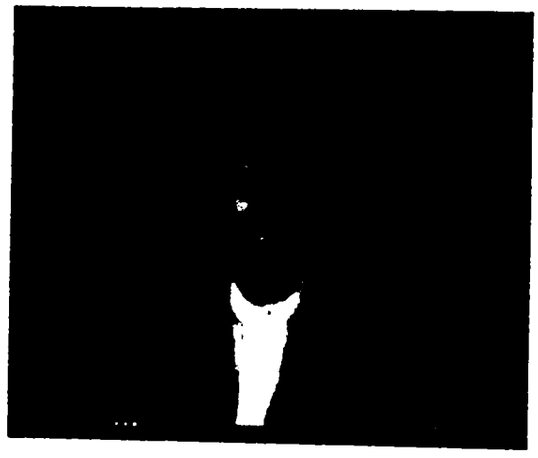
În cele ce urmează se prezintă imaginile de test decodate pentru a facilita aprecierea subiectivă a calității, pe baza criteriului scorului mediu al opiniilor, Anexa 2. Se observă că imaginile reconstruite, din compararea cu imaginile originale, permit distingerea trăsăturilor esențiale ale personajelor prezente în imagini. Ținând cont de factorul de compresie relativ mare obținut, 192,17:1, putem considera că algoritmul propus satisface necesitățile pentru o transmisie la

BUPT

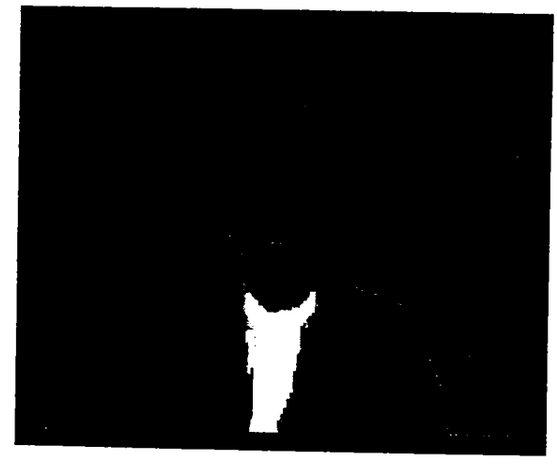
rate foarte scăzute de biţi. Pentru imaginea Claire care reprezintă formatul QCIF, rata de transmisie obţinută este 0,0408 biţi/pixel, ceea ce este echivalent cu o rată de transmisie de 6,204 kbiţi/s pentru o transmisie de 6 cadre/s. Rezultatele obţinute ne îndreptăţesc să considerăm că se pot obţine rezultate acceptabile chiar şi la transmiterea imaginilor color. Pentru celelalte imagini, factorul de compresie mediu obţinut pentru primele 15 cadre din secvenţă este:

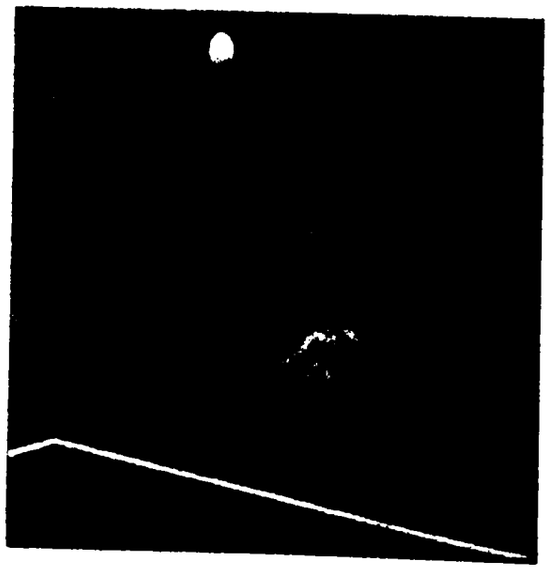
- Tenis: 85,39
- MISSA: 90,56
- Salesman: 114,29.

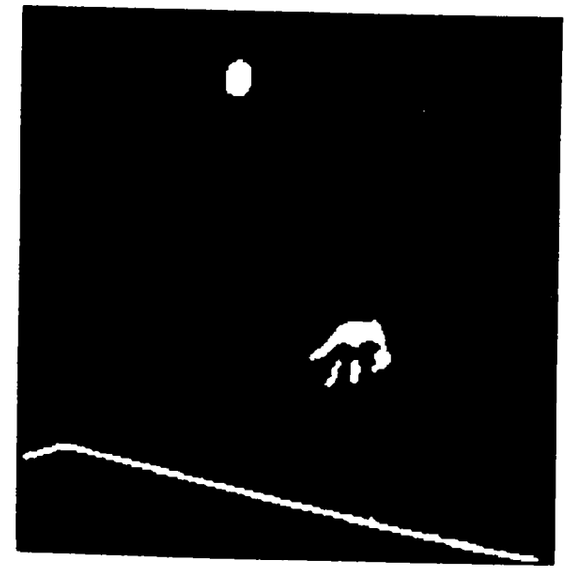
Imagine originală



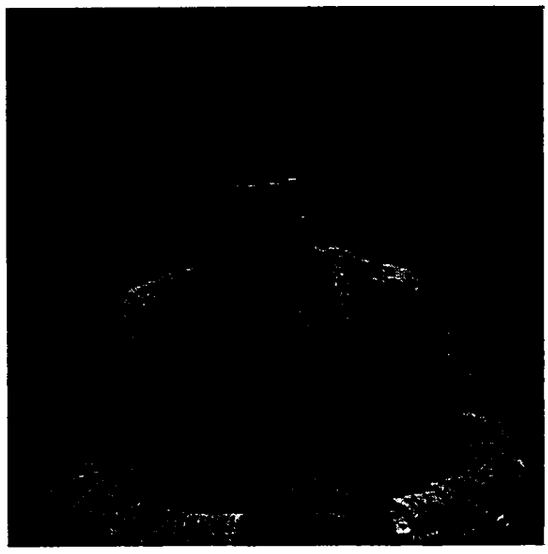
Imagine reconstruită







BUPT



6.6 Concluzii

În concluzie, se poate spune spune că metodele de estimare a mişcării bazate pe regiuni, cu încorporarea proprietăţilor sistemului vizual uman sunt potrivite pentru tehnicile de codare la rate foarte scăzute, care necesită rate de compresie foarte mari. Rata de transmisie obţinută, de 6,204 kbiţi/s, pentru o transmisie cu 6 cadre pe secundă în cazul formatului QCIF este comparabilă cu cele mai bune performanţe raportate în literatură [KON 95], [MOU 97]. Rezultatele obţinute au o calitate comparabilă. Astfel din punct de vedere subiectiv nu se observă o degradare a calităţii imaginii obţinute, iar din punct de vedere calitativ (PSNR) rezultatele sunt mai bune cu 0,12-0,15 dB la rate de transmisie mai scăzute cu 0,025- 0,036 kbiţi/s. Faptul că raportul semnal zgomot se situează în domeniul 13-

BUPT

19 dB nu constituie un dezavantaj, întrucât în cazul transmiterii imaginilor prin reţelele telefonice este foarte importantă transmiterea aspectului emoţional al comunicaţiei.

Rezultatele obţinute cu o complexitate moderată, comparabile cu cele raportate în literatura de specialitate, chiar mai bune, mă încurajează să cred că acest domeniu de cercetare, relativ nou în compresia imaginilor, este unul foarte promiţător.

Consider că rezultatele obţinute sunt viabile şi pot constitui baza unor cercetări în domeniul codării secvenţelor video la rate foarte scăzute. Una din principalele tendinţe o constituie posibila combinaţie a metodei cu tehnicile multirezoluţie pentru segmentare, în vederea eficientizării calculului şi a stabilităţii etapei de segmentare.

BUPT

Capitolul 7

CONCLUZII

7.1 Contribuţii personale

Autorul prezintă rezultatele obţinute în segmentarea imaginilor precum şi a estimării vectorilor de mişcare comparativ cu metodele clasice propuse, din punct de vedere al ratei de transmisie bit/pixel cât şi din punct de vedere al criteriilor obiective de eroare (raportul semnal/zgomot). S-au obţinut valori cuprinse între 0,0408 biţi/pixel (raport de compresie de 192,17:1) şi 0,093 biţi/pixel (raport de compresie de 85,39:1), în condiţiile unei calităţi bune a imaginii reconstruite. Întrucât, aşa cum s-a precizat, întregul sistem se bazează pe considerarea caracteristicilor sistemului vizual uman se constată o calitate a reconstrucţiei similară în condiţiile obţinerii unui factor de compresie cu aproximativ 1% mai bun .

Pe lângă toate aceste avantaje mai pot fi precizate şi câteva dezavantaje şi anume: utilizarea unei segmentări statice, dependenţa factorului de compresie de calitatea obţinută în etapa de segmentare a imaginii.

