

UNIVERSITATEA "POLITEHNICA" DIN TIMIȘOARA
Facultatea de electronică și telecomunicații

Teză de doctorat

**Contribuții la recunoașterea
caracterelor alfanumerice scrise de mână**

BIBLIOTECA CENTRALĂ
UNIVERSITATEA "POLITEHNICA"
TIMIȘOARA

**Conducător științific:
Prof. dr. ing. Corneliu I. Toma**

**Autor:
Ș.l. ing. Dan L. Lacrămă**

teza de doctorat

*1000 071
0000 000*

1998

Contribuții la recunoașterea
caracterelor alfanumerice scrise de mână
Teză de doctorat

Cuprins

Cuvânt înainte	1
1. Probleme generale privind recunoașterea caracterelor alfanumerice	1
1.1. Principalele aplicații ale recunoașterii caracterelor	2
1.2. Recunoașterea caracterelor alfanumerice	4
1.2.1. Sisteme de recunoaștere off-line	7
1.2.2. Sisteme de recunoaștere on-line	14
1.2.3. Identificarea semnăturilor	16
1.3. Psihologia citirii	17
1.4. Concluzii	20
2. Procesări preliminare	23
2.1. Preprocesări	23
2.2. Binarizarea	29
2.2.1. Tehnica YDH	30
2.2.2. Tehnici liniare adaptive	33
2.2.3. Tehnica funcției integrale	34
2.2.4. Tehnica contrastului local	36
2.3. Detecția și separarea regiunilor conținând text	38
2.4. Separarea rândurilor, cuvintelor și caracterelor	40
2.4.1. Separarea textului în rânduri și cuvinte	41
2.4.2. Separarea cuvintelor în litere	43
2.5. Subțierea caracterelor	44
2.6. Concluzii	46
3. Recunoașterea caracterelor alfanumerice	48
3.1. Recunoașterea statistică a caracterelor	49
3.1.1. Funcții discriminante polinomiale	52
3.1.2. Corelații obiect-prototip	52
3.1.3. Metoda histogramelor direcționale	54
3.1.4. Metoda gradientilor	55
3.1.5. Momente invariante	55
3.1.6. Funcțiile Gabor	57
3.2. Recunoașterea sintactică a caracterelor	58
3.2.1. Descriptori structurali	59
3.2.2. Gramatici formale	63
3.3. Rețele neuronale utilizate în analiza și recunoașterea caracterelor	68
3.3.1. Perceptronul multi-strat	68
3.3.2. Rețele Kohonen	70
3.3.3. Rețele cu neuroni RBF	72
3.4. Recunoașterea directă a cuvintelor	73
3.5. Validarea recunoașterii	75
3.5.1. Validarea cu dicționar	75
3.5.2. Lexicoane pentru recunoașterea directă a cuvintelor	77

Contribuții la recunoașterea
caracterelor alfanumerice scrise de mână
Teză de doctorat

3.5.3. Analiza contextuală	78
3.6. Personalizarea echipamentelor de recunoaștere a scrisului de mână	79
3.7. Analiza performanțelor sistemelor și algoritmilor de recunoaștere	81
3.8. Concluzii	83
4. Prezentarea sistemului de segmentare și recunoaștere a caracterelor	85
4.1. Divizarea documentelor prin metoda profilogramelor	85
4.1.1. Separarea documentului în cuvinte	89
4.1.1.1. Separarea textului în rânduri	89
4.1.1.2. Separarea rândurilor în cuvinte	95
4.1.2. Separarea cuvintelor în litere	98
4.1.2.1. Flexibilizarea separării caracterelor	105
4.2. Strategie etapizată de recunoaștere a caracterelor scrise de mână	108
4.2.1. Divizarea verticală	110
4.2.2. Gruparea caracterelor alfabetului românesc	112
4.2.3. Sortarea primară	114
4.2.4. Vectori descriptori micști	115
4.2.5. Rețele neuronale RBF arborescente	117
4.3. Sistem personalizat de recunoaștere a caracterelor	119
5. Rezultate experimentale	124
5.1. Divizarea prin metoda profilogramelor	124
5.2. Recunoașterea caracterelor scrise de mână	129
5.2.1. Recunoașterea caracterelor în cazul general	130
5.2.2. Personalizarea recunoașterii	132
6. Concluzii	136
6.1. Contribuții personale	136
6.2. Posibilități de dezvoltare ulterioară	139
6.3. Perspectivele de evoluție ale citirii automate	141
Bibliografie	143

Cuvânt înainte

Doresc să exprim cele mai respectuoase mulțumiri domnului prof.dr.ing. Corneliu I. Toma, conducătorul științific al tezei, pentru sprijinul și îndrumarea constantă și pentru felul în care m-a încurajat mereu în momentele de căutare ce au apărut pe parcursul acestei teze. Discuțiile avute m-au făcut să conștientizez importanța prezentării materialului de față într-o formă clară și accesibilă.

Mulțumiri speciale domnului prof.dr.ing. Mircea Petrescu, de la Universitatea "Politehnica" din București și domnului prof.dr.ing. Mircea Ivănescu, de la Universitatea din Craiova, pentru sprijinul acordat și pentru acceptul de a deveni referenți ai tezei.

Alese mulțumiri domnilor prof.dr.ing. Ioan Naforniță și prof.dr.ing. Alimpie Ignea, pentru sugestiile oferite pe parcursul elaborării tezei și pentru întregul sprijin acordat în activitatea mea de zi cu zi.

Sincere mulțumiri domnului conf.dr.ing. Vasile Gui, care mi-a deschis drumul spre domeniul prelucrării numerice a imaginilor și care mi-a dat mereu sfaturi extrem de utile pentru cercetarea întreprinsă la elaborarea acestei teze.

Mulțumiri domnului prof.dr.ing. Marius Oteșteanu și domnului conf.dr.ing. Alexandru Isar pentru sugestiile și sursele bibliografice puse la îndemână.

Mulțumiri tuturor membrilor din Departamentul de Comunicații pentru crearea unei atmosfere colegiale de lucru și pentru încurajările permanente.

Domnului dr. Graham Hall de la University of Central Lancashire Preston, Marea Britanie îi adresez mulțumiri speciale pentru îndrumarea și ajutorul acordat la experimentarea unor soluții cuprinse în această lucrare.

Cu siguranță teza de față ar fi fost mult mai dificil de finalizat fără încrederea și susținerea constantă primită din partea familiei mele.

Autorul

Timișoara, septembrie 1998

1. Probleme generale privind recunoașterea caracterelor alfanumerice

Odată cu generalizarea utilizării tehnicii de calcul în industrie, comerț, administrație și în general, în toate domeniile de activitate ale societăților moderne, unii futurologi au emis previziuni categorice cu privire la treptata dispariție a documentelor tipărite sau scrise de mână. S-a apreciat, în acea perioadă, că odată cu mondializarea rețelelor de calculatoare și a mijloacelor mass-mediei electronice, transferul informației se va face direct, fără a mai necesita utilizarea suportului tradițional, hârtia. În cartea sa “Șocul viitorului”, Alvin Toffler prevestea că într-un viitor previzibil, în societatea informațională, “galaxia Gutemberg va fi înlocuită de galaxia Marconi” [TOF 82]. Realitățile actuale se dovedesc însă a fi foarte diferite de aceste profetii. În marea majoritate a domeniilor de activitate, introducerea calculatoarelor a adus, pe lângă incontestabilele avantaje legate de sporirea vitezei și eficienței prelucrării datelor, și o creștere substanțială a cantității de documente tradiționale pe hârtie. Evident, cauzele acestei stări de fapt sunt complexe și nu fac obiectul acestei lucrări.

Comunicarea prin scris este, alături de vorbire, un mod natural de transfer al informației în orice societate omenească. Cu foarte puține excepții, toate civilizațiile cu rol important în istorie au avut scriere proprie și și-au dezvoltat un sistem administrativ propriu [DRA 85, MAT 89, HEL 98]. Lumea modernă este puternic dependentă de transferul informației sub formă scrisă și prin urmare, fiecare membru al ei, dispunând de o instruire elementară, este producător și consumator de documente scrise. Documentele tipărite sunt practic omniprezente și există un mare număr de activități economice și sociale care necesită documente scrise de mână. În diferite grade, funcție de legislația specifică fiecărei țări, unele instrumente de plată financiar bancare sunt obligatoriu completate de mână pe formulare tip, iar altele se pot bate la mașină, necesitând însă o semnătură olografă. Situația este similară în cazul activității de relații cu publicul a administrației locale și guvernamentale sau a firmelor prestatoare de servicii. Având în vedere aceste realități și o tradiție profund înrădăcinată în domenii cum sunt editarea de carte, trimerile poștale sau chiar presa scrisă, este socotit actualmente extrem de probabilă o lungă perioadă de coexistență între modernele documente „electronice” și cele tradiționale [BRE 97].

Există în această problemă o serie de implicații tehnologice serioase pornind de la o incompatibilitate de fond. Deși computerele sunt ele însele mari producătoare de documente scrise, ciclul nu se închide în mod natural fiindcă

reintroducerea automată a textelor, mai ales a celor scrise de mână, nu este încă soluționată cu adevărat satisfăcător. De exemplu, introducerea manuală în calculator a formularelor anuale de declarare a venitului pentru impozitare consumă în țări ca SUA un volum uriaș de muncă și timp. Automatizarea analizei unor asemenea formulare completate cu caractere de tipar sau majuscule scrise de mână, fiecare într-un câmp separat, este intens cercetată [BRE 97]. Problemele similare, care au apărut încă din deceniul trecut, la sortarea volumului mereu în creștere al trimiterilor poștale, au obligat operatorii naționali de comunicații din SUA, Japonia, Franța, Marea Britanie etc. să introducă metode și echipamente automate de interpretare a adreselor destinatarilor [USP 98, LAP 98]. Ratele de respingere și eroare ale acestora sunt încă destul de mari și cercetările pentru perfecționare continuă [GIL 93, WAD 93, WAK 97].

Recunoașterea caracterelor tipărite s-a studiat încă odată cu primele computere și există deja sisteme și echipamente perfecționate care reușesc să realizeze citirea automată cu o bună acuratețe (99.9% clasificări corecte). Aceste sisteme, deși sunt încă departe de performanțele creierului uman, au devenit produse comerciale relativ răspândite și satisfac în mod rezonabil cerințele pieței [PAV 93, GOR 97].

Recunoașterea scrisului de mână cu caractere separate sau cursive este o problemă mult mai complexă datorită variațiilor individuale ale grafiei fiecărui om. Analiza scrisului cursiv este complicată în plus de dificultatea separării corecte a cuvintelor în litere. Rezultatele experimentale raportate de laboratoarele de cercetare din domeniu sunt încă modeste și se poate constata că aceasta direcție de cercetare, prezentând o importanță vitală pentru dezvoltarea în viitor a unor activități umane extrem de diverse, este încă deschisă [BRE 97].

1.1 Principalele aplicații ale recunoașterii caracterelor

Principalele aplicații ale recunoașterii caracterelor sunt legate de introducerea automată în sistemul de calcul a unor date disponibile sub formă scrisă. Există un număr mare de posibile aplicații ale transiterii informației scrise către calculator, dar la ora actuală cercetările sunt canalizate pe rezolvarea următoarelor probleme practice [BRE 97]:

- citirea automată a tipăriturilor și manuscriselor;
- procesarea automată a formularelor completate de mână;
- prelucrarea instrumentelor de plată în domeniul bancar;
- sortarea automată a trimiterilor poștale (în special a scrisorilor);

- interpretarea automată a notițelor scrise pe ecrane sensibile;
- identificarea automată a numerelor de înmatriculare ale vehiculelor.

Practic, toate aceste surse de informație de tip text sunt disponibile în una din următoarele forme [SEN 94]:

- imaginea unor documente tipărite;
- imaginea unor documente olografe;
- imaginea unor documente mixte;
- notițe scrise direct pe ecran.

Documentele tipărite sunt acele documente ce conțin doar texte scrise cu caractere de tipar. În practică, este vorba de produsele tipografice (cărți, reviste, ziare etc.) și de documentele scoase la imprimantele calculatoarelor (documentații, rapoarte etc.). Cazul special al identificării unor numere de înmatriculare, în imaginile achiziționate, este în ultimă instanță cuprins tot în această categorie, deoarece dincolo de suportul diferit este vorba în fapt de recunoașterea unor caractere de tipar (cifre și litere) [ROB 96, GOR 97].

La ora actuală sunt disponibile pe piață programe cu performanțe bune capabile să execute citirea automată a documentelor tipărite. Utilizarea lor economisește munca necesară unei reintroduceri manuale a textului, dar, pentru a evita posibile erori, se recomandă executarea de către operator a unei verificări post-recunoaștere a fișierului text rezultat [KAN 95].

Documentele olografe, scrise integral de mână, de obicei cu scris cursiv, implică dificultăți suplimentare privind separarea textului în rânduri, a rândurilor în cuvinte și mai ales a cuvintelor în literele. Citirea automată a acestor texte este afectată de rate de eroare și respingere mari, fiind necesară utilizarea unor metodologii speciale, de genul recunoașterii cuvintelor în ansamblu, sau adăugarea validării contextuale pe lângă cea uzuală bazată pe lexicon [HO 92]. Deocamdată rezultate rezonabile s-au obținut doar pentru cazul particular al sistemelor de recunoaștere personalizate, unde calculatorul citește scrisul unui singur utilizator după ce, inițial, l-a învățat pe baza unor eșantioane tipice.

Documentele mixte sunt cele ce cuprind atât caractere tipărite cât și caractere de mână. În această categorie se includ formularele completate cu scris de mână, textele tipărite cuprinzând adnotări și trimiterile poștale, care, putând fi scrise cu caractere de tipar sau de mână, necesită sortatoare capabile să se descurce cu ambele tipuri de caractere. Actualmente, sunt în derulare programe intense de cercetare orientate spre rezolvarea problemei procesării automate a formularelor și trimiterilor poștale. În amândouă aceste cazuri este în general valabilă ipoteza că părțile scrise de mână se completează îngrijit cu litere de tipar separate între ele prin spații [TOR 96, LIO 96, GAR 96]. Excepțiile sunt de obicei respinse la clasificarea automată și se reprocesează manual.

Toate aceste cazuri corespund unei proces cuprinzând trei etape distincte:

- scrierea documentului;
- achiziția și memorarea imaginii documentului;
- interpretarea (citirea automată).

Prin urmare, în cea de a treia etapă, indiferent de sursa primară, singura informație disponibilă este imaginea textului ce trebuie citit. Metodele și algoritmii utilizați pentru citirea automată a acestor texte vor fi analizate în detaliu în paragraful 1.2.1 “Sisteme de recunoaștere off-line”.

La începutul acestui deceniu, au apărut pe piață agende electronice personale dotate cu un ecran sensibil la atingere (sensitive touch screen personal organisers) care rețin informația scrisă de mână cu ajutorul unui stilou special direct pe ecran. Inițial aceste agende erau construite să înregistreze imaginea textului scris, dar actualmente sunt disponibile variante capabile de interpretarea notițelor și transformarea imaginii într-o înregistrare de tip text. Spre deosebire de situațiile anterioare, achiziția datelor de intrare se face direct și informația este disponibilă în timp real. Întreaga procedură de recunoaștere se modifică substanțial, deoarece în acest caz la recunoaștere se ține cont și de informația temporală [NIS 95]. Metodele și algoritmii aplicați în recunoașterea scrisului de mână pe ecrane sensibile la atingere se vor analiza mai în detaliu în paragraful 1.2.2 “Sisteme de recunoaștere on-line”.

1.2 Recunoașterea caracterelor alfanumerice

Având în vedere cele de mai sus, indiferent de trăsăturile specifice ale aplicațiilor particulare, dacă se are în vedere întreaga procedură, de la constituirea sursei primare de informație până la afișarea pe monitor a textului “citit”, metodele de recunoaștere utilizate în domeniu se pot grupa pe două mari direcții [WIL 97]:

- recunoașterea on-line;
- recunoașterea off-line.

Prima categorie este legată de agendele electronice cu ecrane sensibile la atingere, în vreme ce a doua este utilizată pentru sarcini de genul prelucrării automate a imaginilor anterior achiziționate ale unor documente cum sunt: textele tipărite, manuscrisele, trimerile poștale, formularele cecurile etc. [ABU 98].

Un caz aparte în cadrul recunoașterii scrisului îl reprezintă verificarea semnăturilor. Această aplicație specială, destinată sistemelor de securitate, presupune analiza grafologică automată a unui cuvânt fix cu puternică specificitate individuală (semnătura personală), în vederea determinării identității autorului. Rezultatul nu este recunoașterea caracterelor sau a cuvântului, ci decizia dacă

autorul eșantionului studiat este sau nu este același cu cel al semnăturii tip, stocată în baza de date. Analiza semnăturilor se poate face atât în timp real [SAB 95, WU 97] cât și pe imaginea unui document scris anterior [SAB 97].

Dacă se consideră etapa centrală a procesului, recunoașterea efectivă a caracterelor, clasificarea sistemelor și metodelor de citire automată devine mai laborioasă. Se poate aprecia că de-a lungul deceniilor de cercetări în domeniu, practic mare parte din algoritmi aplicați în clasificarea formelor au fost utilizați cu rezultate mai mult sau mai puțin bune și în aplicațiile de recunoaștere a caracterelor. Există prin urmare:

- recunoașterea caracterelor prin metode statistice,
- recunoașterea caracterelor prin metode sintactice,

la care se adaugă și metodologii mixte care utilizează combinații ce permit însumarea unor avantaje proprii fiecărei metode în parte. Aceste metode de recunoaștere pot fi utilizate într-o procedură de clasificare tradițională sau implementate într-o rețea neuronală [SCH 92].

Cazul recunoașterii caracterelor este mai puțin favorabil utilizării metodelor statistice datorită numărului mare de clase și variației între formele aceluiași caracter la tipăriți sau între diverși autori, la scrisul de mână [ROB 96]. Totuși, în cele trei decenii de cercetări privind citirea automată n-au lipsit încercări, uneori destul de fructuoase. În literatură sunt pomenite un număr mare de metode de recunoaștere a formelor utilizate de sistemele de citire automată. Printre acestea, cele mai importante sunt:

- corelația caracterului necunoscut cu un șablon al fiecărei clase;
- metoda histogramelor;
- metoda gradientilor;
- comparația codurilor lanț ce descriu contururile;
- metode bazate pe proprietățile topologice - linii, bucle, găuri etc.;
- combinația celor de mai sus prin intermediu unor criterii de încredere;
- comparația momentelor - invariante: Hu, Zernike, Afine etc.;
- metoda coeficienților Gabor.

În ultima perioadă recunoașterea utilizând caracteristicile histogramelor este des utilizată ca una din etapele unei metode complexe, rezultatul ei cu o anumită pondere intrând în decizia finală [SRI 93, CED 97].

Utilizarea rețelelor neuronale a readus în actualitate și clasificările pe baza parametrilor secundari, cum sunt codurile Freeman, transformate, momentele invariante etc. Având în vedere, pe de o parte, că algoritmi statistici sunt în general simpli și, pe de altă parte, că resursele în dezvoltarea rețelelor neuronale sunt departe de a fi total explorate, se poate aprecia că această combinație poate fi în continuare promițătoare [CHA 94, BAR 96, BRE 97].

Recunoașterea prin metode sintactice este frecvent abordată în cercetările mai recente. Avantajul major al acestei abordări constă în aceea că prin analiza părților componente ale caracterelor - linii, bucle deschise sau închise, pauze îmbinări etc. - și prin efectuarea clasificării în baza existenței sau inexistenței acestor elemente într-o anumită inter-relație, se obține o bună insensibilitate la variația tipurilor tipografice sau a grafiilor particulare [YON 93, HUL96, MAR97].

Se pot identifica două direcții de dezvoltare a recunoașterii cu metode sintactice:

- analiza bazată pe descriptori structurali;
- gramatici formale (metode lingvistice).

Prima metodologie se bazează pe divizarea caracterul în elementele sale constitutive. Poziția acestora este apoi determinată funcție de o grilă de divizare aflată în relație cu linia de bază a rândului scris [CED 98a]. Clasificarea se face după criteriul prezenței sau absenței subcomponentelor și a plasării lor în spațiu.

Precizia recunoașterii depinde în mod esențial de complexitatea descriptorului utilizat, adică de modul cum sunt detectate și reprezentate elementele constitutive ale caracterului. Se poate vedea în figura 1.2.1 că, spre exemplu, aproximările poligonale ale cifrelor „4” și „9” sunt foarte apropiate și cu greu pot fi diferențiate de un echipament automat de citire [WAN 95].

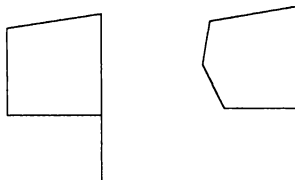


Figura 1.2.1. Aproximarea poligonală a cifrelor „4” și „9” [WAN 95]

Soluția este utilizarea unor reprezentări mai complexe a buclelor prin serii de arce de cerc, de elipsă etc. și, prin urmare, creșterea complexității modelelor și a prelucrării caracterului necunoscut.

În mod similar cu situația anterioară, analiza lingvistică pornește de la identificarea subcomponentelor, dar structura este descrisă ca o inter-relație existentă între aceste componente. Poziția fiecărui element este considerată relativ la poziția vecinilor săi în arhitectura generală a caracterului [SAL 94, SAN 97].

Utilizarea rețelelor neuronale nu introduce elemente principial diferite, dar implementările de acest fel sunt foarte frecvente și au rezultate foarte bune. Aceste rezultate sunt bazate pe eficacitatea rețelelor neuronale de a recunoaște

caracteristicile unei clase chiar dacă forma studiată prezintă variații destul de mari față de modelul ideal al acesteia.

În implementarea directă se aduce la intrarea rețelei imaginea digitală a caracterul necunoscut, eventual binarizată, iar la ieșire se obține răspunsul. Pentru a simplifica însă rețeaua și mai ales, pentru a scădea timpul de învățare, se optează pentru o mai bună procesare a datelor de intrare. În locul imaginilor brute se pot aduce, cum s-a arătat mai sus, un vector conținând o serie de parametri măsurabili sau mărimi secundare rezultate din prelucrarea acestora [HAY 94].

Implementarea modelelor pur sintactice pe rețele neuronale este mai puțin utilizată deoarece, odată realizată separarea în părți componente și stabilit graful care exprimă relațiile existente între acestea, fie recunoașterea este directă fie se apelează la clasificatorii fuzzy. În schimb este mai frecventă utilizarea la intrarea unor rețele neuronale a unor vectori descriptori micști conținând combinarea rezultatelor metodelor de recunoaștere statistice și sintactice.

Există o clară diferențiere între modul în care se abordează diferitele tipuri de documente scrise:

- texte tipărite;
- texte scrise de mână cu caractere de tipar (separate);
- texte scrise cursiv.

Există, așa cum se va arăta în continuare, o strategie diferită de lucru pentru citirea tipăriturilor, o alta pentru citirea scrisului de mână cu caractere de tipar și în sfârșit un proces specific substanțial diferit pentru scrierea cursivă unde problema suprapunerii parțiale a caracterelor componente ale cuvintelor este extrem de acută și generatoare de mari dificultăți [BAI 93].

Cât privește implementarea practică, se poate remarca o impresionantă varietate de modalități: calculatoare standard, rețele de transputere, rețele neuronale simulate pe computere cu arhitectură Von Neuman sau capabile de a procesa datele în paralel MIMD și nu în ultimul rând echipamente specializate pentru aplicații de anvergură cum sunt de exemplu sortatoarele automate de scrisori [PIT 94].

1.2.1 Sisteme de recunoaștere off-line

Sistemele de recunoaștere off-line sunt destinate citirii textelor scrise pe un suport fix: hârtie, plastic, metal etc. Recunoașterea se desfășoară ulterior încheierii scrierii întregului text de analizat. Documentul supus procedurii de recunoaștere poate să fie scris de mână sau tipărit. Acest document este scanat sau fotografiat și

analiza lui se face pe imaginea introdusă și stocată în calculator [ABU 98, CED 97].

Indiferent de situație, utilitatea acestei procesări a documentelor constă în faptul că imaginea textului analizat este înlocuită cu un fișier ASCII conținând în cazul ideal aceeași informație într-un spațiu de stocare mult mai mic.

În practică toate cele trei tipuri de documente scrise (tipărite, olografe sau mixte) pot fi supuse procedurii de recunoaștere off-line. Așa cum s-a pomenit mai sus, specificitatea fiecărui tip presupune schimbări ale metodologiei aplicate la citirea automată. În cele ce urmează se va face o prezentare sumară a caracteristicilor specifice etapelor de lucru pentru cele trei situații distincte, urmând ca stadiile de prelucrare esențiale ale procesului să fie analizate aprofundat în cadrul unor paragrafe special destinate [LIM 96].

Analiza documentelor tipărite se numește recunoașterea optică a caracterelor O.C.R. (Optical Character Recognition) și este deja relativ bine pusă la punct. Există pe piață un număr mare de programe care realizează transformarea imaginilor în fișiere conținând un text ASCII similar cu cel original. Rata de eroare în situații fără dificultăți speciale este de ordinul zecimilor sau chiar sutimilor de procent ceea ce de obicei este suficient [ABU 98].

Uzual aceste programe sunt destinate introducerii rapide pe calculator sub formă de text a unor documente tipărite de interes (cărți, articole, rapoarte etc.), dar există și cazuri speciale cum este cel descris în [YAN 94b, TRI 97] unde se caută și identifică denumirea unor obiective pe hărți geografice.

Schema bloc a unui sistem de recunoaștere optică a caracterelor, propusă de autorul acestei teze în [LAC 97e], este prezentată în figura 1.2.2. Fiecare bloc component corespunde unei etape specifice a procesului. Această structură corespunde situațiilor tipice și poate suferii anumite variații în situația unor cerințe speciale.

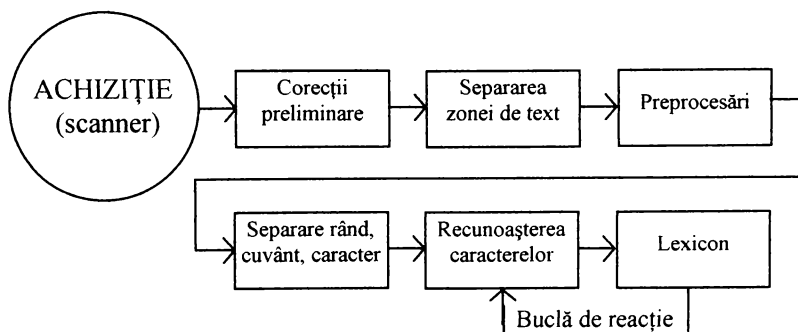


Figura 1.2.2. Schema bloc a unui sistem de recunoaștere optică a caracterelor

În schema prezentată s-a figurat ca modalitate de introducere a datelor de intrare un scanner, ceea ce corespunde realității în foarte multe situații. Este însă posibil ca imaginile să fie preluate și prin intermediul unor camere CCD sau alte echipamente similare. La achiziția cu camere TV trebuie luate toate precauțiile pentru a obține iluminarea uniformă a documentului.

Deoarece, în unele cazuri, nu există garanții ferme privitoare la calitatea imaginilor supuse prelucrărilor, este, în general, de dorit intercalarea la începutul succesiunii de proceduri preliminare a unei etape de prelucrare care să corecteze în măsura în care este posibil deficiențele generale ale imaginii. Aceste măsuri suplimentare sunt menite să mărească rezistența sistemului la diferiți factori perturbatori [BAI 93].

În general, corecțiile primare aplicate acționează asupra factorului de zgomot, distorsiunilor geometrice și deficiențelor de iluminare. Funcție de problemele ce se estimează că pot să impiețeze asupra acurateței imaginii, proiectantul poate să prevadă procesări pornind de la filtrări simple, precum netezirea, până la proceduri complexe de restaurare [GON 93].

În cazurile în care imaginea conține zone de text scris intercalate cu alte porțiuni, sau când există regiuni „albe”, recunoașterea și delimitarea strictă a părților din imagine care cuprind textul este obligatorie, deoarece în caz contrar sistemul încearcă interpretarea inclusiv a părților ce conțin de exemplu figuri, lucru care evident duce la erori și blocaje [KAM 93, LU 98]. Separarea se face pornind de la faptul că în zona de text există o textură specifică având o densitate mare de contururi (corpul literelor față de fond) în contrast cu porțiunile albe sau cu figurile, unde contururile lipsesc sau sunt mai puțin frecvente.

Etapa următoare este prelucrarea specifică a regiunii sau regiunilor conținând text, în vederea evidențierii diferențelor dintre caracterele ce se doresc citite și fond. Procedurile de bază sunt detecția conturilor și binarizarea urmată, în general, de erodarea corpului caracterelor până la reducerea lor la structuri compuse din linii și bucle deschise sau închise cu grosimea de un pixel [MES 96].

Separarea textului în rânduri, cuvinte și respectiv caractere se face printr-o succesiune de prelucrări axiale ale imaginilor. Cea mai frecvent folosită metodă prelucrează histogramele direcționale pe orizontală și verticală ale căror minime locale permit detecția frontierelor ce separă rândurile, cuvintele și respectiv caracterele [SHA 93, LET 95, LIK95].

Abia după ce toate aceste etape au fost parcurse, se poate aborda efectiv sortarea semnelor grafice în categorii corespunzătoare diferitelor litere și cifre. Această operație, reprezentând citirea automată propriu-zisă, se poate aborda cu diverse metode și strategii așa cum s-a pomenit anterior.

În final, din caracterele recunoscute sunt reasamblate cuvintele și se verifică dacă acestea se găsesc într-un lexicon (adică un nomenclator de cuvinte pentru aplicații cu vocabular limitat). Sistemele dotate cu aceste facilități au, într-o anumită măsură, capacitatea de a se auto-controla, bucla de feed-back fiind parcursă de cuvintele ce se dovedesc inexistente în dicționar și sunt trimise înapoi la „citire”. Dacă răspunsul este pozitiv cuvântul este declarat recunoscut, dacă nu, se revine pe bucla de reacție și se încearcă o altă variantă de clasificare a caracterelor componente [SEN 96]. Unele sisteme au și un bloc final de analiză contextuală, în care se verifică asamblarea cuvintelor în propoziții și fraze [CHA 96].

Recunoașterea scrisului de mână cu caractere de tipar este relativ similară cu O.C.R.-ul, dar rezultatele obținute sunt mult mai modeste deoarece în acest caz intervine, cu un substanțial efect negativ, variația formei și dimensiunii caracterelor de la un individ la altul.

Aplicația tipică ce ilustrează acest caz este procesarea automată a datelor înscrise cu mâna în formularele tipărite prezente, așa cum s-a spus mai sus, în practic toate compartimentele de relații cu publicul ale băncilor, firmelor, administrației financiare și locale [SRI 96]. Aceste formulare sunt obligatoriu completate fie cu litere de tipar fie cu majuscule.

Deși, din punct de vedere al blocurilor funcționale, schema este aparent identică cu cea de la O.C.R., în zona corespunzătoare separării și la recunoașterea caracterelor, abordarea concretă este substanțial diferită în acord cu creșterea complexității problemei. Schema bloc a sistemului de recunoaștere a scrisului de mână discret, propusă de autorul acestei teze în [LAC 97e], este prezentată în figura 1.2.3.

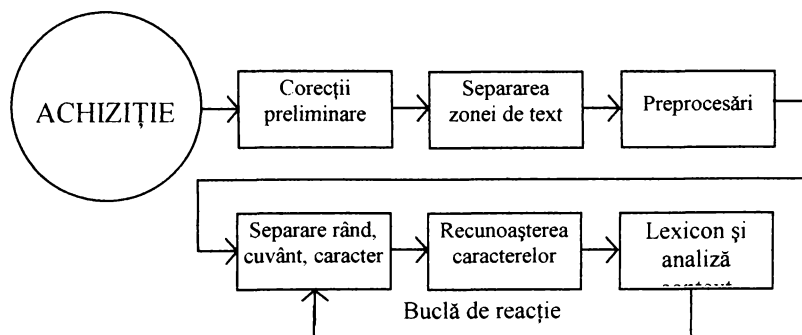


Figura 1.2.3. Schema bloc a unui sistem de recunoaștere a textelor scrise de mână cu caractere de tipar (separate)

În primul rând bucla de reacție cuprinde în acest caz și etapa separării deoarece, chiar în cazul literelor de tipar sau al majusculilor, scrisul de mână este susceptibil să conțină suprapuneri și legături accidentale între caractere. Prin urmare

în asemenea situații este indicată încercarea mai multor variante de separare și păstrarea celei pe care o validează verificarea în lexicon [KIM 97].

Lexiconul are în cazul procesării formularelor tipizate o importanță aparte. Un element specific de mare importanță în acest tip de aplicații este faptul că pentru fiecare formular în parte numărul cuvintelor ce pot să apară este redus și prin urmare există posibilitatea utilizării eficiente a unui dicționar de dimensiune mică pentru verificarea finală [HO 92, YU 96]. În anumite cazuri analiza cuvintelor recunoscute se face ținând cont de contextul în care ele apar. Rezultă de aici o posibilitate suplimentară de validare a recunoașterii. În plus între anumite câmpuri ale unui formular pot să existe o serie de corelații care permit o verificare a interpretării corecte [GAR 96]. Dacă rezultatul recunoașterii nu corespunde, fie se reia procesul pornind cu etapa de separare, fie se refuză clasificarea și documentul se transferă la linia de procesare manuală.

Există deja în funcțiune echipamente de procesare a unor instrumente de plată bancare în care singurele caractere de recunoscut sunt cifre (valori și coduri) plasate în câmpuri speciale ale formularului. Problema interpretării acestora cuprinde două etape distincte [KNE 95]:

- detecția poziției câmpurilor ce trebuie analizate;
- recunoașterea cifrelor.

Pentru scăderea ratelor de eroare și respingere documentele ce trebuiesc analizate se completează cifră cu cifră în căsuțe separate, utilizând o cerneală specială ce conține particule magnetice fine în suspensie. Imaginea preluată este astfel epurată din start de posibilele artefacte rezultate din combinația între liniatura trasată cu tuș a formularului și caracterele scrise cu cerneală [KAW 98].

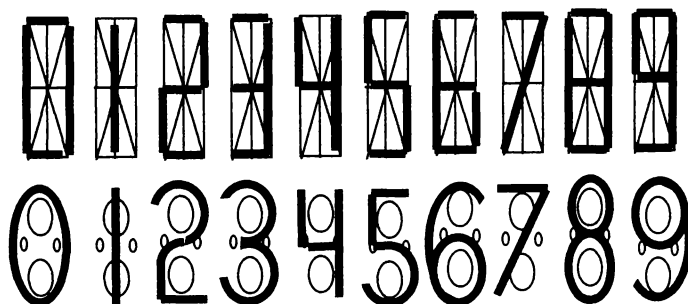


Figura 1.2.4. Grile de restricționare a formei literelor [SCH 92]

Urmare a tuturor acestor măsuri, sistemele sunt relativ mai simple și deci mai eficiente în funcționare. Rezultatele lor sunt cu atât mai bune cu cât formularele pe care se completează datele de analizat sunt mai bine întocmite. Se poate merge

chiar până la introducerea unor restricții de trasare a literelor cu ajutorul unor grile speciale de genul celor din figura 1.2.4. Cele două exemple prezentate diferă nu doar ca formă ci și prin faptul că în prima grilă se impune trasarea caracterelor cu segmente de dreaptă fixe, în vreme ce în cazul celei de a doua este permisă o flexibilitate suplimentară, rezultând un mod de scriere mai apropiat de cel natural.

Recunoașterea caracterelor din cadrul textelor scrise cursiv este mult mai delicată și mai plină de dificultăți. Spre deosebire de recunoașterea caracterelor tipărite, care a devenit practic o componentă uzuală a soft-urilor ce se comercializează odată cu dispozitivele de scannare, și de recunoașterea scrisului de mână cu caractere de tipar unde există progrese importante, recunoașterea caracterelor scrise de mână cursiv este încă la nivelul experimentelor de laborator, relativ departe de atingerea unor performanțe acceptabile pentru uzul public [LAZ 97, HEU 98]. Această situație se datorează unui număr de caracteristici ale scrisului cursiv de mână care modifică substanțial datele problemei față de situația tipăriturilor [PAR 95]. Cele mai importante dintre acestea sunt:

- variația formei literelor;
- înclinarea variabilă a scrisului;
- separarea dificilă a cuvintelor în litere.

Forma literelor variază substanțial de la un individ la altul și chiar în cazul scrisului aceluiași om apar diferențe notabile între un text scris atent și o notiță fugară. Înclinarea scrisului este de asemenea o caracteristică proprie fiecărui individ și produce în mod evident dificultăți suplimentare la separarea caracterelor componente din interiorul unui cuvânt [PAL 96].

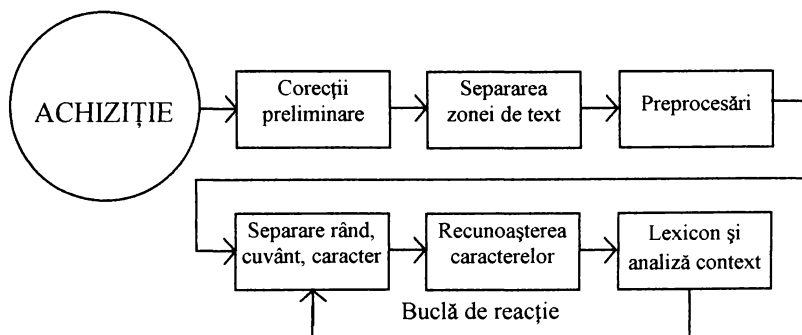


Figura 1.2.5. Schema bloc a unui sistem de recunoaștere a textelor scrise de mână cursiv

Schema bloc a unui sistem de recunoaștere a scrisului cursiv, propusă de autorul acestei teze în [LAC 97e], este prezentată în figura 1.2.5. Unele blocuri funcționale ale schemei corespund unor operațiuni similare cu cele din cazurile

descrise anterior. De exemplu separarea zonei de text prin detecția regiunilor cu densitate mare de contururi este relativ similară cu cea din cazul scrisului cu litere de tipar. Alte blocuri sunt însă similare doar ca nume cu cele din situațiile anterior prezentate. Ele corespund unor proceduri mai complexe deoarece problema recunoașterii caracterelor de mână impune o mult mai mare flexibilitate. În plus, rezultatele experimentale ale unor cercetări recente au condus la concluzia că, pentru a putea examina caracteristicile individuale ale scrisului unui individ, este mai utilă analiza pe o imagine nebinarizată și nesupusă procedurilor de subțiere. În acest fel se pot pune în evidență diferențele de apăsare la trasarea unor subcomponente ale caracterelor [TRI 97, DOR 98]. Separarea rândurilor și cuvintelor în documentele olografe este realizată în mod similar cu separarea din celelalte cazuri, dar trebuie avute în vedere la scrisul de mână următoarele probleme suplimentare [BAE 98]:

- posibilitatea unor suprapuneri parțiale ale zonelor ascendente și descendente ale rândurilor alăturate;
- înclinări accidentale ale rândurilor;
- suprapunerea sau legarea accidentală a cuvintelor consecutive.

În general aceste situații defavorabile nu sunt frecvente în cazul unor documente scrise îngrijit și prin urmare este acceptată ca soluție respingerea recunoașterii la apariția lor.

Problema cea mai complexă și delicată la „citirea” scrisului cursiv de mână, este separarea cuvintelor în caracterele care le compun. În cadrul cercetărilor curente s-au adoptat două tipuri de strategii [MOH 96]:

- analiza simultană a unui set de posibile separări;
- recunoașterea cuvântului ca un întreg.

Ambele situații presupun utilizarea controlului final într-un dicționar și întoarcerea pe bucla de reacție în caz de nevalidare [KIM 97]. Dicționarele sunt diferite conform situației specifice în cele două cazuri. Organizarea internă și caracteristicile acestor dicționare se vor analiza mai pe larg în paragraful 3.5 ”Validarea recunoașterii”. Analiza contextuală este prezentă în marea majoritate a implementărilor practice ale algoritmilor destinați recunoașterii scrisului cursiv. În cazul aplicațiilor de uz general este utilizată de obicei o analiză gramaticală automată, dar pentru situațiile particulare unde există corelații de informație între diferite zone ale textului, sau unde informația cuprinsă în textul descifrat poate să corespundă doar unui număr limitat de cazuri, se pot face analize de context specifice mult mai eficiente [SPI 97].

Analiza comparativă a celor trei tipuri de proceduri de citire automată duce la concluzia că acestea sunt similare doar în formă, fiind substanțial diferite în conținut. Acest fapt introduce complicații mari la rezolvarea unor aplicații în care eșantioanele ce trebuiesc procesate conțin în mod aleator diferite tipuri de scriere.

Această situație, existentă spre exemplu la sortarea scrisorilor, face necesară utilizarea unor echipamente capabile să execute simultan și prelucrările specifice interpretării textelor tipărite și cele pentru scrisul de mână, fie el cursiv sau cu litere de tipar. Așa cum se va arăta ulterior, o importanță esențială în această aplicație particulară o are detecția și analiza codului poștal alfanumeric care cuprinde litere și cifre discrete chiar și în cazul adreselor scrise cursiv de mână [USP 98, LAP 98].

1.2.2 Sisteme de recunoaștere on-line

Sistemele de recunoaștere on-line, denumite și dinamice, sunt sistemele destinate recunoașterii caracterelor scrise cu stilouri electronice pe un ecran sensibil la atingere. Există mai multe tipuri de asemenea dispozitive în care sistemul stilou-ecran funcționează pe baze mecanice, magnetice, sau optice [NIS 95, CHE 97]. Informația de analizat este disponibilă instantaneu în momentul marcării fiecărui pixel component al fiecărui caracter. Prin urmare, fluxul de date conține pentru fiecare pixel, pe lângă coordonatele spațiale, și o coordonată timp care specifică momentul activării sale. În plus, este disponibilă în mod ne-echivoc informația referitoare la ridicarea și coborârea stiloului de pe ecran. Aceste două elemente suplimentare fac posibilă departajarea subcomponentelor care se suprapun, fie că ele aparțin unui același caracter, sau unor caractere diferite [SEN 94].

Informația temporală se obține prin eșantionarea cu o frecvență constantă a fiecărei locații (x,y) a ecranului sensibil. Rezultă un flux de informație unidimensional care este similar cu cel specific sistemelor de recunoaștere a vorbirii și prin urmare tehnicile specifice acestora se pot aplica cu mult succes [SAB 95]. În multe din metodologiile de recunoaștere aplicate în cazul sistemelor dinamice se procedează la reparametrizarea coordonatelor spațiale, preferându-se reprezentarea în termen de lungimi de segmente, unghiuri, curburi. O problemă delicată specifică acestui tip de recunoaștere este cea a “elementelor întârziate”, linii și puncte adăugate de obicei după terminarea cuvântului (în cazul alfabetului românesc punctele pe “i” și “j”, sedilele la “ș” și “ț”, parte ascendentă la “ă”, “î”, linia transversală la “t”). Există trei direcții principale de acțiune [WIL 97]:

- neglijarea acestor date;
- asocierea “elementului întârziat” cu litera peste care se suprapune;
- analiza contextului vizual înconjurător.

Prima metodă este, evident, expusă la erori suplimentare, dar deoarece alfabetul englez nu are decât bara transversală la “t” și punctele pe “i” și “j”, în

lumea anglo-saxonă și nu numai, se apreciază în situații necritice că acesta este un sacrificiu rezonabil pentru reducerea complexității algoritmului.

Asocierea elementului întârziat cu litera peste care se suprapune parțial sau total este metoda folosită curent pentru bara transversală a literelor "ı", dar este utilă și în situația caracterul "ç" al alfabetului francez sau pentru caracterele având semne diacritice adăugate de obicei ulterior în contact direct cu corpul principal al literei.

În sfârșit, analiza "contextului vizual înconjurător" al elementului întârziat, adică a vecinătății apropiate acestuia, permite efectuarea unor asocieri între el și caracterul cel mai apropiat. Astfel se rezolvă situația pentru asocierea punctelor la literele "ı" și "j", respectiv a semnelor diacritice neafiate în contact cu corpul principal. De exemplu un algoritm pentru citirea scrierii arabe, unde apar multe asemenea elemente, trebuie obligatoriu să cuprindă un asemenea bloc funcțional [BAD 95]. Pentru multe din alfabetele țărilor europene trebuiesc concepute metodologii mixte care să se poată comporta flexibil în situațiile echivoce când semnul diacritic este sau nu plasat corect în raport cu corpul principal. Cazul literelor "ș" și "ț" din alfabetul românesc la care sedila poate să fie legată sau nu de corpul principal al literei funcție de stilul de scris al subiectului, este ilustrativ în acest sens [LAC 98b].

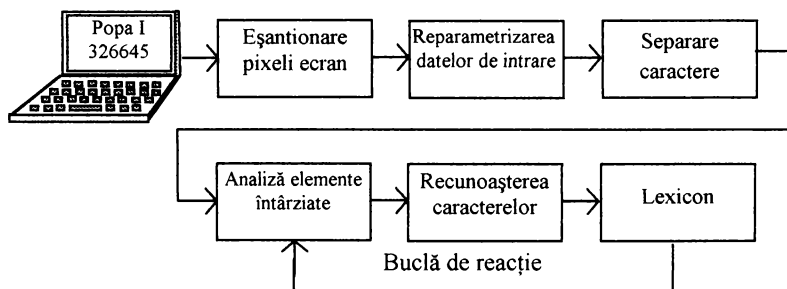


Figura 1.2.6. Schema bloc a unui echipament de recunoaștere on-line a caracterelor scrise de mână

De obicei sistemele on-line sunt sisteme personalizate. Instrucțiunile de folosire cer utilizatorului să parcurgă o etapă inițială de antrenament în care agenda personală învață să-i recunoască scrisul prin analiza unor eșantioane având identitatea precizată. O asemenea procedură cuprinde fie scrierea repetată a fiecărui caracter al alfabetului, fie, mai eficient, scrierea repetată a unui set de cuvinte standard [NIS 95]. Verificarea cuvintelor recunoscute se face în mod uzual prin căutarea într-un dicționar care poate fi, funcție de metoda utilizată, o colecție de descrieri ale cuvintelor sau un lexicon. Primul tip se aplică de obicei în situațiile în care vocabularul posibil este limitat la un număr de cuvinte specifice unui anumit domeniu de interes în aplicația respectivă. Schema bloc a unui sistem de

recunoaștere on-line, propusă de autorul acestei teze în [LAC 97e], este prezentată în figura 1.2.6. Bucla de reacție se întoarce la etapa de analiză a apartenenței elementelor întârziate deoarece alocarea lor la diferite caractere poate să producă diferențe substanțiale în rezultatele recunoașterii caracterelor și implicit a cuvintelor. Lexiconul este de tipul dicționar și are un caracter general având în vedere caracterul general al notițelor ce se fac pe ecran [CHE 97].

1.2.3 Identificarea semnăturilor

Analiza și clasificarea semnăturilor olografe este un caz special al recunoașterii caracterelor scrise cursiv de mână. Așa cum s-a spus mai sus, rezultatul recunoașterii nu este identificarea cuvântului scris ci o certificare a autenticității înscrisului procesat [HAN 96]. Există în practică două moduri de abordare a problemei corespunzând unor situații diferite:

- analiza grafologică automată a unei semnături personale sau a unui eșantion de text olograf;
- validarea accesului într-o zonă de securitate pe baza identificării autorului semnăturii.

Analiza semnăturii în ambele cazuri are în comun faptul că, într-o fază anterioară, se înmagazinează în memorie mai multe semnături ale unei persoane și prin corelare se obține un standard individual [SAB 95, SAB 97].

În cazul analizei unui text oarecare procedura constă în identificarea, în eșantionul analizat, a idioscriptului individual (formele tipice ale caracterelor sau subcomponentelor lor în scrisul unui subiect). Acest idioscript este de asemenea detectat prin analiza comparativă a unor eșantioane (cuvinte tipice sau caractere izolate) scrise de persoana în cauză.

Analiza grafologică automată, face parte din grupul procesărilor off-line deoarece examinarea documentului se face ulterior scrierii lui. Documentul studiat fie el semnătură sau porțiune de text este introdus prin scanare sau un procedeu similar în memorie [WU 97].

Identificarea autorului unui înscris oarecare se face prin identificarea în textul studiat a unor forme standard de caractere sau părți de caractere. Recunoașterea semnăturilor personale, utilizată deja în automatizarea procesării cecurilor bancare, pornește de la compararea anvelopei eșantionului studiat cu cea a semnăturii prototip existentă în memorie. În etapa a doua se poate efectua și căutarea unor drepte sau bucle tipice [SEN 94].

Problema validării unor semnături personale în cazul autorului unic se rezolvă fie prin compararea directă, fie prin determinarea diferenței între vectorii descriptori ai eșantionului prototip și ai celui analizat. Complexitatea algoritmului de implementare crește însă, dacă studiul presupune detectarea autorului dintr-un grup de autori posibili. În această situație minimizarea erorilor cere vectori descriptori cu structură complexă și clasificarea prin metode mai sofisticate (rețele neuronale, tehnici fuzzy) [PAR 95].

Identificarea semnăturilor pentru aplicații de securitate este făcută prin proceduri on-line, subiecții fiind supuși testului semnăturii pe un ecran sensibil la atingere. Validarea presupune identitatea următoarelor caracteristici globale și locale între prototipul din memorie și eșantionul procesat:

- forma caracterelor și subcomponentelor lor;
- ordinea trasării subcomponentelor;
- viteza de scriere;
- durata totală a execuției semnăturii;
- anvelopa semnăturii.

Existența mai multor autori posibili este caracteristică acestui tip de aplicație, dar având în vedere disponibilitatea informației temporale rezolvarea problemei identificării corecte este mult ușurată față de cazul anterior[WU 97].

1.3 Psihologia citirii

Înainte de a efectua o analiză detaliată a metodelor și sistemelor de recunoaștere a caracterelor, este utilă o prezentare a cunoștințelor actuale legate de modul în care creierul uman îndeplinește această sarcină. Analiza atentă a citirii la om este, în mod evident, o sursă importantă de informații care să permită mai buna înțelegere a desfășurării întregului proces și de a evidenția care sunt caracteristicile importante pe care se bazează recunoașterea caracterelor și care sunt fenomenele esențiale ce intervin. De asemenea, se pot trage concluzii utile cu privire la granița dintre variațiile formei caracterelor pe care subiecții umani le acceptă ca nederanjante și cele care în anumite condiții conduc la erori.

Există la ora actuală un număr relativ mare de studii privitoare la acest subiect între care pentru informații utile în domeniul recunoașterii caracterelor scrise de mână se pot cita: [DOW 82, TAY 83, RAY 89]. Studiile lor s-au concentrat pe citirea caracterelor individuale sau a cuvintelor izolate. Cercetările s-au făcut cu ajutorul unor aparate numite tahistoscoape. Acestea permit afișarea unei imagini conținând o literă sau un cuvânt pe un interval foarte scurt de timp în

622.671/
1.36.6.0

fața subiecților experimentului. Apoi este afișată o grilă alb negru concepută special pentru inhibarea memoriei iconice, care ar permite păstrarea în minte a imaginii inițiale pe un timp nedefinit. O concluzie interesantă derivată din aceste cercetări este faptul că recunoașterea cuvintelor ca un întreg și nu ca o sumă a caracterelor componente este o modalitate frecventă și importantă de citire.

În literatura de specialitate [RAY 89] este descris experimentul efectuat de W. Reicher prin care se dovedește importanța contextului în recunoașterea caracterelor individuale. Subiecții au recunoscut mai bine un anumit caracter în interiorul unui cuvânt decât într-un non-cuvânt. În ambele cazuri timpul de expunere al imaginilor s-a redus până la limita la care au apărut erorile. Explicația stă în utilizarea de către creierul uman a unui lexicon intern organizat probabil printr-o dublă ordonare după modul cum sună cuvintele (criteriul fonetic) și după cum arată ele scrise (criteriul vizual) [DOW 82].

În sprijinul acestei afirmații în [DOW 82] sunt citate o serie de cercetări neurologice efectuate în Japonia de Yamadori și Sasanuma, care au relevat că distrugerea unei anumite zone a creierului poate afecta sever abilitatea citirii textelor scrise în alfabetul silabic Kana în vreme ce capacitatea de a recunoaște scrierea de tip hieroglific Kanji este mult mai puțin diminuată. Prin urmare, recunoașterea caracterelor și recunoașterea cuvintelor ca entitate unitară este realizată în creier pe căi diferite și prin urmare este probabilă desfășurarea procesului pe ambele căi simultan. Analiza acestor concluzii precum și rezultatele experimentelor pe diferite grupuri de subiecți au condus la formularea ipotezei că mecanismul citirii poate fi modelat ca o succesiune de trei etape distincte [TAY 83]:

- Recunoașterea întregului cuvânt (primele 50...100 ms). Recunoașterea se bazează pe imaginea de ansamblu a cuvintelor: lungime, anvelopă etc.
- Identificarea extensiilor unor litere (după minimum 50ms) În cazul nereușitei în prima etapă, cititorul utilizează unele detalii pregnante cum sunt extensiile ascendente și descendente ale unor litere componente. Unele cercetări atribuie un rol în această etapă identificării unor părți importante ale cuvintelor cum sunt prefixele și sufixele.
- Recunoașterea literelor individuale. În cazul insuccesului în pașii anteriori se trece la procesul lent de recunoaștere a caracterelor componente una câte una.

Recunoașterea întregului cuvânt se face pe baza caracteristicilor de ansamblu ale acestuia, dar un rol extrem de important îl au aici două elemente suplimentare:

- existența cuvântului în vocabularul cititorului,
- contextul în care apare cuvântul.

Experiențele arată că rata de eroare la recunoaștere a unor subiecți cu nivel mediu de instruire în citirea unor non-cuvinte este de 3,2% pentru caractere tipărite,

4,4% pentru caractere discrete scrise de mână și 28% pentru cuvinte cursive [SUE 93]. Un argument suplimentar în acest sens este dat de diferența dintre specialiști și nespecialiști la citirea și înțelegerea unui text de specialitate dintr-un anumit domeniu. De exemplu rețetele farmaceutice sunt în mod tradițional cifrate pentru neofiți dar clare pentru medici, farmaciști și unii bolnavi cu “experiență”.

Contextul în care se găsește cuvântul analizat joacă de asemenea un rol important în recunoaștere. Cuvântul englez prezentat în figura 1.3.1 poate fi ușor interpretat ca “dump”, “clump” sau chiar “Jump” funcție de propoziția sau fraza în care se găsește [SEN 94].



Figura 1.3.1. Cuvânt interpretabil funcție de context [SEN 94]

Detecția unor detalii pregnante ale unui cuvânt în etapa a doua a citirii reprezintă în unele cazuri cheia identificării cuvântului. Spre exemplu cuvântul “minimum” din figura 1.3.2 devine un “zig-zag” absolut neinteligibil în absența celor două puncte pe “i”.

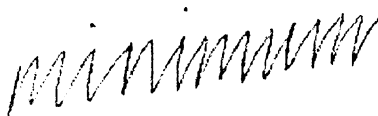


Figura 1.3.2. Cuvânt scris ambiguu [SEN 94]

Investigații aprofundate cu privire la caracteristicile esențiale ale unui cuvânt [MCG 94] au sugerat idea că recunoașterea acestuia se bazează într-o bună măsură pe identificarea unor caracteristici tipice pentru litere în structura lui generală. A treia etapă, recunoașterea caracterelor componente separat, este cea mai lentă din întregul proces. H Bouma a demonstrat cu ajutorul experimentelor descrise în lucrarea sa [BOU 71] că în această etapă, procesul de recunoaștere se bazează, în mod substanțial, pe percepția formei literei și a structurii sale spațiale. Bouma a investigat pe un grup de subiecți care sunt caracteristicile folosite de oameni pentru a diferenția caracterele izolate în situația prezentării acestora de la distanță sau

pentru un interval de timp redus. Erorile comise au fost analizate și a reieșit că se pot identifica șapte grupe de litere ale alfabetului englez “apropiate psihologic” și deci susceptibile de a fi confundate. Aceste grupuri de litere sunt prezentate în tabelul 1.3.1.

Tabel 1.3.1. Clasificarea Bouma a caracterelor [BOU 71]

Contur exterior	Forma Bouma	Cod	Caractere
Centrale	anvelopă rectangulară	1	a, s, z, x
	anvelopă rotundă	2	e, o, c
	extensii oblice	3	r, v, w
	extensii drepte verticale	4	n, m, u
Ascendente	extensii ascendente	5	d, h, k, b
	filiforme	6	t, i, l, f
Descendente	descendente	7	g, j, p, q, z

Urmare a acestei clasificări, autorul studiului a propus codarea numerică a cuvintelor funcție de literele pe care le conțin. Prin urmare cuvântul “pol” va avea codul 726 diferit de “pod” care are codul 725, dar la fel cu “goi”. Rezultă de aici că anvelopa cuvintelor este o caracteristică utilă în clasificare cu toate că păstrează foarte multe ambiguități ca și cea semnalată între “pod” și “goi”. Dacă însă acestui criteriu i se adaugă și recunoașterea primei litere, atunci, în situația unui vocabular limitat la cuvintele compatibile cu contextul, datele sunt adesea suficiente pentru o decizie univocă.

Procedeul are însă în mod evident limitele sale. În primul rând, caracterele scrise de mână variază substanțial de la un subiect la altul. În plus, chiar descrierea unui caracter scris caligrafic poate să fie făcută în moduri diferite pentru unele litere. De exemplu litera “d” poate fi considerată ca o structură complexă cuprinzând:

- o buclă închisă în zona centrală și deasupra o linie verticală scurtă în zona ascendentă;
- o buclă deschisă în zona centrală și adiacent o linie verticală lungă ce ocupă zona centrală și ascendentă;
- o buclă închisă în zona centrală și suprapus în partea dreaptă o linie verticală lungă ce ocupă zona centrală și ascendentă.

Diferențele sunt aparent nesemnificative, dar corespund diferitelor stiluri de scriere și influențează substanțial recunoașterea, mai ales în cazul sistemelor on-line unde contează ordinea trasării subcomponentelor.

Rezultatele testelor psihologice sunt foarte utile în optimizarea algoritmilor utilizați în sistemele de recunoaștere a caracterelor scrise de mână. Cunoașterea și folosirea caracteristicilor esențiale în identificarea literelor și cuvintelor, evidențiate de cercetările citate mai sus, micșorează sensibilitatea la variațiile scrierii umane. Problemele rămân însă dificile deoarece textele care trebuiesc “citite” sunt scrise uneori extrem de criptic, mai ales în cazurile în care autorul și-a făcut adnotări grăbite destinate doar uzului propriu. Cazul acestor notițe diferențiază și complică recunoașterea scrierii de recunoașterea vorbirii, unde elementul cercetat este din start conceput pentru asigurarea unui transfer de informație cu exteriorul. Situația defavorabilă menționată apare mai ales în cazul textelor scrise pe agendele cu ecrane sensibile la atingere [SEN 94].

1.4 Concluzii

Citirea automată a unui text este o procedură complexă de achiziție prelucrare și analiză a imaginii documentului ce conține acel text. În literatura de specialitate este prezentat, un mare număr de soluții concrete de implementare a procesului în ansamblu și a fiecărei etape componente în parte. Diferențele apar funcție de caracteristicile specifice aplicației deservite și mai ales ale tipului de scriere supus prelucrării.

Analiza comparativă, efectuată în cursul acestui capitol, între sistemele on-line și off-line conduce fără echivoc la concluzia că primele au performanțe net superioare rezultate din însuși modul de achiziție al datelor primare. Sistemele on-line au însă o gamă limitată de aplicații și prin urmare, principala direcție de cercetare în domeniu rămâne perfecționarea echipamentelor și metodologiilor off-line care permit citirea indiferent de modul de preluare a imaginii documentelor ce trebuie procesate. Pe parcursul prezentării, s-au evidențiat de asemenea diferențele substanțiale care decurg din necesitatea analizei celor trei tipuri principale de documente: tipărituri, scris de mână cursiv sau discret. Cele trei tipuri de proceduri de citire automată sunt similare doar în formă, fiind substanțial diferite în conținut.

Analiza documentelor tipărite este deja relativ bine pusă la punct. Există pe piață programe care realizează transformarea imaginilor în fișiere conținând un text ASCII similar cu cel original. Rata de eroare în situații tipice este de ordinul zecimilor sau chiar sutimilor de procent ceea ce de obicei este suficient [ABU 98].

În recunoașterea scrisului de mână cu caractere de tipar (discret) performanțele sunt mai puțin spectaculoase deoarece intervine, variația formei și dimensiunii caracterelor de la un individ la altul și chiar între două documente scrise de același individ. Aplicația tipică ce ilustrează acest caz este procesarea

automată a datelor înscrise cu mâna în formularele tipărite prezente în compartimentele de relații cu publicul ale băncilor, firmelor, administrației financiare și locale [BRE 97].

Recunoașterea textelor scrise cursiv este mai delicată și mai plină de dificultăți aflându-se încă la nivelul experimentelor de laborator, relativ departe de atingerea unor performanțe acceptabile pentru uzul public [HEU 98]. Această situație se datorează problemelor specifice scrisului cursiv de mână: variația formei literelor, înclinarea variabilă a scrisului și mai ales separarea dificilă a cuvintelor în litere. Forma literelor variază de la un subiect la altul și chiar în cazul scrisului aceluiași individ apar diferențe notabile între un text scris atent și o notiță fugară. Înclinarea scrisului este de asemenea o caracteristică proprie fiecărei persoane și produce dificultăți suplimentare la separarea caracterelor componente din interiorul unui cuvânt [PAL 96]. Rezultatele obținute la segmentarea cuvântului în litere sunt încă departe de a fi perfecte fiindcă nu se pot elimina cazurile de neseparare a unor caractere vecine sau apariția unor frontiere false care secționează un caracter și îl transformă în două litere/cifre distincte [BAE 98]. În consecință ratele de eroare și respingere ale recunoașterii caracterelor scrise cursiv sunt net mai proaste decât cele din cazul tipăriturilor.

Citirea automată a scrisului cursiv de mână rămâne un domeniu de cercetare deschis, pentru care este necesară o îmbunătățire substanțială a performanțelor și o diversificare a soluțiilor astfel încât să fie satisfăcute cerințele unui număr mereu în creștere de aplicații. În cadrul tezei de față autorul propune o serie de soluții de îmbunătățire a performanțelor obținute la citirea documentelor scrise cursiv de mână. Ele vizează creșterea eficienței și preciziei algoritmilor pentru realizarea separării cuvintelor în literele componente și recunoașterea efectivă a acestora. Procedura de citire este abordată într-o concepție modulară, implementarea fiind făcută etapizat.

2. Procesări preliminare

Analizând schemele bloc din paragraful 1.2 “Metode de recunoaștere a caracterelor alfanumerice” se poate observa că etapa esențială a clasificării caracterelor este în mod obligatoriu precedată de o serie de etape preliminare. Acestea au ca scop transformarea imaginii originale, de exemplu a unei întregi pagini din documentul analizat, într-o serie de sub-imagini fără deficiențe calitative, conținând fiecare câte o singură literă sau un singur cuvânt funcție de metodologia de citire aplicată [CED 97].

Dezideratul enunțat presupune în mod evident un număr de procesări succesive menite să rezolve următoarele aspecte [SCH 92]:

- eliminarea imperfecțiunilor tehnice ale imaginii de intrare (zgomote, defocalizări, contrast insuficient, iluminare neomogenă etc.);
- eliminarea regiunilor imaginii ce cuprind altceva decât text;
- procesarea specifică a sub-imaginilor conținând regiunile de tip text;
- segmentarea succesivă, din sub-imaginile analizate, a rândurilor, a cuvintelor și în final a caracterelor.

În cele ce urmează se va examina utilitatea fiecăreia dintre aceste procesări, condițiile de aplicare în diferite situații și diverse modalități concrete de implementare, comparându-se avantajele și deficiențele fiecăreia în parte.

2.1 Preprocesări

În cazul sistemelor off-line, imaginea achiziționată a documentului prezintă o serie de deficiențe având un efect negativ asupra performanțelor sistemului de recunoaștere [ROS 82, POP 87, JAI 89, PRA 91, TOM 91, GON 93, SON 93]. Aceste probleme apar, fie la preluarea efectivă a imaginilor, fie pe lanțul de transmisie până la echipamentul de procesare. Perturbațiile pe calea de transmisie afectează chiar și sistemele on-line, alimentându-le cu date parțial incorecte. Efectul zgomotului suprapus peste semnalul util poate fi atenuat prin două modalități:

- reducerea zgomotului prin îmbunătățirea hard-ului;
- filtrarea imaginii finale.

O atenție specială în etapă de achiziție trebuie acordată iluminării documentului, deoarece iluminarea neuniformă a documentului preluat creează probleme suplimentare în etapa de segmentare.

În plus, în cazul sistemelor off-line, și dispozitivul de preluare, scanner sau cameră CCD, poate să prezinte disfuncționalități. Astfel, scannerile mai vechi pot să aibă deficiențe datorate uzurii la sistemul mecanic, care duc la schimbarea ratei de transfer, având ca efect distorsiuni geometrice ale imaginii preluate. Camerele CCD reglate inadecvat redau imagini defocalizate, cu contrast redus, mânjite sau „mișcate”, iar în cazul unor deficiențe ale lentilei apar și distorsiuni geometrice.

Aceste situații impun utilizarea unor tehnici și metode de prelucrare destinate restaurării imaginilor, între care cele mai uzuale, în cazul cunoașterii modelului matematic al fenomenului perturbator, sunt filtrele inverse [PRA 91].

Prin urmare, din rațiunile enumerate mai sus, sistemele performante de recunoaștere optică a caracterelor trebuie să aibă în componența lor un bloc de preprocesare în care imaginea inițială este prelucrată astfel încât să fie reduse la maximum efectele erorilor apărute.

De obicei, în sistemele comerciale, proiectanții adoptă o atitudine optimistă, estimând că marea majoritate a problemelor de mai sus sunt de mică amploare și din rațiuni de preț utilizează la intrare doar un filtru simplu de netezire. În practică, există două tipuri de neteziri: cea temporală și cea spațială. Netezirea temporală presupune existența unui număr $N > 2$ de imagini identice, cu excepția zgomotului suprapus peste cadrul analizat [ROS 82]. Acest lucru este ușor de realizat în cazul când dispozitivul de captare este o cameră.

Se presupune o situație în care zgomotul $v_i(k, l)$, suprapus peste pixelii $a_i(k, l)$ ai imaginilor succesive A_i cu $i \in \{1, 2, 3, \dots, N\}$, este o variabilă aleatoare independentă cu media nulă și deviația standard σ . Rezultatul medierii temporale a_m în locația (k, l) este:

$$a_m = \frac{a_1 + a_2 + \dots + a_N}{N} + \frac{v_1 + v_2 + \dots + v_N}{N} \quad (2.1.1)$$

unde cel de al doilea termen este tot un zgomot cu valoare aleatoare independentă de conținutul imaginilor A_i , cu media zero, dar cu dispersia standard σ_m :

$$\sigma_m = \frac{\sigma}{\sqrt{N}} \quad (2.1.2)$$

Deci medierea temporală permite obținerea unei imagini netezite, având zgomotul atenuat, fără reducerea clarității inițiale, adică:

$$a_m(k,l) = \frac{1}{N} \sum_{i=1}^N a_i(k,l) \quad (2.1.3)$$

În cazul în care nu se dispune de mai multe imagini se execută operația de netezire spațială. Nivelul de gri final al fiecărui pixel depinde printr-o relație liniară de nivelul inițial de gri al vecinilor săi și al său. Se implementează prin convoluționarea imaginii cu o matrice numită „mască spațială” [POP 87].

Cea mai simplă metodă de reducere a zgomotului în cazul de față este medierea în interiorul vecinătății pixelului procesat [JAI 89]. Fiecare pixel este înlocuit cu media valorilor pixelilor vecini. Există două variante de lucru: medierea octo-vecinilor, realizată cu matricea h_8 din relația (2.1.4), sau doar a quadro-vecinilor, realizată cu matricea h_4 din (2.1.5):

$$h_8 = \frac{1}{9} \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} \quad (2.1.4)$$

$$h_4 = \frac{1}{5} \begin{vmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{vmatrix} \quad (2.1.5)$$

Matricele pătrate de dimensiuni mici 3×3 sau 5×5 sunt cele mai des utilizate deoarece nu afectează substanțial claritatea imaginii [TOM 91]. Expresia matematică principală a procedurii este prin urmare:

$$b(m,n) = \frac{1}{N} \sum_{k \in V} \sum_{l \in V} a(k,l) \quad (2.1.6)$$

unde $a(k,l)$ și $b(m,n)$ reprezintă nivelul de gri la intrare și respectiv la ieșire al pixelului procesat, iar V este vecinătatea de N pixeli centrată pe acesta. În mod evident, abaterea de la media locală a unor pixeli afectați de zgomot va fi micșorată cu condiția ca numărul pixelilor perturbați din vecinătatea V considerată să fie mic în comparație cu N . Dacă această condiție nu este îndeplinită pentru o anumită dimensiune N a vecinătății de mediere, atunci se poate recurge la mărirea ei la $N_1 > N$. Creșterea lui N are însă ca rezultat nedorit reducerea clarității imaginii, iar acest lucru, în cazul imaginilor unor documente scrise, produce dificultăți suplimentare în etapele ulterioare de segmentare și recunoaștere.

O modalitate de reducere a acestui efect numit de mănșire (blurring) este ponderarea pixelilor mediați funcție de distanța față de centrul vecinătății [PIT 93]:

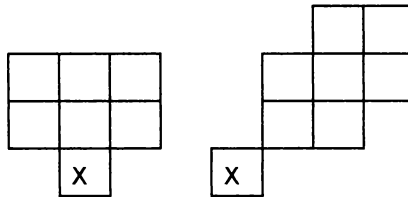
$$b(m,n) = \sum_{(k,l) \in V} c(k,l) \cdot a(k,l) \quad (2.1.7)$$

unde $c(k,l)$, $(k,l) \in V$ sunt coeficienții cu care se face ponderarea pixelilor $a(k,l)$ din vecinătatea V . Valorile acestor coeficienți respectă, în general, o anumită distribuție statistică, cel mai adesea fiind utilizată distribuția gaussiană [GON 93].

Filtrul de netezire cu prag este o variantă îmbunătățită utilizată în practica curentă [SON 94]. Specificitatea sa constă în faptul că înlocuirea pixelului procesat cu media ponderată a vecinilor săi se face doar dacă este satisfăcută condiția:

$$|a(k,l) - b(k,l)| < T \quad (2.1.8)$$

unde T este un prag de decizie stabilit astfel încât să fie protejate tranzițiile caracter/fond din imaginea originală de efectul de „mânjire”. Rezultatele aplicării acestei metode sunt foarte bune, dar apare problema dificilă a automatizării selecției pragului T .



Cu X s-a marcat pixelul supus prelucrării
Figura 2.1.1. Vecinătate direcțională $V(\theta_i)$ [TOM 91]

În literatură [TOM 91] este recomandată utilizarea medierii în vecinătatea cu dispersie minimă dintre cele P vecinătăți direcționale $V(\theta_i)$ definite similar cu cele exemplificate în figura 2.1.1 și cu $P \in \{2,4,8\}$. Calculul dispersiei celor k pixeli din vecinătatea $V(\theta_i)$, $\sigma(\theta_i)$ se face cu relația:

$$\sigma^2 = \frac{1}{k} \left\{ \sum_{(k,l) \in V(\theta_i)} [a(k,l) - \frac{1}{k} \sum_{(k,l) \in V(\theta_i)} a(k,l)]^2 \right\} = \frac{1}{k} \left\{ \sum_{(k,l) \in V(\theta_i)} [a(k,l)]^2 - \frac{[\sum_{(k,l) \in V(\theta_i)} a(k,l)]^2}{k} \right\} \quad (2.1.9)$$

iar setul de valori posibile ale pixelului sunt:

$$b(m,n,\theta_i) = \sum_{(k,l) \in V(\theta_i)} c(k,l) \cdot a(k,l) \quad \text{pentru } i=1,2,\dots,N \quad (2.1.10)$$

unde vecinătatea $V(\theta_i)$ este formată din pixelii din V aflați pe direcția θ_i față de pixelul central sau apropiați de această direcție. După efectuarea calculului pentru toate cele N direcții θ_i considerate, valoarea pixelului $b(m,n)$ se adoptă conform relației:

$$b(m,n) = b(m,n,\theta_p) \quad \text{unde} \quad \sigma(m,n,\theta_p) = \min \{ \sigma(m,n,\theta_i) \} \quad (2.1.11)$$

O soluție relativ simplă și eficientă pentru reducerea zgomotului fără deteriorarea clarității imaginii este filtrul median [PRA 91]. Expresia matematică a transformării este dată de relația:

$$a(k,l) = \text{mediana} \{ a(m,n), (m,n) \in V \} \quad (2.1.12)$$

Valoarea inițială $a(k,l)$ a pixelului este înlocuită cu mediana seriei ordonate crescător de valori $a(m,n)$ corespunzând nivelelor de gri ale pixelilor din vecinătatea V a acestuia. Dezavantajul major al metodei mai sus prezentate este riscul ca punctele izolate componente ale literelor ‘i’, ‘j’, ‘a’ etc. să fie și ele eliminate din imagine.

Netezirea efectuată cu ajutorul măștilor spațiale este utilă în eliminarea zgomotului sub formă de impulsuri sau chiar a degradărilor sub formă de bandă subțire dar, dacă zonele afectate au dimensiuni mai mari, se recomandă reluarea procedurii de preluare a imaginii. În cazul în care acest lucru nu este posibil trebuie să se apeleze la procesări din categoria restaurărilor. Cele mai des utilizate sunt filtrarea inversă și filtrarea Wiener [ROS 82]. Filtrarea inversă se bazează pe constatarea că degradarea unei imagini poate fi modelată matematic de relațiile:

$$a(k,l) = (b * h)(k,l) + v(k,l) \quad \text{domeniul timp} \quad (2.1.13)$$

$$B(u,v) = A(u,v)H(u,v) + N(u,v) \quad \text{domeniul frecvență} \quad (2.1.14)$$

unde $h(k,l)$ este funcția de transfer care produce degradarea, $v(k,l)$ descrie zgomotul suprapus pe imaginea originală, $a(k,l)$, $b(k,l)$ este imaginea degradată, iar majusculele semnifică transformatele lor Fourier.

Rezultă prin urmare că teoretic se poate obține imaginea originală transformând ecuația (2.1.14) în:

$$A(u,v) = [B(u,v) - N(u,v)] \cdot H^{-1}(u,v) \quad (2.1.15)$$

Din păcate, funcțiile h și v nu sunt de obicei cunoscute cu acuratețe și, prin urmare, filtrarea nu are efectul scontat. Pentru determinarea lui h se poate face o preluare a unui cadru standard și din comparație rezultă cu bună aproximație funcția de transfer.

Zgomotul fiind aleator, nu poate fi modelat cu exactitate. Se utilizează modele teoretice ideale care se aleg în funcție de cunoștințele apriorice privitoare la natura zgomotului din aplicația concretă. De exemplu, considerând o cameră defocalizată într-un lanț de transmisie fără zgomot, transformata Fourier a funcției de degradare (mânjire) este conform [SON 93]:

$$H(u, v) = \frac{J_1(ar)}{ar} \tag{2.1.16}$$

în care J_1 este funcția Bessel de ordinul întâi, $r^2 = u^2 + v^2$, iar a este defocalizarea. Restaurarea prin filtrare Wiener dă o estimare a imaginii originale $a(m, n)$ cu o eroare minimă pătratică e^2 garantată:

$$e^2 = \varepsilon\{[a(m, n) - \hat{a}(m, n)]^2\} \tag{2.1.17}$$

Dacă $\hat{a}(m, n)$ este o combinație liniară a valorilor din imaginea degradată $b(m, n)$, estimata e aproape de optim și minimizarea ecuației este ușoară. Estimata \hat{a} este egală cu optimul teoretic dacă imaginile a , b și zgomotul v sunt procese stohastice omogene, iar densitățile lor de probabilitate au distribuții gaussiene. Estimata imaginii originale se poate determina din relația [JAI 89]:

$$\hat{A}(u, v) = H_W(u, v) \cdot G(u, v) \tag{2.1.18}$$

unde $H_W(u, v)$ este filtrarea Wiener definită de ecuația

$$H_W(u, v) = \frac{H^*(u, v)}{|H(u, v)|^2 + \frac{S_{vv}(u, v)}{S_{bb}(u, v)}} \tag{2.1.19}$$

cu H transformata Fourier a funcției de degradare, H^* conjugata complexă a acesteia, iar S_{bb} și S_{vv} densitățile spectrale ale imaginii degradate și respectiv ale zgomotului.

Determinarea funcției de transfer h a degradării și modelarea matematică a zgomotului se face prin metode similare cu cele descrise anterior. În condițiile în care zgomotul este suficient de mic încât să poată fi considerat absent, iar dispersia $S_{vv} = 0$, expresia funcției H_W se simplifică:

$$H_w(u, v) = \frac{H^*(u, v)}{|H(u, v)|^2} \quad (2.1.20)$$

Pe lângă filtrele prezentate mai sus, există multe alte metode mai complexe și mai performante. În general însă, marea majoritate a aplicațiilor permit reluarea achiziției documentului ce trebuie analizat, astfel încât să se elimine problemele de imagine [SCH 92].

Evident că fiecare din aceste proceduri, fie ele mai simple sau mai complicate, consumă un anumit timp. De aceea, dacă este necesară prelucrarea în timp real, este absolut obligatoriu ca imaginile de intrare să fie de bună calitate, așa încât să se poată trece direct la prelucrările specifice recunoașterii sau să fie necesare doar filtrări simple de eliminare a zgomotului [CED 97].

2.2 Binarizarea

În sistemele off-line, în funcție de metodologia aplicată, imaginile documentelor ce trebuie analizate sunt achiziționate binar sau cu mai multe nuanțe de gri. În mod tradițional, înainte de aplicarea procedurilor de separare și recunoaștere a caracterelor, se aplică binarizarea imaginii [YON 94, ZLA 94, OH 95, TRI 95, FAN 95b, LEE 96c, CHE 97, DEB 97, LIU 97, TAN 97, LU 98].

Totuși, unele cercetări mai recente tind să acrediteze ideea că la procesarea scrisului de mână, este mai avantajos să se achiziționeze și analizeze imagini cu mai multe nuanțelor de gri. Se evidențiază astfel apăsarea variabilă exercitată asupra instrumentului de scris rezultând contraste locale diferite față de fond. Nuanțele de gri împreună cu grosimea originală a liniilor permit, prin urmare, analiza mai nuanțată a tușeilor proprii idioscriptului personal al fiecărui individ. Tehnicile ce utilizează imagini nebinarizate sunt, în mod evident, mai laborioase, dar literatura recentă este unanimă în aprecierea că rezultatele sunt mai mult decât promițătoare [WAN 93, YAN 94b, LEE 95, ROC 95, SHI 96].

În cazul tehnicilor on-line, unde preluarea datelor se face direct de pe ecranul sensibil la atingere al echipamentului, problema binarizării imaginii nu se pune. Datele preluate sunt implicit binare deoarece pixelii imaginilor sunt fie marcați de instrumentul de scris, fie nu.

Deoarece însă binarizarea este folosită întotdeauna pentru citirea automată a textele tipărite și în literatura recentă apar metodologii de procesare a scrisului de mână off-line care, din rațiuni de simplitate, utilizează imagini binare [CED 97], includerea prezentării acestei prelucrări în lucrarea de față rămâne utilă.

În mod ideal, se poate considera foaia de hârtie ca o suprafață albă omogenă, iar caracterele negre. Într-o asemenea situație teoretică, binarizarea se face simplu printr-o segmentare cu prag global [GON 93], adică o comparație a nivelului de gri al fiecărui pixel $a(m,n)$ din imaginea A , de dimensiune $N \times N$, cu un prag general în întreaga imagine T și alocarea valorii $b(m,n)$ după cum urmează:

$$b(m,n) = \begin{cases} 1 & a(m,n) \geq T \\ 0 & a(m,n) < T \end{cases} \quad \text{pentru } m,n = 0,1,\dots,N-1 \quad (2.2.1)$$

Aceste supoziții sunt, însă, excesiv de optimiste pentru aplicațiile reale [DES 95, TRI 96, LU 98, DOR 98]. La preluarea documentelor de analizat, fie cu scannerul, fie cu o cameră CCD, imaginile achiziționate sunt afectate de erori ce creează dificultăți suplimentare:

- oricât de bună ar fi iluminarea, foaia apare în imagine cu umbre (preluare cu camere CCD);
- documentele mai vechi nu mai au culoarea constantă din cauza îmbătrânirii hârtiei;
- scrierea suprapusă pe hârti și figuri are fundal de culoare;
- scrierea de mână conține tușee de grosime și nuanță de gri variabilă;
- caracterele de tipar nu au o nuanță de gri uniformă fie datorită scrierii cu diferite culori, fie mai des din cauza unei calități proaste a tiparului.

Pentru a minimiza erorile de binarizare produse de aceste imperfecțiuni ale datelor de intrare, se utilizează variante ale segmentării cu prag cu performanțe îmbunătățite:

- tehnica YDH;
- tehnici liniare adaptive;
- tehnica funcției integrale;
- tehnica contrastului local.

În cele ce urmează se va prezenta pe scurt algoritmul matematic al fiecărei metode în parte.

2.2.1 Tehnica YDH

Tehnica de față a fost inițial propusă de Yasunda, Dubois and Huang [YAS 80] ca etapă preliminară într-o schemă de codare. Algoritmul constă în aplicarea în succesiune a patru procesări simple:

- A. expandarea gamei dinamice;
- B. netezirea;
- C. detecția pragului adaptiv;
- D. segmentarea imaginii.

A. Expandarea gamei dinamice

Se consideră o imagine $A = \{a(m,n), m,n = 0,1,\dots,N-1\}$. Pixelii $a_1(m,n)$ ai imaginii expandate se calculează cu relația:

$$a_1(m,n) = \frac{a(m,n) - \min(A)}{\max(A) - \min(A)} \quad (2.2.2)$$

unde:

$$\begin{aligned} \min(A) &= \min\{a(m,n)\} \\ \max(A) &= \max\{a(m,n)\} \end{aligned} \quad \text{pentru } m,n = 0,1,\dots,N-1 \quad (2.2.3)$$

Imaginea rezultată are gama dinamică mărită, ceea ce înseamnă contururi bine puse în evidență, dar, în mod inevitabil, nivelul de zgomot va fi de asemenea amplificat. Prin urmare este necesară o etapă de filtrare.

B. Netezirea

Se execută prin medierea neponderată a octo-vecinilor:

$$a_2(m,n) = \begin{cases} a_1(m,n) & \text{pentru } r(m,n) > T_1 \\ \sum_{(m',n') \in I(m,n)} \frac{1}{8} \cdot a_1(m',n') & \text{î n rest} \end{cases} \quad (2.2.4)$$

în care T_1 este un parametru predefinit și $r(m,n)$ se calculează cu relația de mai jos:

$$r(m,n) = \max_{(m',n') \in C(m,n)} \{a_1(m',n')\} - \min_{(m',n') \in C(m,n)} \{a_1(m',n')\} \quad (2.2.5)$$

unde $C(m,n)$ este:

$$C(m,n) = \{(m',n'); \text{ cu } |m' - m| \leq 1; |n' - n| \leq 1; i(m',n') \neq (m,n)\} \quad (2.2.6)$$

C. Detectia pragului adaptiv

Se divide imaginea rezultată $A_2 = \{a_2(m, n), m, n = 0, 1, \dots, N-1\}$ în K blocuri $B_i, i = 0, 1, \dots, K-1$ de dimensiunea $M \times M, (N/K < M < N)$, care se suprapun parțial. Pentru un pixel oarecare $a_2(m, n)$ din blocul B_i , este valabilă relația:

$$a_3(m, n) = \begin{cases} 256 & \text{pentru } \max\{B_i\} - \min\{B_i\} < T \\ a_2(m, n) - \frac{\min\{B_i\}}{\text{med}\{B_i\} - \min\{B_i\}} 255 & \text{ori } a_2(m, n) \geq \text{med}\{B_i\} \\ \text{med}\{B_i\} & \text{î n rest} \end{cases} \quad (2.2.7)$$

în care:

$$\begin{aligned} \max\{B_i\} &= \max_{(m,n) \in B_i} \{a_2(m, n)\} \\ \min\{B_i\} &= \min_{(m,n) \in B_i} \{a_2(m, n)\} \\ \text{med}\{B_i\} &= \sum_{(m,n) \in B_i} \frac{a_2(m, n)}{M \times M} \end{aligned} \quad (2.2.8)$$

Odată rezolvată problema selecției pragului, imaginea poate fi segmentată pentru a delimita în mod neechivoc corpul caracterelor alfanumerice cuprinzând nuanțele închise, de fondul de nuanță deschisă.

D. Segmentarea

$$a_{bin}(m, n) = \begin{cases} 1 & \text{pentru } \min_1 \{a_3(m, n)\} < T_2 \\ & \text{ori } \sigma(m, n) > T_3 \\ 0 & \text{î n rest} \end{cases} \quad (2.2.9)$$

unde:

$$a(m, n) = \begin{cases} \min_1 \{a_3(m, n)\} = \min_{\substack{-1 \leq i \leq 1 \\ -1 \leq j \leq 1}} a_3(m - i, n - j) \\ \sqrt{\frac{\sum_{\substack{-1 \leq i \leq 1 \\ -1 \leq j \leq 1}} [a_3(m - i, n - j) - \text{med}_1 \{a_3(m, n)\}]^2}{3}} \\ \text{med}_1 \{a_3(m, n)\} = \frac{\sum_{\substack{-1 \leq i \leq 1 \\ -1 \leq j \leq 1}} a_3(m - i, n - j)}{9} \end{cases} \quad (2.2.10)$$

iar pragurile T_2 și T_3 sunt de asemenea predefinite

2.2.2 Tehnici liniare adaptive

Această metodă propusă de White și Rohrer [WHI 83] utilizează comparația nivelului de gri al fiecărui pixel $a(m,n)$ cu o medie a nivelurilor de gri ai pixelilor din vecinătatea sa, $V(m,n)$. Vecinătatea $V(m,n)$ are o dimensiune aproximativ egală cu mărimea unui caracter. Calculul se face cu consum minim de memorie și timp utilizând relațiile neliniare:

$$\begin{aligned} a_H(m,n) &= a_H(m-1,n) + W_H [a_H(m-1,n) - a_b(m,n)] \\ a_V(m,n) &= a_V(m,n-1) + W_V [a_V(m,n-1) - a_v(m,n)] \\ a_{bin}(m,n) &= \begin{cases} 1 & \text{pentru } Z_b |a(m-L,n)| > a_v(m,n) \\ 0 & \text{în rest} \end{cases} \end{aligned} \quad (2.2.11)$$

unde:

- W_H și W_V sunt două funcții de predicție predefinite. Ambele funcții sunt reprezentate cu aproximație în figura 2.2.1. Ele au valori nule pentru argumente nule și o valoare între 0 și valoarea argumentului în rest;

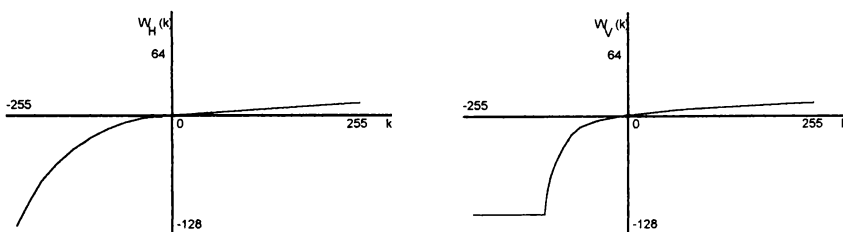


Figura 2.2.1. Funcțiile de predicție W_H și W_V [WHI 83]

- Z_b este un coeficient cu valoare conformă cu cea din graficul din figura 2.2.2;
- L este un factor de predicție;
- $a_H(m,n)$ este o medie pe orizontală calculată cu nivelele de gri ale tuturor pixelilor cu coordonata m' ce satisface relația $0 \leq m' \leq m$;

- $a_H(m,n)$ este o medie locală calculată cu nivelele de gri ale tuturor pixelilor cu $a(m',n')$ ce satisfac condițiile $0 \leq m' \leq m$ și $0 \leq n' \leq n$.

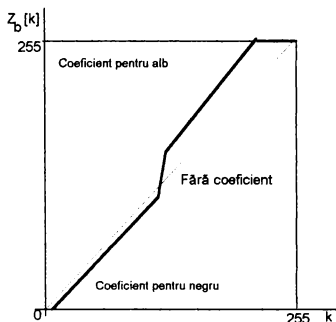


Figura 2.2.2. Coeficientul Z_b [WHI 83]

Analizând graficele celor două funcții se poate observa că nivelul mediu de ieșire nu iese din gama de valori de la intrare și converge la valoarea de intrare pentru arii cu nuanță de gri uniformă. Coeficientul Z_b este utilizat ca offset al nivelului deciziei și pentru a elimina zgomotul suprapus pe fundal.

Dezavantajul major al acestei tehnici este acela că la ieșire se văd „umbre” pe fundal și de asemenea apare partea din stânga sus a petelor întunecate de pe hârtie. Motivul acestor deficiențe este faptul că media locală $a_H(m,n)$ se calculează pe direcția stânga sus.

Prin urmare o îmbunătățire a performanțelor se poate obține aplicând algoritmul iterativ prima oară exact ca mai sus și a doua oară utilizând media locală a pixelilor din dreapta jos. Valoarea finală se determină prin execuția unui ȘI logic între rezultatele celor două etape succesive.

În [MOH 93] se arată însă că implementările experimentale ale acestui algoritim au viteza de lucru redusă, iar introducerea ȘI-ului provoacă erodarea accentuată a unor linii din componența caracterelor.

2.2.3 Tehnica funcției integrale

Algoritmul de față este cunoscut deasemenea ca tehnica celei de a doua derivate propusă de Palumbo, Swaminathan și Srihari [PAL 86]. Procedura cuprinde următoarele trei etape succesive:

- A. derivarea;
- B. etichetarea;
- C. analiza.

A. Derivarea

Se detectează pixelii plasați în apropierea tranzițiilor bruște ale nivelului de gri în imagine prin utilizarea derivatei a II-a :

$$a''(m,n) = \frac{\partial^2 a(m,n)}{\partial m \partial n} \quad (2.2.12)$$

implementată practic prin intermediul unei corelații cu o mască spațială, de obicei Laplacian-ul

$$\begin{vmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{vmatrix} \text{ sau } \begin{vmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{vmatrix} \quad (2.2.13)$$

B. Etichetarea

Se etichetează cu '+' pixelii din partea întunecată a tranzițiilor detectate și cu '-' cei din partea luminoasă. Expresia matematică a acestei acțiuni este:

$$a_1(m,n) = \begin{cases} 0 & \text{dacă } a(m,n) < T \\ - & \text{dacă } a(m,n) \geq T; a''(m,n) < 0 \\ + & \text{dacă } a(m,n) \geq T; a''(m,n) > 0 \end{cases} \quad (2.2.14)$$

C. Analiza

Se scanează orizontal și vertical imaginea $a_1(m,n)$ pentru a detecta și marca secvențele de etichete de tipul:

$$S(0,-,+)=,-,+,\dots[a_1(m,n)=0 \text{ sau } +],\dots,+,- \quad (2.2.15)$$

pentru a evidenția tranzițiile datorate liniilor componente ale caracterelor. Lungimea acestor secvențe reprezintă grosimea acestor linii, grosime ce trebuie să fie cam

aceeași în întreg documentul analizat. Se ia în considerare doar textul obișnuit, deoarece titlurile, sublinierile și alte efecte tipografice speciale se pot abate de la această regulă.

În final se analizează secvențele. Contururile unor pete pe hârtie precum și zgomotele nu dau secvențe de genul celor din relația (2.2.15) sau au o lungime afară din gama acceptată pentru caractere și deci pot fi eliminate.

2.2.4 Tehnica contrastului local

Metoda a fost propusă de Giuliano, Paitra și Stringa [GIU 77]. Fiecare pixel în imaginea de ieșire este calculat folosind următorul operator spațial:

$$O = \begin{vmatrix} O_2 & O_2 \\ O_1 & \\ O_2 & O_2 \end{vmatrix} \quad (2.2.16)$$

unde O_1 și O_2 sunt de forma:

$$O_1 = \begin{vmatrix} x & x & x \\ x & (x,y) & x \\ x & x & x \end{vmatrix}; \quad O_2 = \begin{vmatrix} x & x & x \\ x & x & x \\ x & x & x \end{vmatrix} \quad (2.2.17)$$

În ariile pe care se suprapune O_2 se iau în considerare doar pixelii care satisfac condiția $a(m,n) > T_2$. În continuare se notează acești pixeli cu $c(m,n)$.

$$c(m,n) = \begin{cases} a(m,n) & \text{pentru } a(m,n) > T_2 \\ 0 & \text{pentru } a(m,n) \leq T_2 \end{cases} \quad (2.2.18)$$

Pentru conciziune algoritmul este prezentat prin intermediul unei scheme logice în figura 2.2.4. Parametrii implicați în calcule sunt:

- $a(m,n)$ valoarea inițială a pixelului analizat;
- $s(m,n)$ este media locală, calculată cu valorile $a(m,n)$, pe regiunea pe care s-a suprapus matricea O_1 ;
- q_i sunt mediile locale, calculate cu valorile $c(m,n)$, ale regiunilor peste care s-au suprapus matricele O_2 ;

- T_1 este un parametru predeterminat utilizat pentru extracția tuturor pixelilor al căror nivel de gri este mai mic decât el, adică pixelii aparținând caracterelor (similar cu metoda pragul global);
- T_2 este un parametru predeterminat necesar pentru detecția pixelilor din regiunile peste care se suprapun matricile O_2 care au nivel de gri mai mic decât el;
- T_3, T_4, T_5 sunt parametrii predeterminați utilizați pentru comparația lui $s(m,n)$ cu q_i ;

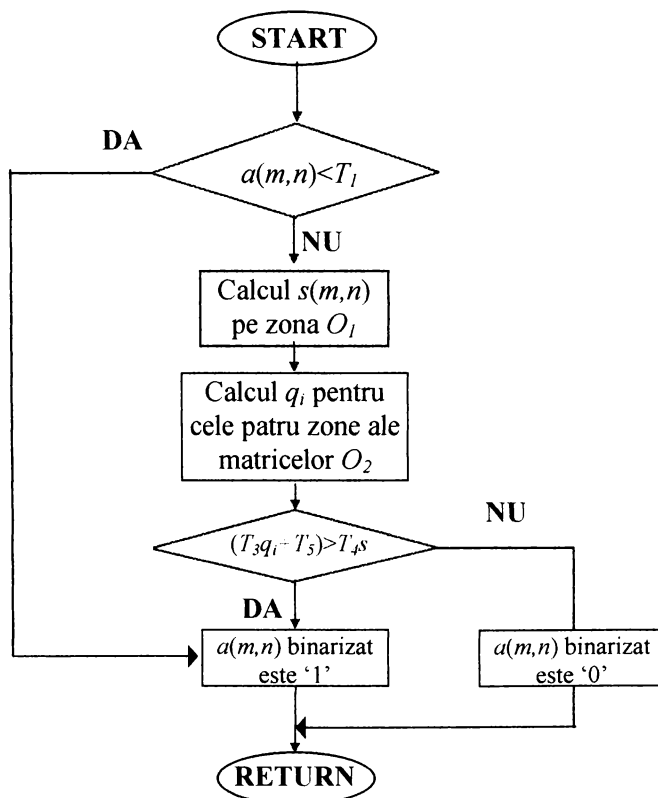


Figura 2.2.3. Schema logică a funcției de binarizare prin tehnica contrastului integral

Implementările algoritmilor descriși anterior și ai altora similari cu aceștia se pot face atât pe echipamente uni-procesor cât și pe rețele de transputere sau alte sisteme cu procesare paralelă. În ultimul timp odată cu dezvoltarea și perfecționarea

tehnicii rețelelor neuronale, au apărut în literatura de specialitate articole care raportează utilizarea cu succes a unor rețele de tip Kohonen pentru realizarea binarizării [ZLA 94]. Valoarea finală a fiecărui pixel în imaginea binarizată este stabilită în corelație cu valoarea vecinilor săi, ponderea influenței fiind invers proporțională cu distanța în baza unei legi gaussiene.