În lucrarea de faţă pot fi evidenţiate un număr de 10 contribuţii personale. Aceste contribuţii având un caracter de noutate în domeniul estimării mişcării pentru obţinerea unor performanţe înalte la codarea secvenţelor de imagini, sunt prezentate, pe scurt, în cele ce urmează:

- Propunerea unei analogii între nivelurile de cunoaştere şi complexitate rezultate din analiza secvenţelor de imagini şi nivelurile utilizate în procesul de codare al acestora, ceea ce conduce la ideea utilizării informaţiilor semantice din imagine pentru obţinerea de rezultate superioare în procesul de codare.
- Un nou algoritm rapid de corespondenţă a blocurilor propus în lucrarea "A New Block Matching Algorithm for Motion Estimation" [ALE 98a] care prezintă rezultate apropiate de cele ale algoritmului cu căutare completă (Full Search) şi mai bune decât algoritmii rapizi propuşi în literatură [LI 94], [KOG 81]. De asemenea algoritmul prezintă o creştere a preciziei de estimare în comparaţie cu ceilalţi algoritmi de estimare rapidă.
- Abordarea sistematică a principalelor metode de codare cunoscute în codarea secvenţelor de imagini şi prezentarea avantajelor utilizării metodelor bazate pe reprezentarea semantică a imaginilor;

BUPT

- Propunerea unei scheme generale pentru un sistem de compresie care structurează un sistem de codare în trei niveluri şi anume: nivelul transformării spaţiului imaginii, care corespunde unui nivel de codare fără pierderi, nivelul cuantizării în spaţiul transformat, care corespunde unei codări cu pierderi a informaţiei şi cel de-al treilea nivel, care corespunde adaptării fluxului de biţi obţinut prin codare la canalul de transmisie.
- Propunerea unei clasificări ierarhice a tehnicilor de codare având la bază schema generală, a sistemului de compresie, propusă.
- Pe baza analizei schemelor de codare existente şi a algoritmului de estimare al mişcării se propune o schemă de codare bazată pe reprezentarea semantică a imaginilor.
- Un nou algoritm de segmentare adaptivă a imaginilor bazat pe o tehnică cunoscută de segmentare, şi anume algoritmii de tip divizare contopire (split and merge), dar care se bazează pe proprietăţile sistemului vizual uman:
  - informaţia referitoare la contur şi muchii este apreciată foarte mult de sistemul vizual uman, fiind responsabila principală în ceea ce priveşte percepţia;
  - textura are o importanţă diminuată, fiind asociată cu informaţia suplimentară, influenţând percepţia numai atunci când este asociată cu informaţia de contur.

  Astfel algoritmul prezentat în lucrarea "Adaptive Split and Merge Algorithm for Region Based Motion Estimation" [ALE 98b] realizează divizarea imaginii în regiuni de formă pătrată ţinând cont de informaţia de contur obţinută prin aplicarea operatorilor Sobel imaginii originale.
- Utilizarea modelului liniar pentru aproximarea regiunii şi a criterului de minimizare a erorii pătratice pentru testarea omogenităţii în cadrul etapei de divizare a imaginii. Pentru implementare a fost definită o matrice caracteristică prezintă o relaţie de recurenţă. Matricea caracteristică a fost determinată pentru cazul modelului ales putând fi modificată pentru modele mai complexe. De asemenea pentru calculul imaginii estimate se utilizează ca şi coordonate centrul de greutate al regiunii pătrate şi nu primul element (elementul din colţul din stânga sus) pentru micşorarea erorilor de aproximare.
- Propunerea unui algoritm de contopire în trei paşi care să ţină cont de prezenţa muchiilor în imagine şi care realizează scăderea numărului de regiuni prin eliminarea în etapa finală a regiunilor având aria mai mică decât o valoare impusă.
- Algoritm de estimare a mişcării pentru codarea bazată pe regiuni a secvenţelor de imagini. Algoritmul original se bazează pe segmentarea statică

BUPT

a imaginilor şi căutarea corespondenţelor între cadrele succesive din imagine. Pe baza corespondenţelor se realizează o atribuire adaptivă a modelului de mişcare fiecărei regiuni, model ce este apoi utilizat la codarea secvenţei. Pentru regiunile ce nu au o corespondenţă în imaginea anterioară se utilizează o metodă de codare cu pierderi pe baza conturului, a centrului de greutate şi a modelului de aproximare a regiunii determinat în etapa de segmentare. Atribuirea adaptivă a modelului de mişcare pentru fiecare regiune se bazează pe presupunerea că scena cuprinde numai obiecte rigide (corpuri nedeformabile) şi există limite maxime ale mişcării între imaginile succesive [ALE 98c], [ALE 98b], [ALE 98d], [ALE 98e].

- Selecţia setului de trăsături caracteristice pentru determinarea transformărilor (rotaţie, translaţie, scalare) din imagine [ALE 98c], [ALE 98e]. Astfel pe baza corespondenţei dintre cadrul curent şi cel precedent translaţia reprezintă deplasarea determinată între centrele de greutate ale regiunilor, rotaţia se determină pe baza momentelor de inerţie sau a unor puncte caracteristice ia scalarea pe baza ariei obiectului.
- Metodă nouă de determinare a rotaţiei şi translaţiei unei regiuni. Aceasta se bazează pe determinarea recursivă a centrului de greutate al regiunii şi utilizarea unei metode de determinare a unghiului de rotaţie pe baza unor puncte aleatoare aparţinând conturului regiunii. În acest scop autorul propune o teoremă ce se bazează pe faptul că în imagine există numai corpuri rigide. Metoda se bazează pe identificarea regiunii şi utilizarea unor puncte determinate de intersecţia unui cerc având raza cuprinsă între distanţa minimă şi maximă a centrului de greutate faţă de conturul regiunii. În acest fel se obţin puncte ce caracterizează regiunea, pe baza cărora prin minimizarea unui criteriu de eroare introdus de autor se determină unghiul de rotaţie al regiunii.

7.2 Posibilităţi de dezvoltare ulterioară

Rezultatele obţinute în cadrul acestei teze îl îndreptăţesc pe autor să considere metodele de codare bazate pe considerarea caracteristicilor sistemului vizual uman, ca având un potenţiel ridicat în ceea ce priveşte codarea la rate foarte scăzute. Din acest motiv prezenta teză este considerată doar o etapă în preocupările mai generale ale autorului în domeniul prelucrărilor de imagini.

Prin cercetările întreprinse de autor în cadrul tezei de doctorat se conturează următoarele direcţii de studiu şi dezvoltare, şi anume:

- îmbunătăţirea calităţii segmentării prin realizarea unei segmentări spaţio-temporale precum şi prin utilizarea metodelor morfologiei matematice şi a algoritmilor ierarhici;

BUPT

- creşterea raportului de compresie fără alterarea calităţii imaginii din punct de vedere subiectiv, prin luarea în consideraţie şi a altor caracteristici ale sistemului vizual uman;
- realizarea estimării mişcării pentru secvenţe de imagini color, prin încorporarea informaţiei de crominanţă la estimarea mişcării.

Motivată de aplicaţiile imediate (video-telefonie, video-conferinţe, televiziune digitală, multimedia, etc.), codarea la rate foarte scăzute reprezintă principalul scop al celor mai multe din cercetările actuale. Întrucât, aşa cum s-a arătat în capitolul 2, tehnicile de codare bazate pe reducerea statistică a redundanţei din imagine conduc la anumite limitări în cazul ratelor scăzute de biţi, s-a trecut la tehnicile generaţiei a doua care utilizează reprezentarea semantică a imaginilor.

Necesitatea unor rapoarte de compresie tot mai mari a determinat evoluţia tehnicilor de codare spre cele ce realizează încorporarea proprietăţilor sistemului vizual uman. Din acest motiv este necesară reprezentarea imaginilor funcţie de conţinutul lor semantic. Segmentarea imaginilor tratată în capitolul 4, are din acest motiv un rol esenţial deoarece conduce la reprezentarea imaginii sub formă de primitive semantice lipsite de ambiguitate (contururi, regiuni).

În aceste condiţii, întrucât estimarea mişcării depinde într-o foarte mare măsură de calitatea segmentării, una dintre direcţiile viitoare este aceea de a îmbunătăţi calitatea acesteia. Dintre metodele de îmbunătăţire, segmentarea spaţio-temporală şi utilizarea morfologiei matematice par să fie dintre cele mai utilizate în ultima perioadă [GUN 98]. Aceasta datorită faptului că în marea majoritate a cazurilor practice apar problme datorate discontinuităţilor ce apar în secvenţele de imagini. Prin aplicarea algoritmilor de estimare-segmentare este posibilă rezolvarea problemei determinată de apariţia dicontinuităţilor.

O altă direcţie o constituie aplicarea morfologiei matematice în segmentarea mişcării, datorită faptului că metodele bazate pe morfologia matematică sunt foarte apropiate de caracteristicile percepţiei vizuale umane.

O direcţie posibilă este şi abordarea ierarhică a estimării mişcării mai ales pentru aplicaţiile cum ar fi codarea şi grafica pe calculator [KWA 98], [STR 98]. De asemenea, în condiţiile în care o cantitate de informaţie din ce în ce mai mare este disponibilă pe mediile de transmisie, este necesară posibilitatea accesării rapide şi eficiente a acesteia. De aceea, la ora actuală interesul cercetărilor este îndreptat spre realizarea unei codări eficiente şi a unei interactivităţi sporite. Standardul MPEG-4 încearcă să rezolve o parte din aceste cerinţe iar propunerea MPEG-7, a cărui denumire comercială este Interfaţă Multimedia cu Descrierea Conţinutului (Multimedia Content Description Intreface) vrea să realizeze standardizarea descriptorilor pentru diferitele tipuri de informaţii multimedia. Astfel principalele tipuri de date video prevăzute a fi traficate în propunerea MPEG-7 [PER 97] sunt:

BUPT

- Secvenţe numerice sau filme în formatele standardelor existente MPEG-1, MPEG-2, MPEG-4
- filme şi secvenţe analogice;
- imagini electronice statice;
- imagini grafice (CAD);
- modele tridimensionale (în special modele faciale);
- date asociate formatelor video;
- alte tipuri de date ce vor fi definite.