2.3 Detecția și separarea regiunilor conținând text

De multe ori documentele ce trebuie procesate cuprind pe lângă caractere scrise și o serie de imagini, desene, grafice etc. Chiar în situația unor documente „pure” conținând doar text fără figuri sau decorații tipografice, este foarte improbabil ca imaginea să încadreze exact zonele tipărite sau scrise de mână. Din acest motiv se impune ca în imaginile procesate să fie inițial delimitate și apoi decupate porțiunile ce conțin grafică și zone „albe” pentru ca ulterior sistemul de recunoaștere a caracterelor să poată lucra corect [KAM 93, KUO 94, KAN 95, FAN 95a, ETE 97, YUE 97, JAI 98]

Metoda de separare se bazează pe proprietățile diferite ale zonelor grafice sau albe față de cele ale zonelor de text. În fapt zona scrisă este o porțiune din textura „text” și are prin urmare o structură cvasi-periodică și o densitatea net mai mare de tranziții bruște ale nuanței de gri a pixelilor în comparație cu alte zone. Indiferent care dintre cele două criterii de mai sus este utilizat, imaginea originală este împărțită inițial în regiuni de formă pătrată sau dreptunghiulară ce se analizează individual [KAM 93].

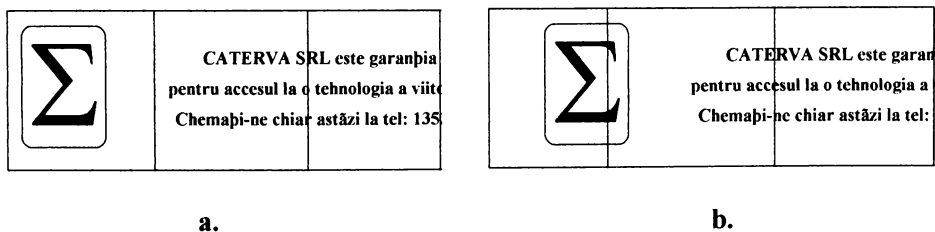


Figura 2.3.1. Împărțirea unei imagini în regiuni [KAM 93]

- a. Cazul favorabil
- b. Cazul nefavorabil

Situația prezentată în figura 2.3.1.a este o situație favorabilă, deoarece granițele regiunilor separă corect desenul emblemei de textul reclamei, ceea ce face ca segmentarea imaginii să fie completă după o singură etapă de analiză separată a sub-imaginilor create prin divizarea originalului. Cazul nefavorabil din figura 2.3.1.b având o probabilitate de apariție mult mai mare este rezolvat prin utilizarea unor etape succesive de împărțire și analiză a diviziunilor. După prima etapă apar trei categorii de sub-imagini:

- segmente de text,
- segmente de grafică inclusiv spații goale,
- segmente mixte.

Regiunile mixte se vor diviza în continuare până la obținerea unor segmente cu structură omogenă. Este posibil ca, în situația unor documente cu figuri „înghesuite” în text, la sortarea automată să rezulte în zonele scrise și porțiuni mici din grafică indiferent cât de judicios s-au stabilit pragurile de decizie la segmentare. În acest caz, fie se optează pentru un algoritm adaptiv, fie se apelează la o rețea neuronală care permite definirea unor funcții de discriminare complexe. Aplicând ultima metodă s-au obținut rezultate foarte bune chiar și în separarea scrisului pe hărți geografice și meteorologice [DES 95, TRI 97]. Prima diferență sesizabilă între regiunile de imagine cuprinzând text și cele albe sau cu figuri este densitatea substanțial mai mare a tranzițiilor abrupte de nivel de gri, adică a pixelilor de contur. Determinarea acestui parametru este, de obicei, precedată de binarizarea imaginii astfel încât contururile devin tranziții de tip treaptă $0 \rightarrow 1$ sau $1 \rightarrow 0$. Numărul tranzițiilor raportat la numărul total de pixeli din zona analizată se compară cu un sistem de praguri și, funcție de rezultatul comparației, se clasifică segmentul ca fiind text, figură, alb sau mixt.

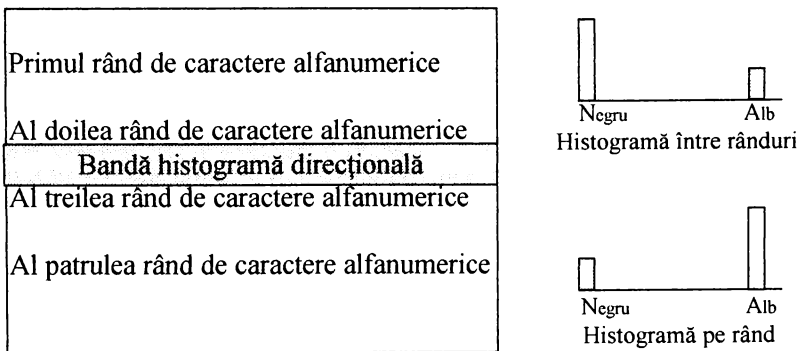


Figura 2.3.2. Utilizarea histogramelor direcționale la detecția structurii periodice a texturii „scris” [SHA 93]

O metodă simplă și eficientă [SHA 93] folosește binarizarea urmată de analiza variației valorilor histogramelor unor benzi înguste numite histogramme direcționale, așa cum se arată în figura 2.3.2. Scannarea perpendicular pe direcția scrisului va produce succesiv histogramme cu mult negru și puțin alb pe rânduri și respectiv cu mult alb și puțin negru pe spațiile dintre acestea.

Posibilitatea utilizării acestor histogramme și în etapa următoare la stabilirea granițelor între rânduri, între cuvinte și în final între litere este un important atu al metodei, dar simplitatea algoritmului este doar aparentă. În plus, exactitatea rezultatelor este puternic dependentă de paralelismul benzilor în care se calculează histogrammele cu direcția scrisului, fapt ce ridică probleme delicate de aliniere.

Este, de asemenea, pomenită în literatură metoda hibridă rezultată din combinarea detecției cu histogramme direcționale și cea utilizând densitatea pixelilor de contur. În speță este vorba despre calculul densității tranzițiilor cu ajutorul histogrammelor direcționale [YUE 97].

Decizia de clasificare a sub-imaginilor se poate lua având proprietățile structurale ale segmentului analizat, și anume în funcție de apropierea între caracteristicile acestuia și cele tipice unei porțiuni de textură „scris”. Acest tip de textură este similar atât la tipărituri cât și la scrisul de mână, având o structură cvasi-periodică datorită repetiției: “rând, spațiu liber, rând” de-a lungul unei axe perpendiculare pe direcția de scriere.

Pe baza acestei caracteristici s-au dezvoltat metode utilizând detecția periodicității prin analiza Fourier a sub-imaginilor [BUR 92]. Se pot utiliza cu succes oricare din transformările sensibile la structuri periodice: transformata Fourier rapidă, Cosinus, Sinus, Haar, Slant etc. Condiția principală de selecție a uneia dintre aceste transformări este viteza de lucru. Segmentele sunt declarate ca având structură tip „scris” dacă este întrunită condiția ca vectorul descriptor format din amplitudinile primelor N componente ale transformatei să fie suficient de apropiat de vectorul descriptor al prototipului zonei cu textură “scris”. Metoda are rezultate foarte bune dar este laborioasă și consumă timp pe echipamente de tip MISD chiar dacă se utilizează algoritmi rapizi. Dacă se dispune de sisteme MIMD, cum ar fi de exemplu rețelele de transputere, acest fapt devine mai puțin deranjant [PIT 94].

2.4 Separarea rândurilor, cuvintelor și caracterelor

Ca urmare a procedurilor prezentate în paragraful anterior s-au obținut din imaginea inițială un număr de sub-imagini cuprinzând zone de text având contrastul

îmbunătățit astfel încât să se evidențieze cât mai pregnant caracterele de fundal. Pentru a permite unui sistem automat să recunoască literele și cifrele din textul analizat este necesară delimitarea fiecăreia în parte. Operația presupune trei etape distincte[VER 92]:

- separarea rândurilor;
- separarea cuvintelor în interiorul rândurilor;
- delimitarea spațiului afectat fiecărui caracter.

Fiecare dintre aceste stadii de procesare a documentului presupune utilizarea unor proceduri de segmentare a textului, având rolul de a diviza succesiv imaginea inițială până la sub-imagini ce conțin câte un singur caracter. Metodele utilizate pentru segmentarea textului în rânduri și cele pentru divizarea rândurilor în cuvinte au un număr important de caracteristici comune și vor fi discutate în comun. Tehnicile de separare a caracterelor componente ale unui cuvânt sunt, de obicei, mult mai complexe și prezintă diferențe importante între variantele destinate textelor tipărite și cele pentru documentele scrise de mână. Prin urmare autorul a optat pentru prezentarea lor distinctă.

2.4.1 Separarea textului în rânduri și cuvinte

Separarea textului în rânduri și cuvinte reprezintă primele două trepte de segmentare spațială a imaginii [SHA 93, DEF 95, LIK 95, CED 97].. O metodă utilizabilă în special în cazul textelor bătute la mașină, este propusă în [SHA 93]. Metoda folosește proiecțiile pe o axă verticală pe direcția rândurilor conform cu cele prezentate în figura 2.4.1. Proiecțiile pe axa paralelă cu rândul indică spațiul aferent cuvintelor. În cazul în care rândurile sunt înclinate se aplică o variantă a metodei care ia în considerare separat părți ale acestora.

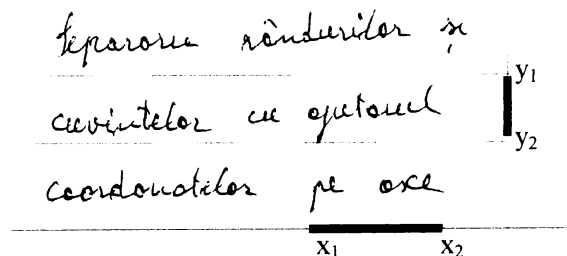


Figura 2.4.1. Determinarea spațiilor aferente unui rând al unui document și al cuvintelor ce-l compun [SHA 93]

În general, metoda dă rezultate bune la documentele tipărite sau la cele scrise de mână dar cu caractere ordonate și constante. În aceste situații se poate realiza chiar și o separare a caracterelor considerând lățimea medie a fiecărei litere.

Metoda cea mai eficientă utilizată la separarea rândurilor scrise de mână sau de tipar este analiza variației valorilor minime și maxime ale unor histograme direcționale, similare cu cele utilizate în paragraful anterior [DEF 95].

În cazul analizei unor documente olografe în care este posibilă o anumită înclinație a rândurilor se recomandă detecția inițială a acestei înclinații. În acest scop se poate utiliza Transformata Hough care este capabilă să pună în evidență orientările liniilor din imaginea analizată [PRA 91]. Expresia matematică a Transformatei Hough este:

$$(x,y) \rightarrow (r, \theta) \quad (2.4.1)$$

unde:

$$r = x \cos \theta + y \sin \theta \quad (2.4.2)$$

Imaginea transformată va prezenta două zone cu densitate mare corespunzătoare direcției rândurilor și înclinării scrierii față de aceasta. Cu această informație preliminară se poate stabili unghiul optim pentru histogramele direcționale și, în final, se pot stabili rândurile de text. Astfel, regiunea scrisă din imagine sau sub-imaginea se împarte în benzi, cuprinzând fiecare câte un rând. În interiorul acestora urmând să se facă separarea cuvintelor.

Literatura de specialitate descrie două tipuri de frontiere de separare a rândurilor:

- granițe drepte [SAP 93];
- granițe ce urmăresc linia de bază a cuvintelor [LIK 95].

Algoritmul pentru trasarea frontierelor drepte este simplu și rapid, dar produce erori în cazul suprapunerii parțiale a părților ascendentă și descendentă a două rânduri consecutive. Frontierele ce urmăresc linia de bază necesită procesări mai laborioase, dar numărul erorilor de segmentare este mult redus. Totuși granițele ce urmăresc linia de bază a cuvintelor, produc “erori de amputare” a extensiilor descendente izolate ale unor litere din alfabetele naționale ale unor țări europene (de exemplu sedilele de la ‘ș’ și ‘ț’ din alfabetul românesc).

Cele mai importante metode de separare a cuvintelor pomenite în literatură [SHA 93, LIK 95, MOH 96, CED97], sunt:

- determinarea numărului de pixeli aparținând caracterelor din vecinătatea unui pixel de fundal;

- analiza densității de pixeli aparținând caracterelor.

Prima abordare pornește de la constatarea evidentă că în vecinătatea unui pixel de fundal aflat în zona ce separă două cuvinte, numărul de pixeli aparținând caracterelor este mai mic decât în vecinătatea unui pixel de fundal aflat în spațiul din interiorul cuvântului [CED 97]. Va rezulta pentru fiecare cuvânt un obiect similar cu învelișul convex al literelor ce-l formează. Ulterior aceste obiecte se erodează și se obține cu o bună aproximație spațiul aferent fiecărui cuvânt.

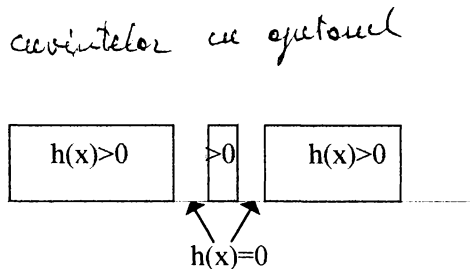


Figura 2.4.2. Separarea cuvintelor în interiorul unui rând prin metoda histogramelor direcționale

A doua tehnică reprezintă reutilizarea metodei histogramelor direcționale pentru separarea cuvintelor, așa cum se vede în figura 2.4.2. În cazul scrisului de tipar se recomandă direcția perpendiculară pe cea a rândurilor, în timp ce pentru scrisul de mână se utilizează cu rezultate mai bune înclinări de aproximativ 15° de la perpendiculară [MOH 96].

2.4.2 Separarea cuvintelor în litere

În etapa următoare, se separă literele fie prin utilizarea rezultatelor scannării cu histogramă direcțională, fie prin delimitarea spațiului aferent caracterului pe baza unei variante dilatate a acestuia. Este necesar să se sublinieze aici că în cazul scrisului de mână rezultatele obținute la segmentarea cuvântului în litere sunt încă departe de a fi perfecte. În ciuda unor intense eforturi financiare și de cercetare nu s-au putut elimina complet cazurile de apariție a unor frontiere false care secționează un caracter și îl transformă în două litere/cifre distincte care evident sunt citite separat.

În [SHA 93] este descris un algoritm care face separarea caracterelor în interiorul cuvintelor prin analiza conturului superior și a celui inferior al grupului de

litere sau cifre privit ca un întreg. Ideea se bazează pe constatarea că cele două contururi prezintă discontinuități în zonele dintre două caractere. În cazul documentelor tipărite discontinuitatea apare atât sus cât și jos, ceea ce permite o separare rapidă și corectă. Dacă metoda este aplicată scrisului de mână, atunci, în majoritatea cazurilor, pauzele din partea superioară și inferioară nu sunt sincrone, astfel încât se consideră mai multe posibilități de segmentare care se analizează separat și abia funcție de rezultatul final se stabilește soluția corectă.

O metodologie relativ nou apărută propune eliminarea separării caracterelor și analiza cuvântului ca un întreg [HO 92, WAD 93, SET 94, GEI 95, GOV 96, PAR 96]. Tehnica este considerată, fie o alternativă de sine stătătoare de citire a scrisului de mână cursiv, fie o etapă utilizată la analiza inițială a cuvintelor greu lizibile. Recunoașterea directă a cuvintelor fără analiza pe litere este utilizată cu precădere în situațiile în care vocabularul aplicației (numărul cuvintelor posibile) este limitat. Cuvântul cercetat este analizat ca un tot unitar rezultând descrierea lui matematică, un vector descriptor, cuprinzând caracteristici statistice (caracteristici Bouma, coeficienții transformatei Walsh, momente etc.) și sintactice (linii, bucle deschise sau închise, pauze). Clasificarea se face prin comparația cu un lexicon cuprinzând vectorii descriptori ai tuturor cuvintelor posibile în aplicația curentă. În cele mai multe situații validarea clasificării se face și pe baza unei analize contextuale [MES 96]. Detaliile privitoare la certificarea finală a corectitudinii clasificării cuvintelor se vor analiza în paragraful 3.5 “Validarea recunoașterii”.

2.5 Subțierea caracterelor

După procedura de segmentare se obține un șir de sub-imagini care cuprind fiecare câte un singur caracter. Fie că aceste porțiuni de imagine au un caracter binar sau conțin o paletă de nuanțe de gri, ele pot fi privite ca fiind formate din două categorii de pixeli: cei aparținând corpului caracterelor alfanumerice conținute și cei care aparțin fundalului. De asemenea, în ambele cazuri, corpul caracterelor este diferit de situația ideală, liniile și curbele componente nu au grosimea standard de un pixel. Indiferent de metodele de recunoaștere utilizate ulterior, nesatisfacerea acestei condiții generează erori de clasificare. Deci se impune cu necesitate o etapă de subțiere a caracterelor. Operația se realizează printr-o serie de erodări succesive efectuate conform unor reguli care garantează nedistorsionarea formei de bază [SAH 93, ZHO 94, DES 98]. În esență aceste reguli aplicate erodărilor succesive sunt:

- nu se erodează pixelul de capăt al unei linii sau curbe deschise - astfel se evită scurtarea ei;

- nu se erodează pixelul care produce deconectarea a două părți conexe ale aceluiași caracter;
- nu se erodează pixelul care prin eliminarea sa din corpul caracterului produce transformarea unei curbe închise într-una deschisă.

Există mai multe metode de erodare aplicabile în cazul caracterelor, utilizând seturi de matrici ce analizează pixelul candidat la erodare în contextul vecinătății înconjurătoare. Dacă structura acestei vecinătăți nu corespunde nici uneia din structurile tip care fac ca erodarea să încalce una din regulile de mai sus, atunci pixelul candidat la erodare este marcat ca fiind erodabil și analiza continuă în același fel. Marcarea unui pixel influențează modul de tratare al vecinilor săi, astfel încât pentru a elimina sensibilitatea algoritmului de analiză la ordinea în care pixelii sunt luați în considerare, procedurile clasice necesită mai multe treceri de stabilire a pixelilor care se marchează.

p3	p2	p1
p4	p	p0
p5	p6	p7

Figura 2.5.1 Vecinătatea 3×3 a pixelului p [SAH 93]

În [SAH 93] se propune metoda *Algoritm de erodare cu o singură trecere de marcare* utilizând un set de matrici care, prin modul elastic de definire al structurii lor, permit efectuarea estimării inițiale într-o singură scanare a pixelilor candidați. Articolul citat demonstrează de asemenea că în condiții similare algoritmul în discuție are performanțe superioare față de alte metode similare. Vecinătatea 3×3 a pixelului p este reprezentată figura 2.5.1. Setul de matrici cu ajutorul cărora se face analiza poziției fiecărui pixel este:

<table border="1" style="border-collapse: collapse;"> <tr><td>*</td><td>≤ b</td><td>x</td></tr> <tr><td></td><td>p</td><td>x</td></tr> <tr><td>x</td><td>x</td><td>x</td></tr> </table>	*	≤ b	x		p	x	x	x	x	<table border="1" style="border-collapse: collapse;"> <tr><td>x</td><td>x</td><td>x</td></tr> <tr><td></td><td>p</td><td>x</td></tr> <tr><td>*</td><td></td><td>x</td></tr> </table>	x	x	x		p	x	*		x	<table border="1" style="border-collapse: collapse;"> <tr><td>x</td><td>x</td><td>x</td></tr> <tr><td></td><td>p</td><td></td></tr> <tr><td>y</td><td>y</td><td>y</td></tr> </table>	x	x	x		p		y	y	y	<table border="1" style="border-collapse: collapse;"> <tr><td>x</td><td></td><td>y</td></tr> <tr><td></td><td>p</td><td>y</td></tr> <tr><td>x</td><td></td><td>y</td></tr> </table>	x		y		p	y	x		y
*	≤ b	x																																					
	p	x																																					
x	x	x																																					
x	x	x																																					
	p	x																																					
*		x																																					
x	x	x																																					
	p																																						
y	y	y																																					
x		y																																					
	p	y																																					
x		y																																					
<table border="1" style="border-collapse: collapse;"> <tr><td>*</td><td></td><td>x</td></tr> <tr><td>≤ b</td><td>p</td><td>x</td></tr> <tr><td>x</td><td>x</td><td>x</td></tr> </table>	*		x	≤ b	p	x	x	x	x	<table border="1" style="border-collapse: collapse;"> <tr><td>x</td><td></td><td>*</td></tr> <tr><td>x</td><td>p</td><td></td></tr> <tr><td>x</td><td>x</td><td>x</td></tr> </table>	x		*	x	p		x	x	x	<table border="1" style="border-collapse: collapse;"> <tr><td>x</td><td></td><td>y</td></tr> <tr><td>x</td><td>p</td><td>y</td></tr> <tr><td>x</td><td></td><td>y</td></tr> </table>	x		y	x	p	y	x		y	<table border="1" style="border-collapse: collapse;"> <tr><td>> c</td><td></td><td>x</td></tr> <tr><td>< t</td><td>p</td><td></td></tr> <tr><td>y</td><td>y</td><td>y</td></tr> </table>	> c		x	< t	p		y	y	y
*		x																																					
≤ b	p	x																																					
x	x	x																																					
x		*																																					
x	p																																						
x	x	x																																					
x		y																																					
x	p	y																																					
x		y																																					
> c		x																																					
< t	p																																						
y	y	y																																					
<table border="1" style="border-collapse: collapse;"> <tr><td>x</td><td>≤ b</td><td>*</td></tr> <tr><td>x</td><td>p</td><td></td></tr> <tr><td>x</td><td>x</td><td>x</td></tr> </table>	x	≤ b	*	x	p		x	x	x	<table border="1" style="border-collapse: collapse;"> <tr><td>x</td><td>x</td><td>x</td></tr> <tr><td>x</td><td>p</td><td></td></tr> <tr><td>x</td><td></td><td>*</td></tr> </table>	x	x	x	x	p		x		*	<table border="1" style="border-collapse: collapse;"> <tr><td>x</td><td>x</td><td>x</td></tr> <tr><td>≤ b</td><td>p</td><td></td></tr> <tr><td>#</td><td>#</td><td>#</td></tr> </table>	x	x	x	≤ b	p		#	#	#										
x	≤ b	*																																					
x	p																																						
x	x	x																																					
x	x	x																																					
x	p																																						
x		*																																					
x	x	x																																					
≤ b	p																																						
#	#	#																																					

x	≤ b	y
x	p	y
x		y

x	x	x
≤ b	p	x
*		x

în care:

- numărul de ordine - un punct final declarat în etapa k, este notat cu k;
- orice punct cu valoarea egală cu l este punct de frontieră stânga;
- orice punct cu valoarea egală cu t este punct de frontieră sus;
- orice punct cu valoarea egală cu r este punct de frontieră dreapta;
- orice punct cu valoarea egală cu b este punct de frontieră jos;
- dintre pixelii marcați cu # cel puțin unul este negru;
- x, y sunt valori oarecare;
- blanc-urile sunt puncte inițial albe;
- * sunt puncte inițial negre.

Prin urmare un pixel este neerodabil dacă vecinătatea sa are una din structurile descrise mai sus. Acesta este deci marcat ca punct final. Configurațiile din primul rând sunt pentru detecția punctelor de contur stânga l, cele de pe rândul doi pentru punctele de contur de sus t, cele de pe rândul trei pentru punctele de contur dreapta r, iar cele de pe ultimul rând pentru punctele de contur de jos b. Algoritmul controlează fiecare pixel și îl marchează, fie ca punct erodabil, fie ca punct aparținând scheletului sau punct final. Este de remarcat faptul că un pixel odată marcat ca aparținând scheletului nu mai este analizat în trecerile ulterioare. În timpul unei treceri pixelii de contur se clasifică așa cum urmează:

- un pixel negru este pixel de frontieră stânga l dacă p₄ este inițial alb,
- un pixel negru este pixel de frontieră de sus t dacă p₂ este inițial alb,
- un pixel negru este pixel de frontieră dreapta r dacă p₀ este inițial alb,
- un pixel negru este pixel de frontieră de jos b dacă p₆ este inițial alb.

Deci fiecare punct negru de frontieră este analizat o singură dată spre a decide dacă este punct erodabil sau final.

2.6 Concluzii

În cursul acestui capitol au fost prezentate principalele etape de procesare anterioare procedurii de recunoaștere a caracterelor alfanumerice. Aceste proceduri preliminare, la care sunt supuse imaginile documentelor analizate, nu reprezintă doar un simplu preambul. Succesul întregului proces depinde esențial de calitatea și precizia lor. De aceea analiza etapelor inițiale de procesare se face ținând cont de

influența pe care o au în desfășurarea ulterioară a procedurii în ansamblu și sunt evidențiate, acolo unde este cazul, diferențele de abordare recomandate de literatură pentru prelucrarea textelor tipărite și respectiv scrise de mână. O problemă este deci rezolvarea unor aplicații complexe care cer realizarea unor echipamente capabile de a trata corespunzător fiecare tip de document în parte.

Între prelucrările prezentate, s-a acordat o atenție specială celor legate de divizarea imaginii inițiale a textului analizat în sub-imagini conținând fiecare câte un singur caracter alfanumeric. Operația presupune trei etape distincte: separarea rândurilor, separarea cuvintelor în interiorul rândurilor și în final, delimitarea spațiului afectat fiecărui caracter.

În cazul textelor tipărite, tehnicile de separare actuale funcționează corect, deoarece rândurile sunt drepte și echidistante, iar caracterele sunt scrise perpendicular pe axa orizontală a acestora. Scrisul de mână este însă neregulat, rândurile sunt uneori inegal distanțate, au o anumită înclinare, cuvintele sunt nealiniat, literele sunt înclinate etc. Prin urmare, rezultatele metodelor actuale sunt afectate de multe erori. Principala deficiență constatată de autor în cursul experimentării acestor tehnici este lipsa de flexibilitate care le micșorează capacitatea de adaptare la neregularitățile scrisului de mână.

Literatura de specialitate descrie două tipuri de frontiere de separare a rândurilor: granițe drepte și granițe ce urmăresc linia de bază a cuvintelor. Algoritmii pentru trasarea frontierelor drepte este simplu și rapid, dar produce erori în cazul suprapunerii parțiale a părților ascendentă și descendentă a două rânduri consecutive. Frontierele ce urmăresc linia de bază a cuvintelor sunt mai precise. Totuși aceste granițe produc “erori de amputare” a extensiilor descendente izolate ale unor litere (de exemplu sedilele de la ‘ș’ și ‘ț’ din alfabetul românesc). Deasemenea, separarea rândurilor în cuvinte cu granițe drepte, fără a ține cont de unghiul de înclinare al scrisului conduce la greșeli de segmentare [CED 97].

Separarea literelor din cuvintele scrise cursiv reprezintă un punct critic al procesului de citire automată. Erorile pot conduce la neșegmenarea unor caractere (frontiere nedetectate) sau dimpotrivă, la divizarea incorectă a unui caracter în mai multe părți (frontiere false). Ulterior, în procesul de recunoaștere, aceste greșeli conduc inevitabil la clasificarea grupului de litere ca non-cuvinte, adică respingerea recunoașterii în etapa de verificare cu dicționarul, sau, mai grav, la acreditarea unor variante eronate de clasificare [HAM 97].

În capitolul 4 “Prezentarea sistemului de segmentare și recunoaștere a caracterelor alfanumerice scrise de mână”, autorul propune un set de soluții flexibile care conduc la îmbunătățirea performanțelor obținute la segmentarea textului în rânduri, cuvinte și caractere.

3. Recunoașterea caracterelor alfanumerice

Urmare a parcurgerii etapelor succesive prezentate în capitolul anterior s-au obținut segmente de imagine ce cuprind câte un singur caracter. În această situație sunt îndeplinite condițiile inițiale pentru a se efectua analiza și recunoașterea literelor/cifrelor scrise, alcătuirea din acestea a cuvintelor și în final estimarea veridicității lor.

Analiza și recunoașterea caracterelor este un caz particular al analizei și recunoașterii formelor, deci metodele, tehnicile și criteriile utilizate în general sunt valabile și în această situație specială. Evident, datorită specificităților citirii automate, eficiența anumitor algoritmi este mărită în raport cu cazul general, în vreme ce unele moduri de abordare a clasificării se dovedesc ineficiente sau greu de implementat practic [ROB 96].

Totuși, în lunga istorie a cercetărilor în domeniu, puține metodologii au fost complet ocolite și nu de puține ori tehnici abandonate pe parcurs au revenit în actualitate. Uneori acestea se regăsesc ca parte componentă a unei metode hibride utilizând combinarea ponderată a rezultatelor mai multor algoritmi utilizați în paralel, alții redevin atractive datorită noilor computere a căror viteză, memorie și capacitate de calcul estompează o serie de dezavantaje anterior considerate insurmontabile [VID 95]. Datorită acestor stări de fapt și frecvenței cu care apar mereu noi soluții și variante de soluții, este dificil să se organizeze materialul prezent sub forma unei înșiruii liniare de metode ierarhizate clar funcție de complexitatea și calitatea rezultatelor finale. Autorul a optat, prin urmare, pentru o grupare a metodelor pe criterii de formă și modalitate de abordare, încercând să examineze comparativ caracteristicile fiecărei categorii considerate.

S-au luat în considerare cele două grupuri fundamentale de metodologii de analiză și recunoaștere a formelor:

- metode statistice;
- metode sintactice (structurale).

La fiecare dintre ele sunt evidențiate specificitățile citirii automate și felul în care acestea influențează tehnicile concrete de recunoaștere. Datorită numărului mare de implementări raportate recent în literatura de specialitate și a rezultatelor experimentale promițătoare [CAR 92, BRE 97], autorul a acordat o atenție specială sistemelor de recunoaștere cu rețele neuronale.

Procesul de recunoaștere este, în general, o succesiune de trei etape:

- *Etapa de învățare.* Definirea din start a prototipurilor pe baza cunoștințelor apriori sau modelarea lor pe parcursul procesului pe baza experienței acumulate în lucrul efectiv.
- *Descrierea caracterului studiat.* Este vorba fie de construcția unui vector descriptor statistic, fie de alcătuirea unui model structural.
- *Decizia de clasificare.* Se ia funcție de rezultatul comparațiilor între modelul literei/cifrei necunoscute și cel al prototipului.

În cazul diferitelor metodologii descrise în literatura de specialitate există multiple variante de implementare a acestor etape. Majoritatea tehnicilor de analiză și recunoaștere respectă, pe lângă succesiunea descrisă mai sus, regula clasificării în baza unui set de descriptori. Aceștia înmagazinează în valorile lor informațiile esențiale privitoare la caracterul analizat, iar stabilirea deciziei de clasificare se face în funcție de poziția determinată de valorile lor în spațiul de decizie [SCH 92].

3.1 Recunoașterea statistică a caracterelor

Metodele statistice reprezintă, așa cum implică și numele lor, o clasificare pe baze statistice a obiectelor analizate. Acestea sunt categorisite într-o clasă, dacă proprietățile lor specifice sunt mai „aproape” de cele ale acelei clase decât de cele ale oricărei alte clase în discuție [YAN 94a, DJO 85, HAN 95, LI 96, LIO 96, HAM 97, HO 98].

Proprietățile specifice se numesc caracteristici (features) f_k , unde $k=0,1,\dots,N-1$. Aceste caracteristici se măsoară prin diverse metode. Valorile lor sunt de obicei prelucrate în vederea obținerii desensibilizării la transformările geometrice (translația, rotația, scalarea), rezultând un set de descriptori invariant. Alegerea acelor parametri ce cuprind esența informațională și construirea descriptorilor are o influență covârșitoare pentru succesul întregului proces de recunoaștere. Deocamdată nu există reguli clare de selectare a setului optim de caracteristici, una din deficiențele metodei constituind-o tocmai faptul că alegerea celei mai eficiente serii de proprietăți analizate este mai mult artă decât logică științifică [LAM 95, KIT 97]. Caracteristicile considerate funcționează ca și criterii de clasificare. Adică, reprezintă axele de coordonate ale unui spațiu de decizie N -dimensional în care cele C clase sunt regiuni disjuncte. În figura 3.1.1 este reprezentat un spațiu de decizie bi-dimensional divizat între patru clase $c_1\dots c_4$. Expresia matematică a granițelor între două clase c_q și c_p este chiar funcția de decizie sau funcția discriminant $D(x)$, având forma generală [SHA 93]:

$$D(x) = g_q(\underline{x}) - g_p(\underline{x}) \quad (3.1.1)$$

în care $g_q(\underline{x})$ și $g_p(\underline{x})$ sunt funcțiile de decizie cu privire la apartenența obiectului \underline{x} la clasele c_q și respectiv c_p .

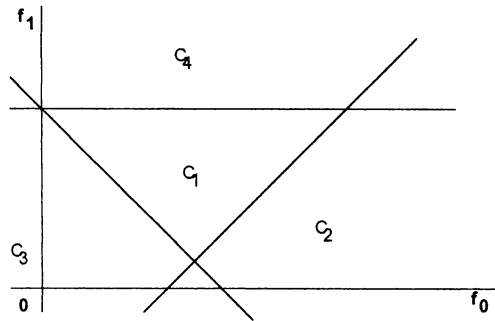


Figura 3.1.1. Spațiu de decizie cu doi descriptori și patru clase

Funcțiile de decizie sunt de obicei distanțe în spațiul de decizie între obiectul necunoscut și modelele ideale ale claselor, numite prototipuri. Există o serie de modalități diferite de definire a acestor prototipuri [JAI 89]:

- Stabilirea unui obiect real ca *reprezentant tipic (prototip)* al clasei din care face parte. Prototipul se alege dintre obiectele preatribuite claselor predefinite cât mai aproape de zona centrală a „teritoriului” acestora.
- Stabilirea apriori pe baze statistice a proprietăților *modelului ideal* al clasei. Prototipul este virtual și se construiește prin medierea caracteristicilor obiectelor reale existente în clasa respectivă la un moment inițial. Numărul și „teritoriile” claselor sunt de asemenea predefinite.
- Estimarea, după fiecare admitere a unui element nou în clasă, a *prototipului virtual* după o procedură similară celei descrise anterior. „Teritoriile” claselor sunt elastice, iar dacă un obiect nu este apropiat nici unui model existent la un moment dat, se generează o nouă clasă corespunzătoare lui.

În general, recunoașterea statistică a formelor utilizează cunoașterea apriorică a probabilităților de apariție a reprezentanților fiecăreia dintre cele C clase. În cazul recunoașterii caracterelor, având în vedere vocabularul imens și numărul mare de clase posibile - litere majuscule, litere mici și cifre, la care se adaugă în anumite situații și semne de punctuație, calcularea probabilităților este practic imposibilă. În rest, se respectă procedura de mai sus. Inițial se efectuează

măsurători asupra literei/cifrei necunoscute extrăgându-se caracteristicile $f_n(k)$ din care se construiește un vector descriptor $V_n = \{f_n(k), k=0,1,\dots,N-1\}$. Acesta este comparat cu vectorii prototip $V_{p_i} = \{fp_i(k), k=0,1,\dots,N-1\}$ ai celor C clase - în cazul de față caracterele alfanumerice. Decizia se ia funcție de distanțele dintre vectorul literei/cifrei analizate și vectorii descriptori ai reprezentanților tipici. Distanța între doi vectori se calculează cu [SCH 92]:

$$D_e(i) = |V_{p_i} - V_n| \quad i=1,2,\dots,C \quad (3.1.2)$$

Se poate adopta, spre exemplu, distanța euclidiană și prin urmare expresia (3.1.2) devine:

$$D(i) = \sqrt{\sum_{k=1}^{N-1} [fp_i(k) - f_n(k)]^2} \quad i=1,2,\dots,C \quad (3.1.3)$$

În final obiectul analizat este asimilat acelei clase j pentru care este îndeplinită condiția de minim:

$$D_j = \min(D_i, \forall i \in \{1,2,\dots,C\}) \quad i=1,2,\dots,C \quad (3.1.4)$$

Aplicația de față necesită utilizarea în general a unui număr mare de caracteristici prin a căror combinare în descriptori se elimină dezavantajele particulare ale fiecăruia și se obțin echipamente robuste de recunoaștere. Deoarece măsurarea acestor caracteristici se face prin intermediul unor algoritmi cu sensibilitate și acuratețe diferită este preferată uneori ponderarea rezultatelor lor în expresia de calcul a distanțelor:

$$D(i) = \sqrt{\sum_{k=1}^{N-1} a_k [fp_i(k) - f_n(k)]^2} \quad i=1,2,\dots,C \quad (3.1.5)$$

unde ponderile a_k sunt stabilite de obicei empiric pe baza experimentelor de calibrare [ROS 82].

Există un mare număr de criterii de discriminare și, în consecință, descriptori aferenți ce se pot utiliza ca și funcții discriminant la recunoașterea caracterelor alfanumerice.

3.1.1 Funcții discriminante polinomiale

Se selectează perechi de pixeli corespondenți din imaginea literei/cifrei analizate $A=\{a(m,n), m,n=0,1,\dots,W-1\}$ și cea a prototipului $P_i=\{p_i(m,n), m,n=0,1,\dots,W-1\}$ unde W este dimensiunea unui segment din imaginea originală ce conține un singur caracter alfanumeric.

Se efectuează produsul valorilor lor și rezultatele sunt caracteristici primare din care se construiește un descriptor de tip funcție polinomială având o expresie matematică de forma [CED 98f]:

$$f_d(C_i) = \sum_{k=0}^{K-1} c_k(i) \cdot a(m,n) \cdot p_i(m,n) \quad \text{cu } a(m,n) \in A \text{ și } p_i(m,n) \in P_i \quad (3.1.6)$$

în care $c_k(i)$ sunt coeficienții funcției discriminant liniare f_d a clasei C_i [POP 87, PRA 91, GON 93].

Experimentele citate în [SRI 93] arată că utilizarea imaginilor binarizate dă o rată de recunoaștere de aproximativ 89%, în vreme ce utilizarea imaginilor cu nuanțe de gri dă o rată ușor superioară - aproximativ 93%.

3.1.2 Corelații obiect-prototip

O metodă clasică de determinare a gradului de asemănare între un obiect și prototipul unei clase este efectuarea corelației între ele [ALT 84, MAT 92, KOP 97].

Se consideră două subimagini de dimensiune $W \times W$, prototipul $P=\{p(m,n), m,n=0,1,\dots,W-1\}$ imaginea caracterului analizat $A=\{a(m,n), m,n=0,1,\dots,W-1\}$. Nepotrivirea dintre cele două se poate exprima matematic prin una din expresiile următoare:

$$\sum_{m=0}^{W-1} \sum_{n=0}^{W-1} |a(m,n) - p(m,n)| \quad (3.1.7)$$

$$\sum_{m=0}^{W-1} \sum_{n=0}^{W-1} (\alpha(m,n) - p(m,n))^2 \quad (3.1.8)$$

A doua ecuație este recomandată de [MAT 92] ca fiind mai exactă, dar algoritmul de calcul este mai lent și prin urmare trebuie simplificat.

$$\sum \sum (a - p)^2 = \sum \sum a^2 + \sum \sum p^2 - 2 \sum \sum ap \quad (3.1.9)$$

Primii doi termeni din partea dreaptă a egalității de mai sus sunt fiși și în consecință diferențele între cele două imagini sunt mici dacă, și numai dacă, cel de al treilea termen este mare. Deci pentru A și P dați, produsul $\sum \sum \alpha(m,n) \cdot p(m,n)$ este o măsură utilizabilă a similarității celor două imagini.

Calculul acestui produs se face prin suprapunerea celor două imagini, efectuarea produselor pixel cu pixel, translatarea relativă a uneia față de cealaltă și repetarea acestei succesiuni până la terminarea tuturor suprapunerilor posibile. Expresia corespunzătoare este:

$$C(u,v) = \sum_{m=0}^{W-1} \sum_{n=0}^{W-1} \alpha(m,n) p(m+u,n+v) \quad u, v = 0, 1, \dots, W-1 \quad (3.1.10)$$

cunoscută ca inter-corelația între A și P .

Conform [PRA 91] inter-corelația normalizată este:

$$C_n(u,v) = \frac{\sum_{m=0}^{W-1} \sum_{n=0}^{W-1} \alpha(m,n) p(m+u,n+v)}{\sqrt{\sum_{m=0}^{W-1} \sum_{n=0}^{W-1} \alpha^2(m,n) + \sum_{m=0}^{W-1} \sum_{n=0}^{W-1} p^2(m,n)}} \quad u, v = 0, 1, \dots, W-1 \quad (3.1.11)$$

Domeniul de variație al inter-corelației C_n este în acest caz definit între 0 și 1, crescând odată cu gradul de similaritate între imaginea cercetată și cea a prototipului clasei considerate. Utilizarea metodei de față presupune efectuarea anterioară a unor procesări asupra imaginilor A și T , care să desensibilizeze rezultatele finale la diferențele inerente datorate poziției, unghiului de rotire și scării diferite ale caracterului cercetat și respectiv prototipului în cele două imagini.

Având în vedere faptul că în recunoașterea caracterelor, variațiile între două exemplare ale aceleiași litere sunt uneori substanțiale, analiza similarității se poate face folosind teoria funcțiilor fuzzy [JAW 95, CHI 97]. Diferența între doi pixeli ce nu coincid este în abordarea clasică 1. În varianta propusă, această diferență este influențată și de valoarea vecinilor celor doi pixeli. În acest fel se poate practic reduce influența unui pixel neconcordanț prin contribuția vecinilor săi.

3.1.3 Metoda histogramelor direcționale

Histogramele direcționale au multiple aplicații în algoritmi de separare a rândurilor, cuvintelor și literelor în interiorul acestora. Încercările de revalorificare a datelor obținute în cursul acestor preprocesări au dus la fundamentarea unei metodologii de analiză și recunoaștere a caracterelor bazată pe scannarea pe orizontală și verticală cu o asemenea histogramă [LAC 96a, CED 98e]. Rezultatele aplicării acestei proceduri sunt prezentate în figura 3.1.2.

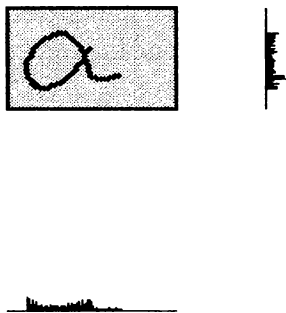


Figura 3.1.2. Utilizarea histogramelor direcționale în analiza și recunoașterea caracterelor alfanumerice

Numărul și distribuția spațială relativă a minimelor și maximelor reprezintă caracteristici utilizabile în clasificare. În cazul unor imagini fără zgomot, bine segmentate, rezultatele sunt rapide și corecte.

Deoarece acest algoritm analizează global spațiul ocupat de literă/cifră rezultatele sunt mai puțin influențate de variațiile minore ale dimensiunii și/sau poziției diverselor segmente liniare ori curbe din care este compusă litera.

Această insensibilitate implică însă și posibilitatea unor erori sistematice, de exemplu la caractere diferite, dar cu formă apropiată, cum sunt literele românești 'ă' și 'â'. Din acest motiv tehnica este folosită de obicei în combinație cu alte metode de recunoaștere în cadrul unor algoritmi complecși [LAC 98e].

3.1.4 Metoda gradientilor

Algoritmul de recunoaștere bazat pe gradienti folosește un descriptor construit pe baza magnitudinii și direcției gradientilor într-o serie prestabilită de locații ale imaginii [CED 98b, CED 98c, CED 98d].

Întâi se divizează imaginea caracterului în regiuni pătrate $k \times k$. Se calculează gradientii după un număr d de direcții și se obține un vector descriptor cu $k^2 d$ elemente, utilizat ulterior la clasificare. Calculul gradientilor se face prin convoluția imaginii literei/cifrei cu măștile spațiale ale operatorului Sobel pe diverse direcții. În continuare se dă exemplul unor matrici Sobel 3×3 având orientări din 45° în 45° .

1	2	1		0	1	2		-1	0	1	(3.1.12)
0	0	0		-1	0	1		-2	0	2	
-1	-2	-1		-2	-1	0		-1	0	1	
0°				45°				90°			

Este esențial ca imaginile astfel analizate să fie neafectate de perturbații sau foarte bine filtrate. Gradientii sunt în fapt derivate spațiale și, prin urmare, sunt extrem de sensibili la orice tip de zgomote [GON 93].

3.1.5 Momente invariante

Momentele invariante sunt una din cele mai rapide și eficiente metode de a transforma o imagine într-un set puțin numeros de caracteristici care să cuprindă în ele toate elementele cu adevărat importante din ea [LEN 92, WAK 93, WON 95, BAY 96]. Clasificarea pe baza momentelor va fi relativ insensibilă la deplasare, mărime și unghi deoarece momentele invariante sunt insensibile la cele trei transformări geometrice în discuție. În plus, deoarece relația de calcul a momentelor

este de fapt o integrală discretizată (3.1.13), sensibilitatea la zgomot va fi mai puțin pronunțată decât în cazul altor metode [LIN93].

Calculul momentelor de ordin (pq) se face cu expresia [HU 61]:

$$m_{pq} = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} u^p v^q f(u,v) \quad (3.1.13)$$

Pentru a obține invarianța la translație se calculează momentele centrale:

$$\mu_{pq} = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} (u - u_0)^p (v - v_0)^q f(u,v) \quad (3.1.14)$$

unde u_0 și v_0 sunt coordonatele centrului de greutate și se calculează cu relațiile:

$$\begin{aligned} u_0 &= \frac{m_{10}}{m_{00}} \\ v_0 &= \frac{m_{01}}{m_{00}} \end{aligned} \quad (3.1.15)$$

iar m_{00} este în cazul imaginilor binare chiar aria obiectului.

Soluția insensibilității la scalare ($u' = \alpha u$ și $v' = \alpha v$) se obține din sistemul de ecuații:

$$\begin{aligned} \eta_{pq} &= \frac{\mu_{pq}}{(\mu_{00})^\gamma} \\ \gamma &= \frac{p+q}{2} + 1 \\ \mu_{pq} &= \frac{\mu_{pq}}{\alpha^{(p+q+2)}} \end{aligned} \quad (3.1.16)$$

Rezultă momentele centrale insensibile la translație și scară având expresia:

$$\mathcal{G}_{pq} = \frac{\mu_{pq}}{(\mu_{00})^\gamma} \quad (3.1.17)$$

Insensibilitatea la rotație se realizează prin combinarea momentelor calculate cu (3.1.17) în expresii de forma:

$$\begin{aligned}
 \varphi_1 &= \mathcal{G}_{20} + \mathcal{G}_{02} \\
 \varphi_2 &= (\mathcal{G}_{20} - \mathcal{G}_{02})^2 + 4 \cdot \mathcal{G}_{11}^2 \\
 \varphi_3 &= (\mathcal{G}_{30} - 3 \cdot \mathcal{G}_{12})^2 + (3 \cdot \mathcal{G}_{21} - \mathcal{G}_{03})^2 \\
 \varphi_3 &= (\mathcal{G}_{30} + \mathcal{G}_{12})^2 + (\mathcal{G}_{21} + \mathcal{G}_{03})^2 \\
 &:
 \end{aligned}
 \tag{3.1.18}$$

Folosind deci un vector descriptor, având ca elemente momentele invariante ale unui caracter din relațiile (3.1.18), se poate face o comparație rapidă și precisă cu prototipurile tuturor caracterelor, comparație insensibilă la transformările geometrice. Pentru imaginile digitizate, care sunt discontinue, această insensibilitate este doar relativă. În literatura studiată sunt de asemenea recomandate ca fiind foarte eficiente momentele Zernike și Afine. Primele se dovedesc foarte utile în cazul unor imagini cu zgomot [KHO 90a, KHO 90b], iar celelalte au o remarcabilă insensibilitate la rotație, translație și scalare în imaginile discrete [FLU 94].

3.1.6 Funcțiile Gabor

Transformata Gabor are proprietatea că maximizează produsul între rezoluțiile din domeniile spațial și frecvență. În plus, cercetările neuro-biologice au demonstrat de asemenea că forma lor este asemănătoare cu caracteristica de transfer a celulelor corticale striate din creierul uman [LIN 91, LIS 92]. Prin urmare, un set de valori locale ale funcțiilor Gabor poate fi selectat ca și grup de caracteristici. Coeficienții lor sunt o reprezentare compactă a imaginii. Funcțiile Gabor și modul lor de utilizare sunt tratate în detaliu în [LIS 92]. În esență se divide imaginea caracterului analizat, având, spre exemplu, dimensiunile inițiale 32×32, în 9 regiuni 12×12 parțial suprapuse. Se calculează patru funcții Gabor pentru fiecare regiune și deoarece fiecare coeficient este complex, rezultă câte 8 caracteristici, adică în total 72 caracteristici utilizabile pentru clasificare.

În mod similar cu transformata Gabor se pot folosi și alte transformări. În literatura de specialitate sunt raportate rezultate bune ale unor experimente de recunoașterea caracterelor tipărite sau scrise de mână bazată pe vectori descriptori alcătuiți din coeficienți Fourier [BUR 92].

În tabelul 3.1.1 se dau ratele de recunoaștere pentru metodele mai sus enumerate în cazul unor cifre și, respectiv, caractere alfanumerice scrise de tipar, conform rezultatelor experimentale din [KHO 90a, SRI 93, SUE 93]. Așa cum se poate observa din datele cuprinse în tabel, pentru toate metodele, rezultatele

obținute la analiza și recunoașterea cifrelor sunt sensibil mai bune. Acest fapt este explicabil prin numărul mai mic de clase și, nu în ultimul rând, prin existența a mai puține asemănări între formele cifrelor arabe decât între formele caracterelor latine [LET 95, RED 98].

Tabel 3.1.1. Performanțele metodelor de recunoaștere statistice

Metode de recunoaștere statistice	Recunoaștere cifre [%]	Recunoaștere caractere alfanumerice [%]
1. Funcții polinomiale (P)	99	91,7
2. Corelații + Fuzzy (C)	96	89
3. Histograme (H)	96,7	93
4. Gradienti (Gr)	97,8	94,5
5. Momente Zerniche (M)	-	96
6. Funcții Gabor (Ga)	98	92,9
7. Combinație (P+C+Ga)	99,6	93.2

3.2 Recunoașterea sintactică a caracterelor

Conform celor arătate anterior, se poate concluziona că recunoașterea statistică a caracterelor este axată pe extragerea unor informații privind ansamblul literei/cifrei cercetate, sau a unei părți componente a acesteia, fără însă a lua în considerare relații existente între ele.

Algoritmii sintactici de recunoaștere a caracterelor se bazează pe analiza structurală, deoarece structura este, extrem de relevantă, în cazul caracterelor pentru natura literelor/cifrelor ce trebuie „citite” [CHE 95a, NIS 96a, NIS 96b, SAR 98a]. Metodele de recunoaștere sintactice sunt fundamentate pe ipoteza că orice caracter scris de mână sau tipărit este o formă complexă, creată prin alăturarea mai multor forme simple, numite primitive, într-o ordine și o dispoziție spațială anume. Descrierea se face pe baza unei cuantificări a structurii posibilă în două moduri:

- descriptori structurali;
- gramatici formale.

Aceste instrumente asigură descrierea într-o formă matematică a caracteristicilor relaționale care permit recunoașterea și clasificarea caracterului analizat.

Procesul de recunoaștere, utilizând descriptori structurali, cuprinde inițial o etapă în care este analizată structura literei sau cifrei cercetate, analiză în urma

căreia se creează o descriere, un model sintactic al acesteia. Clasificarea modelului sintactic se face practic similar cu cea a celui statistic prin comparația cu modelele prototipurilor [HA 97, JAI 97, HU 98].

În situația algoritmilor utilizând gramatici formale, pe lângă această comparație, se poate utiliza și o formă specifică. Clasificarea unui caracter într-o categorie anume este făcută dacă codul ce descrie structura acestuia este o parte a limbajului generat de gramatica ce definește categoria respectivă. Deci, în situația unui aceluiași set de primitive, comparația are loc practic între regulile de producție ale gramaticii și regulile în baza cărora forma complexă a caracterului de clasificat este descompusă într-un șir de asemenea primitive [REV 96]

Avantajul esențial, comun ambelor abordări, este insensibilitatea la micile variații ale liniilor și curbelor ce compun caracterele. Concentrându-se pe relația dintre părți, metodele sintactice nu sunt afectate în mod dramatic de rotiri, translații, scalări și, într-o anumită măsură, nici de micile schimbări la nivelul părților componente etc. [MAR 97].

Așa cum se va vedea în cele ce urmează, implementările practice ale algoritmilor bazați pe descriptori structurali au multe puncte comune cu metodologiile de recunoaștere statistică a caracterelor. Pe de altă parte, analiza și recunoașterea caracterelor în baza teoriei gramaticilor formale are un mod specific de desfășurare legat, în primul rând, de analiza inițială a construcției caracterului cercetat prin prisma regulilor de producție definite apriori.

3.2.1 Descriptori structurali

Descriptorii structurali sunt o modalitate de reprezentare a formei caracterelor care permite reprezentarea matematică a structurii lor spațiale. Această descriere privește ansamblul caracterului fără însă a analiza părțile componente ale acestuia și inter-relațiile dintre ele. Tipurile de descriptorii structurali utilizați pentru recunoașterea caracterelor sunt:

- codurile Freeman;
- descriptorii morfologici;
- descriptorii de contur.

Codurile Freeman, numite și coduri lanț, sunt o modalitate rapidă și eficientă de codare a conturilor [TOM 91, SON 94, CED 98a]. Principiul este extrem de simplu: se numerotează într-o anumită ordine direcțiile posibile de deplasare de-a lungul unui contur, așa cum se arată în figura 3.2.1. Aplicația în discuție necesită utilizarea variantei în care sunt permise opt direcții de mișcare. Utilizarea codurilor Freeman face posibilă eliminarea etapei de erodare din lanțul

procesărilor inițiale și, în consecință, tratează caracterele ca și obiecte având o anumită lățime, deci nu linii și curbe cu grosimea de un pixel [SUE 93]. Implementarea practică presupune extragerea conturului literei/cifrei analizate așa cum se arată în figura 3.2.2.b. Acest contur primar este netezit prin aplicarea metodei aproximărilor poligonale, a metodei „benzii de cauciuc”, ori alta similară și se obține conturul minim codat din. figura 3.2.2.c [PIT 93].

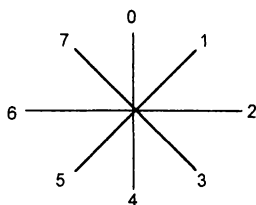
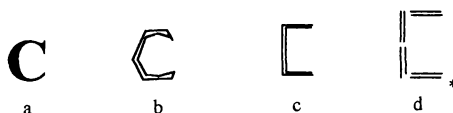


Figura 3.2.1. Numerotarea direcțiilor în vederea scrierii codurilor Freeman



*= Punct de start

Cod Freeman primar = 60026442

Cod Freeman diferențial = 20246064

Număr de formă = 02460642

Figura 3.2.2. Codarea Freeman a literei C mare de tipar

- a.** Caracterul inițial **b.** Conturul primar
c. Conturul minimizat **d.** Codarea Freeman primară

În fapt, procedura devine relativ mai complexă în cazul general, în care litera/cifra cercetată este deplasată, rotită și/sau scalată în raport cu prototipul. Trebuie analizată și eliminată sensibilitatea codului la toate aceste variații posibile. În plus codul se schimbă la schimbarea poziției punctului de start al parcurgerii conturului. Insensibilitatea la translație și rotație provine din însăși proprietatea generală a descriptorilor structurali de a analiza părțile componente și relațiile dintre ele, independent de poziția ansamblului în cadrul imaginii [SCH 92]. În schimb, scalarea poate produce perturbații în sensul creșterii sau reducerii numărului de pași necesari parcurgerii întregului contur. În figura 3.2.2.d este evident că parcurgerea liniei verticale din doi pași sau unul va aduce o schimbare esențială a lungimii și componentei codului lanț.

Soluția propusă este corelarea acestor pași cu scara, pe baza măsurării dimensiunilor caracterului cercetat și compararea cu dimensiunile omologe standard ale prototipurilor. De obicei se iau în calcul înălțimea și lățimea literelor/cifrelor și în baza raportului dintre ele se reeșantionează sub-imaginea inițială, astfel încât să se obțină un număr egal de pași necesari parcurgerii conturilor ce se vor compara în faza de clasificare. Scrierea codului lanț al caracterului pornind din puncte diferite generează coduri diferite. Eliminarea acestui inconvenient se face printr-o procedură de prelucrare suplimentară a codului Freeman primar. Din codul Freeman se calculează codul diferențial prin scrierea diferenței modulo 8 între cifrele succesive. Practic, se numără întotdeauna, în același sens, numărul de direcții din figura 3.2.1 ce separă două orientări consecutive ale conturului, adică două cifre alăturate din codul primar. Numărul de formă, care este cifra minimă obținută din permutarea circulară a codului diferențial, este constant, independent de poziția punctului de start [JAI 89].

Prin metodologia de mai sus se obține un vector V_c ce descrie forma caracterului, descriptor invariant în raport cu toate transformările geometrice. Acesta poate fi comparat cu descriptorii similari ai prototipurilor categoriilor posibile și finalmente se poate face clasificarea printr-o metodă similară celor aplicate în cazul metodelor statistice.

$$V_c = [02460642] \quad (3.2.1)$$

Deoarece s-a considerat ca nejustificat efortul de a coda ca și contur închis un caracter, metodele mai noi propun utilizarea codului Freeman primar ca și descriptor local [SUE 93, SRI 97]. Vectorul rezultat în urma acestei proceduri conține orientările conturului caracterului într-un număr de locații prestabilite în subimaginea acestuia, iar clasificarea decurge în mod absolut identic cu cea descrisă la metoda clasică. Experimentele arată că această abordare este nu doar mai rapidă, ci și mai performantă din punctul de vedere al rezultatelor finale [CED 98a].

Principalul dezavantaj al metodei este legat de sensibilitatea mare a codului de variațiile scrisului de mână. Reducerea complexității conturilor prin relaxare reduce rata erorilor, dar nu suficient pentru a satisface cerințele. Din acest motiv procedura are o aplicabilitate limitată, fiind utilizabilă doar la sistemele personalizate și, de obicei, în combinație cu alte metode [ROB 96].

Descriptorii morfologici sunt considerați de mulți specialiști drept cei mai apropiați de modul uman de percepție al caracterelor alfanumerice, deoarece rezultatul utilizării lor este o descriere a părților componente ale literelor sau cifrelor (linii, arce, găuri etc.) și a pozițiilor lor. Acest motiv a făcut ca în ei să fie puse multe speranțe legate de apropierea sau chiar egalarea performanțelor umane în ce

privește citirea. Din păcate și în acest caz previziunile s-au dovedit prea optimiste [PAV 93]. Se cunosc totuși implementări [WAN 93, WIL 97] ale unor sisteme de recunoaștere morfologice cu rețele neuronale care s-au dovedit capabile de bune performanțe în special în analiza și clasificarea cifrelor.

Metodele de recunoaștere a caracterelor utilizând descriptorii de contur reprezintă, în esență, o combinație între codurile Freeman și descrierea morfologică a caracterelor. În speță, este vorba de detectarea cu ajutorul analizei conturilor bazată pe codare Freeman a unor particularități topologice cum sunt: „golfuri”, spirale, „cioturi”, intrânduri, pauze, „spini”, intersecții, brațe etc [KIM 94, CHE 95b, PEI 95, HU 96, COM 96, WAK 97]

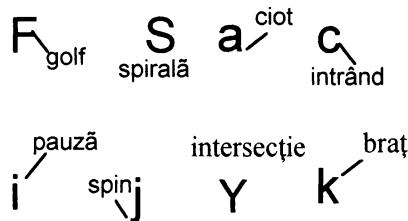


Figura 3.2.3. Elemente topologice detectabile cu codurile Freeman

Aceste caracteristici, odată descoperite, permit fie detecția directă a naturii caracterului, fie, în condițiile în care metoda este un preambul al recunoașterii propriu-zise, îngustarea ariei de cercetat la un număr mai redus de caractere.

În tabelul 3.2.2 se prezintă sintetic rezultatele experimentale (procentajul de clasificări corecte) raportate în literatura de specialitate pentru metodele anterior prezentate.

Tabel 3.2.2. Performanțele metodelor de recunoaștere sintactice

Metodă	Descriptori structurali	Clasificări corecte [%]
Freeman (F)	Coduri Freeman	93
Morfologică (M)	Descriptori morfologici	89,4
Caracteristici specifice contur (C)	Descriptori de contur	83
Combinație F+M+C+P	Descriptori micști	96

S-a notat cu P metoda funcțiilor discriminante polinomiale (statistică), care este utilizată aici, alături de metodele sintactice, pentru a contrabalansa efectul detecției eronate a conturilor. Analizând performanțele se poate constata că cele corespunzătoare combinației de metode statistice și sintactice sunt net superioare [HUA 95, RAH 97, KIT 98, HEU 98].

3.2.2 Gramatici formale

Gramaticile formale s-au dezvoltat ca urmare a necesității înțelegerii și modelării limbajelor naturale în vederea fundamentării teoretice a unor echipamente care să poată înțelege și transpune sub formă de text vorbirea unui operator uman. În fapt, scopul pe termen lung era mult mai ambițios, realizarea unor computere și roboți inteligenți capabili să înțeleagă și să execute comenzi verbale și chiar să converseze cu omul. Speranțele inițiale de rezolvare rapidă a problemei măcar din punct de vedere principal s-au dovedit însă excesiv de optimiste și, în general, cercetările au fost reorientate pe alte direcții. Rezultatele se arată însă foarte utile într-o serie de alte domenii între care la loc de frunte se găsește analiza și recunoașterea formelor unde gramaticile formale se utilizează ca și mod simbolic de a descrie structurile obiectelor [SAL 94, HUL 96, SAN 97]. Definirea unei asemenea gramatici înseamnă în fapt definirea unui set de patru elemente constitutive după cum urmează [JAI 89]:

$$G = \{V_T, V_N, P, S\} \quad (3.2.2)$$

unde:

- V_T este un set de primitive sau terminale, adică elemente constitutive ce nu se mai pot descompune într-o serie de alte primitive. Se notează cu litere latine mici: a, b, c, \dots ;
- V_N este un set de non-terminale, adică elemente intermediare între caractere și primitive. Caracterele se descompun mai întâi în non-terminale, care, la rândul lor, se descompun în primitive. Se notează cu majuscule latine A, B, C, \dots ;
- P este setul de reguli de producție ce dă modul în care se pot compune terminalele și non-terminalele. Practic, setul de primitive și aceste reguli sunt elementele necesare și suficiente pentru descrierea structurilor analizate;

- S este rădăcina, punctul de start al oricărei structuri. Principial $S \in V_N$. Este necesar să se precizeze faptul că mulțimile V_T și V_N sunt disjuncte:

$$V_T \cap V_N = \emptyset \quad (3.2.3)$$

Șirurile cuprinzând aranjamentele oarecare de terminale și non-terminale se notează în mod tradițional cu litere grecești $\alpha, \beta, \gamma, \dots$. Lungimea șirului α se notează cu $|\alpha|$. De asemenea, în cele ce urmează, se vor folosi următoarele simboluri:

$$\begin{aligned} \{\varepsilon\} &= 0 && \text{șirul vid care satisface relația } x \circ \varepsilon = \varepsilon \circ x = x \\ V \circ V &= V^2 && \text{unde 'o' înseamnă concatenare} \\ V^2 \circ V &= V^3 \\ V^* &= V \cup V^2 \cup V^3 \cup \dots \end{aligned} \quad (3.2.4)$$

Prin urmare V^+ , setul închis pozitiv, este mulțimea tuturor structurilor ce se pot descrie cu o anumită gramatică formală. Dacă la acestea se adaugă și elementul vid ε se obține setul închis al producțiilor V^* :

$$V^* = V^+ \cup \{\varepsilon\} \quad (3.2.5)$$

Prin producții se înțeleg relații de forma:

$$\alpha \rightarrow \beta \quad (3.2.6)$$

unde $\alpha \in (V_N \cup V_T)^+ - V_T^+$ iar $\beta \in (V_N \cup V_T)^+$. Deci, în cazul sintezei unor structuri complexe pornind de la primitive, în relația anterioară, α trebuie să conțină cel puțin un membru non-terminal, în vreme ce β este un aranjament oarecare de terminale și non-terminale. În literatură se obișnuiește a se nota aceste aranjamente cu litere grecești. În situația ce face obiectul acestei teze, este interesantă mai ales abordarea inversă, adică descompunerea structurii complexe a caracterului în elementele primitive constitutive. Prin urmare relațiile de sinteza de forma celei din (3.2.6) se folosesc de la dreapta la stânga devenind relații de analiză. Există mai multe tipuri de gramatici diferențiate între ele de restricțiile impuse producțiilor [GON 93]:

- gramaticile libere sau nerestrictive;
- gramaticile senzitive la context;
- gramaticile libere de context;
- gramatici regulate.

Gramaticile libere sau nerestricțive sunt gramatici care nu au nici un fel de restricții impuse regulilor de rescriere. Forma lor este cea generală: $\alpha \rightarrow \beta$

Nu au aplicații prea importante în plan practic, dar prezintă particularitatea interesantă de a șterge părți componente ale structurii, neexistând obligativitatea ca lungimea șirului α să fie mai mică sau cel mult egală cu lungimea șirului β .

Gramaticile senzitive la context au producții de forma (3.2.6) cu condiția suplimentară ca $f \neq \varepsilon$. În general regulile de rescriere sunt de forma:

$$\alpha \mu \beta \rightarrow \alpha \nu \beta \quad (3.2.7)$$

adică ν îl înlocuiește pe μ doar în contextul indicat.

Gramaticile libere de context sunt acele gramatici care au producții de forma:

$$\alpha \rightarrow \beta \text{ cu } \begin{cases} \alpha = A \in V_N \\ \beta \in (V_N \cup V_T)^* - \{\varepsilon\} \end{cases} \quad (3.2.8)$$

Se observă că înlocuirea non-terminalului A cu un aranjament de terminale și non-terminale, simbolizat în expresia de mai sus prin β , este independentă de contextul în care se găsește A în momentul inițial. Deoarece sunt permise expresii de tipul $A \rightarrow \theta A \varphi$, apare posibilitatea auto-stabilirii non-terminalului A în centrul structurii complexe descrise de membrul drept al producției. Analizând gramaticile descrise anterior, se poate constata că restricțiile impuse sunt mai mari în ultimul caz. Aceste gramatici libere de context sunt foarte importante în cazul recunoașterii caracterelor, deoarece sunt versatile și permit descrierea în mod foarte eficient a structurilor complexe [SAN 97].

Gramaticile regulate se caracterizează prin producții ce au forma:

$$\alpha \rightarrow \beta \text{ cu } \begin{cases} \alpha = A \in V_N \\ \beta = B\gamma \text{ cu } B \in V_N \text{ și } \gamma \in [V_N]^k \quad \forall k = 1, 2, 3, \dots \\ |A| \leq |\beta| \end{cases} \quad (3.2.9)$$

Deci sunt admise doar reguli de rescriere de genul:

$$\begin{aligned} A &\rightarrow a \\ &\text{sau} \\ A &\rightarrow aB \end{aligned} \quad (3.2.10)$$

Gramaticile regulate au avantajul că testele de echivalență sunt bine definite și au reprezentări grafice simple [SCH 92]. Se definește o gramatică regulată $G = \{V_T, V_N, P, S\}$ în care:

$$\begin{aligned}
 V_T &= \{a, b\} \\
 V_N &= \{S, A, B\} \\
 P &= \{S \rightarrow aB \\
 &S \rightarrow bA \\
 &A \rightarrow a \\
 &A \rightarrow aA \\
 &B \rightarrow b\}
 \end{aligned} \tag{3.2.11}$$

Transpunerea sub formă de graf a gramaticii definite de ecuațiile (3.2.11) se face pe baza următoarelor reguli:

- Nodurile corespund elementelor N_i din V_T la care se adaugă nodul terminal notat cu T .
- Un arc de la nodul N_l la N_k notat cu un simbol terminal $t \in V_N$ există dacă există o producție de forma $N_l = tN_k$.
- Un arc de la nodul N_k la T notat cu un simbol terminal $t \in V_N$ există dacă există o producție de forma $N_k = aN_k$.

Graful rezultat din aplicarea acestor reguli de corespondență pentru gramatica definită de relațiile (3.2.11) este prezentat în figura 3.2.4.

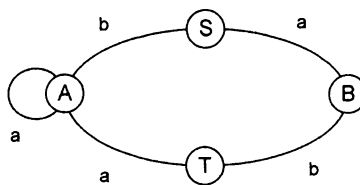


Figura 3.2.4. Reprezentarea sub formă de graf a gramaticii G definite în (3.2.11)

În cele ce urmează se dă un exemplu de descriere sintactică a structurii literelor ‘A’ și ‘C’. În acest scop se definește o gramatică $G = \{V_T, V_N, P, S\}$ în care:

$$\begin{aligned}
 V_T &= \{a, b\} \\
 V_N &= \{A, B\} \\
 P &= \{S \rightarrow A^2; A \rightarrow b+B; B \rightarrow a+b\}
 \end{aligned} \tag{3.2.12}$$

în care, prin pătrat se înțelege $A \times A$, simbolizând concatenarea pe verticală.

Considerând primitivele din figura 3.2.5.a și operațiile din figura 3.2.5.b, descrierile caracterelor din. figura 3.2.5.c sunt:

$$\begin{aligned} D_A &= (b+a+\underline{b})+(b+a+\underline{b}) \\ D_C &= \underline{a}+b+b+a \end{aligned} \tag{3.2.13}$$

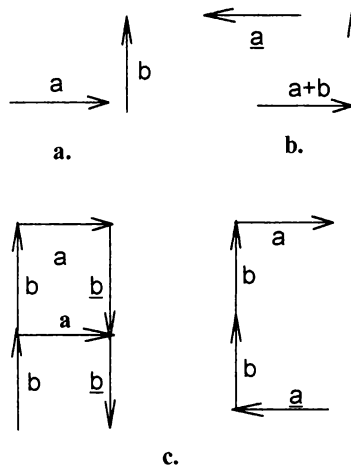


Figura 3.2.5. Descrierea literelor A și C

- a. Setul de primitive;
- b. Operațiile ce se definesc pe acest set
- c. Caracterele A și C.

Ambele corespund întru totul celor două structuri, dar D_A poate fi generată de regulile de producție din (3.2.11) iar D_C nu.

$$S \rightarrow A^2 \rightarrow (b+B)+(b+B) \rightarrow (b+a+\underline{b})+(b+a+\underline{b}) = D_A \tag{3.2.14}$$

Acest fapt face ca sistemul să recunoască o structură tip ‘A’, restul structurilor fiind etichetate ‘non-A’. Rezultă deci, că trebuie definit un set general de primitive și câte o gramatică pentru fiecare caracter separat [GON 93].

3.3 Rețelele neuronale utilizate în analiza și recunoașterea caracterelor

Rețele neuronale de diferite tipuri reprezintă la ora actuală o soluție flexibilă și eficientă pentru un număr foarte mare de aplicații ce necesită rezolvarea unor probleme de recunoaștere a formelor. Utilizarea acestora pentru recunoașterea caracterelor este atât de frecventă încât, această aplicație a devenit un test tradițional pentru evaluarea eficacității rețelelor [CAR 92, CHA 94, GEO 95, SHU 96, LAZ 97, DEC 98]. Uneori imaginea caracterelor de recunoscut este plasată direct la intrare fără procesări speciale [LIN 92].

În aplicațiile citate în literatura de specialitate [BRE 97], vectorul de intrare este un descriptor complex al caracterului conținând de cele mai multe ori în interiorul său rezultatele unor metodologii laborioase cu etape de procesare atât statistice cât și sintactice. Sunt raportate rezultate remarcabile în domeniu obținute prin utilizarea practic a tuturor formelor de rețele neuronale cunoscute. Între ele însă, tipurile de rețele cele mai des utilizate pentru recunoașterea caracterelor sunt:

- perceptronul multi-strat;
- rețelele Kohonen;
- rețelele cu neuroni RBF.

În cele ce urmează, se va face o prezentare a acestor trei tipuri de rețele, punându-se accent în primul rând pe acele structuri utile pentru domeniul de interes al acestei lucrări. Autorul dorește să precizeze că aceste exemple reprezintă doar o parte dintr-o gamă foarte vastă de soluții posibile aflate la ora actuală în atenția cercetătorilor.

3.3.1 Perceptronul multi-strat

Rețelele tip MLP (Multi-Layer Perceptron) sunt, datorită simplității și eficienței lor, cele mai frecvent utilizate rețele neuronale în aplicațiile de recunoașterea caracterelor [VIN 92, WAA 94, HAY 94, BRE 97]. Aceste rețele sunt compuse dintr-un număr de $L > 2$ straturi de neuroni. Există, așa cum se vede în figura 3.3.1, un strat de intrare, unul de ieșire și $L-2$ straturi ascunse. Fiecare strat $i=0, 1, \dots, k-1, k, \dots, L-1$ are un număr N_i de neuroni. Fiecare intrare i_{pq} a neuronului q din stratul k este conectată cu câte o ieșire o_p a neuronului p din stratul anterior $k-1, p=0, 1, \dots, N_{k-1}$. Ieșirea unui neuron oarecare din stratul k este o funcție polinomială de forma [SCH 92]:

$$o_q = F(S_q) = F\left(\sum_{i=0}^{N_{k-1}} o_p w_{pq} + \theta_q\right) \quad q=0,1,\dots,N_k \quad (3.3.1)$$

în care w_{pq} sunt ponderile stabilite în procesul de învățare, θ_q este o polarizare inițială, iar funcția F este o funcție treaptă (figura 3.3.2.a), sau o funcție sigmoidală (figura 3.3.2.b). Procesul de învățare, sau antrenament, constă în aducerea la intrare a vectorilor descriptori ai unui set de caractere cunoscute și modificarea ponderilor w_{pq} până ce la ieșire se obțin răspunsurile corecte [PIT 94].

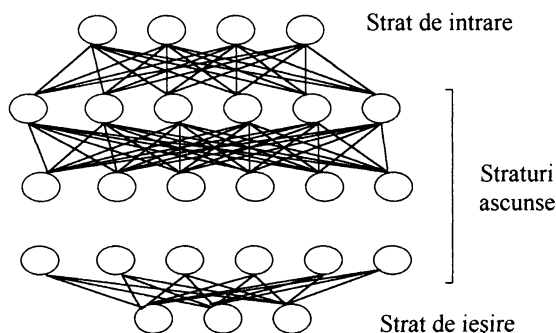


Figura 3.3.1 Structura unei rețele tip MLP.

Rețelele MLP au un algoritm specific de învățare, numit „Propagare înapoi” (Back propagation), prin care ponderile intrărilor neuronului p din stratul k se determină pornind de la ponderile intrărilor neuronilor $q=0,1,\dots,N_{k+1}$ din stratului următor $k+1$.

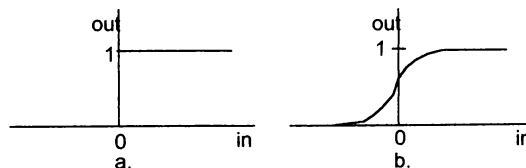


Figura 3.3.2. Funcții de transfer ale neuronilor din rețeaua MLP

a. Funcție treaptă; b. Funcție sigmoidală

Prin urmare determinarea se face pornind de la ieșire spre intrare printr-o procedură recursivă, denumită “Regula Delta”, descrisă în [HAY 94].

- Se determină în a „s”-a trecere valorile i^s de la ieșirile o ale tuturor unităților din rețea
- Se determină ponderile $w_{p,L-1}$ ale stratului de ieșire folosind relația:

$$\Delta^s w_{p,L-1} = \epsilon (t_p^s - o_p^s) F_p^{-1} (net_p^s) o_q^s \quad (3.3.2)$$

în care ϵ este eroare între ieșirea o_p^s și valoarea teoretică corectă t_p^s la iterația s iar net_p este:

$$net_p = \sum_i w_{pq} i_q + \theta_p \quad (3.3.3)$$

de asemenea la iterația s

- Se determină ponderile w_{pq} ale straturilor ascunse cu relațiile

$$\Delta^s w_{q,p} = \epsilon \delta_p^{-s} o_p^{-s} \quad (3.3.4)$$

În care ‘ o ’ semnifică o unificare de notație a ieșirii atât la stratul de intrare cât și la celelalte iar δ_p^s este derivata cu semn schimbat a erorii etapei E_s .

$$\delta_p^s = - \frac{\partial E_s}{\partial net_p^s} \quad (3.3.5)$$

- Se oprește procesul, dacă diferența între vechea și noua pondere Δw_{pq} , este sub un prag prestabilit. În caz contrar se reia procedura cu valorile rectificate ale ponderilor.

După terminarea procedurii de învățare, se trece la testarea performanțelor rețelei cu un set similar, dar diferit de caractere. Procentajul de clasificări corecte obținut la testare este garantat ca fiind procentul de reușite în cazul oricărui set de caractere, având aceleași caracteristici cu cele folosite la antrenament și testare.

La intrarea acestui tip de rețea se pot aduce practic orice tip de descriptori din cei tratați în subcapitolele precedente. De altfel, multe din rezultatele experimentale prezentate anterior au fost obținute cu ajutorul unor asemenea rețele. Creșterea semnificativă a performanțelor și scurtarea timpului de învățare se obține

utilizând cunoștințe apriorice în stabilirea valorii inițiale a ponderilor. În speță, unele legături pot fi tăiate fixând $w_{i,j}=0$ iar altele întărite [TOD 94].

3.3.2 Rețelele Kohonen

Rețelele auto-organizante Kohonen sunt destul de des prezentate în literatură ca soluție naturală de recunoaștere a caracterelor [CAR 92, BAR 96, BRE 97], deoarece structura lor este inspirată din biologie și se bazează pe conceptul de învățare competitivă. Rețeaua este în fapt o hartă neurală n -dimensională, în care fiecare neuron component este ieșire și are intrarea conectată cu toate unitățile de intrare, în mod similar cu felul în care neuronii umani sunt conectați la senzorii vizuali. Legătura între celula j de ieșire și celula i de intrare are o pondere notată w_{ij} . Totalitatea acestor ponderi, pentru toți neuronii de ieșire, formează o matrice numită a conexiunilor exterioare. Matricea conexiunilor interioare cuprinde ponderile w_{kj} ale legăturilor între celulele din stratul de ieșire. Valorile acestor ponderi sunt dependente de distanța dintre neuronul i și k , legătura fiind puternică între vecini. Funcția de variație cu distanța a acestor ponderi este numită „Pălăria mexicană” și este reprezentată în figura 3.3.3. Forma funcției este de asemenea inspirată de rezultatele cercetărilor pe creierul omenesc [HAY 94].

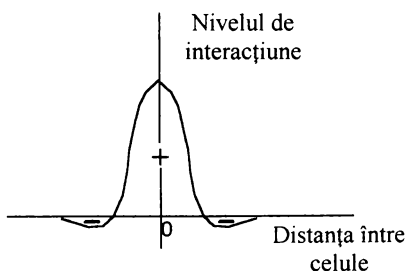


Figura 3.3.3. Funcția de variație a ponderilor cu distanța

Se consideră o rețea Kohonen cu N_i neuroni de intrare și N_o neuroni de ieșire având o funcție de transfer sigmoidă F . Un neuron de ieșire j , aflat în legătură cu toți neuronii din vecinătatea sa V_j , va avea valoarea de ieșire $o(t)$ la timpul t determinată prin relația [TOD 94]:

$$o_j(t) = F\left(\sum_{i=0}^{N_i-1} w_{ij}(t) \cdot x_i(t) + \sum_{k \in J_j} w_{kj}(t) \cdot s_k(t)\right) \quad \text{pentru } \forall j=0, 1, \dots, N_O-1 \quad (3.3.6)$$

unde $X(t) = \{x_i(t); i=0, 1, \dots, N_i-1\}$ este vectorul prezentat la intrare la momentul t . După un timp t_p , necesar propagării calculului și interacțiunilor reciproce, rețeaua ajunge într-o stare stabilă în care o grupare de neuroni de ieșire este activată având $o(t) = 1, t > t + t_p$ restul fiind pe 0. Procesul de învățare constă în a forța ieșirea să fie identică pentru intrări identice. Deoarece antrenamentul este nesupervizat, nu se furnizează configurația corectă a ieșirii. Pentru fiecare caracter din setul de învățare, prezent la intrare la un anumit moment t , se stabilește o celulă și se modifică, pentru ea și vecinii ei, ponderile legăturilor externe, astfel încât ieșirea să reproducă intrarea. Dimensiunea grupului activ scade apoi în timp apropiindu-se de forma dorită [BRE 97].

3.3.3 Rețele cu neuroni RBF

Neuronii RBF (Radial Basis Function) sunt celule având funcții de transfer radiale. Acest fapt permite construirea, cu un număr redus de entități, a unor funcții de decizie între clase cu forme extrem de complexe. Din acest motiv, utilizarea rețelelor RBF în recunoașterea caracterelor conduce la rezultate foarte bune [LEE 91, DEB 92, LEM 93, LIN 93, BRE 97, HWA 97].

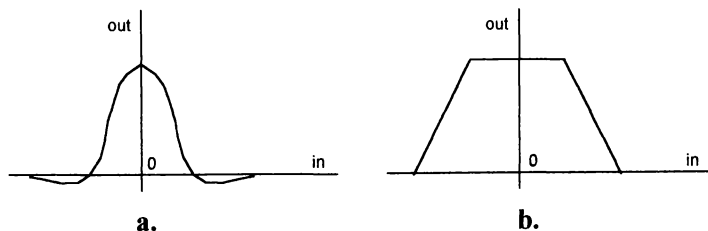


Figura 3.3.4. Funcții de transfer de tip radial

a. Funcția ideală; **b.** Implementare aproximativă.

Arhitectura recomandată în [LEE 91, LEM 93] pentru rețelele de acest tip este similară cu cea a perceptronului multi-strat. Stratul de intrare și cel de ieșire sunt compuse din neuroni având funcții de transfer sigmoideale, neuroni notați LWI (Linear Weighted Input), iar stratul ascuns, unic, este constituit din celule RBF. În

procesul de învățare se creează entități corespunzătoare cu necesităților dictate de corectarea răspunsului de la ieșirea rețelei.

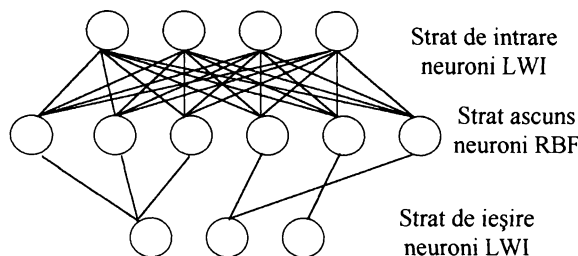


Figura 3.3.5. Rețea cu neuroni RBF

Pe lângă numărul mic de neuroni necesari la implementarea unor rețele capabile să recunoască caracterele, structura prezentată are avantajul că un proces de învățare efectuat ulterior, cu un set diferit, nu produce pierderea cunoștințelor deprinse în antrenamentul anterior. Acest fapt înseamnă că, spre exemplu, rețeaua poate să clasifice corect tipuri foarte diferite de caractere cum ar fi: Times New Roman, Linius și Gotice. O rețeaua având această arhitectură a fost experimentată de compania britanică Nestor Inc și este utilizată cu bune rezultate la recunoașterea caracterelor tipărite și/sau scrise de mână având la intrare, nu un descriptor obținut după procesări sofisticate ci pur și simplu pixelii imaginii caracterului [DEB 92].

3.4 Recunoașterea directă a cuvintelor

Recunoașterea directă a cuvintelor, fără separarea lor în caracterele componente, reprezintă o metodă mai nouă de abordare a analizei documentelor căreia i se acordă din ce în ce mai mult credit în urma rezultatelor promițătoare raportate în literatura de specialitate în ultimii ani [HO 92, SET 94, GEI 95, GOV 96]. Avantajul cel mai important al acestei strategii este evitarea divizării corpului fiecărui cuvânt în literele componente, divizare care, mai ales în cazul scrisului cursiv de mână, este sensibilă la erori și drept urmare rădăcina unui număr foarte mare de clasificări greșite. În plus, experimentele au relevat că, mecanismul citirii se desfășoară în creierul uman pornind de la o primă etapă de identificare a cuvintelor ca entități unice. Abia apoi, în caz de nereușită în această încercare, se trece la recunoașterea separată a literelor componente [WAD 93]. Există două modalități de aplicare a recunoașterii cuvintelor fără separarea pe caracterele componente [PAR 96]:

- ca etapă premergătoare a recunoașterii pe litere;
- ca metodă unică de recunoaștere.

În cazul în care analiza cuvintelor în ansamblu se utilizează ca pas premergător analizei pe litere, procedura decurge astfel. Se stabilește din totalitatea lexiconului un grup de n cuvinte posibile, ale căror caracteristici se apropie de elementele corespunzătoare din vectorul descriptor al cuvântului cercetat. Apoi se pot încerca variante de clasificare a caracterelor componente și se determină care dintre acestea este cuvântul real, în baza unui calcul probabilistic și a unei analize contextuale. Deși această procedură este laborioasă, faptul că este similară cu modul de citire al oamenilor, o recomandă ca o cale promițătoare de urmat [SET 94]. Cuvintele cercetate sunt reprezentate printr-un model complex ce cuprinde o serie de caracteristici globale ale cuvântului, dar și trăsături locale ale unor subcomponente ale acestuia [GOV 96]:

- anvelopa cuvintelor (codul Bouma);
- prima literă, ultima literă;
- prefixe și sufixe ușor detectabile;
- liniile de nivel maxim l_M , minim l_m , de bază l_b și de vârf l_v - determinate ca în figura 3.4.1;
- elemente ale caracterelor componente care se găsesc plasate în zonele ascendentă (l_M, l_v) și descendentă (l_b, l_m);
- elemente proprii metodelor de recunoaștere statistice - coeficienții transformatei Walsh, momente invariante etc.;
- elemente proprii metodelor de recunoaștere sintactice - linii, bucle, pauze etc. a căror poziție este considerată funcție de liniile de nivel.

Așa cum se poate observa, parțial vectorii descriptori ai cuvintelor cuprind elementele relevate ca esențiale de experimentele de psihologia citirii la care se adaugă rezultatul unor procesări matematice capabile să pună în evidență caracteristicile esențiale ale datelor de intrare și să elimine parțial redundanța informațională foarte mare în imaginile brute sau prelucrate doar la un nivel redus de complexitate. Figura 3.4.1 pune în evidență liniile de nivel ale unui cuvânt cuprinzând caractere de toate tipurile. De obicei, aceste linii de nivel se stabilesc înainte de divizarea rândului în cuvinte și rămân aceleași pentru toate cuvintele rândului.

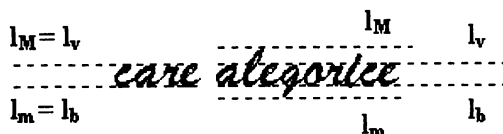


Figura 3.4.1. Stabilirea liniilor de nivel

Se poate concluziona că existența unor litere înalte cum sunt majusculele ori literele “l”, “t”, a unora joase precum literele: “j”, “g”, “p”, “q” și a unor caractere cu puncte, accente ori sedile: “i”, “ă”, “â”, “î”, “ș”, “ț”, reprezintă caracteristici importante de selecție, deoarece determină pozițiile relative ale liniilor de nivel. Numărul și poziția liniilor orizontale sau verticale, a buclelor și a altor elemente constitutive tip din alcătuirea literelor sunt, de asemenea, analizate și luate în considerare la clasificări. Poziția acestor subcomponente este precizată relativ la liniile de nivel sau la o grilă ce împarte imaginea cuvântului în sub-imagini așa cum este cazul cu algoritmul FSGD [CED 98b].

Metoda de recunoaștere directă a cuvintelor, ca singură procedură de identificare, utilizează în general aceleași proceduri, dar vectorii descriptori folosiți sunt mai bogați în componente. În plus analiza contextuală devine neapărat necesară și capătă un rol major în proces, devenind un test de validare hotărâtor în detectarea erorilor și rezolvarea clasificărilor ambigue [SPI 97].

3.5 Validarea recunoașterii

Sistemele performante de citire automată a documentelor conțin în schema lor un bloc final de validare a recunoașterii cuvintelor componente ale textului analizat. Validarea efectivă presupune identificarea cuvântului citit într-o bază de date ce cuprinde un dicționar sau o colecție de vectori descriptori ai cuvintelor. Dicționarul se folosește pentru validarea recunoașterii cuvintelor citite literă cu literă, în vreme ce colecția de descrieri este utilizată în procedurile ce analizează cuvintele în ansamblul lor [MOH 96]. Ulterior, se poate face și o verificare a încadrării cuvântului declarat recunoscut în contextul propoziției, frazei sau câmpului de formular tipizat din care face parte. Aceste teste finale sunt extrem de utile, deoarece permit alegerea variantei corecte dintr-un set de posibile clasificări și elimină un mare număr de erori. De altfel, experimentele psihologice demonstrează că, verificarea existenței cuvântului în vocabularul propriu și validarea contextuală sunt instrumente utile de control și pentru creierul uman [HO 92].

În ambele cazuri însă, un handicap practic încă nesurmontat față de inteligența naturală este imposibilitatea corectării erorilor ortografice ale textului. Un cuvânt scris greșit, deci inexistent în dicționar, este respins ca și “cuvânt nerecunoscut”, sau mai grav clasificat eronat de sistemul automat, în vreme ce citit de om el poate fi recunoscut prin asemănarea relativă cu varianta gramaticală corectă sau prin deducere din analiza contextul semantic în care se află plasat. Echipamentele și metodologiile actuale nu pot, deocamdată, să înlocuiască din acest punct de vedere imensa capacitate de procesare paralelă de care dispune scoarța

cerebrală a omului și nici să imite pe echipamentele de calcul procesul complex de învățare reprezentat de educația și experiența de viață a subiecților umani. În lipsa acestor două elemente testele finale de verificare în dicționar și validare contextuală rămân utile, dar inevitabil limitate ca performanțe [WAD 93].

3.5.1 Validarea cu dicționar

Validarea cu dicționar a cuvintelor citite de sistemele automate de analiză a documentelor se realizează prin identificarea acestor cuvinte într-un fișier care cuprinde toate cuvintele posibile într-o anumită aplicație. Această verificare pornește de la faptul că, fiecare cuvânt al textului supus procedurii de recunoaștere trebuie să fie un cuvânt existent în limba în care a fost scris documentul analizat. Prin urmare, șirurile de caractere recunoscute sunt identificate ca și cuvinte doar în condiția în care sunt găsite într-un dicționar electronic al limbii respective [HO 92, SAR 96b, SEN 96, KIM 97]. În funcție de caracteristicile diferitelor aplicații, există diferite tipuri de dicționare. Se folosesc:

- dicționare generale;
- lexicon (vocabular de aplicație).

Dicționarele generale sunt baze de date ce cuprind toate cuvintele unei anumite limbi, inclusiv articularele, sufixele și prefixele posibile ale acestora. Ele sunt utilizate în aplicațiile de uz general unde nu există posibilitatea stabilirii unui grup limitat de posibilități. Indiferent de limbă aceste dicționare sunt voluminoase și utilizarea lor presupune disponibilități de hardware pe măsură și un timp de căutare mai mare [PAV 93]. Procedura de verificare presupune, de obicei, existența mai multor variante posibile. În ordinea probabilității acestora, șirurile de caractere recunoscute sunt căutate în dicționarul electronic. Variantele care nu sunt găsite, sunt declarate cuvinte necunoscute (non-cuvinte) și abandonate. Dacă nici una din soluțiile propuse nu este validată, se revine într-o etapă anterioară, de obicei la separarea cuvintelor în litere. În caz că, după parcurgerea buclei de un număr N_{max} de ori, erorile persistă, sistemul decide abandonarea recunoașterii, șirul de litere fiind declarat "cuvânt nerecunoscut".

Limbajele naturale, rezultat al unor evoluții culturale lungi și complexe, conțin foarte multe perechi sau grupuri de cuvinte asemănătoare ca formă dar diferite ca înțeles semantic. Dicționarele generale trebuie să conțină toate aceste cuvinte și la validare riscul ca o clasificare eronată să fie acceptată este foarte mare. Drept urmare, structura rezultatelor experimentale ale testării unor asemenea sisteme se caracterizează printr-un număr mic de respingeri, dar un număr relativ mai mare de cuvinte clasificate eronat. În plus este foarte posibil ca, în condițiile

mai sus enunțate, pentru un același cuvânt să fie validate mai multe variante posibile de recunoaștere, alegerea uneia dintre ele devenind o operație riscantă. Pentru reducerea numărului de erori generate de similaritate se apelează la o etapă suplimentară de validare, analiza contextuală [CHA 96].

Lexicoanele conțin doar un grup limitat de cuvinte proprii vocabularului specific al aplicației respective. Limitarea numărului de posibilități reduce dimensiunile bazei de date, ducând la creșterea vitezei și la o eficientizare a procedurii de validare. Aceste avantaje sunt argumente serioase ca peste tot unde este posibil vocabularul de aplicație să fie preferat dicționarului general. Funcționarea sistemelor dotate cu lexicon specific aplicației este foarte asemănătoare cu cea descrisă anterior la validarea cu dicționar general. Eficiența sporită nu provine dintr-o abordare diferită ci din faptul că, vocabularul fiind mai redus ca volum, șansa existenței unor cuvinte foarte asemănătoare este substanțial redusă și deci soluțiile alternative incorecte supuse procedurii de validare sunt cu mai mare probabilitate respinse ca non-cuvinte. Mărirea acurateței recunoașterii este deci un efect cumulat al creșterii procentului de clasificări corecte și al celui de recunoașteri respinse în dauna ratei de eroare [GAR 96].

Bazele de date de tip dicționar general sau lexicon de aplicație sunt organizate în formula proprie dicționarelor lingvistice electronice. Căutarea în ele se poate face, fie secvențial pentru vocabulare restrânse, fie mai eficient, prin îngustarea zonei de interes pe baza analizei rând pe rând a literelor componente în ordine de la începutul cuvântului pentru lexicoane de dimensiuni mari și dicționare generale [GOR 97].