BUPT

ANEXE

A.1 Modele de mişcare

A.1.1 Echivalenţa dintre deplasare şi viteză în estimarea mişcării

Pentru a arăta echivalenţa dintre abordările din punct de vedere al vitezei instantanee şi deplasării, putem rescrie ecuaţia 2.9 sub forma:

$$\vec{r}' - \vec{r} = \begin{pmatrix} x'-x \\ y'-y \end{pmatrix} = \begin{pmatrix} \dfrac{-\Omega_x xy + \Omega_y(x^2+1) - \Omega_z y + \dfrac{(D_x - D_z x)}{Z}}{1 - \Omega_y x + \Omega_x y + \dfrac{D_z}{Z}} \\ \dfrac{-\Omega_x(y^2+1) + \Omega_y xy - \Omega_z x + \dfrac{D_y - D_z y}{Z}}{1 - \Omega_y x + \Omega_x y + \dfrac{D_z}{Z}} \end{pmatrix} \quad \textbf{(A.1)}$$

Dacă facem presupunerea că componenta Z a mişcării de translaţie este mică în raport cu distanţa dintre obiect şi cameră ($|D_z / Z| << 1$) şi că unghiul de rotaţie precum şi cel de observaţie sunt mici ($|-\Omega_y x + \Omega_x y| << 1$), dezvoltarea în serie Taylor de ordinul I a relaţiei A.1 ne conduce la o formă similară cu cea dată de relaţia 2.7. Din acest motiv cele două formulări în termeni ai vitezei instantanee şi deplasării sunt echivalente în condiţiile presupuse mai sus.

A.1.2 Modelele parametrice ale suprafeţei plane.

Pentru o suprafaţă plană, viteza instantanee în cadrul imaginii este dată de:

$$\begin{pmatrix} \dfrac{dx}{dt} \\ \dfrac{dy}{dt} \end{pmatrix} = \begin{pmatrix} a_1 + a_2 x + a_3 y + a_7 x^2 + a_8 xy \\ a_4 + a_5 x + a_6 y + a_7 xy + a_8 x^2 \end{pmatrix} \quad \textbf{(A.2)}$$

în care:

$$a_1 = \omega_y + T_x k_z \quad \textbf{(A.3)}$$

$$a_2 = T_x k_x - T_z k_z \quad \textbf{(A.4)}$$

$$a_3 = -\omega_z + T_x k_y \quad \textbf{(A.5)}$$

$$a_4 = -\omega_x + T_y k_z \quad \textbf{(A.6)}$$

BUPT

$$a_5 = \omega_z + T_y k_x \quad \textbf{(A.7)}$$

$$a_6 = T_y k_y - T_z k_z \quad \textbf{(A.8)}$$

$$a_7 = \omega_y - T_z k_x \quad \textbf{(A.9)}$$

$$a_8 = -\omega_x - T_z k_y \quad \textbf{(A.10)}$$

În mod similar se poate obţine deplasarea în imaginea plană, a proiecţiei imaginii tridimensionale:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \dfrac{a_1 + a_2 x + a_3 y}{a_7 x + a_8 y + a_9} \\ \dfrac{a_4 + a_5 x + a_6 y}{a_7 x + a_8 y + a_9} \end{pmatrix} \quad \textbf{(A.11)}$$

unde:

$$a_1 = \Omega_y + D_x k_z \quad \textbf{(A.12)}$$

$$a_2 = 1 + D_x k_x \quad \textbf{(A.13)}$$

$$a_3 = -\Omega_z + D_x k_y \quad \textbf{(A.14)}$$

$$a_4 = -\Omega_x - D_y k_z \quad \textbf{(A.15)}$$

$$a_5 = \Omega_z + D_y k_x \quad \textbf{(A.16)}$$

$$a_6 = 1 + D_y k_y \quad \textbf{(A.17)}$$

$$a_7 = -\Omega_y + D_z k_x \quad \textbf{(A.18)}$$

$$a_8 = \Omega_x + D_z k_y \quad \textbf{(A.19)}$$

$$a_9 = 1 + D_z k_z \quad \textbf{(A.20)}$$

A.2 Aprecierea calităţii imaginii

A.2.1 Aprecierea calităţii prin măsurări cantitative

Toate măsurătorile care apreciază calitatea unei imagini compresate au la bază eroarea reziduală, adică diferenţa pixel cu pixel dintre imaginea originală şi cea refăcută după compresie. Pentru a "condensa" informaţia cuprinsă în această imagine reziduală într-un singur număr, se utilizează diferite mărimi calculate folosind această informaţie. Numărul care va reprezenta această măsură a erorii trebuie să satisfacă următoarele condiţii:

BUPT

- să fie un număr pozitiv
- să fie zero în cazul în care imaginile originală şi compresată sunt identice
- să crească monoton odată cu înrăutăţirea calităţii imaginii.

Mărimile de măsură a erorii cele mai utilizate sunt:

- *eroarea medie absolută* (*Mean Absolute Error - MAE*) , sau eroarea / pixel, care este definită ca:

$$MAE = \frac{1}{mn}\sum_{x=1}^{m}\sum_{y=1}^{n}|f(x,y)-g(x,y)| \qquad \textbf{(A2.1)}$$

unde $f(x,y)$ reprezintă imaginea originală, iar $g(x,y)$ imaginea refăcută, ambele de dimensiuni $m \times n$.

- *eroarea medie pătratică* (*Mean Square Error - MSE*), dată de relaţia:

$$MSE = \frac{1}{mn}\sum_{x=1}^{m}\sum_{y=1}^{n}[f(x,y)-g(x,y)]^2 \qquad \textbf{(A2.2)}$$

Pentru ca cele două rezultate ale relaţiei de mai sus să fie comparabile, ultima se înlocuieşte cu rădăcina sa pătrată, rezultând aşa numita RMSE (*Root Mean Square Error*):

$$RMSE = \frac{1}{mn}\sqrt{\sum_{x=1}^{m}\sum_{y=1}^{n}[f(x,y)-g(x,y)]^2} \qquad \textbf{(A2.3)}$$

Se observă imediat că aceste mărimi satisfac evident primele două criterii, în timp ce pe al treilea nu. De exemplu, considerând două imagini identice din toate punctele de vedere mai puţin faptul că luminanţa diferă cu 40 pentru fiecare pixel, ar rezulta o eroare RMSE egală cu 20, ceea ce indică o calitate slabă. De asemenea, dacă considerăm două imagini identice, dar având o mică translaţie una faţă de cealaltă, eroarea obţinută va fi mare deoarece nu se va mai realiza diferenţa între pixeli identici.

Din aceste motive aceste mărimi nu oferă decât o indicaţie relativă pentru calitatea unei imagini, decizia finală fiind luată doar din punct de vedere al ochiului uman.

În afară de mărimile prezentate anterior, se mai utilizează ca măsură a erorii şi aşa numitele rapoarte semnal-zgomot care arată raportul între puterea imaginii originale şi puterea erorii. În literatură există mai multe variante de definire a acestui parametru, cele mai întâlnite fiind prezentate mai jos:

- raportul semnal / zgomot în care se consideră ca eroare MSE:

BUPT

$$SNR_{mse} = \frac{\sum_{x=1}^{m}\sum_{y=1}^{n} f(x,y)^2}{\sum_{x=1}^{m}\sum_{y=1}^{n} [f(x,y) - g(x,y)]^2} \quad \textbf{(A2.4)}$$

- raportul care utilizează ca eroare RMSE:

$$SNR_{rmse} = \sqrt{\frac{\sum_{x=1}^{m}\sum_{y=1}^{n} f(x,y)^2}{\sum_{x=1}^{m}\sum_{y=1}^{n} [f(x,y) - g(x,y)]^2}} \quad \textbf{(A2.5)}$$

- varianta logaritmică:

$$SNR_{\log} = 10 \cdot \log_{10}\left[\frac{\sum_{x=1}^{m}\sum_{y=1}^{n} f(x,y)^2}{\sum_{x=1}^{m}\sum_{y=1}^{n} [f(x,y) - g(x,y)]^2}\right] \quad \textbf{(A2.6)}$$

- rapoarte semnal/zgomot de vârf:

$$PSNR_{\log} = 10 \cdot \log_{10}\left[\frac{\max[f(x,y)]^2}{\frac{1}{mn}\sum_{x=1}^{m}\sum_{y=1}^{n} [f(x,y) - g(x,y)]^2}\right]$$

(A2.7)

sau:

$$PSNR_{\log} = 10 \cdot \log_{10}\left[\frac{[\max[f(x,y)] - \min[f(x,y)]]^2}{\frac{1}{mn}\sum_{x=1}^{m}\sum_{y=1}^{n} [f(x,y) - g(x,y)]^2}\right]$$

(A2.8)

BUPT

sau:

$$PSNR_{\log} = 10 \cdot \log_{10} \left[\frac{Maximum \quad Dynamic \quad Range^{2}}{\frac{1}{mn} \sum_{x=1}^{m} \sum_{y=1}^{n} [f(x,y) - g(x,y)]^2} \right] \tag{A2.9}$$

unde Maximum Dynamic Range reprezintă domeniul dinamic maxim al semnalului şi are de obicei valoarea 255, corespunzător a 256 nivele de gri.