3.5.2 Lexicoane pentru recunoașterea directă a cuvintelor

Recunoașterea cuvintelor, ca entități unitare și nu ca ansambluri de litere distincte, are ca rezultat nu șiruri de caractere ce trebuiesc validate ca prezente în vocabular sau respinse ca non-cuvinte, ci descrieri matematice ale unor caracteristici proprii structurilor generale ale acestora. Prin urmare, lexiconul utilizat la validare, nu mai este o bază de date de tip dicționar electronic, ci o colecție de vectori descriptori ai cuvintelor care servesc ca prototipuri pentru clasificare. Practic, în această situație lexiconul devine componentă activă a procesului de recunoaștere, păstrându-și însă și caracterul de instanță primară de validare prin aceea că, insuccesul la clasificare produce întoarcerea pe bucla de feed-back la din etapele anterioare ale procesului, adică la separarea rândului de text în cuvinte sau la analiza caracteristicilor cuvântului [SAR 96b]. Organizarea internă a acestor prototipuri este făcută în relație directă cu structura componentelor vectorului

descriptor. Se stabilește o ordine a importanței caracteristicilor și pe baza valorii lor se ordonează cuvintele în interiorul dicționarului. De exemplu, pentru o anumită aplicație se poate stabili că descriptorii relevanți sunt: codul Bouma, litera inițială și litera finală.

Codul Bouma este reflectarea matematică a anvelopei cuvântului și prin urmare împarte vocabularul aplicației în grupe având aceeași lungime și aceeași dispunere a extensiilor ascendente și descendente. Următoarele două criterii enumerate introduc o regulă de ordonare în interiorul acestor grupe.

Selectarea primei și ultimei litere ca elemente descriere relevante s-a făcut deoarece ele sunt clasificabile cu mai mare certitudine decât caracterele din interior și corespund tendinței existente și la om de a identifica prefixele și sufixele într-o etapă timpurie a citirii [SEN 94].

Fraza de mai sus are un număr de 41 de cuvinte distincte. Ele se împart în 23 de grupe cu același cod Bouma. Deci, utilizarea doar a acestui criteriu produce 13 posibile confuzii. Criteriul al doilea elimină 12 dintre ele cel de al treilea o elimină și pe ultima. Lexiconul unei aplicații, al cărui vocabular este limitat la setul de cuvinte ce compun fraza menționată, se poate structura în mod univoc pe baza celor trei criterii de clasificare de mai sus. Grupele se ordonează în seria crescătoare a numărului reprezentat de coduri, iar în interiorul lor ierarhizarea se face alfabetic după prima și apoi după ultima literă a lor. Se poate extrage concluzia că găsirea unei reguli de organizare univocă a dicționarului este implicit echivalentă cu eliminarea soluțiilor ambigue la clasificare. În practică, însă, soluția este dependentă de toate etapele procesului de recunoaștere. Dacă vocabularul aplicației este mai bogat, numărul de criterii trebuie să crească și odată cu ele crește și complexitatea procesului [KIM 97].

3.5.3 Analiza contextuală

Validarea contextuală poate fi utilizată în principiu pentru orice tip de document, fie el tipărit sau olograf. Ea este însă, cel mai adesea folosită, ca test final în echipamentele de analiză a documentelor scrise cursiv și a formularelor tipizate completate de mână. Integrarea cuvântului cercetat în contextul propoziției în cazul unui text oarecare se poate analiza pornind de la regulile gramaticale generale. Astfel, dintre posibilele cuvinte se elimină cele care distonează cu restul din punct de vedere al identității ca parte de vorbire sau al acordurilor substantiv-verb, substantiv-adjectiv, verb-adverb etc. În limbile în care structura propozițiilor este conformă unor tipare relativ fixe (engleza, germana etc.), o asemenea analiză este foarte eficientă [CHA 96, MES 96, SPI 97]. Analiza semantică generală pentru

limbile naturale constituie subiectul unei ramuri importante a cercetărilor în probleme de inteligență artificială și reprezintă o componentă esențială în realizarea unei comunicări directe pe cale scrisă sau verbală între operatorul uman și calculator. Soluțiile existente la ora actuală sunt încă parțiale și nu satisfac deocamdată cerințele unor aplicații nerestrictive. În schimb există câteva cazuri particulare [GAR 96, SRI 96, YU 96, MAD 97] importante unde, datorită îngustării gamei de opțiuni posibile, semnificația cuvintelor poate fi utilizată ca un criteriu de validare a recunoașterii.

Cuvintele scrise în câmpurile unui formular tipizat bine conceput sunt în mod normal un răspuns strict la o întrebare precisă, adică o variantă dintr-un set limitat de posibilități. Rezultă că, pentru poziția respectivă, există un vocabular strict mărginit la câteva cuvinte acceptabile în câmpul respectiv. Deci, analiza contextuală este realizată prin intermediul unei distribuții spațiale de lexicoane locale reduse ca dimensiuni [YU 96]. O situație similară apare și la sortarea automatizată a scrisorilor. Aici limitarea vocabularului provine din analiza separată a codului poștal, care, în țările unde automatizarea poștală este în uz curent, are o structură complexă și permite limitarea la unități a numărului de adrese și destinatari posibili pentru un același cod poștal [SRI 96]. Echipamentele de sortare tratează deci separat zona adresei destinatarului și cea a codului, rezultând câte o soluție de recunoaștere a caracterelor propusă pentru fiecare din ele. Validarea constă în confruntarea celor două soluții între ele în condițiile extragerii din baza de date a listei adreselor și numelor persoanelor arondate aceluși cod.

3.6 Personalizarea echipamentelor de recunoaștere a scrisului de mână

Un impediment major în atingerea unor performanțe superioare la recunoașterea scrisului cursiv de mână este variația existentă între forma aceleiași litere scrise de diverși subiecți. Așa cum se demonstrează în studiile grafologice, modul în care fiecare om trasează subcomponentele caracterelor alfabetului este o caracteristică individuală a fiecărei persoane și prin urmare diferențele sunt suficient de mari pentru ca echipamentele de analiză a documentelor să fie induse în eroare.

Funcție de numărul de indivizi al cărui scris trebuie să știe să-l interpreteze automat un anumit sistem de recunoaștere, se poate defini un grad de personalizare al echipamentului și algoritmului de recunoaștere. Prin urmare se pot diferenția trei tipuri de aplicații de citire automată a scrisului cursiv de mână [SEN 94]:

- recunoașterea scrisului cursiv al unui număr mare de indivizi, număr considerat reprezentativ pentru uzul general;
- recunoașterea scrisului cursiv al unui grup restrâns de subiecți;
- recunoașterea scrisului cursiv al unui singur subiect.

Echipamentele și procedurile proprii fiecărei categorii prezintă specificități aparte și se pot remarca diferențe substanțiale ale valorilor ratelor de reușită, eroare și abandon pentru fiecare din ele. În acest context, trebuie subliniat faptul că, recunoașterea caracterelor, ca orice proces de recunoaștere a formelor, presupune, indiferent de tipul clasificatorului utilizat, o primă etapă de învățare utilizând un set reprezentativ de eșantioane. Alcătuirea, adecvată tipului de aplicație, a acestui set este esențială pentru succesul învățării și implicit pentru procentul de reușită în exploatarea propriu-zisă a sistemului [SCH 92].

Primul dintre cele trei cazuri enumerate mai sus corespunde situației celei mai dificile. Rezolvarea necesită echipamente complexe și algoritmi capabili să elimine variațiile de idioscript ale scrisului diferiților subiecți, realizând recunoașterea pe baza unor caracteristici general valabile. Setul de eșantioane pentru instruire trebuie să fie voluminos și variat, astfel încât să poată asigura adaptabilitatea ulterioară a sistemului la recunoașterea unor texte cu o grafie sensibil diferită de cele învățate anterior. În literatura de specialitate [USP 98, LAP 98] se recomandă utilizarea bazelor de date standard create de serviciile naționale de poștă din SUA, Franța, Canada etc. cuprinzând eșantioane reprezentative pentru un număr mare de categorii. Procedura de recunoaștere utilizată este complexă, cuprinzând mai multe căi paralele de prelucrare, proprii atât metodelor statistice cât și celor sintactice. Rezultă vectori descriptori de dimensiuni mari, având elemente corespunzătoare rezultatelor procesărilor paralele anterior pomenite. În aceste condiții, utilizarea unor metode de clasificare standard este dificilă preferându-se, unanim la ora actuală, rețelele neuronale. Cu toate eforturile însă, procentul de succes al unor asemenea sisteme este încă mult sub limitele minime admise în diferite aplicații concrete. În general, în implementările practice verificările cu lexiconul și analiza contextuală reduc marja de eroare dar, provoacă un număr mare de abandonuri ale recunoașterii. Eșantioanele respinse sunt transmise pe calea de procesare manuală [SRI 93, WAA 94, GIL 95].

O situație mult mai favorabilă este cea a sistemelor destinate recunoașterii scrisului unei singure persoane. Un asemenea caz este cel al agendelor personale cu ecran sensibil la atingere (echipamente de recunoaștere on-line) care de obicei sunt puse să învețe și apoi să recunoască doar scrierea cursivă a proprietarului. Gradul înalt de personalizare nu elimină total problema variațiilor pentru că scrisul unui aceluiași individ cunoaște schimbări importante funcție de viteză, stare psihică și fizică, de moment, de calitatea instrumentului de scris și a materialului pe care se scrie, de poziția relativă la direcția de scris etc. [BAI 93] În cazul particular al scrisului cursiv de mână, o problemă o constituie alăturarea în cuvinte a unor

anumite litere care implică nu doar inserarea unor segmente de legătură ci și modificarea unor subcomponente ale acestora [SIN 97]. Este, de exemplu, cazul combinațiilor “af” și “bf” din figura 3.6.1.

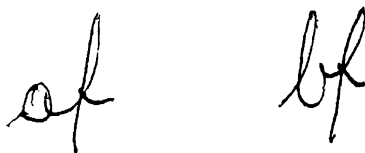


Figura 3.6.1. Legături inter-caractere ce influențează subcomponentele acestora

Sistemele destinate recunoașterii scrisului unui grup restrâns de subiecți sunt concepute similar cu cele destinate uzului general (număr mare de utilizatori), dar, având un set de instruire mai puțin variat, eficiența învățării este mai bună. Există și posibilitatea ca o rețea neuronală cu neuroni RBF să învețe pe rând scrisul fiecărui component al grupului fără însă a “uita” cele învățate anterior [LIN 92].

3.7 Analiza performanțelor sistemelor și algoritmilor de recunoaștere

Analiza eficienței unui sistem sau a unui algoritm de recunoaștere a caracterelor se poate face având în vedere un set numeros de caracteristici. Privitor însă la aspectul ierarhizării importanței criteriilor, punctele de vedere nu sunt întru totul concordante. În cadrul oricărei evaluări, trebuie avut în vedere faptul că situațiile concrete diferă de la o aplicație la alta uneori în mod esențial. Există însă, un număr de caracteristici de apreciere a performanțelor general valabile pentru orice situație și criterii specifice unui anumit domeniu particular. Un set complet de criterii de performanță cuprinde [PAV 93, ROB 96]:

- rata de eroare;
- rata de respingere;
- flexibilitatea și versatilitatea în funcționare;
- simplitatea interfeței de utilizator;
- durata timpului de învățare pentru clasificator;
- insensibilitatea la perturbații;
- fiabilitate și mentenanță;

- viteza de procesare;
- raportul calitate preț al echipamentului;

În opinia autorului, rata de eroare și cea de respingere sunt criteriile esențiale, valabile în evaluarea rezultatelor oricărui echipament sau algoritm de recunoaștere a caracterelor. Rata de eroare exprimă procentajul de clasificări greșite iar rata de respingere procentul de neclasificări. În aplicațiile legate de procesarea unor formulare, instrumente financiar bancare de plată, trimiteri poștale etc. este preferată utilizarea unor proceduri de recunoaștere “precaute” având o rată de eroare scăzută și una de respingere mai mare. Acest lucru se datorează faptului că eșantioanele eronate ce se înregistrează sau sortează greșit, produc pagube, spre deosebire de cele respinse, care se trimit ulterior la prelucrarea manuală unde pagubele sunt evitate.

Considerarea ambelor procentaje de nereușită este utilă deoarece ele exprimă împreună într-un mod corect și obiectiv capacitatea reală a echipamentului analizat de a produce la ieșire răspunsul adecvat. Estimarea celor două rate se poate face pe cuvinte sau caractere eronate sau respinse, funcție de varianta de recunoaștere pe caractere sau cuvinte întregi. În cazul unor sisteme performante însă, unde este mică probabilitatea ca două caractere ale aceluiași cuvânt să fie eronate sau respinse, cifrele sunt sensibil egale fie că se consideră rezultatele clasificării literelor sau cuvintelor. Echivalența devine cu atât mai evidentă în cazul sistemelor cu buclă de reacție unde se face verificarea pe un dicționar a fiecărui cuvânt pentru a fi declarat recunoscut [KNE 98].

Flexibilitatea și versatilitatea în funcționare sunt parametri cu pondere în cazul sistemelor destinate uzului general. Ideal, procentele de eroare și respingere ar trebui să rămână mici indiferent de situație. Citirea unui text tipărit cu diferite tipuri de caractere sau a unui text scris de mână unor autori diverși ar trebui să se înscrie în cote de performanțe similare [SRI 93]. Rezultatele raportate în literatură [BRE 97] evidențiază însă serioase diferențe între performanțele obținute în cazul celor trei tipuri de scriere. Recunoașterea caracterelor de tipar este în general cotate ca o reușită, iar sistemele specializate pe acest sector sunt în bună măsură capabile să se adapteze la diferite tipuri de caractere tipografice.

Recunoașterea scrisului de mână cu caractere de tipar este, de asemenea, un caz de relativă reușită, echipamentele fiind capabile de o anumită adaptabilitate la variațiile între diverși autori. Acest progres este de natură să le facă apte pentru utilizarea într-un viitor nu prea îndepărtat în procesarea automată a formularelor completate manual. Citirea automată a scrierii cursive indiferent de autor este însă încă un obiectiv al cercetărilor de laborator. Performanțe notabile s-au obținut doar la sistemele cu grad mare de personalizare unde versatilitatea este redusă, adică pentru analiza textelor scrise de un singur autor sau un număr restrâns de autori.

O interfață de utilizator prietenoasă, împreună cu simplitatea și eficiența procedurilor de învățare pentru rețelele neuronale sau alte tipuri de clasificatori

utilizați la echipamentele de recunoaștere destinate publicului larg sunt, alături de criteriile economice și de fiabilitate, condiții obligatorii de îndeplinit pentru asigurarea succesului de piață. Respectând aceste reguli, agendele electronice cu ecrane sensibile la atingere, având încorporate sisteme de recunoaștere on-line, au reușit să se impună chiar în condițiile unor performanțe tehnice încă doar satisfăcătoare [SEN 94].

Viteza de procesare este importantă pentru aplicații precum sortarea trimiterilor poștale unde eficiența este un parametru esențial de activitate și la sistemele on-line unde, în anumite cazuri, este necesar un răspuns rapid (identificarea semnăturilor) [VID 95, WAD 98].

3.8 Concluzii

Recunoașterea caracterelor alfanumerice scrise de mână reprezintă o problemă dificilă care nu poate fi soluționată la un nivel satisfăcător prin utilizarea unei metode simple ce implică un singur tip de descriptor. Articolele recent apărute în literatura de specialitate [CED 97, KIT 98] accentuează faptul că ratele de eroare și respingere se reduc substanțial dacă în analiza formelor sunt utilizate metode complexe care implică combinarea mai multor feluri de descriptori. Sunt raportate rezultate bune în situațiile în care vectorii descriptori conțin combinații ale unor descriptori statistici cu descriptori sintactici [NIS 96a, RAH 97, HEU 98].

Utilizarea unor asemenea vectori descriptori, având dimensiuni mari, în cadrul unor proceduri de clasificare tradiționale, este foarte dificilă. O soluție mult mai avantajoasă este folosirea unor rețele neuronale a căror arhitectură este mai bine adaptată unor asemenea sarcini complexe. Succesul recunoașterii caracterelor alfanumerice este garantat în cazul utilizării unui set de eșantioane de învățare reprezentativ pentru aplicația ce va fi ulterior deservită de echipamentul implementat [TOD 94].

Perceptronul multistrat și rețelele de tip Kohonen sunt utilizate cu foarte bune rezultate în cadrul aplicațiilor de recunoașterea caracterelor tipărite. Performanțele acestor rețele în cazul caracterelor scrise de mână sunt însă modeste. Rata de clasificări eronate crește mai ales în situațiile care implică recunoașterea literelor din eșantioane scrise de un număr mare de subiecți [AHI 97].

Rețelele neuronale cu neuroni RBF sunt prezentate în literatură ca fiind o soluție avantajoasă de rezolvare a clasificării caracterelor alfanumerice scrise de mână deoarece au o capacitate sporită de adaptare la variațiile de idioscript. Soluțiile propuse de [LEE 91, LEM 93, LIN 93, HWA 97] se rezumă însă la rezolvarea sarcinii mai puțin complexe a recunoașterii cifrelor. Abordarea

clasificării tuturor caracterelor alfanumerice conduce la implementarea unei rețele cu un strat ascuns conținând un număr foarte mare de neuroni. Acest fapt aduce cu sine serioase dificultăți la implementare și îngreuiază controlul sistemului [BRE 97].

În capitolul 4 “Prezentarea sistemului de segmentare și recunoaștere a caracterelor alfanumerice scrise de mână”, autorul propune utilizarea unei rețele RBF arborescente adaptată unui proces de recunoaștere etapizat. De asemenea, se analizează modalitățile de implementare ale unor echipamente personalizate de recunoaștere, care valorifică relativa stabilitate a idioscriptului individual al utilizatorului.

Validarea recunoașterii prin căutarea cuvintelor recunoscute într-un dicționar sau într-un lexicon, funcție de particularitățile aplicației, reprezintă o modalitate foarte eficientă de detectare și reducere a numărului de erori. Această verificare, efectuată în buclă închisă, dă sistemului de recunoaștere posibilitatea de a selecta soluția corectă dintre mai multe variante posibile. În mod similar, analiza contextuală, acolo unde este posibilă, permite realizarea unui control suplimentar și implicit, creșterea credibilității întregului proces. Avantajele unei posibile implementări ale analizei contextuale pentru textele în limba română sunt evidențiate de autor în capitolul final al acestei teze.

4. Prezentarea sistemului de segmentare și recunoaștere a caracterelor alfanumerice scrise de mână

În cadrul acestui capitol se prezintă un sistem de segmentare și recunoaștere a caracterelor alfanumerice scrise de mână. Acest sistem reprezintă implementarea unor metode și algoritmi concepuți de autor. Segmentarea presupune divizarea succesivă a textului considerat în rânduri, cuvinte și caractere. Toate aceste separări se execută prin utilizarea minimelor unor profilograme înclinate cu unghiul mediu al scrisului [LAC 98d]. Ulterior literele rezultate din segmentare sunt analizate fiecare în parte. Recunoașterea lor se face cu o rețea neuronală RBF la intrarea căreia se plasează un vector descriptor complex cuprinzând informația despre structura spațială a caracterului și o combinație de momente invariante. Considerațiile teoretice referitoare la fiecare dintre etapele anterior menționate sunt însoțite de organigramele algoritmilor implementați de autor în limbajul C.

4.1 Divizarea documentelor prin metoda profilogramelor

În cele ce urmează se va prezenta o metodă de segmentare a imaginilor binare care conțin o zonă compactă de text, propusă de autor. Această procedură este concepută astfel încât să se poată aplica atât imaginilor care conțin texte scrise de mână, cât și imaginilor care conțin documente tipărite.

La baza algoritmului prezentat stă utilizarea unor histograme calculate după o anumită direcție, specificată printr-un unghi de înclinare (unghiul mediu de înclinare al scrisului). Pentru a se asigura operativitatea prezentării, acest tip de histograme vor fi denumite în continuare “profilograme”. Algoritmul de calcul al acestor profilograme este prezentat în figura 4.1.1 [LAC 95].

Ideea utilizării profilogramelor este sugerată de faptul că în cadrul scrisului de mână și uneori chiar și al tipăriturilor apare atât înclinarea literelor, cât și înclinarea rândurilor în text. De asemenea, imposibilitatea asigurării unei alinieri exacte a documentului la achiziție conduce la apariția unei înclinări a rândurilor față de axa orizontală a imaginii. Profilograma este un grafic al numărului de prezențe al punctelor aparținând obiectului analizat (rând, cuvânt, caracter) pe o linie înclinată cu un unghi predefinit. Prin urmare, cu ajutorul profilogramelor sunt posibile

analize ale distribuției spațiale a pixelilor scrisului ținând cont de unghiurile de înclinare medii ale rândurilor față de marginile hârtiei și ale literelor față de rânduri.

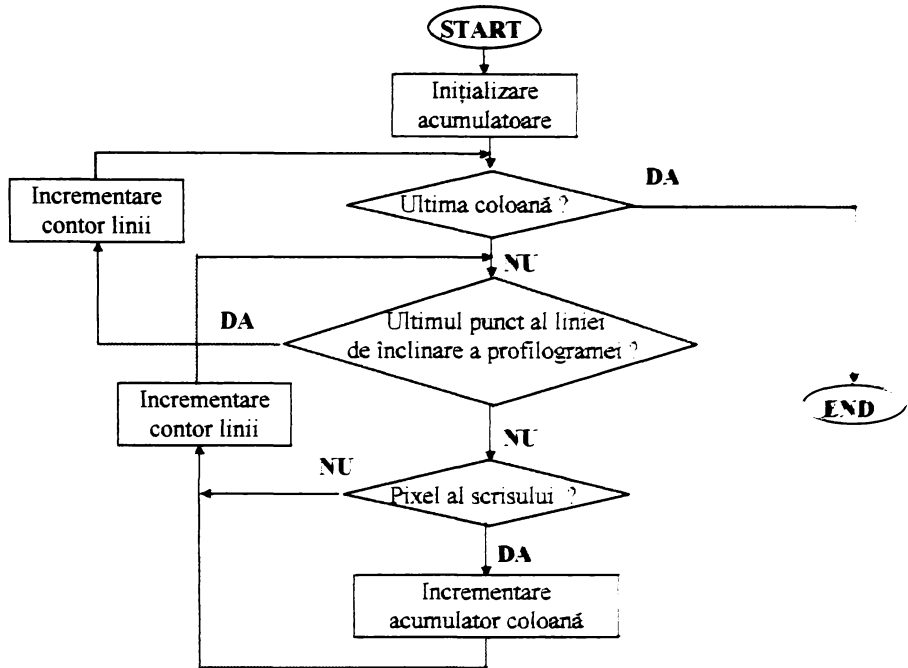


Figura 4.1.1. Metodologia de calcul a unei profilograme

În principal programul care execută separarea are două componente:

- separarea unui text în cuvinte (are ca rezultat un set de fișiere temporare în care se află imaginile cuvintelor din text);
- separarea cuvântului în literele componente (se obțin la ieșire un set de fișiere imagine ale literelor).

Așa cum s-a arătat în cadrul capitolului anterior cele două componente ale procedurii trebuie abordate distinct deoarece prezintă diferențe substanțiale de implementare și grade de complexitate diferite.

Metodologiile generale de lucru propuse de autor pentru cele două etape de segmentare sunt prezentate în schemele logice din figurile 4.1.2 și respectiv 4.1.3. Sunt relevate aici doar etapele majore ale procedurii urmând ca fiecare dintre ele să fie detaliate în paragrafele ce urmează.

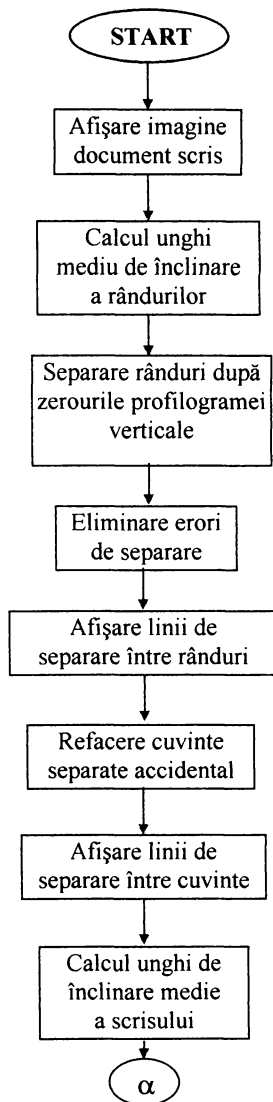


Figura 4.1.2. Metodologia de separare a rândurilor și a cuvintelor

Calculul unghiului mediu de înclinare a rândurilor este o etapă obligatorie pentru a efectua analiza textului în funcție de poziția față de axele imaginii. Separarea rândurilor se face prin detecția spațiilor libere dintre rânduri cu ajutorul zerourilor profilogramei verticale a imaginii.

Erorile datorate rândurilor false (artefacte și părți ascendente și descendente de litere greșit interpretate ca rânduri separate) precum și rândurilor neseperate se detectează printr-o analiză a lățimii fiecărui rând în raport cu media documentului. Separarea lor se face cu ajutorul unor granițe flexibile așa cum se va arăta în paragraful 4.1.1 “Separarea documentului în cuvinte”. Faza de încheiere a acestei etape este trasarea granițelor pe imagine și salvarea în fișiere distincte a fiecărui rând pentru analiza separată.

Divizarea rândurilor în cuvinte se bazează tot pe analiza punctelor de anulare ale profilogramelor în vederea evidențierii pauzelor dintre ele. Detecția erorilor de separare va fi în acest caz mai dificilă deoarece, spre deosebire de rânduri, dimensiunea cuvintelor poate varia într-o gamă foarte largă. În mod similar cu etapa anterioară cuvintele sunt despărțite prin granițe trasate pe imagine și fiecare cuvânt este salvat într-un fișier imagine separat, în scopul divizării finale în literele componente.

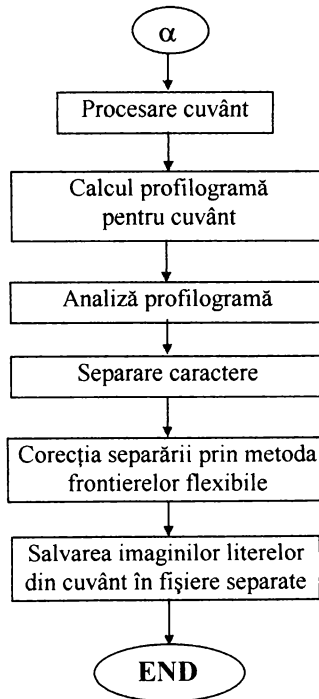


Figura 4.1.3. Metodologia de separare a literelor în cadrul fiecărui cuvânt

Separarea literelor presupune o procedură mai complexă, care cuprinde în fapt un set de procesări paralele desfășurate astfel încât, prin corelarea rezultatelor,

să se poată extrage una sau mai multe variante posibile de divizare. Pentru a se elimina erorile datorate “tăierii” unor părți ascendente sau descendente ale literelor, se aplică o corecție de urmărire a conturului așa cum se va detalia la paragraful 4.1.2.1 “Flexibilizarea separării caracterelor”.

4.1.1 Separarea documentului în cuvinte

Separarea imaginii documentului în imagini individuale ale cuvintelor componente ale textului presupune de asemenea două etape de divizare distincte:

- separarea textului în rânduri;
- separarea rândurilor în cuvinte.

Procedurile utilizate pentru cele două separări au un mare număr de caracteristici comune dar și o serie de elemente specifice. Acest fapt motivează tratarea lor separată în paragrafele următoare.

4.1.1.1 Separarea textului în rânduri

În cazul scrisului de mână este posibil ca rândurile să fie înclinate față de orizontală. De aceea este necesar ca înainte de a trece la separarea rândurilor să se calculeze această înclinare. O soluție a problemei este aplicarea unei transformate Hough asupra imaginii textului. Această transformată are proprietatea de a detecta curbe parametrizabile în imagine (drepte în cazul de față). Ca rezultat al aplicării transformatei se poate obține un unghi mediu de înclinare a rândurilor. Dezavantajul transformatei Hough rezidă în volumul mare de calcule pentru fiecare pixel din imaginea prelucrată, rezultând astfel un timp mare de prelucrare.

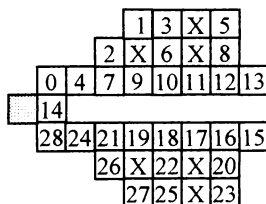


Figura 4.1.4 Masca de prelucrare utilizată pentru determinarea înclinării medii a rândurilor dintr-o zonă compactă de text

Soluția aleasă pentru determinarea înclinării medii a rândurilor constă în utilizarea unei măști speciale, care să parcurgă întreaga imagine. În figura 4.1.4 este prezentată o variantă a acestei măști, structurată astfel încât să fie utilă în determinarea înclinării rândurilor față de orizontală [LAC 95].

Pixelul hașurat este pixelul curent în imagine. Acesta, împreună cu pixelii notați de la 0 până la 28, va determina drepte în imagine. Față de orizontală, unghiul de înclinare al acestor drepte variază în intervalul $(-45^\circ; +45^\circ)$.

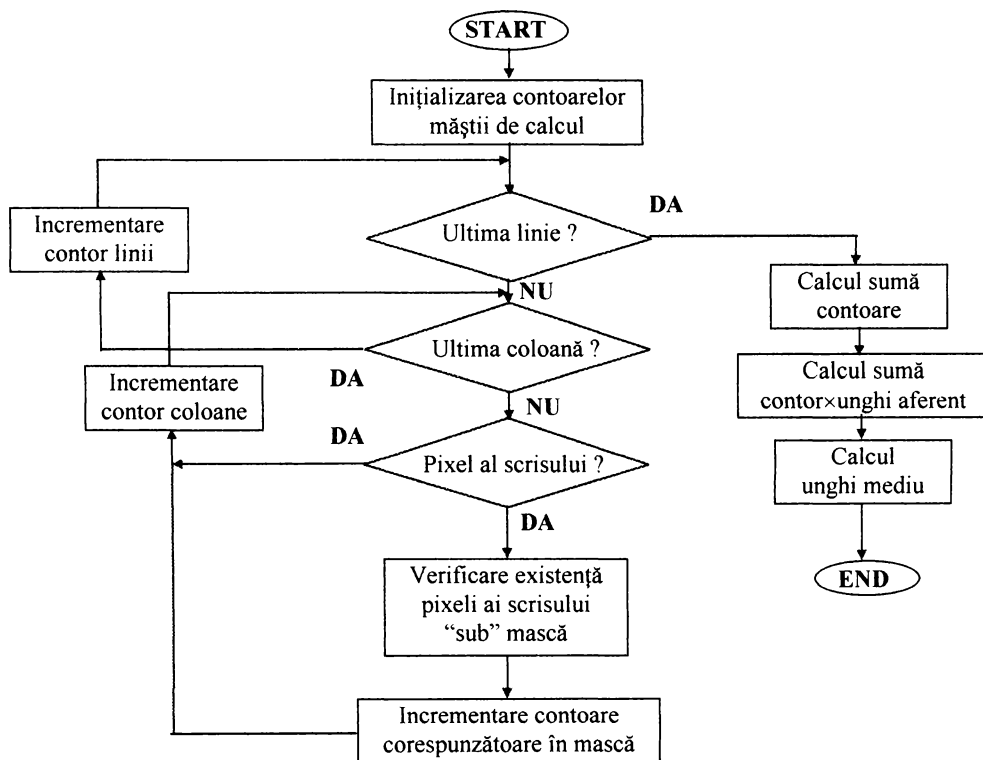


Figura 4.1.5. Schema logică a funcției de calcul a unghiului mediu de înclinare a rândurilor

Se obține astfel o corespondență pixel \leftrightarrow dreaptă \leftrightarrow unghi. Pixelii notați cu "X" în mască aparțin unor drepte deja definite în imagine și de aceea acești pixeli pot sau nu să fie folosiți în continuare în calcule. Cu masca definită în figura 4.1.4 se parcurge întreaga imagine astfel: dacă pixelul curent aparține scrisului, se caută care dintre pixelii 0÷28 aparțin scrisului. În caz afirmativ se incrementează contorul corespunzător direcției găsite ca aparținând scrisului

(fiecare direcție din mască are asociat un contor). Unghiul mediu de înclinare al scrisului pe un rând se determină cu algoritmul din figura 4.1.5, în care se analizează înclinarea segmentelor de caractere de pe acel rând.

După parcurgerea întregii imagini se analizează numărul de apariții aferente fiecărui unghi de înclinare al dreptelor definite de mască. Unghiul mediu de înclinare al rândurilor se poate calcula din valorile contoarelor corespunzătoare înclinărilor din intervalul considerat ($-45^\circ; +45^\circ$) cu formula de mai jos:

$$\text{unghi mediu} = \frac{\sum (\text{unghi} \times \text{contor aferent})}{\sum \text{contoare}} \quad (4.1.1)$$

După determinarea unghiului mediu de înclinare al rândurilor, se poate calcula profilograma pe verticală, după direcția unghiului mediu. Succesiunea rând-pauză din scris va apare pe această profilogramă sub forma unor benzi compacte diferite sau egale cu zero. Prin parcurgerea acestei profilograme se pot determina ușor coordonatele de separare între rânduri.

Algoritmul programului care execută separarea textului în rânduri este prezentat în figura 4.1.6. Se poate observa că separarea este practic efectuată în două trepte. Prima data se execută o divizare ținând cont doar de punctele unde se anulează profilograma verticală și apoi pe baza calculului unor valori medii maxime și minime pentru grosimea rândurilor se detectează și se elimină anomaliile.

Aceste anomalii ce au ca rezultat trasarea unor granițe de separare între rânduri eronate sunt prezente în următoarele două situații:

- Orice sistem de achiziție este afectat de zgomote. Rezultatul acțiunii acestora este apariția unor "defecte" în imagine. Defectele respective nu se pot elimina complet deoarece eliminarea lor ar putea fi însoțită de pierderea unor semne utile din text (de exemplu punctele de la i, j sau punctele de sfârșit de propoziție). Dacă zgomotele apar în intervalul dintre rânduri, ele vor determina "rânduri false" la parcurgerea profilogramei. Eliminarea "rândurilor false" se poate face utilizând un criteriu de medie. În mod normal, într-un text, lățimea rândurilor este aproximativ constantă. De aceea se calculează media lățimii rândurilor din text și dacă există rânduri cu lățimea mai mică decât un anumit procent din medie, acestea se consideră "rânduri false" și nu se iau în considerare.

- De multe ori, în cazul scrisului de mână, rândurile se pot întrepătrunde, datorită literelor mari sau a literelor cu prelungiri (l, g, p, t etc.). Separarea acestor rânduri se poate face tot după un criteriu de medie. Astfel, rândurile care sunt mai mari decât media ponderată cu un anumit coeficient urmează a fi separate. Separarea se face folosind minimul local în zona de interes a profilogramei (acest minim corespunde zonei dintre cele două rânduri).

După separare, se elimină rândurile false și apoi se corectează suprapunerile prin verificarea criteriului de medie cu ajutorul relației:

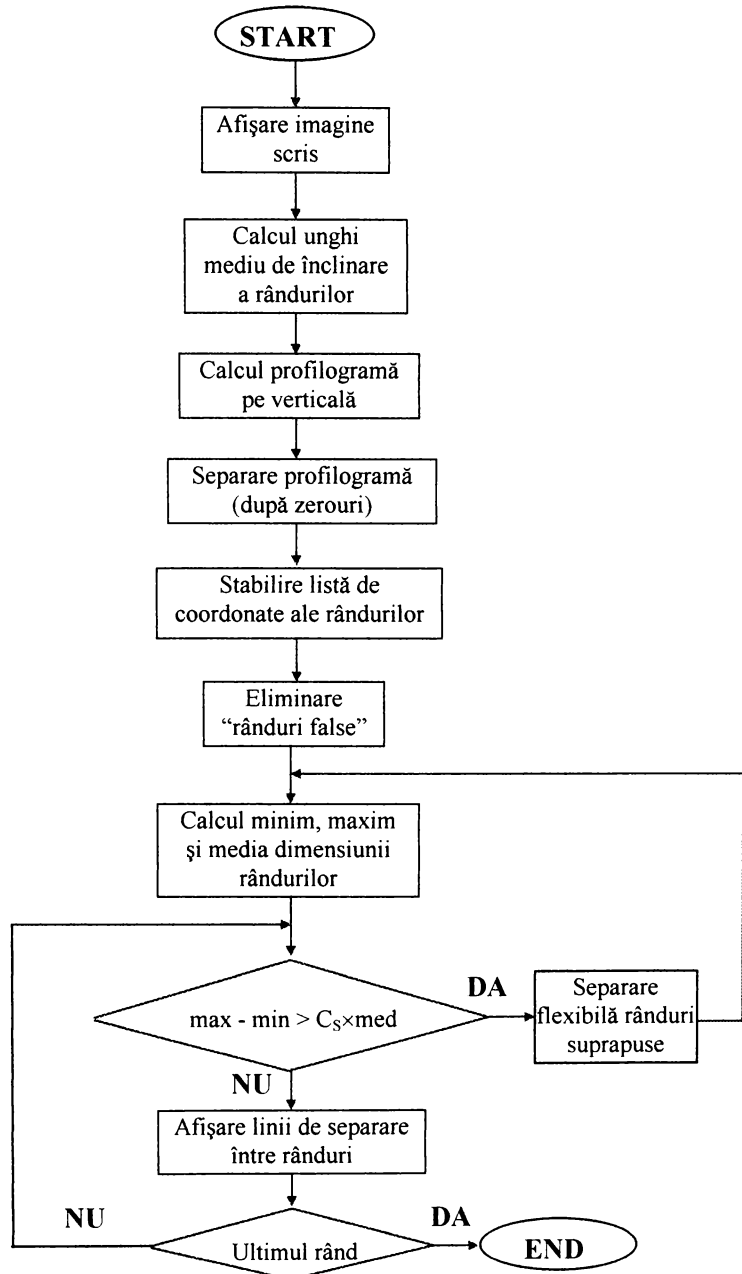


Figura 4.1.6. Schema logică a programului de separare a textului în rânduri

$$\max - \min < C_S \times \text{med} \quad (4.1.2)$$

unde max este dimensiunea maximă obținută pentru un rând, min este cea minimă iar med este media grosimii rândurilor. Coeficientul de proporționalitate C_S a fost stabilit experimental prin prelucrări succesive asupra unui set de eșantioane foarte variat, extras din baza de date “NIST Standard Reference Form Base Handprint Recognition System” [GAR 97].

Inițial s-a încercat separarea rândurilor parțial suprapuse cu frontiere drepte, dar rezultatele au dovedit că segmentarea zonelor descendente sau ascendente ale caracterelor poate afecta substanțial recunoașterea ulterioară a caracterelor și cuvintelor. Testările au dovedit că în general pericolul cel mai mare de erori nu privește caracterul tăiat ci provine de la adăugarea la vecinul de pe verticală a segmentului rezultat din secționare. Un rest de coadă de la ‘p’ sau ‘q’, un capăt de buclă de la ‘y’ sau ‘l’, adăugate la un caracter de pe rândul alăturat, pot modifica fundamental în anumite situații clasificarea acestuia.

Aplicarea metodei de urmărire a bazei cuvintelor prezentată în bibliografie [LIK 95] dă rezultate foarte bune pentru textele în limba engleză, dar pentru limba română, unde există semne diacritice în zona descendentă, de cele mai multe ori le separă pe acestea de rândul de care aparțin și le atașează de vecinul de jos.

În scopul eliminării acestui dezavantaj, autorul propune separarea cu segmente de dreaptă pe porțiunile unde acest lucru este posibil și efectuarea urmăririi ori de câte ori apare o intersecție a frontierei cu un caracter [LAC 98b]. Urmărirea se încearcă întotdeauna mai întâi spre exteriorul rândului, adică în jos când intersecția apare în zona descendentă și în sus când intersecția este detectată în zona ascendentă. În cazul în care segmentul de frontieră care face urmărirea conturului ajunge în zona mediană a rândului analizat sau a celui vecin procedura de urmărire se reia în sens invers. Se evită astfel segmentarea incorectă a semnelor diacritice.

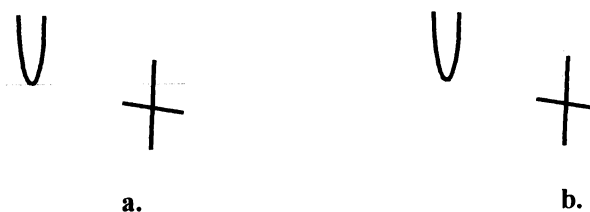


Figura 4.1.7. Segmentarea unei zone de suprapunere parțială

- a.** Segmentarea cu frontieră dreaptă
- b.** Segmentarea cu frontieră flexibilă

În figura 4.1.7.este prezentată o zonă de suprapunere parțială a două rânduri. Eșantioanele scrise de mână cuprind relativ frecvent asemenea cazuri. Segmentarea prin linii drepte reprezentată în figura 4.1.7.a conduce inevitabil la erori, în timp ce segmentarea prin frontiere flexibile soluționează problema corect și eficient. Schema logică a rutinei care execută divizarea textului cu granițe flexibile este prezentată în figura 4.1.8.

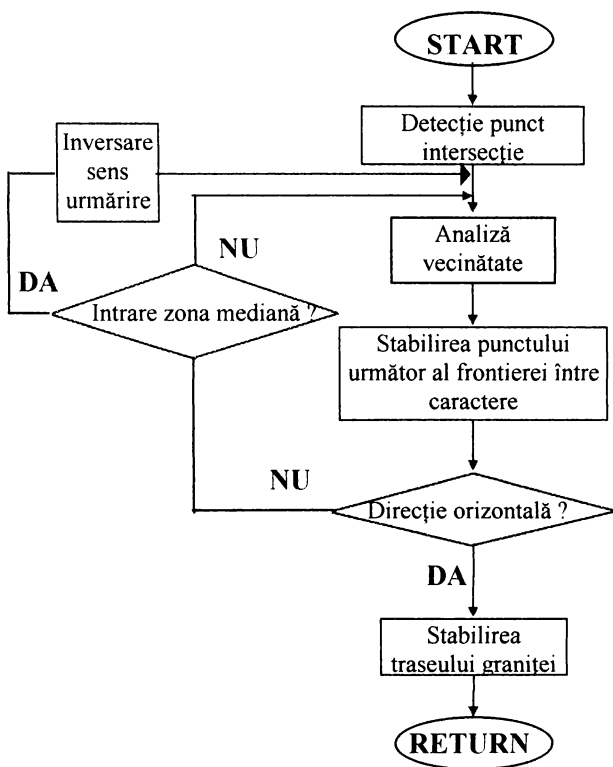


Figura 4.1.8. Schema logică a rutinei de separare flexibilă

Utilizarea procedurii prezentate mai sus nu conduce la scăderea vitezei de lucru deoarece rutina de urmărire intră în funcțiune doar atunci când apar suprapuneri. În rest segmentarea se face normal cu frontiere drepte.

Calculul valorilor minimă, maximă și medie se reia după eliminarea fiecărei erori. Procedura este finalizată atunci când toate rândurile satisfac relația (4.1.2), prin trasarea granițelor între ele și salvarea fiecărui rând în câte un fișier separat.

4.1.1.2 Separarea rândurilor în cuvinte

Cunoscând coordonatele de separare a rândurilor se poate trece la împărțirea fiecărui rând în cuvinte. Operațiunea se execută prin calculul și analiza profilogramei pe orizontală pentru fiecare rând. Unghiul de înclinare al profilogramei este unghiul mediu de înclinare al scrisului.

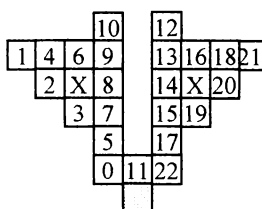


Figura 4.1.9. Masca de prelucrare utilizată pentru determinarea înclinării medii a literelor de pe un rând

Scrisul de mână este caracterizat de o anumită înclinare a literelor, înclinare care este necesar să fie cunoscută pentru segmentarea corectă a cuvintelor. Calculul înclinării medii a literelor se poate face folosind masca din figura 4.1.9 și urmând o procedură similară cu cea utilizată în cazul determinării unghiului de înclinare al rândurilor. Unghiul mediu de înclinare al scrisului pe un rând se determină cu algoritmul din figura 4.1.10 în care se analizează înclinarea segmentelor de caractere de pe acel rând

Se parcurge întreaga imagine a rândului folosind masca de prelucrare din figura 4.1.9. Pixelul curent (hașurat) va parcurge fiecare linie a imaginii. Dacă pixelul curent al măștii se suprapune peste un pixel al scrisului se verifică dacă mai există și alți pixeli ai scrisului “sub” mască. Dacă da, se incrementează contoarele măștii corespunzătoare direcțiilor de scris din mască. Existența pixelilor aparținând scrisului “sub” mască se verifică căutând în imaginea cuvântului pixeli de culoarea caracterelor.

După parcurgerea întregii imagini a rândului, se estimează unghiul de înclinare medie a scrisului pe rând. Formula de calcul utilizează valorile contoarelor aferente unghiurilor de înclinare din gama acceptată (-45° , $+45^\circ$):

$$\text{unghi mediu scris pe rand} = \frac{\sum (\text{unghi} \times \text{contor aferent})}{\sum \text{contoare}} \quad (4.1.3)$$

În final, unghiul mediu de înclinare al scrisului pentru întregul document se determină cu ajutorul mediei ponderate din relația (4.1.4):

$$\text{unghi mediu scris} = \frac{\sum (\text{unghi mediu scris pe rand} \times \text{lungime randuri})}{\sum \text{lungime randuri}} \quad (4.1.4)$$

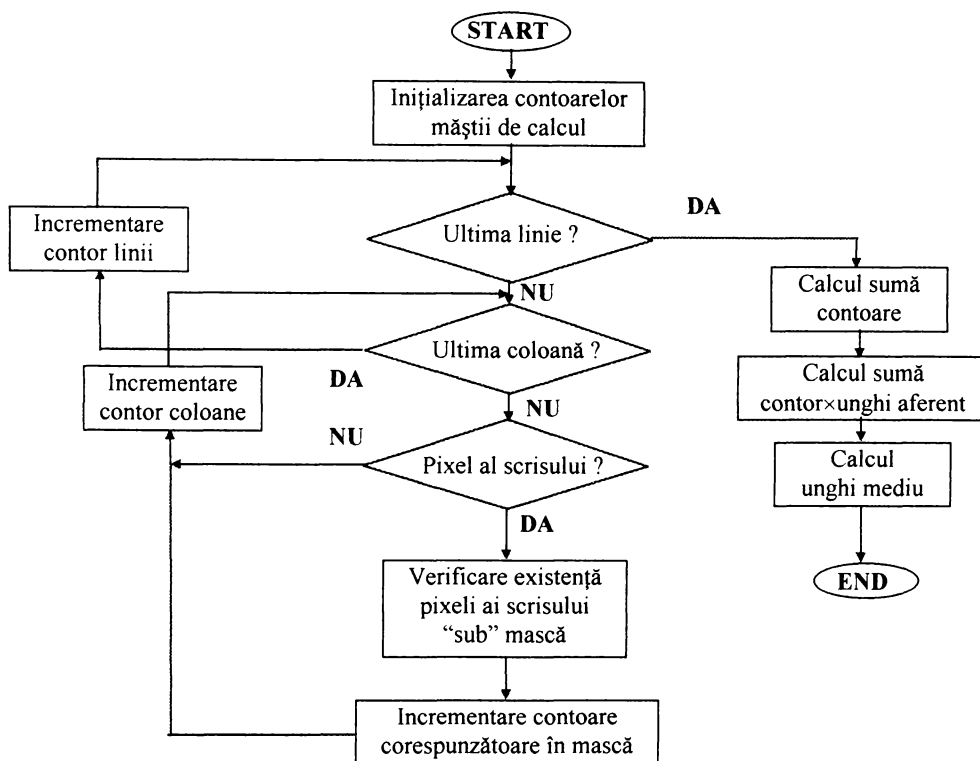


Figura 4.1.10. Schema logică a rutinei de calcul al unghiului mediu de înclinare a scrisului

Motivul pentru care se utilizează media ponderată în locul celei simple este diminuarea efectului perturbator dat de rândurile incomplete (capetele de frază) care uneori fiind scurte și atipice pot introduce valori perturbatoare în calculul anterior.