A.2.2 Aprecierea subiectivă a calităţii

Aprecierea din punct de vedere cantitativ a calităţii prin mărimile definite în paragraful anterior nu poate substitui o evaluare subiectivă, adică impresia pe care o produce imaginea asupra unui observator uman. Există două tehnici principale de măsurare a calităţii răspunsului observatorului uman la diferiţi stimuli: metode primare sau explicite şi metode secundare sau implicite [NET 91]. Metodele primare au la bază examinarea unui set de imagini de către un grup de subiecţi care iau decizii subiective cu privire la calitatea imaginilor. Metodele secundare se bazează pe măsurarea calitativă a calităţii imaginii şi convertirea rezultatelor în măsuri ale calităţii subiective.

Metodele primare sunt mai utile în cazul în care distorsiunile introduse sunt foarte complexe ca aparenţă, aşa cum este cazul prelucrării digitale a imaginilor. Aceste evaluări primare se bazează pe două categorii diferite de metode şi anume: metode de clasificare pe categorii şi metode comparative.

Pentru clasificarea calităţii, subiecţilor li se prezintă, în condiţii normale de observare, o secvenţă de imagini de test obţinute prin prelucrare în modalităţi predeterminate, urmând ca aceştia să asocieze fiecărei imagini (sau secvenţe de imagini) o categorie dintr-un set dat de categorii posibile. Categoriile se bazează în general pe calitatea totală, scara utilizată fiind:

5. Excelentă
4. Bună
3. Medie
2. Slabă
1. Foarte slabă

O altă scară utilizată poate fi cea bazată pe vizibilitatea imperfecţiunilor:

5. Imperceptibile
4. Perceptibile dar nesupărătoare
3. Uşor supărătoare

BUPT

2. Supărătoare

1. Foarte supărătoare

Deoarece răspunsurile subiecţilor depind de foarte mulţi factori, ca de exemplu: luminanţa imaginii, contrastul imaginii, luminozitatea mediului ambiant, dimensiunile imaginii, distanţa de vizionare, experienţa şi motivaţia personală a subiecţilor, trebuiesc alese cu grijă grupul de persoane precum şi testele pentru eliminarea factorilor menţionaţi anterior [BOJ 96]. Rezultatele procedurilor de clasificare sunt prezentate prin calcularea aşa numitului Scor Mediu al Opiniilor – MOS (Mean Opinion Score)

$$MOS = \frac{\sum_{i=1}^{k} n_i C_i}{\sum_{i=1}^{k} n_i} \qquad \textbf{(A2.10)}$$

unde C_i este valoarea numerică corespunzătoare categoriei i, iar n_i este numărul rezultatelor obţinute în acea categorie, iar k numărul de categorii ce formează scara.

Cea de-a doua categorie de metode, cele comparative, utilizează o tehnică de comparare a două imagini: una rezultată în urma compresiei şi cealaltă rezultată în urma distorsionării unei imagini de referinţă cu semnale standardizate (de exemplu zgomot alb). În acest caz observatorul răspunde la întrebarea "cu cât este mai bună sau mai proastă imaginea de test în raport cu imaginea de referinţă?". Datele obţinute fiind apoi prelucrate pentru determinarea aşa numitului punct de egalitate subiectivă.

Scara de comparare utilizată este în acest caz următoarea:

+3 Mult mai bună
+2 Mai bună
+1 Puţin mai bună
0 La fel
-1 Puţin mai proastă
-2 Mai proastă
-3 Mult mai proastă.

BUPT

BIBLIOGRAFIE

1. [ADA 94] R. Adams and L. Bischof, "*Seeded Region Growing*", IEEE Trans on PAMI, vol.16, no.6, p. 575-588, June 1994
2. [ADI 85] G.Adiv, "*Determining three dimensional motion and structure from optical flow generated by several moving objects*", IEEE Transaction on PAMI, vol.7, p.384-401, July 1985.
3. [AGG 81] J.K. Aggarwal, L.S. Davis and W.N. Martin, "*Corespondence Processes in Dynamic Scene Analysis*", Proceedings of the IEEE, vol.69, no.5, p.562-572, May 1981
4. [AGG 88] J.K. Aggarwal and N. Nandhakumar, "*On the Computation of Motion from Sequences of Images – A Review*", IEEE Proceedings, vol.76, no.8, p.917-935, Aug. 1988
5. [AHU 96] N. Ahuja, "*A Transform for Multiscale Image Segmentation by Integrated edge and Region Detection*", IEEE Trans. on PAMI, vol.18, no.12, p.1211-1235, Dec. 1996
6. [AIZ 93] K. Aizawa, C.S. Choi, H. Harashima and T.S. Huang, "*Human facial Motion ans Synthesis with Application to Model-Based Coding*", Motion Analysis and Image Sequence Processing, Kluwer Academic publishers, Boston, 1993
7. [AIZ 95] K. Aizawa and T.S. Huang, "*Model Based Image Coding: Advanced Video Coding Techniques for Very Low Bit Rate Applications*", Proceedings IEEE, vol.83, no.2, p.259-271, Febr. 1995
8. [ALE 95] Fl. Alexa, "*Stadiul actual şi perspective în dezvoltarea televiziunii digitale*", Referat nr.1 în cadrul pregătirii pentru doctorat, Timişoara, 1995
9. [ALE 96] Fl. Alexa, "*Analiza mişcării pentru codarea secvenţelor de imagini*", Referat nr.2 în cadrul pregătirii la doctorat, Timişoara, 1996
10. [ALE 98] Fl. Alexa, "*Adaptive Split and Merge Algorithm for Region Based Motion Estimation*", Proceedings of the Symposium on Electronics and Telecommunications, vol.II , p.195-198, Timişoara, 1998
11. [ALE 98a] Fl. Alexa and Z. Bojković, "*A New Block Matching Algorithm for Motion Estimation*", Proceedings of the IASTED International Conference Signal and Image Processing – SIP '98, Las Vegas, Nevada – USA, October 27-31 1998, p.201-205
12. [ALE 98b] Fl. Alexa, Z. Bojkovici, C.I. Toma and V. Gui "A New Algorithm for Motion Estimation in Image Sequence Coding", Proceedinngs

BUPT

of the VIth Telecommunication Forum TELFOR'98, November 24-26, 1998, Belgrad, p.578-581

13. [ALE 98c] Fl. Alexa, " *Motion Estimation Based on Region", Proceedings of the Symposium on Electronics and Telecommunications "ETC.'98"*, p.199-203, Timişoara, Septembrie 1998
14. [ALE 98d] Fl. Alexa and Z. Bojković, "*Motion Estimation for Region Coding*", Proceedings of the 3rd International Conference on Multimedia Technology and Digital Telecommunication Sevices ICOMT'98, p.133-139, Budapest, October 1998
15. [ANA 89] P. Anandan, "*A Computational Framework and An Algorithm for the Measurement of Visual Motion*", International Journal Computer Vision, vol.2, p.283-310, 1989
16. [ANA 93] P. Anandan, R.J. Bergen, K.J. Hanna and R. Hingorani, "*Hierarhical Model Based Motion Estimation*", in M.I. Sezan and R.L. Lagendjik editors, Motion Analysis and Image Sequences Processing, p.1-22, Kluwer Academic Publishers, 1993
17. [ARA 95] R. Aravind, G. Cash, D. Duttweiller, H.M. Hang, B. Haskel and A. Puri, "*Image and Video Coding Standards*", in Digital Video, IEEE Press, 1995
18. [BAL 82] D.H. Ballard and C.M. Brown, "*Computer Vision*", Prentice-Hall, Englewood Cliffs, New Jersey, 1982
19. [BAR 88] Y. Bar-Shalom and T.E. Fortmann, "*Traking and Data Association*", Academic Press, New York 1988
20. [BAR 94] J.L.Barron, D.J. Fleet and S.S. Beauchemin, "*Performance of Optical flow Techniques*", Int. Journal Computer Vision, Vol.12(1), p.43-77, 1994
21. [BEV 89] J.R. Beveridge ş.a., "*Segmenting Images Using Localizing Histograms and Region Growing*", Intrenational Journal Computer Vision, vol.9, no.2, p.167-178, Feb. 1989
22. [BIE 87] J. Biemond, L. Looijenkga and D. Boekee, "*A Pel Recursive Wiener-Based Displacement Estimation Algorithm*", Signal Processing, vol.13, p.399-412, Dec. 1987
23. [BLA 87] R.E. Blahut, "*Principles and Practice of Information Theory*", Adison-Wesley, Reading, MA, 1987
24. [BOJ 96] Z. Bojković, D. Milanović and B. Reljin, "*Subjective Image Quality Estimation in Some Subband Coding Techniques*", Proceedings of the Third IEEE International Conference on Electronics, Circuits and Systems, Rodos, Greece, p.582-585, Oct. 13-16, 1996
25. [BOJ 97] Z. Bojković, C.I. Toma, V. Gui and R.Vasiu, "*Advanced Topics in Digital Image Compression*", Editura Politehnica, Timişoara, 1997

BUPT

26. [BOR 91] L. Böröczky, J. Driessen and J. Biemond, "*Pel-Recursive Motion Field Estimation from Image Sequences*", Visual Communications and Image Reprezentation, vol.2, p.259-280, 1991
27. [BOR 97] G.D. Borshukov, G. Bozdagi, Y. Altunbasak and A.M. Tekalp, "*Motion Segmentation by Multistage Affine Classification*", IEEE Trans. on Image Processing, vol.6, no.11, p.1591-1594, Nov. 1997
28. [BOR 98] A.G. Borş and I. Pitas, "*Optical Flow Estimation and Moving Object Segmentation Based on Median Radial Basis Function Network*", IEEE Trans. on Image Processing, vol.7, no.5, p.693-702, May 1998
29. [BOU 93] P. Bouthemy and E. Francois, "*Motion Segmentation and Qualitative Dynamic Scene Analysis from an Image Sequence*", Intrenational Journal of Computer Vision, vol.10, no.2, p.157-182, 1993
30. [CAF 76] C. Cafforio and F. Rocca, "*Methods for Measuring Small Displacements of Televisions Images*", IEEE Trans on Information Theory, vol. IT-22, no.5, p.573-579, Sept. 1976
31. [CAF 83] C. Cafforio and F. Rocca, "*The Diferential Method for Motion Estimation*", in Image Sequence Processing and Dynamic Scene Analysis, T.S.Huang ed, p.104-124, Springer-Verlag, New York, 1983
32. [CAN 86] J.F. Canny, "*A Computational Approach to Edge Detection*", IEEE Trans. on PAMI, PAMI-8, no.6, p.679-698, June 1986
33. [CHA 90] M.C.Chan, Y.B.Yu and A.G.Constantinides, "*Variable size block matching motion compensation with applications video coding*", IEE Proceedings, vol.137, No.4, p.205-212, Aug. 1990.
34. [CHA 96] R.F. Chang and W.M. Chen, "*Adaptive Edge-Based Side-Match Finite-State Classified Vector Quantisation with Quadtree Map*", IEEE Transactions On Image Processing, vol.5, no.2, february 1996
35. [CHA 96] F.H. Chan, F.K. Lam, P.W.F Poon, H. Zhu and K.H. Chan, "*Object Boundary Detection by Region and Contour Deformation*", IEE Proc. – Visual Image Signal Processing, vol.143, no.6, p. 353-360, Dec. 1996
36. [CHA 97] Y.L. Chan and W.C. Sin, "*Variable Temporal Length 3D Discrete Cosine Transform Coding*", IEEE Trans. on Image Processing, vol.6, no.5, p.758-763, May 1997
37. [CHA 97b] M.M. Chang, A.M. Tekalp and M.I.Sezan, "*Simultaneous Motion Estimation And Segmentation*", IEEE Trans. on Image Processing, vol.6, no.9, p.1326-1333, September 1997
38. [CHA 98] F.H. Chan, F.K. Lam and H. Zhu, "*Adaptive Thresholding by Varaitional Method*", IEEE Trans. on Image Processing, vol.7, no.3, p.468-463, March 1998
39. [CHA 98] S. Chanfxian and M. Yulong, "*Morphological Thinning Based on Image's Edges*", "Proceedings on International Conference on

BUPT

Communication technology ICCT'98, Beijing, China, October 22-24, 1998, p.S16-14.1-S16-14.5

40. [CHE 96] Y.K. Chen, Y.T. Lin and S.Y. Kung, " *A Feature Traking Algorithm Using Neighborhood Relaxation with Multi-Candidate Pre-Screening*", in Proc. of Intrenational Conference on Image Processing, vol.II, p.513-516, September 1996
41. [CHE 98] P. H. Chen and S.C. Huang, "*A Subsampling Hierarchical Search Algorithm for Fast Block Motion Estimation*", Proceedings of the IASTED International Conference Signal and Image Processing – SIP '98, Las Vegas, Nevada – USA, October 27-31 1998, p.206-209
42. [CLA 89] J.J. Clark, "*Autheticating Edges Produces bz Zero-Crossing Algorithms*", IEEE Trans. On PAMI, vol. PAMI-12, no.8, p.830-833, August 1989
43. [CLA 95] R.J. Clarke, "*Digital Compression of Still Images and Video*", Academic Press, 1995
44. [COR 98] P. Correira and F. Pereira, "*The Role of Analysis in Content Based Video Coding and Indexing*", Signal Processing, vol.66, no.2, p.125-142, April 1998
45. [CRI 94] R.J. Crinon and W.J. Kalodziej, "*Adaptive Model Based Motion Estimation*", IEEE Trans. on Image Processing, vol.3, no.5, p.469-481, Sept. 1994
46. [DOW 94] A.C. Downton, R.W.S. Tregidgo and A. Cuhadar, "*Top-Down Structured Parallelisation of Embedded Image Processing Applications*", IEE Proc.-Visual Image Signal Processing, vol.141, no.6, p.431-437, Dec. 1994
47. [DUF 94] F. Dufaux, I. Moccagata, F. Moscheini and H. Nicolas, "*Vector Quantization Based Motion Field Segmentation Under Entropy Criterion*", Journal of Visual Representation and Image Representation, vol.5, no.4, p.356-369, Dec. 1994
48. [DUF 95] F. Dufaux and F. Mosheini, "*Motion Estimation Techniques for Digital TV: A Review and a New Contribution*", Proceedings of IEEE, vol.83, no.6, p.858-876, June 1995
49. [DUF 95b] F.Dufaux, F. Moscheini and A. Lippman, "*Spatio-Temporal Segmentation Based on Motion and Static Segmentation*", in IEEE Proceeding of ICIP'95, vol.I, p.306-309, Washington, October 1995
50. [EGG 97] H. Eggers, F. Mosheini and R. Castagno, "*Robust Object Traking Based on Spatial Characterization of Objects by Additive Invariants*", in IEEE Proceeding ICIP'97, vol.I, p.450-454, November 1997, Toronto
51. [ENK 88] W. Enkelmann, "*Investigation of Multigrid Algorithms for the Estimation of Optical Flow Fields in Image Sequences*", Computer Vision, Graphics and Image Processing, vol.43, p.150-177, 1988

BUPT

52. [FLE 90] D.J. Fleet and A.D. Jepson, "*Computation of Component Image Velocity from Local Phase Information*", Int. Journal Computer Vision, vol.4, p.77-104, 1990
53. [FLE 91] D.J. Fleet, A.D. Jepson and M.R.M. Jerkin, "*Phase-Based Disparity Measurement*", Computer Vision, Graphics, and Image Processing: Image Understanding, vol.53, p.198-210, March 1991
54. [FRE 77] W. Frei and C.C. Chen, "*Fast Boundary Detection: A Generalization and a New Algorithm*", IEEE Trans. On Computers, vol. C-26, nr.10, p.988-998, Oct. 1977
55. [FUK 81] T. Fukuhara, A. Umahashi and T. Murakami, "*3D Motion Estimation for Model Based Coding*", Proceedings ICASP'81, p.69-72, 1981
56. [GEM 84] S. Geman and D. Geman, "*Stohastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images*", IEEE Trans. on PAMI, vol.6, p.721-741, 1984
57. [GER 92] A. Gersho and R.M. Gray, "*Vector Quantization and Signal Compression*", Kluwer Academic Publisher, Boston,1992
58. [GER 94] P. Gerken, "*Object Based Analysis Synthesis Coding of Image Sequences at Very Low Bit Rates*", IEEE Trans. on Circuits ans Systems for Video Technology, vol.4, no.3, p.228-235, June 1994
59. [GIR 93] B. Girod, "*Motion Compensating Prediction with Fractional Pel Accuracy*"" IEEE Trans. on Communications, vol.41, no.4, p.604-612, Apr. 1993
60. [GIR 94] B. Girod, "*Rate Constrained Motion Estimation*", in SPIE Proceedings Visual Communication and Image Processing '94, vol 2308, p.1026-1033, Chicago, September 1994
61. [GON 87] R.C. Gonzales and P Wintz, "*Digital Image Processing*", Adison Wesley, New York, 1987
62. [GON 93] R.C. Gonzales and R.E. Woods, " *Digital Image Processing* ", Adison- Wesley Publishing Company, London, 1993.
63. [GRA 90] R.M. Gray, "*Entropy and Information Theory*", Springer-Verlag, Berlin, 1990
64. [GRE 95] D. Grennhill and E.R. Davies, "*A New Approach to the Determination of Unbiased Thresholds for Image Segmentation*", IEE Proc. on Image Processing and its Applications Conference, 4-6 July 1995, p.519-523
65. [GUN 98] B. Gunsel, A.M. Tekalp and P.J.L. van Beek, "*Contents Based Access to Video Objects: Temporal Segmentation, Visual Summarization, and Feature Extraction*", Signal Processing, vol.66, no.2, p.261-280, April 1998