Unghiul obținut din ecuația (4.1.4) este unghiul de înclinare al profilogramei orizontale cu care se detectează spațiile libere (pauzele) dintre cuvinte. Aceste spații sunt utilizate pentru separarea cuvintelor, așa cum se poate vedea în figura 4.1.11.

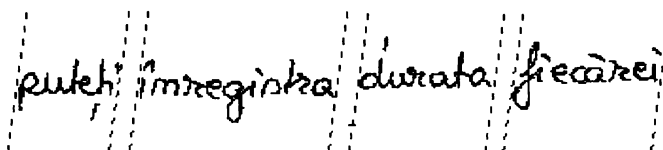


Figura 4.1.11. Separarea cuvintelor prin detecția spațiilor libere

Profilograma orizontală cuprinde zone nule corespunzătoare spațiilor dintre cuvinte și zone nenule în porțiunile ocupate de cuvinte. Prin urmare separarea rândurilor în cuvinte se face prin detecția zonelor nule. Organigrama segmentului de program care execută această divizare este prezentată în figura 4.1.12.

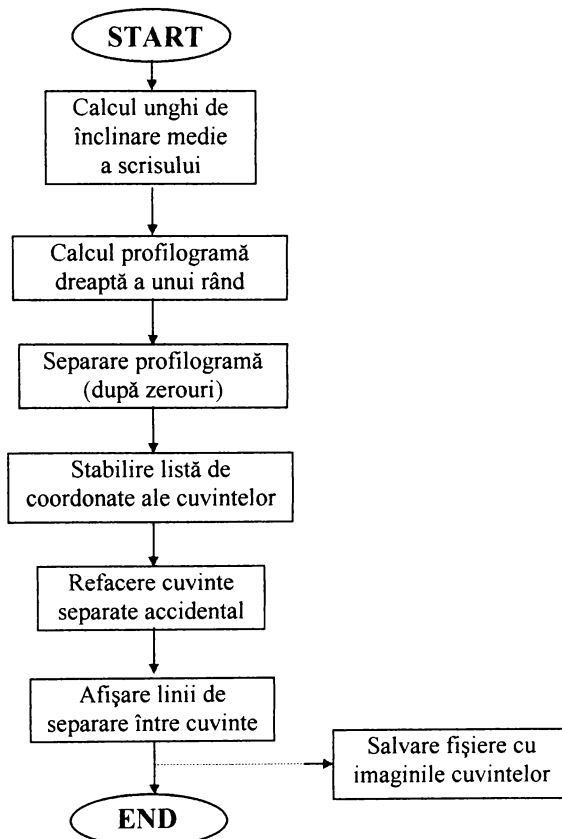


Figura 4.1.12. Schema logică a programului de separare a unui rând în cuvinte

La segmentarea rândurilor trebuie avut în vedere faptul că există situații în care, în mod eronat, un cuvânt poate fi împărțit în mai multe părți, datorită unor pauze ce apar accidental între caracterele lui componente. Numărul erorilor de acest gen poate fi redus astfel: se calculează media pauzelor dintre cuvinte și dacă distanța dintre două cuvinte este mai mică decât un procent din această medie se consideră că cele două cuvinte găsite sunt, de fapt, două părți ale aceluiași cuvânt. O altă soluție este aceea de a se stabili o distanță minimă între două cuvinte și dacă din profilogramă se obține o distanță mai mică, atunci se consideră că este vorba de două părți ale unui singur cuvânt.

Ca și în cazul separării rândurilor, la separarea cuvintelor pot să apară "cuvinte false" datorită zgomotelor dintre cuvinte sau datorită semnelor de punctuație. Acestea nu se elimină în faza de separare deoarece un criteriu de medie nu este satisfăcător în condițiile lungimii variabile a cuvintelor. Ele vor fi ulterior catalogate corespunzător în faza de recunoaștere a literelor.

Pentru a nu încărca în mod inutil memoria, cuvintele odată separate pot fi salvate în fișiere proprii și procesarea lor se poate face în continuare secvențial. Această opțiune, figurată în organigrama algoritmului cu săgeată punctată, a fost adoptată de autorului pentru a preveni blocarea calculatorului în cazul textelor mai voluminoase.

4.1.2 Separarea cuvintelor în litere

Cunoscând coordonatele pentru separarea textului în rânduri, cele pentru separarea rândurilor în cuvinte și unghiul de înclinație al rândului se poate "decupa" din text fiecare cuvânt, pentru ca apoi să fie împărțit în litere. Faza de împărțire în litere este foarte importantă în cadrul algoritmului și de aceea ea cuprinde mai multe prelucrări, care se aplică asupra fiecărui cuvânt în parte. Procesările asupra cuvintelor vor ține cont de unghiul mediu de înclinare al scrisului, calculat anterior [LAC 98d]. Organigrama generală a rutinei de segmentare a cuvintelor în litere este prezentată în figura 4.1.13.

După ce se afișează imaginea cuvântului, se testează dacă este vorba despre un cuvânt multi-caracter, o literă singulară sau un semn de punctuație. Încadrarea într-o categorie sau alta se face pe baza raportului dintre dimensiunea pe orizontală și dimensiunea pe verticală a imaginii cuvântului. Dacă imaginea are o dezvoltare mai mare pe orizontală este vorba despre un cuvânt; dacă imaginea are o dezvoltare mai mare pe verticală este o literă singulară sau un semn de punctuație.

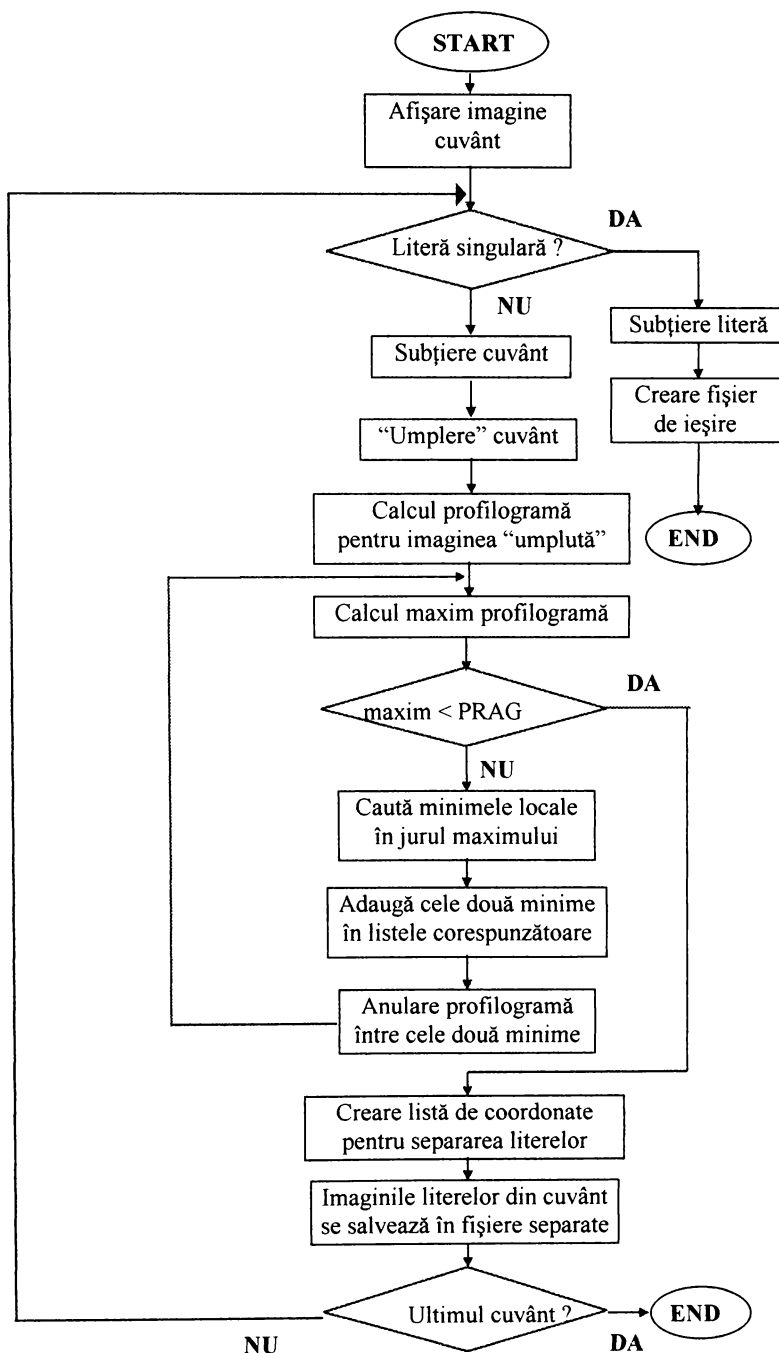


Figura 4.1.13. Schema logică a programului de separare a unui cuvânt în litere

Dacă se decide că este vorba despre o literă singulară sau un semn de punctuație atunci imaginea este subțiată și scrisă în fișierul de ieșire corespunzător. În acest caz nu mai este nevoie de nici o altă prelucrare (calcul unghi mediu de înclinare, separare caractere etc.). Dacă este vorba despre un cuvânt format din două sau mai multe caractere separarea lor se va face ținând cont de unghiul mediu de înclinare al scrisului determinat la paragraful anterior.

Apoi se execută subțierea corpului caracterelor până ce liniile și curbele ce le compun ajung la grosimea de un pixel. Subțierea ajută la separarea literelor deoarece ea determină creșterea distanței între literele succesive. Organigrama procedurii de subțiere este prezentată în figura 4.1.14.

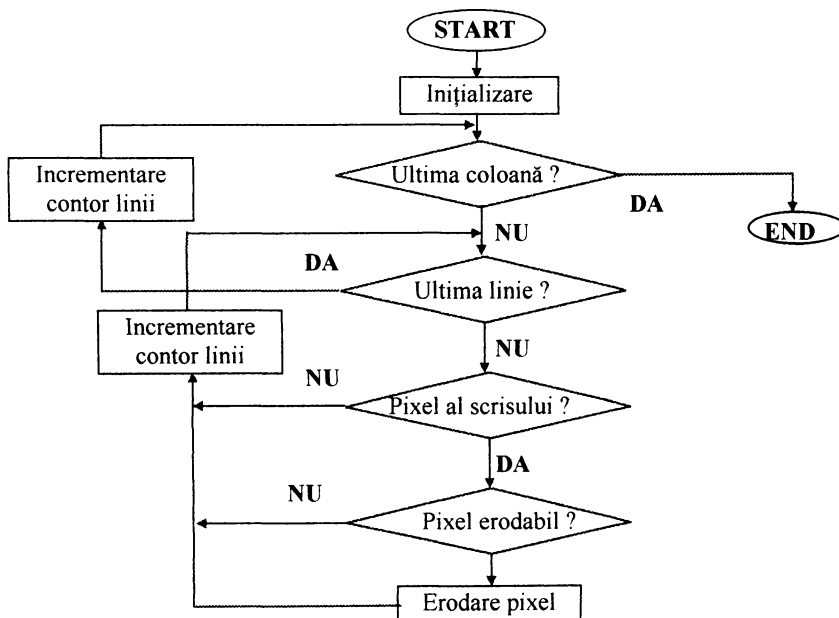


Figura 4.1.14. Schema logică a algoritmul utilizat pentru erodare

Algoritmul de erodare poate fi reprezentat sub forma rețelei cu reacție din organigrama prezentată în figura 4.1.15. Operațiunea se realizează prin prelucrarea imaginii cu ajutorul următoarelor patru măști:

$$\begin{array}{cccc}
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
 1 & 3 & -1 & 0 & 3 & 0 & -1 & 3 & 1 & 0 & 3 & 0 & (4.1.5) \\
 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0
 \end{array}$$

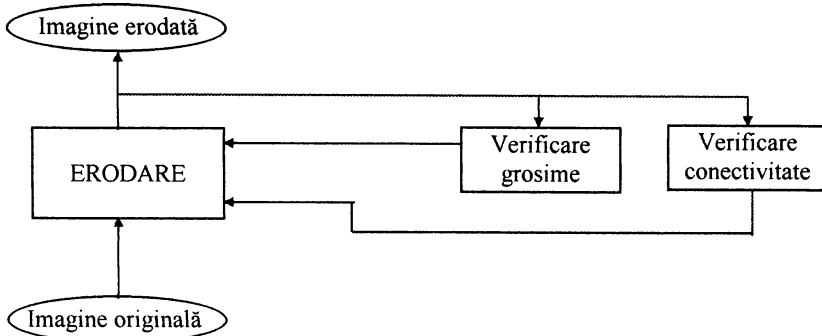


Figura 4.1.15. Reprezentarea algoritmului de erodare ca o rețea cu reacție

Aceste măști se aplică succesiv imaginii, efectul fiind erodarea din patru direcții a imaginii (de la stânga la dreapta, de sus în jos, de la dreapta la stânga și respectiv de jos în sus). În fiecare ciclu de erodare se verifică păstrarea conectivității și a continuității literelor. Algoritmul de erodare se repetă până când toate liniile din imagine vor ajunge la dimensiunea de un pixel, moment în care erodarea este încheiată. Se testează fiecare pixel al imaginii de intrare, verificându-se dacă este posibil sau nu să fie erodat. Decizia de erodare a unui pixel se ia pe baza analizei imaginii obținute după erodare. Dacă prin erodare se întrerupe o linie sau se pierde conectivitatea liniilor din imagine, atunci pixelul respectiv nu va fi erodat.

Dacă, în acest moment, se calculează profilograma pe orizontală a cuvântului, frontierele de separare sunt foarte greu detectabile, deoarece forma histogramei este foarte neregulată având maxime și minime locale foarte apropiate, ceea ce determină frecvente erori de segmentare. Din această cauză următorul pas de prelucrare constă într-o "umplere" a cuvântului. "Umplerea" se realizează cu următoarea procedură:

- Se caută punctele de intersecție între "linia de unghi mediu" și cuvântul scris. "Linia de unghi mediu" este o linie având înclinarea unghiului mediu al scrisului. Ea parcurge întregul cuvânt, de la o extremitate la cealaltă.
- Se unesc punctele de intersecție cu un segment de culoare albă (culoarea corpului caracterelor). În acest fel se obține umplerea spațiilor goale care apar în buclele închise sau deschise ale unor litere.

Efectul obținut va fi apariția unor maxime mai bine puse în valoare în cadrul profilogramei imaginii "umplute" [LAC 95]. Organigrama algoritmului de umplere a unui cuvânt este prezentată în figura 4.1.16.

Etapa următoare constă în calculul profilogramei pe orizontală a imaginii "umplute" la unghiul mediu de înclinare al scrisului. Structura noii profilograme

(conturate) este substanțial diferită, prezentând maxime pronunțate acolo unde scrisul are o dezvoltare mare pe verticală sau acolo unde apar litere în formă de buclă închisă.

Rezultatul returnat de funcția de calcul al profilogramei conturate este șirul acumulatorilor corespunzători coloanelor din imagine (numărul de pixeli de culoarea scrisului care sunt găsiți pe fiecare "linie de unghi mediu"). Profilograma obținută, în care maximele și minimele locale sunt bine evidențiate, permite o mai ușoară segmentare în litere a cuvântului fiind, prin urmare, mai bogată în informație utilă pentru segmentare.

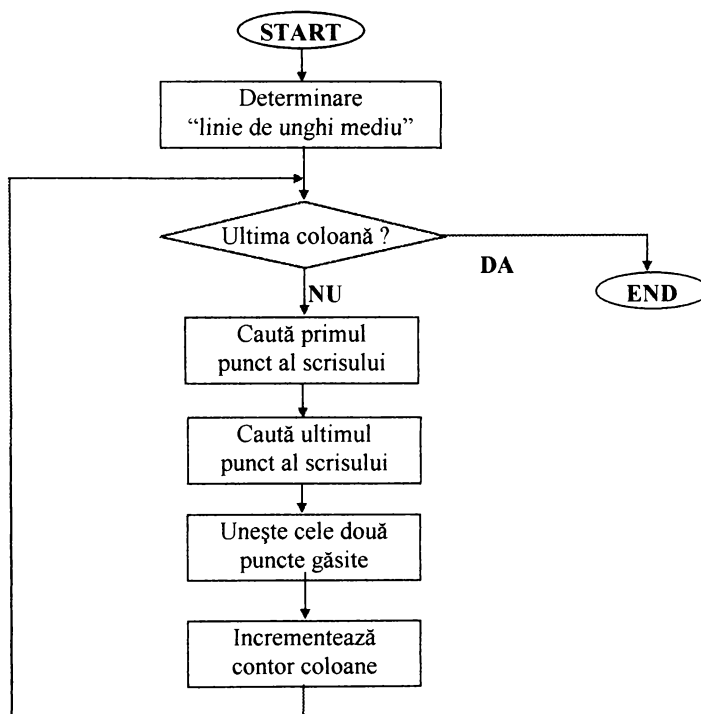


Figura 4.1.16. Schema logică a funcției de "umplere" a cuvintelor

Odată obținută profilograma cuvântului "umplut" se execută o primă separare a acestuia, în zonele în care valorile profilogramei sunt nule, deoarece acolo va fi sigur o zonă între cele două litere.

Fiecare maxim local semnificativ al profilogramei (care depășește pragul de decizie P_f) corespunde câte unei litere, iar coordonatele frontierelor ei de separare, față de caracterele adiacente, sunt dependente de minimele locale vecine în stânga și

în dreapta maximului determinat. Aceste minime corespund legăturilor dintre litere și în aceste zone se poate face divizarea cuvântului procesat. Strategia de analiză a maximelor și minimelor pentru detecția granițelor între litere cuprinde următoarele etape:

- Se caută maximul global al profilogramei.
- Se caută minimele locale în stânga și în dreapta acestui maxim (acestea se pot defini ca un anumit procent p din valoarea maximului). Odată determinate, minimele se introduc în listele în care se memorează coordonatele minimelor locale din stânga, și respectiv din dreapta maximelor determinate în buclă. Construirea acestor serii se face conform procedurii standard [BUR 88] de adăugare ordonată a unor noduri într-o listă.
- Se elimină din profilogramă valorile dintre minimele determinate anterior. Anularea tuturor valorilor profilogramei dintre cele două minime face ca o literă, odată separată, să nu mai fie luată în considerare la următoarele divizări.
- Se calculează apoi maximul profilogramei nou obținute și se compară cu un prag de decizie pentru terminarea divizării P_f . Dacă pragul este depășit se parcurge încă o dată bucla de separare, adică, se reia procesul, căutându-se noul maxim din profilograma rămasă. Dacă nu, se consideră că separarea a fost terminată și se trece la determinarea efectivă a coordonatelor de segmentare a cuvântului.

Organigrama procedurii de separare, descrisă mai sus, este prezentată în figura 4.1.17.

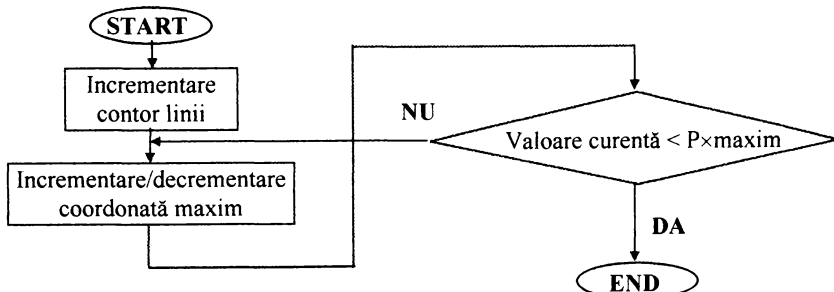


Figura 4.1.17. Analiza de minim și maxim a profilogramelor caracterelor umplute

Pragul de decizie și procentul din valoarea maximului sub care se caută un minim local sunt cei doi parametri care determină segmentarea corectă a literelor. Alegerea adecvată a acestor parametri influențează fundamental calitatea

segmentării. Dacă sunt aleși prea mari, atunci vor rămâne grupuri de litere nedespărțite; dacă sunt prea mici, atunci unele litere sunt împărțite excesiv, adică sunt “rupte” în mai multe părți.

O procedură paralelă, utilizată pentru validarea separării, determinată pe calea descrisă mai sus, pleacă de la ideea că orice cuvânt poate fi împărțit, pe verticală, în trei zone: ascendentă, centrală și descendentă. În zona superioară se găsesc părțile superioare ale literelor (t, l etc.) sau accentele (căciula de la ă, â, punctul de la i, j etc.). În zona inferioară se găsesc, de asemenea, părți ale literelor (p, g etc.) sau sedile (virgula de la ș, ț etc.). Zona centrală sau mediană conține corpul majorității literelor. Ținând cont de acestea, este util să se calculeze o profilogramă pe verticală a cuvântului respectiv. Structura acestei profilograme este divizată la rândul ei într-o zonă de valori mari, corespunzătoare regiunii mediane a cuvântului și două zone mai reduse, corespunzătoare regiunilor inferioară și superioară ale caracterelor. Cele trei domenii se prelucrează independent adică se calculează profilogramele corespunzătoare. Profilogramele zonelor inferioară și superioară vor avea multe valori egale cu zero, valorile nule indicând posibilitatea segmentării în regiunile respective. Acuratețea rezultatelor segmentării crește corelând rezultatele obținute prin utilizarea celor două variante de abordare într-unul din modurile de mai jos:

- Se caută maximul profilogramei întregului cuvânt în regiunea în care profilograma unei zone (inferioară sau superioară) este nenulă. În stânga și în dreapta acestui maxim se caută minimele locale și în aceste minime se face segmentarea.
- Se caută minimul profilogramei întregului cuvânt în zonele în care profilograma unei zone (superioară sau inferioară) este nulă.

Prin combinarea rezultatelor obținute la prelucrarea celor trei zone se obțin coordonatele de separare a întregului cuvânt.

Având listele cu minimele locale găsite în stânga și respectiv în dreapta maximelor, se pot determina coordonatele de separare ale literelor. Coordonata de separare, reprezentând punctul unde se trasează frontiera între două caractere adiacente, se obține ca medie a două valori: coordonata curentă din lista minimelor din dreapta și coordonata următoare din lista minimelor din stânga. Acest ciclu continuă până când sunt eliminate din profilogramă toate valorile mai mari decât un prag predefinit. Lista acestor coordonate va constitui parametrul de intrare al funcției care creează fișierul cu imaginile literelor din cuvânt.

Rezultatul final al segmentării este un set de imagini ale literelor aflate în componența cuvântului. Aceste imagini se procesează și vectorii descriptori rezultați se aplica la intrarea unei rețele neuronale care să facă recunoașterea caracterelor.

În cazul în care se dorește prelucrarea unui text tipărit apar o serie de simplificări în algoritm. Astfel printr-o achiziție îngrijită a imaginii se poate asigura

relativa orizontalitate a scrisului, așa încât să nu mai fie nevoie de etapa de calcul a unghiului de înclinare a rândurilor. De asemenea, nu mai este necesară, în acest caz, etapa de calcul a unghiului de înclinare a literelor în cuvânt. La separarea literelor în cuvânt problema se rezolvă mai ușor, întrucât după subțiere apar între litere spații goale bine conturate. Aceste pauze permit segmentarea simplă a cuvântului, deoarece se folosește o singură profilogramă dreaptă pe orizontală.

Algoritmul asigură segmentarea corectă a regiunilor de text. Performanțele lui depind de alegerea adecvată a coeficienților de prag. Această alegere se face prin rularea algoritmului pe un set reprezentativ de eșantioane. Autorul a efectuat experimentele de calibrare a acestor parametri utilizând baza de date “NIST Standard Reference Form Base Handprint Recognition System” [GAR 97].

4.1.2.1 Flexibilizarea separării caracterelor

Datorită faptului că unghiul mediu de înclinare al scrisului se stabilește ținând cont de ansamblul întregului document analizat, în cazul unui scris mai dezordonat (neregulat) este posibil să se producă secționarea unor părți ale caracterelor. Acest lucru se întâmplă de obicei în zonele ascendente sau descendente ale literelor afectate așa cum se poate observa în figura 4.1.18.



Figura 4.1.18. Intersecție a graniței de separare a caracterelor cu extensia descendentă a unui caracter

Efectul unei asemenea erori de “interacțiune între vecini” poate să afecteze substanțial procesul ulterior de recunoaștere a caracterelor și să vicieze în mod grav întreaga procedură de citire automată. Clasificarea corectă a literei secționată este, în general, nederanjată, deoarece, principial, metodologia separării după minimele locale ale profilogramelor nu permite apariția unor situații cu adevărat critice. Prezența însă a unui rest de porțiune ascendentă sau descendentă în zonele corespunzătoare literelor vecine poate ușor să conducă la interpretări eronate ale

acestora. Pericolul reprezentat de fenomenul de mai sus este cu atât mai mare cu cât limba în care este scris textul analizat are un alfabet mai bogat în semne diacritice. De exemplu, într-un text românesc, secționarea incorectă a lui p din cuvântului “sport” poate conduce la apariția unui segment în zona descendentă a caracterului ‘s’ și acesta va fi cu mare probabilitate recunoscut drept ‘ș’, adică cuvântul este omologat drept “șport”

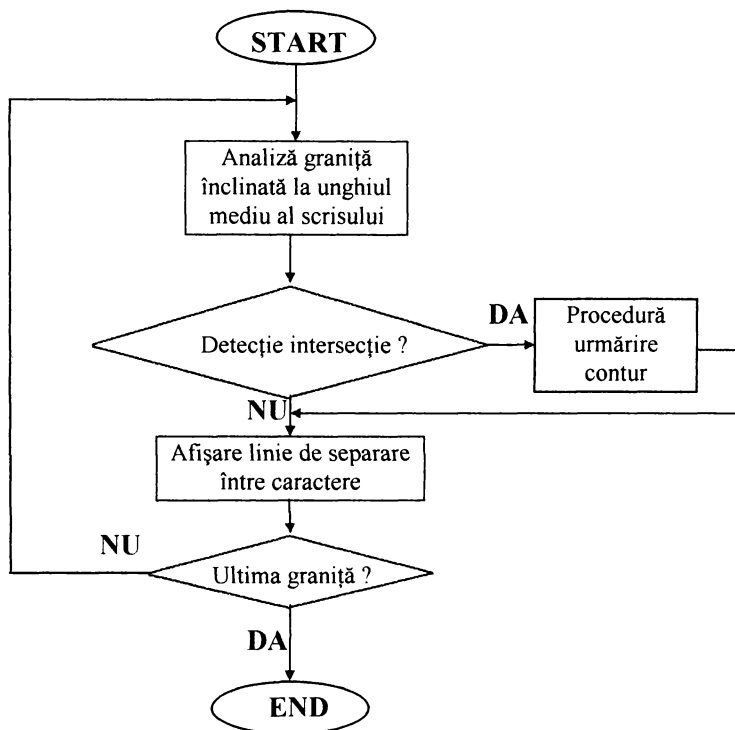


Figura 4.1.19. Schema logică a algoritmului de corecție al granițelor de separare între caracterele unui cuvânt

Având în vedere dezavantajul menționat, autorul propune o soluție a acestei probleme prin includerea în procesul de separare a cuvintelor în caractere a unei proceduri suplimentare de urmărire a conturului caracterului atunci când sunt detectate intersecții în zona extensiilor (ascendentă sau descendentă) ale acestuia [LAC 98b]. Așa cum se poate observa din figura 4.1.19, algoritmului inițial i se adaugă un bloc de testare a existenței intersecțiilor, care, în cazul unui răspuns pozitiv, provoacă rularea unei funcții. Funcția reprezintă adaptarea la condițiile

aplicației de față a unei proceduri recunoscute pentru eficiența și simplitatea ei: “Algoritmul de urmărire a conturului cu mâna pe perete”. Implementarea acestui algoritm se face în forma clasică recomandată în bibliografie [SON 96] ținându-se însă cont de câteva diferențe specifice situației concrete și anume:

- procedura intră în funcțiune pornind din punctul de intersecție;
- conturul se trasează în sus în zona ascendentă și respectiv în jos în zona descendentă;
- procedura se încheie în momentul când se poate reîncepe trasarea frontierei pe direcția dată de unghiul mediu de înclinare al scrisului.

Schema logică a algoritmului de corecție al granițelor de separare între caracterele unui cuvânt este prezentată în figura 4.1.20.

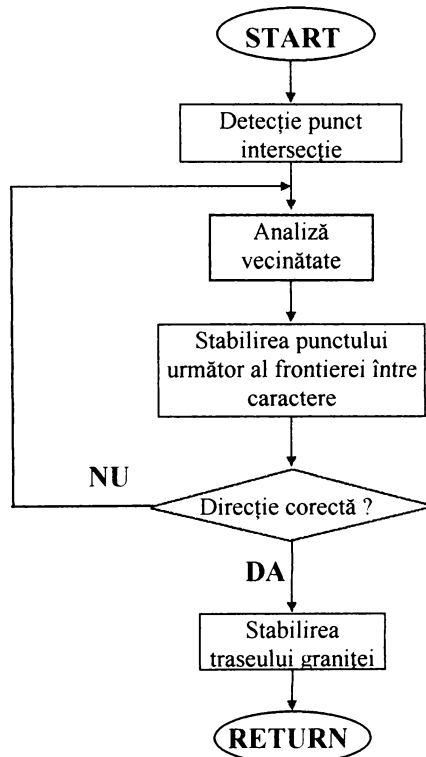


Figura 4.1.20. Schema logică a algoritmului de corecție al granițelor de separare între caracterele unui cuvânt

Figura 4.1.21 exemplifică practic efectele produse de procedura propusă pentru corecția liniilor de separare dintre caracterele unui cuvânt. Granița eronată

este reprezentată cu linie continuă în vreme ce granița corectată este desenată cu linie punctată la fel ca și celelalte frontiere unde nu au apărut probleme. Se poate observa că eroarea de segmentare din figura 4.1.18 a fost corectată. Procedura de urmărire intră în funcțiune începând din punctul de intersecție al frontierei cu extensia descendentă a literei ‘p’ și produce o separare flexibilă, care permite evitarea erorilor ulterioare în faza de recunoaștere a literei ‘s’.



Figura 4.1.21. Granița de separare a caracterelor corectată cu procedura de urmărire a conturului

Ulterior aplicării corecției descrise mai sus granița dintre caractere este retrasată în noile condiții. Rezultatele aplicării în practică conduc la o reducere a ratei de erori provocate de “interacțiunea” dintre caracterele vecine fără a încetini sau complica în mod excesiv metodologia de ansamblu. Procedura prezentată în această secțiune poate fi aplicată cu foarte bune rezultate și în cazul în care documentul este scris de mână cu litere de tipar (handprinted). În această situație, orice intersecție, inclusiv din zona centrală se detectează și se ocolește. Dacă în această situație, urmare a unei suprapunerii parțiale a două caractere alăturate, urmărirea conturului produce o frontieră ce depășește ca dimensiune pe orizontală lățimea medie a unui caracter, se semnalizează eroare.

4.2 Strategie etapizată de recunoaștere a caracterelor scrise de mână

Citirea automată a unui text scris de mână constituie, în abordările clasice, o succesiune de etape în care imaginea inițială a documentului este segmentată până când se obțin sub-imagini ce pot fi apoi clasificate utilizându-se una din metodele folosite curent în recunoașterea formelor (metodele statistice, metode sintactice) sau o combinație a lor așa cum se va arăta în lucrarea de față. În acest subcapitol autorul propune o metodologie nouă, obținută prin îmbinarea mai multor proceduri descrise în literatura de specialitate. Este în esență vorba despre o strategie de lucru

etapizată care permite determinarea caracterelor pe baza unui proces de analiză în trei trepte [LAC 96a, LAC 96c]:

- divizarea pe verticală;
- sortarea primară;
- recunoașterea caracterului.

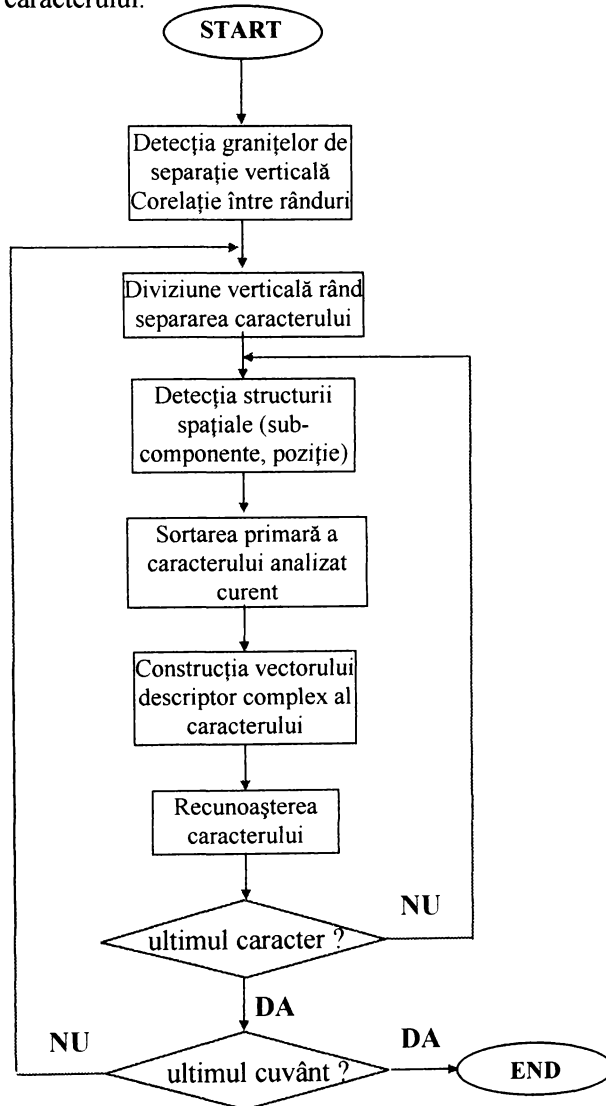


Figura 4.2.1. Schema logică de ansamblu a procedurii de recunoaștere a caracterelor componente ale unui cuvânt

Metodologia presupune, în plus, existența unei sortări inițiale a literelor alfabetului în grupe conținând litere cu structuri spațiale similare. Aceste grupe sunt similare ca idee cu grupele Bouma (vezi paragraful 1.3. “Psihologia citirii”), dar, prin suplimentarea criteriilor de analiză, sunt mai restrictive, cuprinzând fiecare doar două, maxim trei litere [LAC 97d].

Schema logică ce descrie, în ansamblu, algoritmul de lucru pentru recunoașterea caracterelor din cuvintele componente ale unui rând este prezentată în figura. 4.2.1. Evident, procedura se repetă pentru fiecare rând în parte până la terminarea textului analizat și permite atașarea unor blocuri de validare prin control lexical și gramatical. Descrierea detaliată a procedurilor componente prezentate sub formă de blocuri în figura 4.2.1 se va face în paragrafele următoare.

Trebuie subliniat că analiza scrisului cu majuscule impune o procedură similară, dar separată, deoarece structura spațială a acestora cuprinde întotdeauna subcomponente în zona ascendentă. Majusculele izolate (litere de început ale propozițiilor, nume proprii etc.) nu au din punct de vedere statistic un efect perturbator major la nivelul detecției corelate a granițelor de diviziune verticală pentru rândurile unui document.

4.2.1 Divizarea verticală

Autorul propune utilizarea diviziunii pe verticală pentru analiza structurii spațiale a caracterelor [LAC 98a]. Metoda, folosită la recunoașterea cuvintelor în ansamblu în [CED 97], a fost adaptată și implementată pentru necesitățile recunoașterii individuale a caracterelor. Împărțirea pe verticală se face pentru întregul rând din care face parte cuvântul analizat și se aplică, dacă este cazul, corecții locale pentru unele cuvinte așa cum se arată în 4.2.2. Urmare a acestei separări literele componente sunt “tăiate” în trei diviziuni verticale, așa cum se arată în figura 4.2.3:

- zona mediană (centrală),
- zona ascendentă,
- zona descendentă.

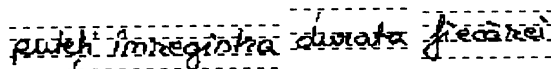


Figura 4.2.2. Corecția granițelor de separație în plan vertical pentru cuvintele de pe un rând

Deși, în alfabetul românesc, datorită literelor ă, â, î, ș și ț, zonele ascendentă și descendentă sunt ceva mai “aglomerate” decât la alte alfabete, textele scrise în limba română păstrează regula generală:

- zona centrală conține majoritatea corpului cuvântului;
- zonele ascendentă și descendentă conțin porțiuni restrânse din litere.

Rezultă de aici că există posibilitatea determinării suficient de exacte a granițelor de diviziune verticală prin analiza densității în plan vertical.



Figura 4.2.3. Definirea zonelor ascendentă, descendentă și mediană

Detecția granițelor de separare ale celor trei zone se realizează cu o precizie foarte bună prin analiza profilogramei verticale a întregului rând. Zona mediană, ce cuprinde componenta principală a corpului majorității literelor, are în profilogramă o amplitudine mult mai mare decât celelalte două zone.

Pentru creșterea preciziei de determinare a acestor granițe analiza este făcută pe întregul rând nu doar pe câte un singur cuvânt iar rezultatele sunt corelate între mai multe rânduri ale textului. Această corelare este o etapă utilă de verificare deoarece raportul central - ascendent și central - descendent este o cvasi-constantă a fiecărui idioscript.

-executanul se obligă să presteze următoarea lucrare (serviciu).....cu r.s.m.f. inițiere
-timpul de lucru.....3 h / zi
-termenul de executare (predare a lucrării).....15.07.1998
-onorariul convenit pentru serviciul contractat este de.....300.000.....lei,
piățibil eșalonat/integral la data de15.07.1998.....
-alte clauze asupra cărora am convenit.....nu e cazul

Figura 4.2.4. Formular tipizat completat de mână

Un caz special îl reprezintă, în acest context, formularele tipizate pentru relațiile cu publicul din domeniul financiar-bancar, administrație etc. Ele conțin în

interiorul aceluiași rând cuvinte tipărite (întrebările formularului) și cuvinte completate cu scris de mână (răspunsurile celui care completează formularul). Granița dintre domeniul median și cel descendent este în general aceeași din cauza liniilor punctate dar probabilitatea ca celelalte granițe de diviziune verticală să coincidă este minimă. În figura 4.2.4. este prezentat un asemenea formular și se poate observa că tipărirea și scrisul de mână au granițe verticale superioare diferite. Așa cum s-a arătat în Capitolul 1 literatura recomandă două soluții de adoptat pentru a permite procesarea automată a formularelor:

- impunerea unor restricții de formă pentru litere și cifre (căsuțe, grile etc.);
- completarea cu cerneală specială.

În ambele situații, este, prin urmare, posibilă separarea celor două tipuri de scris și deci, detecția separată a granițelor proprii porțiunilor scrise de mână. În acest caz procedura de corelare între rânduri devine esențială pentru asigurarea unei detecții corecte a zonelor de divizare verticală.

4.2.2 Gruparea caracterelor alfabetului românesc

Aplicarea strategiei etapizate presupune împărțirea literelor alfabetului în grupe, pe baza poziției lor în raport cu cele trei domenii de separare pe verticală descrise anterior și a formei subcomponentelor caracterelor.

S-a luat în considerare, inițial, criteriul diviziunii pe verticală pentru alfabetul românesc. Grupele rezultate sunt similare cu clasele Bouma, diferența provenind doar din adăugarea literelor specifice românești: ă, â, î, ș și ț. Situații de acest fel există practic în toate alfabetele derivate din alfabetul latin.

Prin urmare, se obțin următoarele grupuri de caractere ale alfabetului românesc în cazul scrisului de mână:

1. litere având doar zonă centrală: *a, c, e, m, n, o, r, s, u, v, w, x, z*;
2. litere cu ascendență: *a, a, b, d, h, k, i, i, l, t*;
3. litere cu descendență: *g, p, q, y, v*;
4. litere cu ascendență și descendență: *f, j, t*.

Odată executată divizarea verticală, se stabilesc zonele ocupate de caracterul studiat. Detecția prezenței unor subcomponente precum liniile, buclele deschise, buclele închise și a pauzelor (în cazul unor caractere ca ă, â, i, î, j, ș, ț) se face utilizând:

- profilogramele verticale și orizontale $p(f_k)$ ale diviziunilor caracterelor în varianta umplută care este mai puțin sensibilă la variațiile de formă. Analiza “formeii umplute” a literelor este mai puțin sensibilă la variațiile

de formă și pune mai clar în evidență existența unor componente structurale de tipul bucle închise sau deschise (vezi figura 4.2.5);

- momentele din relațiile (4.2.1)...(4.2.4.) ale acestor profilograme.

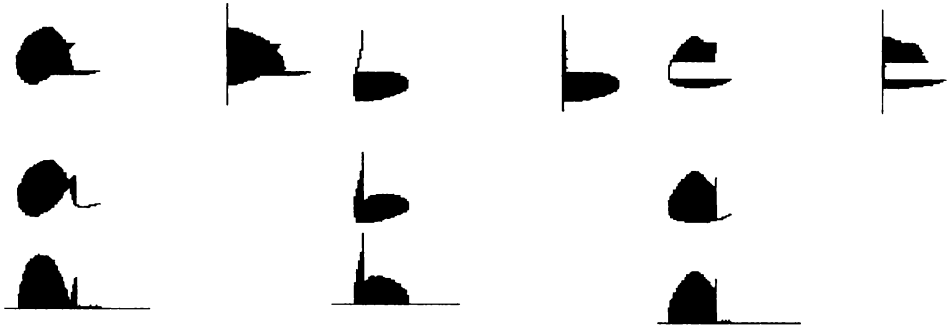


Figura 4.2.5. Formele “umplute” cu linii orizontale și respectiv verticale ale literelor a, b și c

$$\mu = \sum_{k=1}^N f_k p(f_k) \quad \text{Media} \quad (4.2.1)$$

$$\sigma^2 = \sum_{k=1}^N (f_k - \mu)^2 p(f_k) \quad \text{Dispersia} \quad (4.2.2)$$

$$\mu_3 = \frac{1}{\sigma^3} \sum_{k=1}^N (f_k - \mu)^3 p(f_k) \quad \text{Torsiunea} \quad (4.2.3)$$

$$\mu_4 = \frac{1}{4} \sum_{k=1}^N (f_k - \mu)^4 p(f_k) \quad \text{Kurtosis} \quad (4.2.4)$$

În etapa următoare, pentru fiecare literă se evaluează componența structurală, adică se construiește o descriere a formei lor în termen de drepte, bucle închise, bucle deschise și pauze, evidențiindu-se existența acestora într-una sau alta din zonele de diviziune verticală. În final, se obține o reprezentare a structurii spațiale a literei. Pe baza celui de-al doilea criteriu, în cadrul fiecărei grupe se formează sub-grupele cuprinzând doar câteva caractere similare ca formă și aspect. De exemplu, sub-grupul conținând ‘a’ și ‘o’ are descrierea “bucle închise în zona mediană” iar sub-grupul conținând ‘u’ și ‘v’ are descrierea “bucle deschise în zona mediană cu deschiderea orientată în sus”. Structura grupelor și sub-grupelor de

litere ale alfabetului românesc este prezentată în tabelul 2.1. Pentru un echipament de recunoaștere de uz general această sortare a caracterelor standard ale alfabetului limbii în care este scris textul trebuie efectuată o singură dată. Grupele rămân astfel fixate pentru orice alt text scris în limba respectivă, deoarece structura clasificatorului se construiește pe baza acestei informații și rămâne stabilă. În cazul unui echipament de uz individual, în conformitate cu cele descrise la paragraful 4.3 “Sistem personalizat de recunoaștere a caracterelor”, procedura este flexibilă și așa cum se va arăta în capitolul următor performanțele sunt sensibil îmbunătățite.

Tabelul 2.1. Structura grupelor și sub-grupelor de litere ale alfabetului românesc

Grup 1		Grup 2		Grup 3		Grup 4	
a, o, c, e, r, s, x, z, m, n, u, v		ă, â, b, d, h, k, i, î, l, t		g, q, y, p, ș		f, ț, j	
1.1.	<i>a, o</i>	2.1.	<i>a, a</i>	3.1.	<i>g, q, y</i>	4.1.	<i>f, ț, j</i>
1.2.	<i>c, e</i>	2.2.	<i>b, d, h, k</i>	3.2.	<i>p</i>		
1.3.	<i>r, s</i>	2.3.	<i>i, î</i>	3.3.	<i>s</i>		
1.4.	<i>x, z</i>	2.4.	<i>l, t</i>				
1.5.	<i>m, n</i>						
1.6.	<i>u, v</i>						

4.2.3 Sortarea primară

Sortarea primară este clasificarea caracterului necunoscut în subgrupa anterior definită, care are descrierea structurii spațiale similară cu el. Analiza se desfășoară pe baza aceluiași două criterii care au dus la crearea grupelor de litere din tabelul 2.1:

- zonele verticale ocupate de caracter,
- structura spațială a caracterului.

Rezultatul final al acestei etape este un mesaj de genul: “Caracterul analizat este litera ‘a’ sau ‘o’ ”. Prin urmare în cea de a treia etapă decizia trebuie luată doar între aceste două clase. Experimentele arată că rezultatele cele mai bune în recunoașterea caracterelor se obțin în cazul în care tipurile de descriptori utilizați în a doua și a treia etapă sunt diferite. Având în vedere că în cea de a doua etapă, sortarea primară s-a făcut pe baza unei variante simplificate de descriere sintactică a structurii, autorul a efectuat analiza finală de clasificare din cea de a treia etapă prin metoda statistică a momentelor invariante.

4.2.4 Vectori descriptori micști

Momentele invariante sunt una din cele mai rapide și eficiente metode de a transforma o imagine într-un set puțin numeros de caracteristici care să cuprindă toate elementele cu adevărat importante din ea. Lucrarea de față utilizează următoarele trei familii de momente invariante: Hu [HU 61], Zernicke [KHO 90] și Afine [FLU 94].

În urma calculului valorilor primelor R momente invariante pentru un caracter alfanumeric se poate determina un vector descriptor \vec{X} :

$$\vec{X} = \{m_1, m_2, m_3, \dots, m_R\} \quad (4.2.5)$$

care definește un punct în spațiul R -dimensional al caracteristicilor sau al formelor.

La începutul procesului de clasificare se reprezintă în spațiul caracteristicilor un set de eșantioane (caractere) cunoscute (preclasificate). Se notează cu $\vec{v}^{(K)}$ vectorul prototipului clasei K .

$$\vec{v}^{(K)} = \{m_1^{(K)}, m_2^{(K)}, \dots, m_R^{(K)}\} \quad (4.2.6)$$

Pentru un alfabet cu P litere componente rezultă o mulțime Y de clase $Y = \{C_1, C_2, C_3, \dots, C_P\}$, în care există un număr de caractere preclasificate. Un caracter necunoscut descris de vectorul \vec{X} se află în spațiul caracteristicilor la distanța $D(\vec{X}, \vec{v}^{(K)})$ de prototipul clasei K . Distanțele sunt măsura gradului de similitudine al vectorului în cauză cu vectorii aparținând fiecărei clase. Cea mai uzuală este distanța euclidiană:

$$D(\vec{X}, \vec{v}^{(K)}) = \|\vec{X} - \vec{v}^{(K)}\| \quad \text{unde} \quad \|\vec{a}\| = \sqrt{\vec{a}^T \cdot \vec{a}} \quad (4.2.7)$$

Pentru orice caracter de clasificat se determină un șir de P distanțe (funcții discriminant) D_i . Caracterul se atribuie clasei j cu proprietatea:

$$D_j(\bar{X}) > D_i(\bar{X}) \quad \text{pentru } \forall i \neq j \quad (4.2.8)$$

Este important de specificat că prototipul unei clase este “idealul” determinat matematic din grupul de eșantioane aparținând la un moment dat acelei clase. Acest sistem este mai elastic [GON 93] și permite în cazul analizei unui anumit scris de mână o mai bună adaptare la stilul personal.