BUPT

66. [HAJ 98] S.A. Hajjatolesami and J. Kittler, "*Region Growing: A New Approach*", IEEE Trans. on Image Processing, vol.7, no.7, p.1084-1088, July 1998
67. [HAL 79] E.L. Hall, "*Computer Image Processing and Recognition*", Academic Press, New York, 1979
68. [HAR 92] R.M. Haralik and L.G. Saphiro, "*Computer and Robot Vision*", Addison-Wesley, New York, 1992
69. [HAS 74] B.G. Haskell, "*Frame to Frame Coding of Television Pictures Using Two Dimensional Fourier Transform*", IEEE Trans. on Information Theory, vol.IT-20, p.119-120, Jan. 1974
70. [HEE 88] D.J. Heeger, "*Optical Flow Using Spatio-Temporal Filters*", Int. Journal Computer Vision, p.279-302, Jan. 1988
71. [HOF 96] T. Hofmann, J. Puzicha and J.M. Buhmann, "*Unsupervised Segmentation of Textured Images by Pairwise Data Clustering*", Proceedings of IEEE International Conference on Image Processing, Lausanne, 1996, vol. III, p.137-140
72. [HOO 96] A. Hoover ş.a., "*An Experimental Comparision of Range Image Segmentation Algorithms*", IEEE Trans. On PAMI, vol.18, no.7, July 1996
73. [HOR 81] B.K.P. Horn and B.G. Schunk, "*Determining optical flow*", Artificial inteligence, vol.17, p.85-203, Aug. 1981.
74. [HOR 86] B.K.P. Horn, "*Robot Vision*", MIT Press, 1986
75. [HOT 90] M. Hotter, "*Object Oriented Analysis-Synthesis Coding Based on Moving Two Dimensional Objects*", Image Communications, vol.2, nr.4, p.409-428, Dec. 1990
76. [HOT 94] M. Hotter, "*Optimisation and Efficiency of an Object-Oriented Analysis-Synthesis Coder*", IEEE Trans. on Circuits and Systems for Video Technology, vol.4, no.2, p.181-194, April 1994
77. [IRA 92] M. Irani, B. Rousso and S. Peleg, "*Detecting and Traking Multiple Moving Objects Using Temporal Integrations*", in European Conference on Computer Vision, p.282-2876, Santa Margherita, Italy, May 1992
78. [IRA 94] M. Irani, B. Rousso and S. Peleg, "*Computing Occluding and Transparent Motions*", International Journal of Computer Vision, vol.12, no.1, p.5-16, January 1994
79. [ISA 87] J. Isailović, "*Videodisc Systems Theory and Applications*", Prentice Hall, New York, 1987
80. [ITU 95] ITU-T Recommendation H263,"*Video Coding for Narrow Telecommunication Channels at Less Than 64 kbit/s*", Tehnical report, ITU, July, 1995
81. [JAC 87] L. Jacobson and H. Wechsler, "*Derivation of Optical Flow Using a Spatiotemporal Frequency Aproach*", Computer Vision, Graphics, and Image Processing, vol.38, p.29-65, 1987

BUPT

82. [JAH 95] B. Jahne, "*Digital Image Processing: Concepts, Algorithms, and Scientific Applications*", Springer Verlag, Berlin, 1995
83. [JAI 81] A.K. Jain, "*Image Data Compression: A Review*", Proceedings of IEEE, vol.69, no.3, p.349-389, March 1981
84. [JAI 81a] J.R. Jain and A.K. Jain, "*Displacement Measurement and its Applications in Interframe Image Coding*", IEEE Trans. on Communications, vol. COM-29, no.12, p.1799-1808, Dec. 1981
85. [JAI 89] A.K. Jain, "*Fundamentals of Image Processing*", Prentice Hall, New York, 1989
86. [JEP 89] A. Jepson and M.R.M. Jerkin, "*The Fast Computation of Disparity from Phase Differences*", in Conference on Computer Vision and Pattern Recognition, p.398-403, IEEE 1989
87. [KER 95] R. Kermode and A. Lippman, "*Coding for Content: Enhanced Resolution from Coding*", in IEEE Proceeding of ICIP'95, vol.I, p.206-210, Washington, October 1995
88. [KIM 98] J. Kim and J.W. Woods, "*3D Kalman Filter for Image Motion Estimation*", IEEE Trans on Image Processing, vol.7, no.1, p.42-52, January 1998
89. [KIT 85] J. Kittler and J. Illingworth, "*On Threshold Selection Using Clustering Criteria*", IEEE SMC, vol. 15, p.652-655, 1985
90. [KOG 81] T. Koga, K. Iinuma, A.Hirano, Y. Iijima and T.Ishiguro, "*Motion Compensated Interframe Coding of Video Conferencing*", in Proc. National Telecommunications Conference, p.G5.3.1-G5.3.5, New Orleans, Dec. 1981
91. [KOK 96] A.C.Kokaram and S.J. Godsill, "*A System for Reconstruction of Missing Data in Image Sequences Using Sampled 3D AR Models and MRF Motion Priors*", in Computer Vission-ECCV'96, volume II, p.613-624, Springer Lecture Notes in Computer Science, April 1996
92. [KOM 89] T. Komarek and P. Pirsch, "*Array Architectures for Block Matching Algorithms*", IEEE Transsactions on Circuits and Systems, vol.CAS-36, p.1301-1308, Oct. 1989
93. [KON 88] J. Konrad and E. Dubois, "*Estimation of Image Motion Fields: Bayesian Formulation and Stochastic Solutions*", in Proc. ICASSP, p.1072-1075, IEEE 1988
94. [KON 92] J. Konrad and E. Dubois, "*Bayesian Estimation of Motion Vector Fields*", IEEE Trans. on PAMI, vol.14, no.9, p.910-927, Sept. 1992
95. [KON 95] J. Konrad, A.R. Mansouri and E. Dubois, "Towards Region-Based Coding for Very Low Bit Rate Video Services", Rapport technique de l'INRS-Telecommunications no.95-13, Universite du Quebec, Canada
96. [KRU 96] S.A. Kruger and A.D. Calway, "*A Multiresolution Ferequency Domain Method for Estimating Affine Motion Parameters*", in Proceedings International Conference on Image Processing, p.113-116, IEEE, 1996

BUPT

97. [KUN 85] M. Kunt, A. Ikonomopoulos and M. Kocher, "*Second-Generation Image-Coding Techniques*", Proceedings of the IEEE, Vol.73, no.4, p.549-574, April 1985
98. [KUN 87] M. Kunt, M. Bernard and R. Leonardi, "*Recent Results in High Compression Image Coding*", IEEE Trans. on Circuits and Systems, vol. CAS-34, no.11, p.1306-1336, Nov. 1987
99. [KUR 92] T. Kurita, N. Otsu and N. Abdelmalek, "*Maximum likelihood thresholding based on population mixture models*", Pattern Recognition, vol. 25, p.1231-1240, 1992
100. [KWA 98] O.J. Kwan and R. Chelleppa, "*Region Adaptive Subband Image Coding*", IEEE Trans. on Image Processing, vol.7, no.5, p.632-648, May 1998
101. [LAN 98] D.A. Langan, J.W. Modestino and J. Zhang, "*Cluster Validation for Unsupervised Stohastic Model Based Image Segmentation*", IEEE Trans. on Image Processing, vol.7, no.2, p.180-195, Feb. 1998
102. [LEE 90] S.U. Lee, S.Y. Chung and R.H. Pak, "*A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation*", Computer Vision, Graphiscs and Image Processing, vol. 52, p.171-190, 1990
103. [LEE 94] M.H. Lee and G. Grebbin, "*Image Sequence Coding Using Quadtree Based Block-Matching Motion Compensation and Classified Vector Quantisation*", IEE Proc.-Vis. Image Signal Process., vol.141, no.6, december 1994
104. [LEG 82] G.R. Legters and Y.Y. Tzay, "*A Mathematical Model for Computer Image Traking*", IEEE Trans on PAMI, vol.4, no.6, p.583-594, November 1982
105. [LeG 91] D. LeGall, "*MPEG: A Video Compression Standard for Multimedia*", Communication ACM, vol.34, p.47-58, April 1991
106. [LeG 92] D. LeGall, "*The MPEG Video Compression Algorithm*", Signal Processing: Image Communication, vol.4, no.2, p.129-140, April 1992
107. [LEO 95] A. Leonardis, A. Gupta and R. Bajcsy, "*Segmentation of Range Images as the Search for Geometric Parametric Models*", International Journal Computer Vision, vol.14, no.3, p.253-277, March 1995
108. [LEV 85] M.D. Levine, "*Vision in Man and Machine*", McGraw-Hill, New York, 1985
109. [LI 94] R. Li, B. Zeng and M.L. Liou, "*A New Three-step Search Algorithm for Block Motion Estimation*", IEEE Trans. on Circuits and Systems for Video Technology, vol.4, p.438-442, Aug. 1994
110. [LI 94a] H. Li, A. Lundmark and R. Forchheimer, "*Image Sequebnce at Very Low Bit Rates: A Review*", IEEE Trans. on Image Processing, vol.3, no.5, p.589-608, Sept. 1994

BUPT

111. [LI 96] H. Li and R. Forchheimer, "*Extended signal-Theoretic Techniques for Very Low Bit Rate Video Coding*", in Video Coding: The Second Generation Approach, Kluwer Academic Publisher, p.383-428, 1996
112. [MAL 89] S. Mallat, "*Multyfrequency Channel Decomposition of Images and Wavelet Models*", IEEE Trans. on Acoustics, Speech and Signal Processing, vol.ASSP-37, no.12, p.2091-2110, Dec. 1989
113. [MAR 82] D. Marr, "*Vision*", Freeman, San Francisco, 1982
114. [MAR 96] F. Marques, M. Padras and P. Salembier, "*Coding Oriented Segmentation of Video Sequences*", in L.Torres and M.Kunt editors, Video Coding: The Second Generation Approach, Kluwer Academic Publisher, Boston, 1996
115. [MAR 96a] P. Maragos, R.W. Schafer and M.A. Butt, "*Mathematical Morphology and Its Applications to Image and Signal Processing*", Kluwer Academic Publisher, Boston, 1996
116. [MAT 84] A. Mateescu, I. Băncilă, S. Popescu şi E. Burcoci, "*Manualul inginerului electronist*", vol II Transmisiuni de date, Editura Tehnică, Bucureşti, 1984
117. [MEM 98] E. Memim and P. Perez, "*Dense Estimation and Object Based Segmentation of the Optical Flow with Robust Techniques*", IEEE Trans on Image Processing, vol.7, no.5, p.703-719, May 1998
118. [MEY 90] F. Meyer and S. Bencher, "*Morphological segmentation*", Visual Communication, vol.1, p.21-46, 1990.
119. [MIT 86] G. Mitrofan, " *Televiziunea digitală* ", Editura Academiei , Bucureşti, 1986
120. [MOS 93] F. Mosheini, F. Dufaux and H. Nicolas, "*Entropy Criterion for Optimal Bit Allocation Between Motion and Prediction Error Information*", in SPIE Proceedings Visual Communication and Image Processing '93, vol 2094, p.235-242, Cambridge, November 1993
121. [MOS 95] F. Mosheini, F. Dufaux and M. Kunt, "*A New Two Stage Global/Local Motion Estimation Based on a Background/Foreground Segmentation*", IEEE Proceedings ICASSP'95, Detroit, May 1995
122. [MOU 97] P.Moulin and R. Krishanamurthy, "Multiscale Modeling and Estimation o Motion Fields for Video Coding", IEEE Transaction on Image Processing, vol.6, no.11, p.1525-1532, November 1997
123. [MUS 85] H.G. Musmann, P. Pirsch and H.J. Grallert, "*Advances in Picture coding*", Proceeding of the IEEE, vol.73, no.4, p.523-548, Apr. 1985.
124. [NAF 95] I. Naforniţă, A. Câmpeanu şi A.Isar, "*Semnnale, Circuite şi Sisteme*", Universitatea "Politehnica" Timişoara, 1995
125. [NAG 83] H.H. Nagel, "*Displacement Vector Derived from Second Order Intensity Variations in Image Sequences*", Computer Graphics and Image Processing, vol.21, p.85-117, 1983

BUPT

126. [NAG 82] H.H. Nagel and W. Enkelmann, "*Investigation of Second Order Greyvalue Variations to Estimate Corner Point Displacements*", Proceedings of 6th International Conference on Pattern Recognition, p.768-773, Munich, Germany 1982
127. [NAG 83] H.H. Nagel, "*Constraints for the Estimation of Displacement Vector Fields from Image Sequences*", in Proceedings International Joint Conference on Artifficial Inteligence, Karlsruhe, Germany, p.945-951, 1983
128. [NAK 85] K. Nakayama, "*Biological Image Motion Processing: A Review*", Vision Research, vol.75, p.626-660, 1985
129. [NEL 91] M.R.Nelson, "*Arithmetic Coding and Statistical Modelling*", Dr. Dobb's Journal, p. 16-29, February 1991
130. [NET 79] A. Netravali and J.D. Robins, "*Motion Compensated Television Coding part I*", Bell Systems Technical Journal, vol.58, no.3, p.629-668, Apr.1979
131. [NET 88] A.N. Netravali and B.G. Haskell, "*Digital Pictures: representation and compression*", Plenum Press, New York, 1988
132. [NET 91] A.N. Netravali and B.G. Haskel, "*Digital Pictures. Representation and Compression*", Plenum Press, New York, 1991
133. [NIC 91] H. Nicolas and C. Labit, "*Global Motion Identification for Image Sequence Analysis and Coding*", Picture Coding Symposium, p.248-251, Sept. 1991
134. [ORC 93] M.Orchad, "*Predictive Motion Field Segmentation for Image Sequence Coding*", IEEE Trans. on Circuits and Systems for Video Technology, vol.3, p.54-70, February 1993
135. [OTS 79] N. Otsu, "*A Threshold Selection Method from Gray-Level Histograms*", IEEE SMC, vol. 9, p.62-66, 1979
136. [PAR 94] M.Pardas and P.Salembier, "*3D Morphological Segmentation and Motion Estimation for Image Sequences*", Signal Processing, vol.38, p.31-43, January 1994.
137. [PAR 94a] M. Pardas, P. Salembier and B. Gonzales, "*Motion and Region Overlapping Motion Estimation for Segmentation-Based Video Coding*", in IEEE Proceedins of ICIP'94, vol.II, p.428-432, Austin, November 1994
138. [PEA 95] D.E. Pearson, "*Developments in Model Based Video Coding*", Proceedings IEEE, vol.83, no.6, p.892-906, June 1995
139. [PEL 90] S. Peleg and H. Rom, "*Motion Based Segmentation*", IEEE Proceedings on International Conference on Pattern Recognition, p.109-113, Atlantic City, June 1990
140. [PEN 91] A. Pentland, B. Horowitz and S. Sclaroff, "*Non-Rigid Motion and Structure from Contour*", IEEE Proceedings Workshop on Visual Motion, p.288-293, Princetown, October 1991

BUPT

141. [PER 97] F. Pereira, "*First proposals for MPEG-7 Visual Requirements*", Doc ISO/MPEG M1941, MPEG Meeting, Bristol, April 1997
142. [POP 86] E. Pop, I. Naforniţă s.a, " *Metode în prelucrarea numerică a semnalelor* ",Editura Facla, Timişoara, 1986
143. [PIE 90] M.A. Piech, "*Decomposing the Laplacian*", IEEE Trans. on PAMI, vol.PAMI-13, no.8, p.834-837, August 1990
144. [PIT 90] I. Pitas and A.N. Vanetsanopoulos, "*Nonlinear Digital Filters: Principles and Applications*", Kluger, Boston, 1990
145. [POP 98] M. Popović, "*Coding of Video Signal at Low and Very Low Bit Rates: Research and Applications*", Proceedings on 6th International Telecommunication Forum, TELFOR'98, Belgrade, Yugoslavia, 24-26 November 1998, p.553-560
146. [POR 80] M.R. Portnoff, "*Time-Frequency Represenations of Digital Signals and Systems Based on Short-Time Fourier Analysis*", IEEE Trans. on Acoustics, Speech, and Signal Processing, vol.ASSP-28, p.55-69, Feb. 1980
147. [PRA 74] W.K. Pratt, "*Correlation Techniques of Image Regiatration*", IEEE Trans. Aerosp. Electron. Systems, AES-10, n0.5, p.353-358, May 1974
148. [PRA 78] W.K. Pratt, "*Digital Image Processing*", John Wiley, New York, 1978
149. [PRE 70] J.M.S. Prewitt, "*Object Enhancement and Extraction*" in Picture Processing and Psychopictorics, B.S. Lipkin and A. Rosenfeld eds., Academic Press, New York, 1970
150. [QIA 96] R.J. Qian and T.S. Huang, "*Optimal Edge Detection in Two Dimensional Images*", IEEE Trans. on Image Processing, vol.5, no.7, p.1215-1220, July 1996
151. [RAB 75] L.R. Rabiner and B. Gold, "*Theory and Applications of Digital Signal Processing*", Englewood Cliffs, Prentice Hall, 1975
152. [RAB 91] M. Rabbani and P.W. Jones, "*Digital Image Compression Technique*", SPIE Optical Engineering Press, 1991
153. [RAJ 97] R. Rajogopolan, M.T. Orchard and R.D. Brandt, "*Motion Field Modeling for Video Sequences*", IEEE Trans on Image Processing, vol.6, no.11, p.1503-1516, November 1997
154. [RAM 94] K. Ramchandran, A. Ortega and M. Vetterli, "*Bit Allocation for Dependent Quantization with Applications to Multiresolution and MPEG Video Coders*", IEEE Trans. on Image Processing, vol.3, no.5, p.533-545, Sept. 1994
155. [ROS 84] A. Rosenfeld editors, "*Multiresolution Image Processing and Analysis*", Springer-Verlag, Berlin, 1984

BUPT

156. [SAB 65] Gh. Şabac, "*Matematici speciale*", Editura Didactică şi Pedagogică, Bucureşti, 1965
157. [SAL 95] P. Salambier, L.Torres, F. Meyer and C. Gu, "*Region Based Video Coding Using Mathematical Morphology*", Proceedings IEEE, vol.83, no.6, p.843-857, June 1995
158. [SAL 96] P. Saembier, P. Brigger, J.R. Cassas and M. Pardas, "*Morphological Operators for Image and Video Compression*", IEEE Trans. on Image Processing, vol.5, no.6, p.881-897, June 1996
159. [SAL 97] P. Salembier ş.a., "Segmentation Based Video Coding System Allowing the Manipulation of Objects", IEEE Trans. on Circuit and Systems for Video Technology, vol.7, no.1, p.60-74, Feb. 1997
160. [SAM 90] H. Samet, "*A Tutorial on Quadtree Research*", IEEE Transactions On Signal Processing, vol.30, no.3, july 1990
161. [SAN 88] T.D. Sanger, "*Stereo Disparity Computation Using Gabor Filters*", Biological Cybernetics, vol.59, p.405-418, 1988
162. [SAN 90] C.P. Sandbank, "*Digital Television*", John Wiley & Sons, London, 1990
163. [SCH 56] W.F. Schreiber, "*The Measurement of Third Order Probability Distributions of Television Signals*", IRE Transaction on Information Theory, vol.2, p.94-105, Sept.1956
164. [SCH 84] B.G. Schunk, "*The Motion Constraint Equations for Optical Flow*", Proceedings of 7th International Conference on Pattern Recognition, p.20-22, Montreal, Canada 1984
165. [SCH 97] G.M. Schuster and A.K. Katsaggelos, "*Rate Distorsion Based Video Compression*", Kluwer Academic Publishers, Boston, 1997
166. [SHE 94] D. Sheinwald, "*On the Ziv-Lempel Proof and Related Topics*", Proceedings of the IEEE, vol.82, no.6, p.876-881, June 1994
167. [SEF 93] V.Seferidis and M.Ghanbari, "*General Approach to Block-Matching Motion Estimation*", Optical Engineering, vol.32, no.7, p.1464-1474, July 1993
168. [SEF 94] V. Seferidis and M. Ghanbari, "*Generalised Block-Matching Motion Estimation Using Quadtree Structured Spatial Decomposition*", IEE Proc.-Vis. Image Signal Process., vol.141, no.6, December 1994
169. [SEZ 93] M.I. Sezan and R.L. Lagendjik, editors, "*Motion Analysis and Image Sequence Processing*", Kluwer Academic Publishers, Boston, 1993
170. [SHA 97] L. Shafarenko, M. Petrou and J. Kittler, "*Automatic Watershed Segmentation of Randomly Textured Color Images*", IEEE Trans. on Image Processing, vol.6, no.11, p.1530-1544, Nov. 1997
171. [SIK 97] T. Sikora, "*MPEG Digital Video – Coding Standards*", IEEE Signal Processing Magazine, vol.14, p.82-100, September 1997