Rezultatele experimentelor, efectuate de autor, au arătat că utilizarea oricăruia dintre cele trei seturi de momente invariante (considerând numai ordine mai mici decât patru), conduce la rezultate nesatisfăcătoare [LAC 96b]. Conform așteptărilor, deciziile eronate sunt prezente mai ales la discriminarea unor litere apropiate ca formă și în special pentru scrisul mai neglijent sau neregulat.

Examinând atent erorile obținute pe un același set de eșantioane, autorul a observat că “greșelile” diferă de la un set de momente la altul. Această constatare a condus la ideea de a combina momentele Hu, Afine și Zernike în clasificare, adică alegerea acelor momente care accentuează cel mai mult caracteristicile literelor și realizează o diferențiere cât mai netă în spațiul momentelor selectate [LAC 96e].

Studiind valorile momentelor invariante obținute pentru diferitele litere și încercând mai multe variante de combinare pentru obținerea unor performanțe cât mai bune, s-a conturat un procedeu de selectare a momentelor celor mai relevante pe considerentul îndeplinirii simultane a următoarelor două condiții:

- *Dispersia* momentelor de același tip și de același ordin corespunzând unei singure clase de litere să fie mică.
- *Distanța* dintre momentele aparținând unor clase diferite să fie mare.

Un moment este cu atât mai relevant, dacă dispersia lui este mică pentru toate clase de literele considerate. Un astfel de moment asigură o grupare bună a diferitelor clase. În plus, clasele trebuie să fie cât mai distanțate în spațiul de decizie (distanță cât mai mare între valorile corespunzând unor clase diferite). În urma calculării momentelor Hu: H_1 - H_4 ; Zernicke: Z_1 - Z_4 și Afine: A_1 - A_4 pentru cinci seturi diferite de eșantioane și analizând rezultatele prin prisma criteriilor anterior enunțate, s-au ales ca fiind relevante: Z_1 , Z_2 , H_1 , A_1 și A_2 . Prin urmare, clasificarea se face prin comparația unor vectori descriptori cuprinzând doar valorile corespunzătoare celor cinci momente de mai sus, vectori pe care i-am denumit :

- vectorii descriptor micști ai prototipurilor celor P clase $\vec{I}_m^{(K)}$;
- vectorul descriptor mixt al caracterului analizat \vec{X}_m .

Deci, caracterul este atribuit acelei clase C_k pentru care are distanța $D(\vec{X}_m, \vec{I}_m^{(K)})$ este minimă. Unde distanțele $D(\vec{X}_m, \vec{I}_m^{(K)})$ se calculează cu relația:

$$D(\vec{X}_m, \vec{I}_m^{(K)}) = \|\vec{X}_m - \vec{I}_m^{(K)}\| \quad (4.2.9)$$

Testările pe eșantioane de caractere scrise de mână, așa cum se va arăta în Capitolul 5 “Rezultate experimentale” arată o creștere substanțială a ratei de clasificare corectă datorită combinării avantajelor unor momente diferite. În plus, deoarece calculele se reduc ca volum, viteza de recunoaștere crește.

4.2.5 Rețele neuronale RBF arborescente

Rețelele neuronale cu neuroni cu funcții radiale (RBF Radial Basis Function) au fost relativ des folosite pentru recunoașterea caracterelor, fie ele scrise de mână sau de tipar [RIP 96]. Justificarea stă în teorema lui Cover, care stipulează că o problemă de clasificare se poate rezolva prin transferul ei neliniar într-un spațiu multidimensional de un ordin superior.

Această transformare neliniară este realizată cu o rețea neuronală RBF având următoarea structură:

- Stratul de intrare, cuprinzând un număr de noduri sursă egal cu dimensiunea vectorului descriptor de intrare.
- Stratul ascuns, conținând de obicei un număr mare de neuroni RBF fiecare conectat direct la toate nodurile sursă. Transformarea de la intrare la acest strat este neliniară.
- Stratul de ieșire, cuprinzând neuroni liniari conectați fiecare la toți neuronii RBF de pe stratul anterior. Transformarea între straturile doi și trei este liniară.

Strategia de învățare [HAY 94] depinde de tipul rețelei RBF (regulată sau generalizată). Rețelele regulate au centrul funcțiilor radiale ale unităților fixați în centrele de aglomerare ale eșantioanelor din setul de antrenament. Uneori din dorința de simplificare se practică alegerea aleatoare a centrilor. Vectorul ponderilor W se calculează cu formula matricială:

$$W = G^+ \times D \quad (4.2.10)$$

unde vectorul D este răspunsul dorit iar G^+ este pseudoinversa matricei $G=[g_{ji}]$, g_{ji} fiind funcțiile Green:

$$g_{ji} = \exp\left(-\frac{\bar{X}}{\sigma^2} \|X_j - T_i\|^2\right) \quad j=1,2,\dots,N; i=1,2,\dots,M \quad (4.2.11)$$

cu X_j intrarea de ordin j a setului de antrenament, T_i poziția centrului fix, \bar{X} media și σ dispersia.

Deși rezultatele raportate la utilizarea rețelelor RBF în recunoașterea scrisului de mână sunt promițătoare, există serioase probleme legate de dimensiunea prea mare a stratului ascuns [LEM 93]. Construirea unei rețele neuronale capabilă să soluționeze problema recunoașterii tuturor literelor mici ale alfabetului latin necesită un strat intermediar cu foarte multe unități și prin urmare și calcule matriciale pe măsură. Analizând rezultatele experimentale obținute de autor la testarea unei rețele RBF standard, a reieșit clar că mare parte din erorile de clasificare (peste 80%) apar între litere similare.

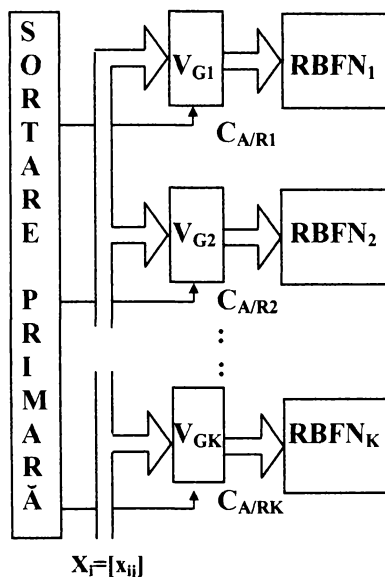


Figura 4.2.6. Rețea neuronală arborescentă RBF

În contextul metodologiei etapizate anterior descrisă, neajunsurile semnalate pot fi remediate folosind o rețea RBF arborescentă [LAC 97a, LAC97c, LAC 98a]. O asemenea rețea va cuprinde o rădăcină unde se execută sortarea primară și ramurile unde se ia decizia finală cu privire la clasificarea în interiorul sub-grupelor. Experimentele au evidențiat oportunitatea blocării ramurilor inactivate. Parametrii de control C_{AR} ai porților de validare V_{Gi} , care realizează obstrucțiile parțiale, sunt ieșirile blocului rădăcină "Sortare Primară". În acest bloc se fac primii doi pași ai clasificării caracterelor, deci coeficienții de atracție C_{AR} sunt rezultatul deciziilor din etapele anterioare.

Schema bloc a acestei structuri este dată în figura 4.2.6. Dacă C_{AR} este '+1' ramura este validată și decizia finală va fi extrasă de la ieșirea ei (atracție către

grupul corespunzător). Dacă $C_{A/R}$ este '-1' ramura este dezafectată și decizia finală nu va avea nici o componentă din acest grup (repulsie). Arhitectura unei ramuri a rețelei arborescente RBF propuse este reprezentată în figura 4.2.7.

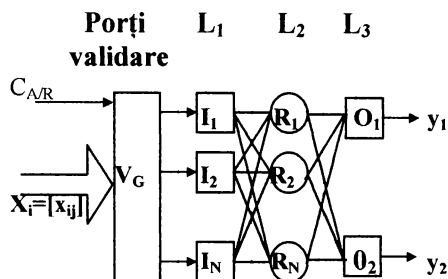


Figura 4.2.7. Ramură a unei rețele neuronale RBF arborescente

Datorită separării făcută cu porțile de validare V_G și coeficienții $C_{A/R}$ parametrii interni ai fiecărei ramuri se determină independent și calculele legate de determinarea centrilor și razelor neuronilor RBF se simplifică considerabil. Poziționarea centrelor neuronilor RBF s-a făcut cu metoda prezentată în [GOP 95]. Testările efectuate pe această rețea neuronală cu vectorii descriptori complecși ai unor seturi de caractere scrise de mână au dovedit că această soluție îmbunătățește rata de recunoaștere corectă (vezi Capitolul 5 "Rezultate experimentale").

4.3 Sistem personalizat de recunoaștere a caracterelor

Scrisul fiecărei persoane are caracteristici individuale specifice, ceea ce duce la o substanțială variație a unui caracter așa cum se vede din eșantioanele prezentate în figura 4.3.1. În plus, funcție de o situație sau alta chiar scrisul aceluiași subiect poate fi relativ diferit, deoarece viteza cu care se scrie și starea fizico-psihică influențează substanțial forma și dispunerea spațială a caracterelor. Totuși, statistic vorbind, pentru una și aceeași persoană diferențele de scriere în condiții similare sunt rezonabil de mici și deci se pot stabili formele individuale tipice ale fiecărui caracter prin medierea unui set de câteva eșantioane. Prin urmare, performanțele sistemelor de recunoaștere sunt mai slabe în cazul unor echipamente de uz general față de echipamentele de uz personal, fapt confirmat și de experiențele făcute de autor [LAC 97b].



Figura 4.3.1. Forme diferite ale aceleiași litere scrise de mână

Pentru a contrabalansa riscul destul de mare de erori datorate unor variații apreciabile a formei literelor și cifrelor de la un individ la altul, marea majoritate a metodologiilor de recunoaștere actuale respectă două reguli fundamentale:

- utilizarea parametrilor insensibili la mici variații de formă;
- mărirea ratei de respingere pentru micșorarea ratei de eroare.

Complementar cu această orientare, există studii intensive pentru perfecționarea unor instrumente personalizate de recunoaștere a scrisului individual, capabile să învețe scrisul unei persoane și să poată apoi traduce în cod ASCII pentru calculator documentele olografe ale acesteia. Aplicabilitatea unor asemenea sisteme a fost inițial considerată ca foarte restrânsă, dar ele au devenit actualmente foarte utilizate deoarece corespund întru totul cerințelor aplicațiilor de recunoaștere “on line”, agendele cu ecran sensibil fiind obiecte de uz aproape exclusiv individual.

Paragraful de față propune utilizarea unei tehnici flexibile, cu caracteristici personalizate și în cazul unui sistem “off line” destinat unui utilizator unic. Experimentele au scos în evidență faptul că recunoașterea caracterelor scrisului de mână individual este sensibil îmbunătățită în situația în care alături de informația privind forma și compoziția literei este disponibilă și distribuția spațială a subcomponentelor acesteia. Procesul efectiv de recunoaștere constă în identificarea subcomponentelor principale: puncte, linii, bucle deschise, bucle închise, spații libere ale fiecărui caracter scris de utilizator și clasificarea lui funcție de prezența și poziția acestora. Practic se recurge la divizarea verticală a caracterelor după regulile și metodologia de la paragraful 4.2 “Strategie etapizată de recunoaștere a caracterelor scrise de mână”. Apoi, în mod similar, se face analiza multicriteriu a caracterului:

- zonele verticale ocupate;
- detecția subcomponentelor din fiecare zonă;
- descriptorul mixt simplificat (momente invariante).

Vectorul descriptor rezultat se compară cu vectorii descriptori medii tipici ai caracterelor scrise de utilizatorul soft-ului de recunoaștere personalizat, aflați deja în memoria calculatorului. Acești vectori descriptori medii tipici sunt procesați automat dintr-un set de eșantioane ale fiecărui caracter extras dintr-un text standard

scris de utilizator. În fapt se execută pentru idioscriptul individual procedura de grupare descrisă la paragraful 4.2.2 “Gruparea caracterelor alfabetului românesc”.

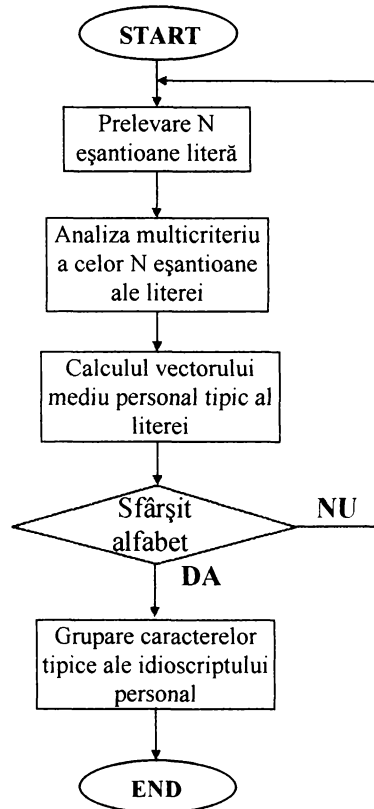


Figura 4.3.2. Algoritmul etapei de învățare a idioscriptului

Condițiile aplicației au impus în acest caz automatizarea procedurii de analiză astfel încât gruparea caracterelor scrisului individual să se producă în blocul “Grupare caractere tipice ale idioscriptului personal”. Schema logică a algoritmului de analiză a idioscriptului individual este dată în figura:4.3.2.

Rezultatele experimentelor arată că media tipică a caracterului se poate stabili în condiții optime la un număr de aproximativ cinci eşantioane pe caracter [LAC 97b]. Pentru a putea scoate în evidență poziția subcomponentelor fiecărui caracter, li s-a cerut subiecților experimentelor ca fiecare eşantion să fie scris pe un formular special unde au fost trasate punctat liniile de diviziune verticală.

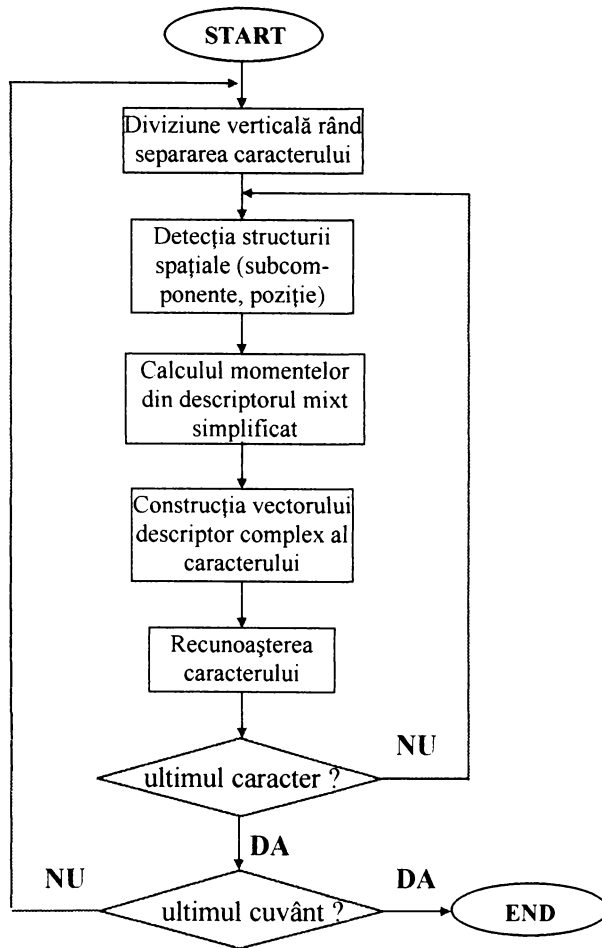


Figura 4.3.3. Algoritmul de procesare al vectorului mediu tipic al unei litere

Odată executată divizarea verticală se pot stabili zonele ocupate de caracterul studiat. Detecția prezenței unor subcomponente precum liniile, buclele deschise, buclele închise și a pauzelor se poate face utilizând:

- profilogramele verticale și orizontale $p(f_k)$ ale diviziunilor caracterelor în varianta umplută care este mai puțin sensibilă la variațiile de formă;
- momentele considerate la paragraful 4.2.2. "Gruparea caracterelor alfabetului românesc" ale acestor profilograme.

Procesarea vectorului mediu tipic al unei litere se face conform schemei logice prezentate în figura 4.3.3.

Deci după cum se poate observa și în acest caz se aplică proceduri similare cu operațiile de la cazul general. Analiza elementelor de formă conduce la descrierea structurii spațiale a literei. La această descriere se adaugă grupul de descriptori mixt al momentelor invariante și rezultă vectorul descriptor complex al caracterului analizat. Algoritmul de lucru pentru fiecare literă în parte este cel din figura 4.3.3. Vectorul descriptor complex rezultat se utilizează în procesul de clasificare cu ajutorul rețelei neuronale RBF descrisă în paragraful anterior.

5. Rezultatele experimentale

Acest capitol cuprinde, pe lângă prezentarea unor rezultate experimentale ilustrative pentru performanțele și modul efectiv de lucru al algoritmilor descriși anterior și o analiză comparativă cu performanțele unor metode similare descrise în literatura recentă. Rezultatele corespunzătoare aplicării procedurilor prezentate în articolele de specialitate sunt obținute prin implementarea și testarea algoritmilor respectivi de către autor și corespund ca performanțe cu cele declarate în articolele respective.

Situațiile prezentate în continuare pun în evidență limitările pe care le au aceste proceduri și evidențiază cazurile în care soluțiile propuse în această teză sunt avantajoase. Așa cum se va vedea mai jos, unele dintre aceste situații sunt legate de problematica specifică textelor în limba română, considerarea lor fiind un primat al activității de cercetare prezentate în lucrarea de față.

5.1 Divizarea prin metoda profilogramelor

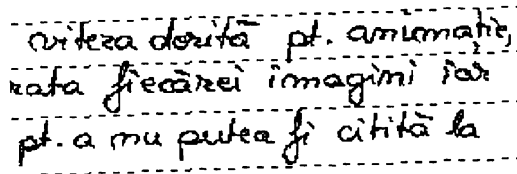
Experimentele efectuate pentru divizarea textelor în rânduri, cuvinte și în final litere au arătat că metoda profilogramelor în varianta ei flexibilă, așa cum a fost propusă în capitolul anterior, are capacitatea de a trata cu succes o serie de situații dificile pe care metodele cunoscute nu sunt capabile să le soluționeze în mod corect.

Executantul se obligă a răspunde material și civil, suportând totodată consecințele răspunderii penale, după caz, pentru pagubele aduse patronului beneficiar, ca urmare a nerealizării lucrărilor sau serviciilor și pentru nerespectarea celorlalte condiții stipulate în prezentul contract.

Pentru lucrările executate sau prestarea serviciilor executantul va răspunde pentru

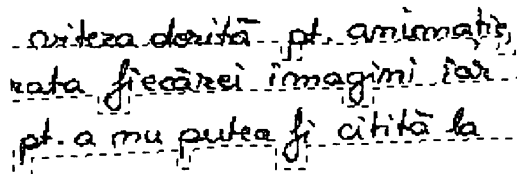
Figura 5.1.1. Separarea rândurilor printr-o metodă standard într-un document cu structură ordonată

Astfel, separarea textului în rânduri, prin frontiere drepte, este foarte eficientă pentru textele tipărite sau scrise de mână având o structură ordonată. În aceste cazuri, rândurile sunt drepte și au înălțimea cvasi-constantă, pauzele dintre ele fiind de asemenea aproximativ egale ca lățime. O astfel de situație este prezentată în figura 5.1.1. Pentru situații mai puțin favorabile, neregularitatea rândurilor și întrepătrunderea lor parțială produc erori substanțiale, care au un puternic impact negativ în continuare în procesul de recunoaștere. Rezultatele a două variante neflexibile de separare [SHA 93, LIK 95] sunt figurate în figura 5.1.2.a și b iar cele ale metodei flexibile [LAC 98] s-au reprezentat în figura 5.1.2.c.



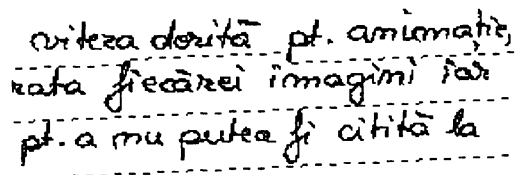
viteza dorită pt. amiază,
rata fiecărei imagini iar
pt. a nu putea fi citită la

a.



viteza dorită pt. amiază,
rata fiecărei imagini iar
pt. a nu putea fi citită la

b.



viteza dorită pt. amiază,
rata fiecărei imagini iar
pt. a nu putea fi citită la

c.

- a. Separarea cu linii drepte (secționare eronată a literei 'p')
- b. Separarea cu urmărirea liniei de bază (separarea sedilei la 't')
- c. Separarea cu frontiere flexibile (divizare corectă)

Figura 5.1.2. Prezentarea comparativă a rezultatelor metodei flexibile în raport cu metodele clasice

Separarea cu linii drepte, prezentată în figura 5.1.2.a, este metoda cea mai simplă și mai rapidă, dar, în situația unor suprapuneri parțiale ale zonei descendente

și respectiv ascendente a două rânduri consecutive, apar inevitabil erori. În figura 5.1.2.b este prezentat rezultatul unei metode de lucru mai laborioase. Separarea rândurilor se face pe baza unor frontiere ce urmăresc conturul inferior al cuvintelor. Rezultatele sunt foarte bune pentru textele scrise în limbile având alfabetul fără semne diacritice (de exemplu engleza). Pentru texte în limba română sau în orice altă limbă în care există litere cu segmente izolate în zona descendentă (sedile) metoda produce amputarea acestora. În situația dată, sunt ulterior afectate recunoașterea corectă atât a cuvântului trunchiat din rândul superior (“animație”), cât și a cuvântului de dedesubt (“iar”), unde sedila conduce cu mare probabilitate la interpretări greșite.

După cum se poate observa, separarea cu o graniță cu formă flexibilă din figura 5.1.2.c este net avantajoasă, deoarece permite atât evitarea secționării acolo unde există suprapuneri parțiale de rânduri cât și păstrarea sedilelor atașate de literele corespunzătoare. Dacă în drumul ei frontiera flexibilă întâlnește o sedilă, în conformitate cu cele prezentate în capitolul anterior, ocolirea acesteia se face pe dedesubt așa cum se arată în figura 5.1.3.

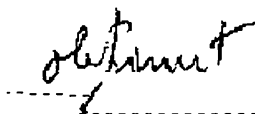
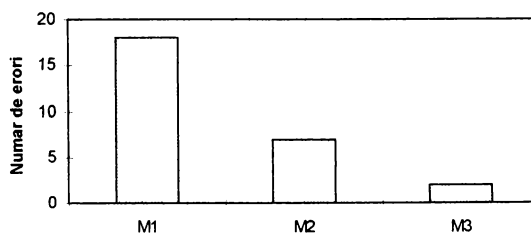


Figura 5.1.3. Eliminarea erorilor de amputare prin ocolirea pe dedesubt a sedilelor



- M1 segmentarea cu linii drepte
- M2 segmentarea cu urmărirea liniei de bază
- M3 segmentarea cu frontiere flexibile

Figura 5.1.4. Numărul de erori de segmentare a textului în rânduri pe setul de eșantioane de test

Datorită metodologiei aplicate, autorul a reușit reducerea substanțială a numărului de erori la segmentarea textelor în rânduri. Statistica acestor erori pe un set de 24 de eșantioane de test este prezentată în figura 5.1.4.

Pe lângă avantajele anterior prezentate, metoda propusă este mai rapidă deoarece introduce procedura de urmărire doar acolo unde este cu adevărat nevoie de ea, în rest granița fiind trasată drept fără complicații suplimentare.

Separarea cuvintelor în caracterele componente în cazul scrisului de mână este prezentată în întreaga literatură de specialitate ca o problemă delicată pentru care încă nu există o soluție suficient de precisă. Una dintre cele mai importante cauze ale erorilor de segmentare este înclinarea variabilă a scrierii. În articolele din bibliografie [PAL 96, CHA 97, JIA 97] această înclinare este tratată ca o abatere de la verticalitate și este compensată prin intermediul unei corecții de unghi. Aceasta însă poate să genereze la rândul ei deformări și ulterior erori de recunoaștere a caracterelor.

Teza de față propune tratarea înclinării variabile a scrisului de mână ca pe o componentă naturală a acestuia și, în această perspectivă, tratează întreaga procedură de separare (umplere, analiză, trasare frontiere) ținând cont de unghiul de înclinare mediu al scrisului [LAC 98d].

În figura 5.1.5. se evidențiază că utilizarea metodei profilogramelor conduce la o separare corectă a cuvintelor în litere pentru cele două eșantioane având o înclinare medie diferită. Separarea cuvintelor în literele lor componente se execută cu granițe având înclinarea egală cu unghiul mediu al scrisului determinat prin metoda de lucru propusă de autor în articolele [LAC 95, LAC 97e] și descrisă în subcapitolul 4.1.2 “Separarea cuvintelor în litere”.

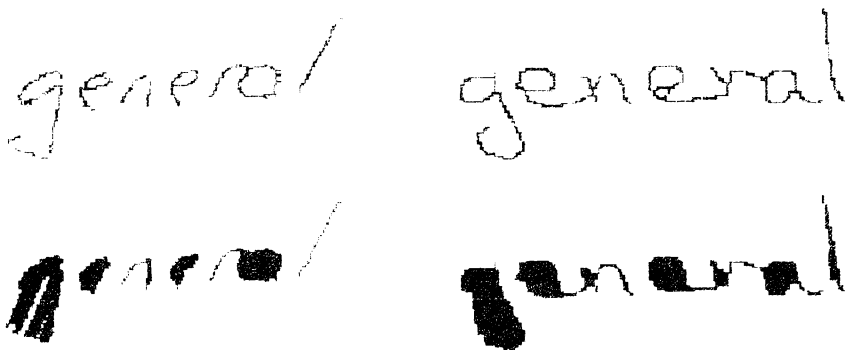


Figura 5.1.5. Separarea cuvintelor în litere pentru două eșantioane scrise cu unghiuri medii de înclinare diferite

Formula de determinare a unghiului mediu de înclinare a scrierii pentru întregul eșantion de text (4.1.4) este o medie ponderată a unghiurilor determinate pentru fiecare rând. În situația în care eșantionul analizat este scris într-o manieră neregulată (unghi variabil), apare posibilitatea ca unghiul mediu de înclinare al scrisului determinat conform metodologiei citate anterior, să nu fie același cu unghiul de înclinare local. Rezultă de aici erori similare cu cea din figura 5.1.6.a. Aplicarea metodei flexibile propusă de autor în articolul [LAC 98d] și descrisă în paragraful 4.1.2.1 “Flexibilizarea separării caracterelor” permite evitarea erorilor așa cum se arată în figura 5.1.6.b.

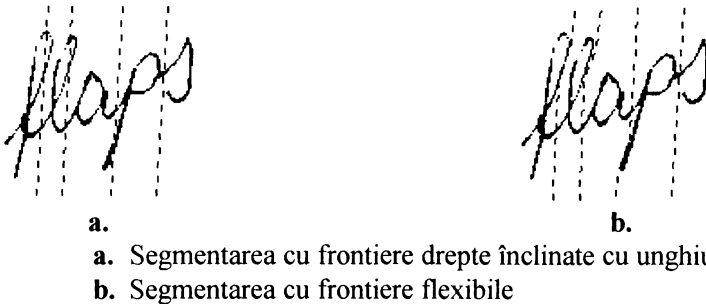
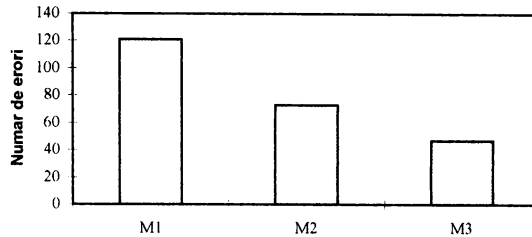


Figura 5.1.6. Separarea cuvintelor în litere în cazul unui scris cu unghi variabil

În figura 5.1.6.a segmentarea cuvântului “flaps”, înclinat cu un unghi diferit de media generală, este segmentat corect, dar frontierele fiind linii drepte se intersectează în zonele ascendente și descendente cu extensiile respective ale caracterelor. Varianta flexibilă de separare prezentată în figura 5.1.5.b evită tăierea caracterelor prin procedura de urmărire a conturului din punctul de intersecție și până în punctul în care continuarea trasării cu linie dreaptă este posibilă din nou.

Metoda de separare a cuvintelor în litere cu frontiere flexibile a fost testată în paralel cu metoda de separare cu frontiere drepte perpendiculare și respectiv cu metoda ce utilizează frontiere drepte înclinate cu unghiul mediu al scrisului. Statistica numărului de erori obținute pe un număr de 10 eșantioane de text este redată în figura 5.1.7.

Așa cum s-a arătat la descrierea metodei, procedura corectivă nu intră în funcțiune decât atunci când este cazul (intersecție între frontieră și caracter în zonele ascendente sau descendente). Prin urmare viteza de lucru a sistemului nu este afectată substanțial. Scăderea în continuare a numărului de erori la segmentarea cuvintelor în litere se poate obține prin considerarea inițială a mai multor variante de separare, examinarea separată a fiecăreia dintre ele și validarea finală a celei corecte prin utilizarea controlului ortografic și eventual a celui gramatical al textului rezultat după citirea automată.



M1 segmentarea cu frontiere drepte perpendiculare
M2 segmentarea cu frontiere drepte înclinate cu unghiul mediu
M3 segmentarea cu frontiere flexibile

Figura 5.1.7. Statistica erorilor de segmentare a cuvintelor în litere pe setul de eșantioane de test

5.2 Recunoașterea caracterelor scrise de mână

Experimentele privind recunoașterea caracterelor scrise de mână s-au realizat în două etape distincte. La început a fost utilizată metoda celor mai apropiați k vecini (distanță “City Block”), iar clasificarea s-a efectuat pe baza unui vector descriptor mixt rezultat din combinarea celor trei tipuri de momente invariante H_u , Zernicke și A fine. Ulterior, odată cu adoptarea soluției unor vectori descriptori complecși cuprinzând pe lângă momente și informația privitoare la structura spațială a caracterelor pe cele trei niveluri, s-a trecut la folosirea unei rețele neuronale RBF arborescente. Rezultatele experimentale s-au obținut pe seturi de eșantioane prelevate din baza de date “NIST Standard Reference Form Base Handprint Recognition System” [GAR 97] considerată una din referințele standard pentru evaluarea performanțelor echipamentelor și soft-urilor de recunoaștere a caracterelor. Totodată în cursul experimentărilor s-au conturat două direcții de lucru în privința recunoașterii caracterelor scrise de mână:

- recunoașterea în cazul general (pe seturi de eșantioane cu autori multipli);
- recunoașterea cu sistem personalizat.

Prin urmare, acest subcapitol a fost organizat pe două paragrafe distincte cuprinzând rezultatele experimentelor pentru cele două situații menționate mai sus.

5.2.1 Recunoașterea caracterelor în cazul general

Testările preliminare s-au efectuat pentru a evalua comparativ eficiența utilizării unui vector descriptor complex cuprinzând o combinație de momente față de vectorii descriptori simpli cuprinzând un singur tip de momente. Din motive de simplitate s-au analizat rezultatele obținute cu metoda celor mai apropiați k vecini utilizată pentru zece eșantioane ale fiecărei litere mici a alfabetului scrise de același subiect. Rezultatele clasificării sunt date în tabelul 5.2.1.

Tabel 5.2.1. Clasificarea cu vectori descriptori simpli

Set momente	Clasificări corecte [%]
Hu	53,08
Affine	69,23
Zernike	72,31

Pentru îmbunătățirea performanțelor, autorul propune clasificarea eșantioanelor utilizând un vector descriptor mixt conținând momente Hu (H_1), Zernicke (Z_{00} , Z_{31}) și Afine (A_1 , A_2). Acuratețea clasificării a crescut la 93,75% pentru același set de eșantioane. Selecția acestor cinci momente din totalul celor 15 utilizate s-a făcut pe baza analizei parametrilor precizați în paragraful 4.2.4 “Vectori descriptori micști”: dispersia și distanța [LAC 96b, LAC 96e].

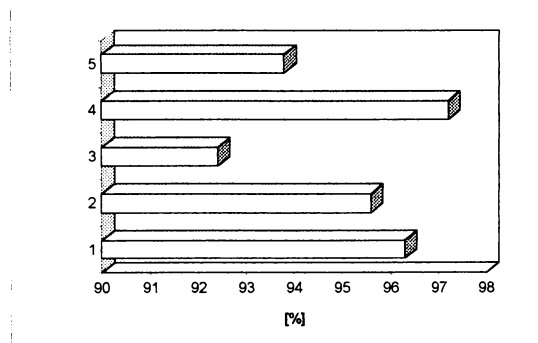


Figura 5.2.1. Rezultatele experimentelor pe un număr de cinci seturi de eșantioane cu autor unic

Experimentarea pe alte cinci seturi de eșantioane având în componență imaginile unor caractere scrise de același subiect au dat rezultate similare. Aceleași

descriptor mixt, iar rezultatele au rămas în aceeași marjă de valori. Testările efectuate pe un set de eșantioane provenind de la zece persoane diferite au dus la reducerea performanțelor și autorul a optat pentru includerea în vectorul descriptor a unor elemente de informație suplimentară legată de arhitectura caracterului analizat[LAC 96a].

Astfel, vectorul descriptor al unui caracter conține:

- combinația de momente H_u , Zernicke și Afine,
- codul zonelor ocupate de caracter,
- identitatea subcomponentelor din fiecare zonă.

Experimentele s-au făcut cu rețeaua neuronală RBF arborescentă propusă de autor în [LAC 97a, LAC 98a] și a cărei arhitectură a fost descrisă în paragraful 4.2.5. Evaluarea comportării și performanțelor s-a făcut pentru următoarele trei seturi de date:

- A. Câte zece eșantioane ale fiecăruia din cele 26 de caractere ale alfabetului latin plus cele cinci specific românești, scrise de mână de același subiect. Testarea s-a făcut cu set egal ca magnitudine și compoziție cu primul.
- B. Caractere scrise de mână ale unui număr de zece subiecți diferiți, câte cinci pentru fiecare literă a alfabetului (adică 1530 de eșantioane). Testul s-a executat cu un set cu aceeași compoziție conținând 500 de eșantioane din setul C.
- C. Caractere scrise de mână provenind de la 30 de subiecți diferiți de cei din setul B. Cele 1500 de eșantioane au fost selectate aleator cu singura restricție că fiecare literă a alfabetului românesc să fie reprezentată. Setul de test conține 500 de eșantioane selectate în aceleași condiții.

În fiecare caz selecția centrelor neuronilor RBF s-a făcut pe baza detecției centrelor de aglomerare ale eșantioanelor revelate de o preclasificare a setului cu metoda celor mai apropiați k vecini, utilizată în mod similar cu etapa întâi. Rezultatele experimentelor pe cele trei seturi sunt date în tabelul 5.2.2 și sunt în ansamblu mai bune decât cele raportate în articolele recente [KUD 96, AHI 97].

Tabel 5.2.2. Rezultatele clasificării cu rețeaua neuronală

Setul de caractere	Clasificări corecte [%]	Reducerea erorilor între litere similare[%]
A	95,36	89,21
B	88,60	47,42
C	78,40	23,84

Nu s-au evidențiat erori sistematice la recunoașterea unor caractere, dar a putut fi detectată o anumită sensibilitate a sistemului la discriminarea între literele

‘ă’ și ‘â’. Pentru reducerea ratei acestor erori, în cazul setului de învățare C se poate asigura o gamă mai bogată de eșantioane pentru cele două caractere, astfel încât celula corespunzătoare recunoașterii lor să fie mai bine populată cu neuroni RBF.

Testările au relevat de asemenea că performanțele sistemului nu sunt afectate de calculul separat al ramurilor. Un alt aspect pozitiv este faptul că stratul RBF al celei mai populate ramuri conține în cazul C, cel mai dificil, doar 37 de neuroni. Adică recunoașterea caracterelor este obținută cu o rețea neuronală având, datorită ramificării, un număr mai mic de neuroni RBF pe stratul interior decât dacă implementarea s-ar fi făcut în modul standard.

5.2.2 Personalizarea recunoașterii

Așa cum se vede în paragraful anterior, rezultatele obținute în cazul efectuării învățării și testării pe un set de eșantioane prelevate de la același subiect sunt substanțial mai bune și, prin urmare, conduc spre ideea utilității realizării unui soft capabil să se personalizeze, adică să învețe scrisul unui singur autor și să lucreze doar cu acesta. Problema principală în aceste cazuri este obținerea unor rezultate cât mai bune cu utilizarea unui număr cât mai mic de eșantioane pentru fiecare literă în parte. Experimentele au dovedit că numărul optim de eșantioane este variabil în principal funcție de două elemente:

- individ,
- literă.

Ca și în cazul agendelor electronice cu ecran sensibil la atingere, autorul a cerut subiecților experimentului să scrie eșantioanele prelevate pentru învățare în modul lor natural, fără grabă, dar și fără un exces de atenție la trasarea caracterelor. S-au cerut zece eșantioane pentru fiecare literă, obținându-se de la un număr de 24 de subiecți tot atâtea seturi similare cu setul A de eșantioane din paragraful anterior.

Rezultatele testelor efectuate separat pe câte un set individual complet dau rate de recunoaștere foarte bune variind între 94,5% și 99,7%. Media de reușite pentru toate aceste zece teste este de 98,47%.

În continuare, autorul a analizat posibilitățile de reducere a numărului de eșantioane cerute pentru fiecare test supraveghindu-se cu atenție rata de recunoaștere medie astfel încât să nu se piardă în mod esențial din acuratețea sistemului.

În mod evident, atât diferențele dintre subiecții supuși testului, cât și structurile caracterelor alfanumerice au avut o influență destul de sensibilă asupra rezultatelor, dar chiar și pe acest număr relativ mic de subiecți avut la dispoziție a fost posibilă evidențierea următoarelor concluzii:

- Setul de învățare a unui caracter este mai redus pentru: caracterele simple cum este litera 'i' și caracterele conținând mai ales segmente de dreaptă cum este 'z'. În figura 5.2.2 sunt reprezentate rezultatele experimentului de reducere a numărului de eșantioane de învățare pentru litera 'i'.

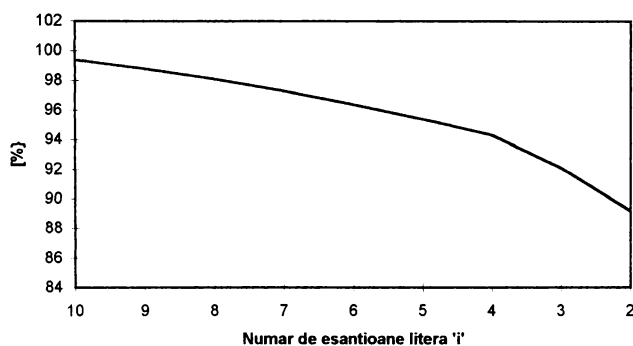


Figura 5.2.2. Rezultatele experimentului de reducere a numărului de eșantioane de învățare pentru litera 'i'

- Setul de învățare a unui caracter este mai mare pentru: caracterele complexe cum sunt literele 'm', 'w' și caracterele conținând mai ales segmente de curbă cum sunt 'a' sau 'b'. În figura 5.2.3 sunt reprezentate rezultatele experimentului de reducere a numărului de eșantioane de învățare pentru litera 'm'.

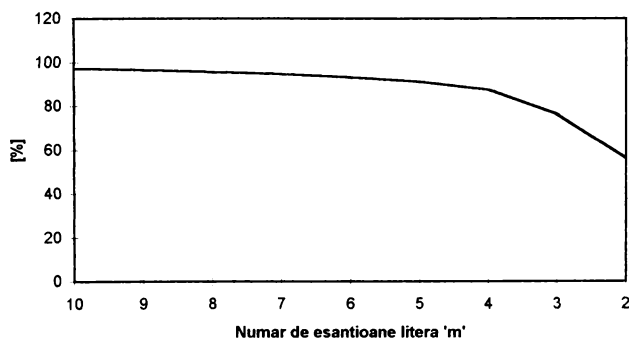


Figura 5.2.3. Rezultatele experimentului de reducere a numărului de eșantioane de învățare pentru litera 'm'

- Setul pentru învățarea cifrelor este mai mic deoarece diferențele dintre aceste sunt mai ușor de evidențiat. În figura 5.2.4 sunt reprezentate

rezultatele experimentului de reducere a numărului de eşantioane de învățare pentru cifre.

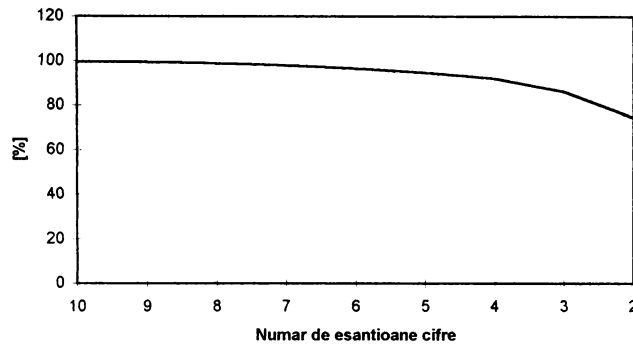


Figura 5.2.4. Influența numărului de eşantioane de învățare pentru recunoașterea cifrelor

- Setul minim necesar pentru învățare poate să varieze funcție de regularitatea idioscriptului individual. În acest sens sunt relevante diferențele înregistrate la experimentele pe seturi individuale reprezentate în figura 5.2.1.

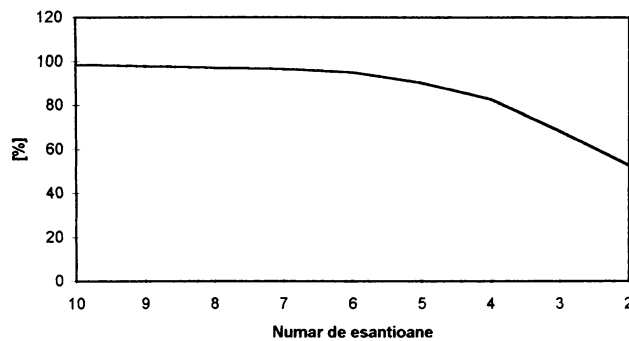


Figura 5.2.5. Influența reducerii numărului de eşantioane pentru acuratețea recunoașterii (media pentru întregul lot de subiecți)

Totuși diferențele față de medie pentru cazul favorabil (literă simplă, subiect cu scris stabil) și respectiv pentru cazul defavorabil (literă complexă, subiect cu scris instabil) nu s-au dovedit substanțiale și s-a putut continua efectuarea analizei pe baza statisticii de ansamblu prezentată în figura 5.2.5.

Pe baza acestor rezultate se poate concluziona că, în situații tipice, este suficientă prelevarea a câte cinci eșantioane de învățare pentru fiecare caracter alfanumeric. Totuși, în interesul creșterii capacității de adaptare a sistemului, în partea de inițializare a algoritmului este cuprinsă și opțiunea de mărire a acestui prag. În acest fel se pot crește performanțele sistemului chiar și în situația unor utilizatori cu grafie foarte neregulată.

6. Concluzii

6.1 Contribuții personale

În lucrarea de față se pot evidenția un număr de 11 contribuții personale. Aceste contribuții, având caracter de noutate în domeniul recunoașterii caracterelor alfanumerice, sunt prezentate sintetic mai jos, fiind însoțite de referințele bibliografice la articolele publicate de autor:

- Considerarea unghiului de înclinație a scrisului de mână ca o componentă naturală a acestuia și nu ca o abatere de la verticalitate, ce trebuie compensată.

În acest fel este eliminată etapa de corecție, generatoare de erori și întreaga procedură propusă pentru separare și recunoaștere conține unghiul de înclinare a scrisului ca și parametru. Rezultatele cercetărilor au dovedit că această abordare permite o creștere a acurateții segmentării cuvintelor în caracterele componente și o rată sensibil mai mare de succes la recunoaștere [LAC 98d].

- Modificarea transformatei Hough pentru determinarea unghiului de înclinare a rândurilor.

Varianta propusă este concepută astfel încât să includă în analiză doar aspectele relevante ale problemei în discuție. La scrisul de mână rândurile de text nu sunt întotdeauna paralele cu orizontala colii de hârtie și în plus la preluarea imaginii sau scanarea documentului analizat nu se pot face de obicei alinieri precise. Cunoașterea unghiului permite o separare corectă a rândurilor dintr-un text și influențează pozitiv etapele ulterioare de divizare a fiecărui rând în cuvinte și a acestora în litere [LAC 95].

- Metoda de determinare a unghiului mediu al scrisului de mână.

Procesarea constă într-o transformare similară cu cea anterioară, adaptată însă la necesitățile aplicației curente. Testele experimentale au arătat o insensibilitate ridicată a rezultatelor prelucrării la zgomotul din imagine și au validat acuratețea determinării în cazul unor eșantioane cu scris relativ ordonat, unde înclinarea nu are abateri mari de la valoarea medie pe întregul document. Pentru cazul unei grafii mai neregulate s-au prevăzut posibilități de corecție locală. Unghiul mediu al scrisului, determinat în cadrul acestei

etape, devine un parametru esențial pentru analizele din stadiile ulterioare asupra cuvintelor și caracterelor [LAC 95, LAC 97e].

- Evitarea amputării extensiilor din zona descendentă, izolate de corpul central al caracterului, prin separarea rândurilor de text cu frontiere flexibile.

Aceste extensii, numite în cadrul tezei semne diacritice inferioare, nu apar la literele alfabetului englez (cel mai frecvent analizat în articolele din literatura de specialitate), dar există în alfabetul românesc și în alte alfabete. Textele scrise în acele limbi nu pot fi segmentate corect în rânduri componente decât prin procedura propusă, procedură care este capabilă să detecteze și să aloce aceste extensii izolate. Soluția de față prezintă, de asemenea, avantajul vitezei, întrucât utilizează segmente de dreaptă ca și granițe în zonele unde nu există probleme. Prezentarea metodei frontierelor flexibile și analiza comparativă a performanțelor acesteia în raport cu procedeele clasice a fost făcută de autor în [LAC 98b].

- Separarea caracterelor componente ale cuvintelor cu frontiere flexibile.

Metoda propusă conduce la reducerea erorilor datorate tăierii extensiilor ascendente și descendente ale caracterelor. Divizarea cuvintelor în literele componente cu frontiere drepte, perpendiculare pe orizontala rândului, poate conduce la secționări nedorite ale extensiilor. Utilizarea unor frontiere drepte, înclinate cu unghiul mediu al scrisului rezultat din analiza întregului document, reduce substanțial pericolul apariției unor asemenea amputări în cazul unui scris relativ regulat. În schimb, pentru cazul unor eșantioane de scris mai puțin ordonat este necesară adoptarea, ca măsură suplimentară de precauție, a procedurii de urmărire a conturului caracterului acolo unde apare intersecția caracter - frontieră de separație. Rutina de ocolire nu periclitează substanțial viteza de lucru a algoritmului general deoarece ea intră în acțiune doar în caz de necesitate [LAC 98c].