BUPT

172. [SIM 91] E.P. Simoncelli, E.H. Adelson and D.J. Heeger, "*Probability Distributions of Optical Flow*", in Conference on Visual Pattern Recognition, p.310-315, IEEE 1991
173. [SKL 88] B. Sklar, "*Digital Communications, Fundamentals and Applications*", Prentice Hall, New York, 1988
174. [SPA 71] Al. Spătaru, "*Teoria informaţiei şi a codării*", Editura Tehnică, Bucureşti, 1971
175. [SRI 85] R.Srinivasan and K.Rao, "*Predictive coding based on motion estimation*", IEEE Transaction on Communications, vol.COM-33, p.888-896, Aug. 1985.
176. [STR 91] P. Strobach, "*Quadtree Structured Recursive Plane Decomposition Coding of Images*", IEEE Transactions On Signal Processing, vol.39, no.6, June 1991
177. [STR 98] M.G. Strintzis, "*Optimal Pyramidal and Subband Decompositions for Hierarchical Coding of Noisy and Quantized Images*", IEEE Trans on Image Processing, vol.7, no.2, p.155-166, February 1998
178. [TAB 97] M. Tabb and H. Ahuja, "*Multiscale Image Segmentation by Integrated Edge and Region Detection*", IEEE Trans. on Image Processing, vol.6, no.5, p.642-655, May 1997
179. [TAG 90] H.D. Tagare and R.J.P. De Figueiredo, "*On the Localisation Performance Mesure and Optimal Edge Detection*", IEEE Trans. on PAMI, vol.12, no.12, p.1186-1190, Dec. 1990
180. [TO 96] T.P.J. To and H. Yuen, "*Efficient Variable Rate Vector Quantisation Using Quadtree Segmentation*", IEEE Transactions On Consumer Electronics, vol.42, no.2, may 1996
181. [TRE 84] O. Tretiak and I. Pastor, "*Velocity Estimation from Image Sequences with Second Order Diferential Operators*", Proceedings of 7th International Conference on Pattern Recognition, p.16-19, Montreal, Canada 1984
182. [TZI 94] G. Tziritas and C.Labit, "*Motion Analysis for Image Sequence Coding*", Elsevier, Amsterdam, 1994
183. [ULL 79] S. Ullman, "*The Interpretation of Visual Motion*", MIT Press, Cambridge, 1979
184. [URA 88] S. Uras, F. Verri and V. Torre, "*A Computational Approach to Motion Perception*", Biological Cybernetics, vol.60, p.79-97, 1988
185. [VAI 90] P.P. Vaidyanathan, "*Multirate Digital Filters, Filter Banks, Polyphase Network and Applications*", Proceedings of the IEEE, vol.78, no.1, p.56-93, Jan.1990
186. [VAS 98] R.A. Vasiu, "*Studiul metodelor de compresie utilizate în prelucrarea şi transmiterea numerică a imaginilor*", Teză de doctorat, 1998

BUPT

187. [VEG 89] J.F. Vega-Riveros and K. Jabbour, "*Review of Motion Analysis Techniques*", IEE Proceedings, vol.134, no.6, p.397-404, Dec. 1989
188. [VET 90] M. Vetterli, "*Wavelets and Filter Bank; Relationship and New Results*", Proceedings of the International Conference on Acoustic, Speech and Signal Processing, ICASSP'90, p.1273-1276, Albuquerque, April 1990
189. [VET 92] M. Vetterli and K.M. Uz, "*Multiresolution Coding Techniques for Digital Television: A Review*", Multidimensional Systems and Signal Processing, Kluwer Academic Publishers, vol.3, p-161-187, 1992
190. [VIN 97] K.L. Vinken, A.S.E. Koster and M.A. Viergever, "*Probabilistic Multiscale Image Segmentation*", IEEE Trans. on Pattern Analysis and Machine Inteligence, vol.19, no.2, p.109-120, Febr. 1997
191. [VIR 74] K. Virupaksha and J.B. O'Neal, "*Entropy Coded Adaptive Differantial Pulse Code Modulation (DPCM) for Speech*", IEEE Trans. on Communications, vol.COM-22, no.6, p.777-787, June 1974
192. [VRA 98] Z. Vranyecz and K. Fazekas, "*Video Coding Methologies Achieving Very Low Bitrate for Multimedia Systems*", Proceedings of the 3rd International Conference on Multimedia Technology and Digital Telecommunications Services ICOMT'98, Budapest 28-30 October 1998, p.75-80
193. [WAL 84] D.R. Walker and K.R. Rao, "*Improved Pel-Recursive Motion Compensation*", IEEE Trans. on Communications, vol. COM-32, no.10, p.1128-1134, Oct. 1984
194. [WAN 92] Q. Wang and R.J. Clarke, "*Motion Estimation and Compensation for Image Sequences Coding*", Signal Processing: Image Communications, vol.4, p.161-174, 1992
195. [WAN 94] J.Y.A. Wang and E.H. Adelson, "*Representing Moving Images With Layers*", IEEE Trans on Image Processing, vol.3, p.625-638, Septeber 1994
196. [WAT 86] A.B. Watson, "*Temporal Sensitivity*", in Handbook of Perception and Human Performance, 1986
197. [WEL 84] T. Welch, "*A Technique for High-Performance Data Compression*", IEEE Computer, vol.17, no.6, p.8-19, June 1984
198. [WEN 93] J. Weng, "*Image Matching Using the Windowed Fourier Phase*", International Journal Computer Vision, vol.11, no.3, p.211-236, March 1993
199. [WIT 87] I.H. Witten, R.M. Neal and J.G. Cleary, "*Arithmetic Coding for Data Compression*", Communications of the ACM, vol.30, no.6, p.520-540, June 1987
200. [WOO 91] J. Woods, "*Subband Image Coding*", Kluwer Academic Publishers, Boston, 1991

BUPT

201. [WU 90] S.F. Wu and J. Kittler, "*A Differential Method for Simultaneous Estimation of Rotation, Change of Scale and Translation*", Signal Processing: Image Communications, vol.2, no.1, p.69-80, May 1990
202. [VIN 97] K.L. Vinken, A.S.E. Koster and M.A. Viergever, "*Probabilistic Multiscale Image Segmentation*", IEEE Trans. on Pattern Analysis and Machine Inteligence, vol.19, no.2, p.109-120, Febr. 1997
203. [WU 93] Z. Wu and R. Leahy, "*An Optimal Graph Theoretic Approach in Data Clustering: Theory and Its Application to Image Segmentation*", IEEE Trans. on PAMI, vol.15, no.11, p.1101-1113, Nov 1993
204. [WU 90] S.F. Wu and J. Kittler, "*A Differential Method for Simultaneous Estimation of Rotation, Change of Scale and Translation*", Signal Processing: Image Communications, vol.2, no.1, p.69-80, May 1990
205. [WU 93] Z. Wu and R. Leahy, "*An Optimal Graph Theoretic Approach in Data Clustering: Theory and Its Application to Image Segmentation*", IEEE Trans. on PAMI, vol.15, no.11, p.1101-1113, Nov 1993
206. [YOK 92] H. Yokoo, "*Improved Variations Relating the Ziv-Lempel and Welch-Type Algorithms for Sequential Data Compression*", IEEE Trans. on Information Theory, vol.38, no.1, p.73-81, Jan.1992
207. [YOU 93] R.W. Young and N.G.Kingsbury, "*Frequency Domain Motion Estimation Using a Complex Lapped Transform*", IEEE Transactionon Image Processing, vol.2, p.2-17, Jan. 1993
208. [ZAF 93] S. Zafar, Y.Q. Zhang and B. Jabbari, "*Multiscale Video Reprezentation Using Multiresolution Motion Compensation and Walvelet Descomposition*", in Journal on Selected areas in Communications, vol 11, nr1, pag 24 – 35, Jan. 1993
209. [ZHU 96] S.C. Zhu and A. Yuille, "*Region Competition: Unifying Snakes, Region Growing and Bayes/MDL for Multiband Image Segmentation*", IEEE Trans. on PAMI, vol.18, no.9, p.884-900, Sept. 1996
210. [ZIV 77] J. Ziv and A. Lempel, "*A Universal Algorithm for Sequential data Compressin*", IEEE Trans. on Information Theory, Vol.IT-23, no.5, p.337-343, May 1977
211. [ZIL 97] F. Ziliani and F. Mosheini, "*Kalman Filtering Motion Prediction for Recursive Spatio-Temporal Segmentation and Traking*", IEEE Proceedings Conference on Image Processing ICIP'97, Santa Barbara, October 1997

BUPT