- Metodologia etapizată de recunoaștere a caracterelor, care asigură clasificarea unui caracter necunoscut printr-un șir de etape succesive.

Caracterul necunoscut este analizat funcție de caracteristicile sale și plasat întâi într-o grupă și apoi într-o subgrupă de litere posibile. Decizia finală este luată doar între variantele din interiorul subgrupeii și poate avea caracter probabilistic. Divizarea deciziei în etape a permis utilizarea unor clasificatori specifici pentru fiecare etapă în parte și realizarea unor optimizări de implementare cu efecte pozitive atât pentru precizia rezultatelor cât și pentru reducerea complexității [LAC 96c, LAC 96a].

- Analiza structurală a literelor alfabetului românesc.

Clasificarea caracterelor alfabetului românesc după structura lor spațială a fost o condiție esențială pentru aplicarea procedurii descrise anterior, deoarece grupele și subgrupele de litere trebuie fixate în preambulul procesului de

recunoaștere. Categoriile și subcategoriile au fost fixate printr-o selecție efectuată cu ajutorul unei combinații a criteriilor Bouma și a caracteristicile de formă ale subcomponentelor literelor alfabetului românesc [LAC 97d, LAC 98a].

- Detecția și descrierea subcomponentelor caracterelor în zonele de separație verticală.

Metoda a fost implementată și optimizată astfel încât să permită crearea unei descrieri a structurii caracterelor scrise de mână cât mai insensibilă la inconstanța formei acestora. Extragerea caracteristicilor definiției pentru o literă se face detectând prezența părților ei în zonele de diviziune verticală și analizând componența acestor părți. Umplerea s-a executat ținând cont de unghiul mediu de înclinare a scrierii. A rezultat în final o descriere a structurii spațiale a caracterului, accentuând elementele general valabile și nu variațiile de idioscript individual ale subiecților cercetării [LAC 96a].

- Descrierea caracterelor printr-un vector descriptor mixt de momente invariante.

Vectorii descriptori micști au fost utilizați deoarece experimentele inițiale au arătat că vectorii simplii, cuprinzând momente de un singur tip, nu au dat rezultate mulțumitoare în cazul scrisului de mână. Inițial s-au utilizat seturi de câte cinci descriptori, momente Hu, Zernicke și Afine. În urma testărilor, a reieșit că se poate renunța la o parte dintre aceștia (cei care nu sunt relevanți pentru aplicația dată) și a rezultat un vector mixt redus, dar care s-a dovedit foarte eficient [LAC 96b, LAC 96e].

- Utilizarea unei structuri arborescente de rețea neuronală RBF pentru recunoașterea caracterelor alfanumerice.

Structurile arborescente reprezintă o implementare naturală ce păstrează etapizarea proprie procedurii anterior pomenite. În cazul de față însă, această implementare a creat și un substanțial avantaj prin evitarea unui strat unic de neuroni RBF, cu dimensiuni foarte mari, care este greu controlabil. Divizarea rețelei în ramuri autonome ca funcționare a permis proiectarea fiecăreia în parte și a făcut posibilă optimizarea lor individuală acolo unde analiza rezultatelor testărilor a dovedit că recunoașterea este mai dificilă [LAC 97a, LAC 97c, LAC 98a].

- Sistem personalizat de recunoaștere a caracterelor proprii idioscriptului individual al utilizatorului.

Sistemul a fost conceput și implementat ca urmare a analizei comparative a rezultatelor experimentale pentru seturi de eșantioane de scris individuale și, respectiv, de grup. Constatarea că rata de succes pentru prima categorie este sensibil mai bună, a condus la ideea de a crea un algoritm capabil să învețe și apoi să recunoască scrisul personal al câte unui singur utilizator. Algoritmul de execuție a fost structurat pe etape similare cu algoritmul

pentru cazul general, dar, având în vedere avantajele specifice, a putut fi optimizată partea de învățare, a putut fi automatizată procedura de grupare a caracterelor și în final s-au obținut rezultate sensibil mai bune la recunoașterea caracterelor [LAC 97b].

6.2 Posibilități de dezvoltare ulterioară

Prin cercetările întreprinse de autor în cadrul tezei, se conturează două direcții de studiu și dezvoltare în domeniul recunoașterii caracterelor scrise de mână:

- Creșterea credibilității separării textului în rânduri, cuvinte și caractere.
- Scăderea ratelor de eroare la recunoașterea caracterelor.

Rezultatele experimentale au arătat, destul de elocvent, că mărirea flexibilității sistemului duce la îmbunătățirea performanțelor în domeniul divizării textului, începând cu separarea rândurilor și încheind cu izolarea fiecărui caracter. Prin urmare, considerarea mai multor alternative de divizare și selecția rezultatelor corecte doar la final este o cale posibilă pentru reducerea în continuare a erorilor. Într-un asemenea sistem, este posibilă implementarea unui bloc logic care să detecteze erorile corelate pe rânduri vecine sau în cuvinte vecine, erori care semnalizează ca foarte probabilă o greșeală de separare. Tratarea unei asemenea situații se face prin apelarea unei rutine speciale care reintroduce în lucru o eventuală variantă de divizare, inițial cotate ca mai puțin probabilă. Este, de asemenea, posibil ca o separare, abordată în paralel, prin mai multe metode de lucru alternative, urmată de o corelare finală a rezultatelor parțiale diferite pentru validarea soluției corecte, să ridice pragul de credibilitate la valori mai apropiate de ideal.

Limba română, având un alfabet ce conține un număr de caractere cu semne diacritice, prezintă o serie de situații specifice, cu avantaje, dar și dezavantaje în raport cu alfabetul anglo-saxon standard. Aprofundarea analizei acestor situații specifice este cu siguranță o potențială resursă de soluții utile.

Scăderea în continuare a ratelor de eroare la recunoaștere este, cu mare probabilitate, posibilă prin creșterea complexității rețelei neuronale utilizată în cercetările descrise anterior. Un prim pas în acest sens este introducerea unor parametri variabili la poziționarea centrilor neuronilor din rețeaua RBF. De asemenea, reducerea unor erori datorate impreciziei la nivelul deciziei inițiale din rădăcina rețelei este probabilă în condițiile utilizării unor tehnici fuzzy, care să permită decizii probabilistice multiple de genul “caracterul este din grupa 1.1 cu

probabilitate 81% sau din grupa 1.3 cu probabilitate 16% sau din grupa 2.2 cu probabilitate 3%” (vezi paragraful 4.2.2). Într-o asemenea situație, coeficienții de atracție / respingere nu mai sunt binari, ci devin practic niște ponderi de încredere care permit activarea parțială a mai multor sub-rețele de decizie în interiorul grupelor. Răspunsul final devine și el, prin urmare, un rezultat probabilistic multiplu: “caracterul este ‘ă’ cu probabilitatea 63% sau ‘d’ cu probabilitatea 37%”. Un asemenea tip de abordare permite ulterior o analiză la nivel de cuvânt sau propoziție care să selecteze soluția bună dintre mai multe variante posibile. Intervine aici validarea cu dicționar și, în limbile unde acest lucru este posibil, validarea gramaticală. Aceste proceduri, deși sunt mai complexe și mai mari consumatoare de timp și resurse hard, se pot dovedi rentabile deoarece au capacitatea să îmbunătățească cu câteva procente rata de reușită a recunoașterii caracterelor.

Similar cu alte aplicații de recunoaștere a formelor, și în cazul recunoașterii caracterelor, aplicarea cu consecvență a principiului combinării mai multor metode într-un complex unitar, utilizat pe parcursul acestei teze, poate fi continuată prin desfășurarea unei analize paralele pe căi alternative și combinarea finală a rezultatelor într-o formulă de intervalidare.

Lucrarea de față nu și-a propus o analiză a limbii române în vederea implementării pe lângă validarea cu lexicon și a unei validări gramaticale. Studiul tematicii propozițiilor în limba noastră, realizat printr-o colaborare inter-disciplinară cu experții în gramatică, ar putea, eventual, să descopere câteva reguli utile de analiză automată a propozițiilor și frazelor, permițând implementarea ulterioară a unui control final al recunoașterii pe această bază.

Datorită creșterii continue a volumului trimiterilor poștale, în mai puțin de un deceniu sortarea automată a corespondenței va deveni necesară și în țara noastră [IPU 98]. În baza celor citite în literatura de specialitate și a experienței acumulate în cursul elaborării tezei, autorul dorește să sugereze câteva măsuri tehnice și organizatorice:

- analiza aprofundată a facilităților și problemelor alfabetului românesc care să permită dezvoltarea unor aplicații soft specifice României;
- modificarea sistemului de codare poștală;
- reorganizarea distribuției și dotarea cu echipamente specifice;
- popularizarea avantajelor și cerințelor limitative ale automatizării la nivelul cetățenilor.

În ce privește primul punct, rezultatele teoretice cuprinse în paragraful 4.2.2 “Gruparea literelor alfabetului românesc” și îmbunătățirile la separare obținute prin flexibilizare pot fi utile. Studiul trebuie însă aprofundat prin analiza de către o echipă cu compoziție multidisciplinară a unor seturi de eșantioane de scris

reprezentative pentru gradul de cultură, variațiile zonale și locale ale populației din România.

Sistemul actual de codificare poștală a localităților este departe de a îndeplini cerințele unei procesări automate. Este necesar să se ajungă la situația ca un același cod să corespundă unui număr redus de destinatari posibili, astfel încât la sortare, analiza adresei propriu-zise și cea a codului să ofere soluții ce se pot verifica reciproc cu ajutorul bazei de date preexistente în calculatorul poștei.

Soluția de ansamblu pentru implementarea unui sistem național modern de preluare și distribuție a corespondenței nu depinde doar de factori de natură tehnică și va necesita, de asemenea, un studiu pluridisciplinar privind o serie de condiții, pornind de la structura demografică și încheind cu necesitățile de compatibilizare cu țările din Uniunea Europeană.

6.3 Perspectivile de evoluție ale citirii automate

Scierea este, alături de vorbire, un mod natural de comunicare inter-umană, cu o tradiție de câteva mii de ani și aflat în strânsă legătură cu cvasi-totalitatea manifestărilor politice, sociale și culturale ale civilizației moderne. Prin urmare, se poate estima că oamenii vor fi în continuare mari producători și consumatori de documente tipărite sau scrise de mână. Deși dezvoltarea rețelei Internet reprezintă probabil începutul unei restructurări profunde în domeniul telecomunicațiilor, literatura de specialitate conține previziunea că, preț de cel puțin o generație, trimiterile poștale clasice nu vor fi înlocuite de variantele lor electronice. Ele vor continua chiar să crească, o anumită perioadă, în special în noile țări industrializate. Rezultatul este că serviciile poștale naționale vor fi nevoite să investească în construirea și perfecționarea unor echipamente de sortare automată a corespondenței, capabile cel puțin să reducă în mod eficient presiunea asupra serviciilor de sortare manuală [LAP 98, USP 98].

Deși calculatoarele sunt extrem de prezente în activitatea financiar bancară, multe din procedurile specifice acesteia sunt profund legate de controlul de autenticitate pentru înscrisurile și formularele completate sau măcar semnate în original de mâna clientului. Această realitate ține nu atât de conservatorismul propriu mediilor financiare, cât mai ales de lipsa unor soluții infailibile pentru securizarea informațiilor în rețelele de calculatoare. De altfel, automatizarea prelucrării formularelor este deja o realitate la câteva din marile bănci americane și fără îndoială, echipamentele și programele respective vor cunoaște, în anii imediat următori, substanțiale perfecționări și totodată o mult mai largă răspândire.

La ora actuală, în SUA, Japonia și Franța, se desfășoară cercetări intense îndreptate spre realizarea unor sisteme de procesare automatizată a declarațiilor anuale de impozit ale cetățenilor și firmelor [BRE 97]. Succesul acestora va fi, cu siguranță, punctul de pornire al unei restructurări majore a serviciilor de relații cu publicul ale instituțiilor din administrația locală, regională și națională.

Se poate aprecia în general că strategiile de marketing actuale ale majorității prestatorilor de servicii sunt atente, cu prioritate, la satisfacerea celor mai sofisticate cereri ale clienților și, prin urmare, preferă mai degrabă să aloce eforturi mari în creșterea capacității de adaptare a echipamentelor de analiză a scrisului decât să impună limitări de genul “formularul trebuie obligatoriu bătut la mașină”, “nu se sortează decât plicuri standard completate cu litere de tipar” etc.

La un moment dat, se părea că problema citirii directe de către calculator a unor texte scrise de mână sau tipărite își pierde din actualitate, datorită progreselor în domeniul recunoașterii vorbirii și mai ales orientării marilor edituri spre furnizarea cărților în formate multimedia. Apariția și marele succes comercial al agendelor electronice pe al căror ecran se pot scrie direct cu mâna nume, adrese, numere de telefon sau scurte adnotări a relansat însă interesul și pentru acest domeniu. Pentru aceste agende, sarcina de descifrare este mult ușurată de faptul că au de “învățat” și apoi de “citi” doar scrisul unui singur utilizator, posesorul agendei. Pe de altă parte însă, adnotările grăbite în agendă pot fi uneori, chiar și în cazul acestei specializări, “înțelese” cu greșeli. Oricum, rezultatele par să fie mulțumitoare și numărul de utilizatori crește cu o dinamică care încurajează aprofundarea cercetărilor în domeniu [SEN 94].

Dincolo, însă, de o modă mai mult sau mai puțin trecătoare și, în general, de aplicațiile momentan interesante, recunoașterea scrisului reprezintă, în mod evident, o componentă importantă a cercetărilor pentru realizarea unui vechi și important vis al științei și tehnologiei - crearea structurilor artificiale având inteligență comparabilă și compatibilă cu cea umană.

Bibliografie

1. [ABU 98] Abuhaiba I.S.I., Holt M., Datta J., "Recognition Of Off-line Cursive Handwriting", Computer Vision And Image Understanding, vol.71, No.1, July 1998, pp.19-38.
2. [ALT 84] Altmann J., Reitbock H.J.P., "A Fast Correlation Method For Scale And Translation Invariant", Pattern Recognition, I.E.E.E. Trans. on PAMI No.1, January 1984, pp.46-57.
3. [AMI 97] Amit Y., Geman D., Wilder K., "Joint Induction Of Shape Features And Tree Classifiers", IEEE Trans. on PAMI, vol.19, No.11, November 1997.
4. [ANQ 95] Anquetin E., Lorette G., "Reconnaissance en ligne de lettres manuscrites cursives par chaines Markov cachees", Traitement du signal, vol.12, No.4-6, Avril 1995.
5. [AVI 95] Avi-Itzhak H.I., Diep T.A., Garland H., "High Accuracy Optical Character Recognition", IEEE Trans. on PAMI, vol.17, No.2, February 1995.
6. [BAD 95] Badr A., Sabri A.M., "Survey And Bibliography Of Arabic Optical Text Recognition", Signal Processing, No.41, 1994, pp.49-77.
7. [BAE 98] Bae J.H., Jung K.C., Kim J.W., Kim H.J., "Segmentation Of Touching Characters Using An MLP", Pattern Recognition Letters, North Holland, vol.19, No.8, June 1998, pp.701-709.
8. [BAI 93] Baird H.S. "Recognition Technology Frontiers", Pattern Recognition Letters, North Holland, vol.14, No.4, April 1993, pp.327-334.
9. [BAR 96] Barge M., Haggarty K., Idan Y., "64-Chanel Correlator Implementing A Kohonen-Like Neural Network For Handwritten-Digit Recognition", Applied Optics, vol.35, August 1996, pp.4655-4665.
10. [BAR 96] Barkakati N., "Borland C++ 4 Ghidul programatorului" ed. Teora, București, 1996.
11. [BAY 96] Bailey R.R., Srinath M., "Orthogonal Moment Features For Use With Parametric And Non-Parametric Classifiers", IEEE Transactions on PAMI, vol.18, April 1996, pp.389-399.
12. [BOJ 97] Bojcovici S.Z., Toma C.I., Gui V., Vasii R., "Advanced Topics In Digital Image Compression", ed. Politehnica Timișoara, 1997.
13. [BOU 71] Bouma H. "Visual Recognition Of Isolated Lower Case Letters" Vision Research, vol.7, No.11, November 1971.
14. [BRE 97] Breuel T.M., "Handwritten Character Recognition Using Neural Networks" IOP Publishing Ltd. and Oxford University Press, 1997.
15. [BUR 88] Burnap S., "Advance Turbo C Programming", Compute! Publication Inc. Greensboro North Carolina, 1988.

16. [BUR 92] Burdin V. Ghorbel F. ș.a. "A Three-Dimensional Primitive Extraction Of Long Bones Obtained From Bi-Dimensional Fourier Descriptors", Pattern Recognition Letters, North Holland, vol.13, 1992, pp 213-217.
17. [CAR 92] Carpenter S.A., "Neural Networks For Vision And Image Processing", MIT Press, 1992.
18. [CED 97] CEDAR, "Off-line Handwritten Text Recognition Research" www.cedar.buffalo.edu.
19. [CED 98a] CEDAR "The Chaincode Algorithm" www.cedar.buffalo.edu.
20. [CED 98b] CEDAR "The FSGD Algorithm" www.cedar.buffalo.edu.
21. [CED 98c] CEDAR "The Gradient Algorithm" www.cedar.buffalo.edu.
22. [CED 98d] CEDAR "The GSC Algorithm" www.cedar.buffalo.edu.
23. [CED 98e] CEDAR "The Histogram Algorithm" www.cedar.buffalo.edu.
24. [CED 98f] CEDAR "The Polynomial Classification Algorithm" www.cedar.buffalo.edu.
25. [CHA 94] Chang H.D., Wang F.J. & Kuo S.C., "A Bayesian Neural Network For Separating Similar Complex Handwritten Chinese Characters", Pattern Recognition Letters, North Holland, vol.15, No.4, April 1994, pp.403-409.
26. [CHA 96] Chang C.H., "Simulated Annealing Clustering Of Chinese Words For Contextual Text Recognition", Pattern Recognition Letters, North Holland, vol.17, No.1, January 1996, pp.57-66.
27. [CHA 97] Chaudhuri B.B. Pal U. "Skew Angle Detection Of Digitized Indian Script Documents", IEEE Trans. on PAMI, vol.19, No.2, February 1997.
28. [CHE 93] Cheriet M. & Suen C.Y., "Extraction Of Key Letters For Cursive Script Recognition", Pattern Recognition Letters, North Holland, vol.14, No.12, December 1993, pp.1009-1024.
29. [CHE 95a] Chenevoy Y., Belaid A., "Une approche structurelle pour la reconnaissance de notices bibliographiques", Traitement du signal, vol.12, No.4-6, Avril 1995.
30. [CHE 95b] Cheng R.H., Lee C.W., Chen Z., "Preclassification Of Handwritten Chinese Characters Based On Basic Stroke Substructures", Pattern Recognition Letters, North Holland, vol.16, No.10, October 1995, pp.1023-1032.
31. [CHE 95c] Cherttez J.P. "Premier degre de caracterisation des ecriture manuscrits", Traitement du signal, vol.12, No.4-6, Avril 1995.
32. [CHE 97] Chen J.L., "A Simplified Approach To The HMM Based Texture Analysis And Its Application To Document Segmentation", Pattern Recognition Letters, North Holland, vol.18, No.10, October 1997, pp.993-1007.
33. [CHI 97] Chiu H.P., Tseng D.C., "Invariant Handwritten Chinese Character Recognition Using Fuzzy Min-Max Neural Networks", Pattern Recognition Letters, North Holland, vol.18, No.5, May 1997, pp.481-491.
34. [COM 96] Commen B.J., Zhang K., "The Normalised String Editing Problem Revised", IEEE Transactions on PAMI, vol.18, June 1996, pp.669-672.

35. [DEB 92] Debenham R.M. & Garth S.C.J., " The Detection Of Eyes In Facial Images Using Radial Basis Functions", in "Neural Network For Vision Speech And Natural Language", editor Linggard R., Cambridge, 1992.
36. [DEB 97] deBruqc D., Amara M., Ruiz V., "On-line Handwritten Segmentation In Linear Drawings", Pattern Recognition Letters, North Holland, vol.18, No.1, February 1997, pp.193-202.
37. [DEC 98] de Carvalho A., Fairhurst M.C., "Applying Adaptive Logic Networks To Character Recognition", Pattern Recognition Letters, North Holland, vol.19, No.5-6, April 1998, pp.469-473.
38. [DEF 95] Deforges O., Piquin P., Viard -Gaudin C., Barba D., "Segmentation d'images de documents par une approche multiresolution. Extraction precise des lignes de texte", Traitement du signal, vol.12, No.4-6, Avril 1995.
39. [DES 95] Deseilligny M.P., Le Men H., Stamon G., "Character String Recognition On Maps, A Rotation-Invariant Recognition Method", Pattern Recognition Letters, North Holland, vol.16, No.12, December 1995, pp.1297-1310.
40. [DES 95] Deseilligny P., Le Men H., Stamon G., "Lecture automatique des ecritures sur cartes scannees", Traitement du signal, vol.12, No.4-6 Avril 1995.
41. [DES 98] Deseilligny M.P., Stamon G., Suen C.Y., "Veinerization: A New Shape Description for Flexible Skeletonization", IEEE Trans. on PAMI, vol.20, No.5, May 1998.
42. [DEW 95] De Waard A., "An Optimised Minimal Edit Distance For Handwritten Word Recognition", Pattern Recognition Letters, North Holland, vol.16, No.10, October 1995, pp.1091-1096.
43. [DJO 97] Djouadi A., Bouktache E., "A Fast Algorithm For The Nearest-Neighbour Classifier", IEEE Trans. on PAMI, vol.19, No.3, March 1997.
44. [DOR 98] Dori D., Velkovitch Y., "Segmentation And Recognition Of Dimensioning Text From Engineering Drawings", Computer Vision and Image Understanding, vol.69, No.2, February 1998, pp.196-201.
45. [DOW 82] Downing J., Leong C.K. "Psychology of Reading", Macmillan, 1982.
46. [DRA 85] Drâmba O., "Istoria culturii și civilizației" ed. Științifică și tehnică București, 1985.
47. [ETE 97] Etemard K., Doerman D., Chellappa R., "Multiscale Segmentation Of Unstructured Document Pages Using Soft Decision Integration", IEEE Trans. on PAMI, vol.19, No.1, January 1997.
48. [FAN 95a] Fan K.C., Wang L.S., "Classification Of Document Blocks Using Density Feature And Connectivity Histogram", Pattern Recognition Letters, North Holland, vol.16, No.9, September 1995, pp.955-962.
49. [FAN 95b] Fan K.C., Lu J.M., Wang L.S., Liao H.Y., "Extraction Of Characters From Documents By Feature Point Clustering", Pattern Recognition Letters, North Holland, vol.16, No.9, September 1995, pp.963-970.
50. [FLU 94] Flusser J. & Suk T., " Affine Moment Invariants A New Tool For Character Recognition", Pattern Recognition Letters, North Holland, vol.15, No.4, April 1994, pp.435-436.

51. [GAD 96] Gader P.D., Mohamed M.A., Keller J.M., "Fusion Of Handwritten Word Classifier", *Pattern Recognition Letters*, North Holland, vol.17, No.6, May 1996, pp.577-584.
52. [GAR 96] Garris M.D., Dimmick D.L., "Form Design For High Accuracy Optical Character Recognition", *IEEE Transactions on PAMI*, vol.18, June 1996, pp.653-656.
53. [GAR 97] Garris M.D., Blue J.L., Candela G.T., Grother P.J., Janet S.A., Wilson C.L., "NIST Standard Database", U.S. Department of Commerce, Information Technology Laboratory, Gaithersburgh MD 20899-0001.
54. [GEI 95] Geiger E.D., "Word-Level Recognition Of Small Sets Of Handwritten Words", *Pattern Recognition Letters*, North Holland, vol.16, No.10, October 1995, pp.999-1009.
55. [GEO 95] Geok S.N., Erdogan S.S., Pan W.N., "Contender's Network: A New Competitive-learning scheme", *Pattern Recognition Letters*, North Holland, vol.16, No.11, November 1995, pp.1111-1118.
56. [GIL 93] Gilloux M., "Research Into The New Generation Of Character And Mailing Address Recognition System At The French Post Office Research Centre", *Pattern Recognition Letters*, North Holland, vol.14, No.4, April 1993, pp.267-276.
57. [GIL 95] Giloux M., "Reconnaissance de chiffres manuscrits par model Makov pseudo-2D ", *Traitement du signal*, vol.12, No.4-6, Avril 1995.
58. [GON 93] Gonzales R., Wintz P., "Digital Image Processing", Addison-Wesley, 1993.
59. [GOP 95] Gopalakrishnan V., Sridhar V., Krishnamurthy H., "Some Applications Of Clustering In The Design Of Neural Networks", *Pattern Recognition Letters*, North Holland, vol.16, No.1, January 1995, pp.59-65.
60. [GOR 97] Gorsky I., Gorskaya L.M., "Estimation Of Prior Probabilities For Numeral Recognition", *Pattern Recognition Letters*, North Holland, vol.18, No.1, January 1997, pp.97-103.
61. [GOV 96] Govindaraju V., Krishnamurty R.K., "Holistic Handwritten Word Recognition Using Temporal Features Derived From Off-line Images", *Pattern Recognition Letters*, North Holland, vol.17, No.5, May 1996, pp.523-527.
62. [HA 97] Ha T.H., Bunke H., "Off-line, Handwritten Numeral Recognition by Perturbation Method", *IEEE Trans. on PAMI*, vol.19, No.5, May 1997.
63. [HAL 87] Hall G. & Terrell T.J., "Low-cost Microprocessor-Based Image Processing System", *Microprocessors And Microsystems*, vol.11, No.10, December 1987, pp.534-540.
64. [HAM 97] Hamamoto Y., Uchimura S., Tomita S., "A Bootstrap Technique For Nearest Neighbor Classifier Design", *IEEE Trans. on PAMI*, vol.19, No.1, January 1997.
65. [HAN 95] Han C.C., Tseng Y.L., Fan K.C., Wang A.B. "Coarse Classification Of Chinese Characters Via Stroke Clustering Method", *Pattern Recognition Letters*, North Holland, vol.16, No.10, October 1995, pp.1079-1089.
66. [HAN 96] Han K., Sethi I.K. "Handwritten Signature And Identification", *Pattern Recognition Letters*, North Holland, vol.17, No.1, January 1996, pp.83-90.

67. [HAS 96] Hastie T., R. Tibshirani R., "Discriminant Adaptive Nearest Neighbour Classification", IEEE Trans. on PAMI, vol.18, No.6, June 1996.
68. [HAY 94] Haykin S., "Neural Networks A Comprehensive Foundation", MacMillan College Publishing Company, Englewood Cliffs, NJ 1994.
69. [HEL 98] Hellemans A., Bunch B. "Istoria descoperirilor științifice" Ed. Orizonturi București 1998.
70. [HEU 98] Heutte L., Paquet T., Moreau J.V., Lecourtier Y., Olivier C, "A Structural/Statistical Feature Based Vector For Handwritten Character Recognition", Pattern Recognition Letters, North Holland, vol.19, No.7, May 1998, pp.629-641.
71. [HO 92] Ho T.K., Hull J.J., Srihari S.N., "A Word Shape Analysis Approach To Lexicon Based Word Recognition", Pattern Recognition Letters, North Holland, vol.13, 1992, pp.821-826.
72. [HO 97] Ho T.K., Baird H.S., "Large-scale Simulation Studies In Image Pattern Recognition", IEEE Trans. on PAMI, vol.19, No.10, October 1997.
73. [HO 98] Ho T.K., Baird H.S., "Pattern Classification With Compact Distribution Maps", Computer Vision And Image Understanding, vol.70, No.1, April 1998, pp.101-110.
74. [HOC 97] Hochberg J., Kelly P., Thomas T., Kerns L., "Automatic Script Identification From Document Images Using Cluster-Based Templates", IEEE Trans. on PAMI, vol.19, No.2, February 1997.
75. [HU 62] Hu M.K., "Visual Pattern Recognition by Moment Invariants", I.E.E.E., Trans. on Information Theory, vol.IT 8, 1962, pp.1798-1807.
76. [HU 96] Hu J., Pavlidis T., "A Hierarchical Approach To Efficient Curvilinear Object Searching" Computer Vision and Image Understanding, vol.63, March 1996, pp.206-220.
77. [HU 98] Hu J., Yan H., "A Model-Based Segmentation Method For Handwritten Numeral Strings", Computer Vision And Image Understanding, vol.70, No.3, June 1998, pp.383-403.
78. [HUA 95] Huang T.S., Suen C.Y., "Combination Of Multiple Experts For The Recognition Of Unconstrained Handwritten Numerals", IEEE Trans. on PAMI, vol.17, No.1, January 1995.
79. [HUL 96] Hull J.J., "Incorporating Language Syntax In Visual Text Recognition With A Statistical Model", IEEE Trans. on PAMI, vol.18, No.12, December 1996.
80. [HWA 97] Hwang Y.S., Bang S.Y., "Recognition Of Unconstrained Handwritten Numerals By A Radial Basis Function Neural Network Classifier", Pattern Recognition Letters, North Holland, vol. 18, No.7, July 1997, pp.657-664.
81. [IPU 98] International Postal Union, "Global and Regional Statistics" www.upu.int.
82. [JAI 89] Jain A.K., "Fundamentals Of Digital Image Processing", Prentice Hall, Englewood Cliffs NY, 1989.
83. [JAI 97] Jain A.K., Zongker D., "Representation And Recognition Of Handwritten Digits Using Deformable Templates", IEEE Trans. on PAMI, vol.19, No.12, December 1997.
84. [JAI 98] Jain A.K., Yu B., "Document Representation And Its Application To Page Decomposition", IEEE Trans. on PAMI, vol.20, No.3, March 1998.
85. [JAW 95] Jawahar P.K., Biswas A.K. "Detection Of Clusters Of Distinct Geometry: A Step Towards Generalised Fuzzy Clustering", Pattern Recognition Letters, North Holland, vol.16, No.11, November 1995, pp.1119-1123.

86. [JIA 97] Jiang H.F., Han C.C., Fan K.C., "A Fast Approach To The Detection And Correction Of Skew Documents", *Pattern Recognition Letters*, vol.18, No.7, July 1997, pp.675-686.
87. [KAH 96] Kashi R.S., Turin W., Nelson W.L., "On-line Signature Verification Using Stroke Direction Coding", *Optical Engineering*, vol.35, September 1996, pp.2526-2533.
88. [KAM 93] Kamel M., Zhao A., "Extraction Of Binary Character/Graphics Images From Grayscale Document Images", *CVGIP Graphical Models And Image Processing*, vol.55, No.3, May 1993, pp.203-217.
89. [KAM 96] Kam A.C., Kopec G.E., "Document Image Decoding By Heuristic Search", *IEEE Trans. on PAMI*, vol.18, No.9, September 1996.
90. [KAN 95] J. Kanai, S.V. Rice, T.A. Nartker, G. Nagy "Automated Evaluation Of OCR Zoning", *IEEE Trans. on PAMI*, vol.17, No.1, January 1995.
91. [KAW 93] Kawatani T. & Miyamoto N., "Verification Of Personal Handwriting Characteristics For Numerals And Its Application To Recognition", *Pattern Recognition Letters*, North Holland, vol.14, No.4, April 1993, pp.335-343.
92. [KHO 90a] Khotanzad A., Lu H.J., "Classification Of Invariant Image Representation Using A Neural Network", *IEEE Trans on ASSP*, vol.38, No.8, June 1990.
93. [KHO 90b] Khotanzad A., Hong H.Y., "Invariant Image Recognition By Zernicke Moments", *IEEE Trans on PAMI*, vol.12, No.5, May 1990.
94. [KIM 94] Kim P.K., Kim H.J., "Off-line Handwritten Korean Character Recognition Based On Stroke Extraction And Representation", *Pattern Recognition Letters*, North Holland, vol.15, No.12, December 1994, pp.1245-1253.
95. [KIM 96] Kim H.J., Jung J.W., Kim S.K., "On-line Chinese Character Recognition", *Pattern Recognition Letters*, North Holland, vol.17, No.12, October 1996, pp.1311-1322.
96. [KIM 97] Kim G., Govindaraju V., "A Lexicon Driven Approach To Handwritten Word Recognition For Real-Time Applications", *IEEE Trans. on PAMI*, vol. 19, No. 4, April 1997.
97. [KIT 97] Kittler J., Hojjatoleslami A., Windeatt T., "Strategies For Combining Classifiers Employing Shared And Distinct Pattern Representations", *Pattern Recognition Letters*, North Holland, vol.18, No.11 13, November 1997, pp.1373-1377.
98. [KIT 98] Kittler J., Hatef M., Duin R.P.W., Matas J., "On Combining Classifiers", *IEEE Trans. on PAMI*, vol.20, No.3, March 1998.
99. [KNE 98] Knerr S., Augustin E., Baret O., Price D. "Hidden Markov Model Based Word Recognition And Its Application To Legal Amount Reading On French Checks", *Computer Vision And Image Understanding*, vol.70, No.3, June 1998, pp.404-419.
100. [KOP 97] Kopec G.E., Lomelin M., "Supervised Template Estimation For Document Image Decoding", *IEEE Trans. on PAMI*, vol.19, No.12, December 1997.
101. [KUD 96] Kudo M., Mizukami K., Nakamura Y., Shimbo M., "Realisation Of Membership Queries In Character Recognition", *Pattern Recognition Letters*, North Holland, vol.17, No.1, January 1996, pp.77-82.
102. [KUO 94] Fan K.C., Liu C.H., Wang Y.K., "Segmentation And Classification Of Mixed Text/Graphic/Image Documents", *Pattern Recognition Letters*, North Holland, vol.15, No.12, December 1994, pp.1201-1209.

- 103.[LAC 94] Lacrămă L. D., "Template Matching in Optical Character Recognition", Proceedings of the Symposium on Electronics and Telecommunications, Timișoara, 1994, vol. III, pp.173-179.
- 104.[LAC 95] Lacrămă L. D., Șnep I., Oaie V. "Segmentarea textelor scrise de mână", Revista Facultății de economie, Universitatea de Vest, Timișoara, vol. XVI, 1995, pp.23-31.
- 105.[LAC 96a] Lacrămă L. D., Alexa F. "A Three Step Strategy In Handwritten Character Recognition", Proceedings of the Symposium on Electronics and Telecommunications, Timișoara, 1996.
- 106.[LAC 96b] Lacrămă L. D. "Mix Moment Vector Used For The Handwritten Character Recognition", Proceedings of the Symposium on Electronics and Telecommunications, Timișoara, 1996.
- 107.[LAC 96c] Lacrămă L. D., Alexa F. "Strategie etapizată în recunoașterea caracterelor alfanumerice scrise de mână", Revista Facultății de economie, Universitatea de Vest, Timișoara, vol. XVII, 1996.
- 108.[LAC 96d] Lacrămă L. D., Alexa F. "Utilizarea momentelor invariante în recunoașterea caracterelor alfanumerice scrise de mână", Revista Facultății de economie, Universitatea de Vest, Timișoara, vol. XVII, 1996.
- 109.[LAC 97a] Lacrămă L. D., Hall G. "Handwritten Character Recognition Using A Neural Network With Attraction/Repulsion Built-in Knowledge", Proceedings of the Fourth Seminar on Neural Network Applications in Electrical Engineering Neurel'97, September 1997, pp.131-135, Belgrad Yugoslavia.
- 110.[LAC 97b] Lacrămă L. D., "Detectia automată a caracteristicilor individuale ale scrisului de mână", Revista Facultății de economie, Universitatea de Vest, Timișoara, vol. XVIII, 1997, pp.324-327.
- 111.[LAC 97c] Lacrămă L. D., "Recunoașterea caracterelor scrise de mână cu o rețea neuronală de tip RBF", Revista Facultății de economie, Universitatea de Vest, Timișoara, vol. XVIII, 1997, pp.328-334.
- 112.[LAC 97d] Lacrămă L. D., "Psihologia citirii și aplicarea ei în recunoașterea caracterelor scrise de mână", Revista Facultății de economie, Universitatea de Vest, Timișoara, vol. XVIII, 1997, pp.320-323.
- 113.[LAC 97e] Lacrămă L. D., "Separation And Recognition Of Handwritten Characters", Seminar of the Electronics Department, University of Central Lancashire, Preston, United Kingdom, May 1997.
- 114.[LAC 98a] Lacrămă L. D., Hall G. "Off-line Handwritten Character Recognition Using A RBF Neural Network", Proceedings Of The First International Symposium On Communications Systems And Signal Processing, Sheffield, United Kingdom, April 1998, vol. III, pp.125-130.
- 115.[LAC 98b] Lacrămă L. D., "Flexible Frontiers For Text Division Into Rows", acceptată pentru publicare în Buletinul Universității "Politehnica" Timișoara, 1998.
- 116.[LAC 98c] Lacrămă L. D., "Separarea flexibilă a cuvintelor în caracterele lor componente", Sesiunea de comunicări științifice a Universității de Vest, Timișoara, Mai 1998.
- 117.[LAC 98d] Lacrămă L. D., "Efectul unghiului mediu de înclinare a scrisului de mână asupra separării și recunoașterii caracterelor", acceptată pentru publicare în Revista Facultății de economie, Universitatea de Vest, Timișoara, vol. XIX, 1998.
- 118.[LAC 98e] Lacrămă L. D., "Analiza structurii caracterelor alfabetului românesc", acceptată pentru publicare în Revista Facultății de economie, Universitatea de Vest, Timișoara, vol. XIX, 1998.

- 119.[LAM 95] Lam L., Ching L., Suen Y, "Optimal Combination Of Pattern Classifiers", Pattern Recognition Letters, North Holland, vol.16, No.9, September 1995, pp.945-954.
- 120.[LAP 98] La Poste Francaise, "Distribution automatise de cheques postaux" www.laposte.fr.
- 121.[LAZ 97]Lazzerini B., Marcelloni F., Reyneri L.M., "Beatrix A Self-Learning System For Off-line Recognition Of Handwritten Texts", Pattern Recognition Letters, North Holland, vol.18, No.6, June 1997, pp.583-594.
- 122.[LEE 91] Lee H.Y., Lee Y.C., Chen H.H., "Handwritten Letter Recognition With Neural Networks" Pattern Recognition Letters, North Holland, vol.12, No.7, July 1997, pp.673-678.
- 123.[LEE 95] Lee S.W., Kim Y.J., "Direct Extraction Of Topographic Features For Gray Scale Character Recognition", IEEE Trans. on PAMI, vol.17, No.7, July 1995.
- 124.[LEE 96a] Lee L.L., Berger T. Aviczer E. "Reliable On-line Human Signature Verification Systems", IEEE Transactions on PAMI, vol.18, June 1996, pp.643-647.
- 125.[LEE 96b] Lee S.W. "Off Line Recognition Of Totally Unconstrained Handwritten Numerals Using Multilayer Cluster Neural Network", IEEE Transactions on PAMI, vol.18, No.5, June 1996, pp.648-652.
- 126.[LEE 96c] Lee S.W., Lee D.J., Park H.-S., "A New Methodology For Gray-Scale Character Segmentation and Recognition", IEEE Trans. on PAMI, vol.18, No.10, October 1996.
- 127.[LEM 93] Lemarie B., "Practical Realisation Of A Radial Basis Function Network For Handwritten Digit Recognition" in " New Trends In Neural Computing", Springer, Berlin, 1993.
- 128.[LEN 92] Lenz R., Osterberg M., "A New Method For Unsupervised Linear Feature Extraction, Using Forth-Order Moments", Pattern Recognition Letters, North Holland, vol.13, No.12, December 1992, pp.827-836.
- 129.[LET 95] Lethelier M., Leroux M., Giloux M. "Traitement des montants numerique de cheque postaux. Une methode de segmentation basee sur la reconnaissance", Traitement du signal, vol.12, No.4-6, Avril 1995.
- 130.[LI 92] Li R.Y., Xu M., "Character Recognition Using A Fast Neural-Net Classifier", Pattern Recognition Letters, North Holland, vol.13, 1992, pp.369-374.
- 131.[LI 96] Li Y., Lopresti D., Nagy G., Tomkins A., "Validation Of Image Defect Models For Optical Character Recognition", IEEE Trans. on PAMI, vol.18, No.2, February 1996.
- 132.[LIA 93] Liao H.Y., Haung J.S. & Huang S.T., "Stroke Based Handwritten Chinese Characters Recognition Using Neural Networks", Pattern Recognition Letters, North Holland, vol.14, No.10, October 1993, pp.833-840.
- 133.[LIK 95] Likforman-Sulem G.L. Faure C., "Une methode de resolution des conflits d'alignements pour la segmentation des documents manuscrits", Traitement du signal, vol.12, No.4-6 Avril 1995.
- 134.[LIM 96], Lim J.H., Teh H.H., Lui H.C., Wang P.Z. "Stochastic Topology With Elastic Matching For Off-line Handwritten Character Recognition", Pattern Recognition Letters, North Holland, vol.17, No.1, January 1996, pp.149-154.
- 135.[LIN 92] Linggard R., "Neural Network For Vision Speech And Natural Language", Cambridge, 1992.

- 136.[LIN 93] Lin F. & Brandt D.R., "Toward Absolute Invariance of Images Under Translation, Rotation, and Dilatation", *Pattern Recognition Letters*, North Holland, vol.14, No.4, May 1993, pp.369-379.
- 137.[LIO 96] Liou C.Y., Yang H.C., "Handprinted Character Recognition Base On Spatial Topology Distance Measurement", *IEEE Transactions on PAMI*, vol.18, No.9, September 1996, pp.941-945.
- 138.[LIS 92] Lisboa P.J.G., "Image Classification Using Gabor Representations With A Neural Net", In "Neural Network For Vision Speech And Natural Language", editor Linggard R., Cambridge, 1992.
- 139.[LIU 97] Liu Y., Srihari N.S., "Document Image Binarization Based On Texture Features", *IEEE Trans. on PAMI*, vol.19, No.5, May 1997.
- 140.[LU 92] Lu S.W., Xu H., "False Stroke Detection And Elimination For Character Recognition", *Pattern Recognition Letters*, North Holland, vol.13, 1992, pp.745-755.
- 141.[LU 98] Lu Z., "Detection Of Text Regions From Digital Engineering Drawings", *IEEE Trans. on PAMI*, vol.20, No.4, April 1998.
- 142.[MAD 97] Maderlechner G., Suda P., Brückner T., "Classification Of Documents By Form And Content", *Pattern Recognition Letters*, North Holland, vol.18, No.11-13, November 1997, pp.1225-1231.
- 143.[MAR 97] Marcelli A., Likhareva N., Pavlidis T., "Structural Indexing for Character Recognition", *Computer Vision and Image Understanding*, vol.66, No.3, June 1997, pp.330-346.
- 144.[MAT 89] Matei C.H., Constantiniu F., Popa D.M., Nicolescu N.C., Rădulescu G., "Istoria lumii în date", ed. Enciclopedică Română, București, 1989.
- 145.[MAT 92] Matias A., "Template Matching on Transputer Network", technical report UCLA, Preston, United Kingdom 1992.
- 146.[MCG 94] McGraw G., Rehling J., Goldstone R. "Roles In Letter Perception: Human Data And Computer Reading", *Computer Speech And Language*, No.5, June 1994, pp.259-274.
- 147.[MES 96] Messelodi S., Modena C.M., "Context Driven Text Segmentation And Recognition", *Pattern Recognition Letters*, North Holland, vol.17, No.1, January 1996, pp.47-56.
- 148.[MOH 96] Mohamed M., Gader P., "Handwritten Word Recognition Using Segmentation-Free Hidden Markov Modeling And Segmentation-Based Dynamic Programming Techniques", *IEEE Transactions on PAMI*, vol.18, No.5, May 1996, pp.548-554.
- 149.[NEG 94] Negrescu L., "Limbajul C++" ed. Microinformatica, Cluj, 1994.
- 150.[NIS 95] Nishida H., "An Approach To Integration Of Off-line And On-line Recognition Of Handwriting", *Pattern Recognition Letters*, North Holland, vol.16, No.11, November 1995, pp.1213-1219.
- 151.[NIS 96a] Nishida H., "Shape Recognition By Integrating Structural Descriptions And Geometrical/Statistical Transforms" *Computer vision and Image Understanding* vol.64, September 1996, pp.248 262.

- 152.[NIS 96b] Nishida H., "Automatic Construction Of Structural Models Incorporating Discontinuous Transformations", IEEE Transactions on PAMI, vol.18, No.4, April 1996, pp.400-411.
- 153.[NIS 97] Nishida H., "Analysis And Synthesis Of Deformed Patterns Based On Structural Models", Computer Vision And Image Understanding, vol.68, No.1, October 1997, pp.59-71.
- 154.[OH 95] Oh I.S., "Document Image Binarization Preserving Stroke Connectivity", Pattern Recognition Letters, North Holland, vol.16, No.7, July 1995, pp.743-748.
- 155.[PAL 96] Pal U., Chaudhuri B.B., "An Improved Document Skew Angle Estimation Technique", Pattern Recognition Letters, North Holland, vol.17, No.8, July 1996, pp.899-904.
- 156.[PAR 95] Parizeau M., Plamondon R., "A Fuzzy-Syntactic Approach To Allograph Modelling For Cursive Script Recognition" IEEE Trans. on PAMI, vol.17, No.7, July 1995.
- 157.[PAR 96] Parisse C., "Global Word Shape Processing In Off-line Recognition Of Handwriting", IEEE Transactions on PAMI, vol.18, April 1996, pp.460-464.
- 158.[PAV 93] Pavlidis T., "Recognition Of Printed Text Under Realistic Conditions", Pattern Recognition Letters, vol.14, No.4, April 1993, pp. 317-328.
- 159.[PEI 95] Pei S.C., Horg J.H., "Circular Arc Detection Based On Hough Transform", Pattern Recognition Letters, North Holland, vol.16, No.6, June 1995, pp.615-625.
- 160.[PIT 93] Pitas I., "Digital Image Processing Algorithms", Prentice Hall, Englewood Cliffs NY, 1993.
- 161.[PIT 94] Pitas I., "Parallel Algorithms For Digital Image Processing", Prentice Hall, Englewood Cliffs NY, 1994.
- 162.[POP 89] Pop I., Naforniță I., Tiponuț V., Mihăescu A., Toma L., "Metode în prelucrarea numerică a semnalelor", ed. Facla, Timișoara, 1989.
- 163.[PRA 91] Pratt W.K., "Digital image processing", Wiley, 1991.
- 164.[RAH 97] Rahman A.F.R., Fairhurst M.C., "A New Hybrid Approach In Combining Multiple Experts To Recognise Handwritten Numerals", Pattern Recognition Letters, North Holland, vol.18, No.8, August 1997, pp. 781-790.
- 165.[RAY 89] Rayner K., Pollatsek A., "Psychology Of Reading", Prentice-Hall, 1989.
- 166.[RED 98] Reddy N.V.S., Nagabhushan P., "A Connectionist Expert System Model For Conflict Resolution In Unconstrained Handwritten Numeral Recognition", Pattern Recognition Letters, North Holland, vol.19, No.2, February 1998, pp.161-169.
- 167.[REV 96] Revow M., Williams C.K., Hinton G.E., "Using Generative Models For Handwritten Digit Recognition", IEEE Transactions on PAMI, vol.18, June 1996, pp.592-606.
- 168.[ROB 96] Robert O., Duin P.W., "A Note On Comparing Classifiers", Pattern Recognition Letters, North Holland, vol.17, No.5, May 1996, pp.523-527.
- 169.[ROC 95] Rocha J., Pavlidis T., "Character Recognition Without Segmentation", IEEE Trans. on PAMI, vol.17, No.9, September 1995.

- 170.[ROS 82] Rosenfeld A., Kak A.C., "Digital Picture Processing ", New York Academic Press, 1982.
- 171.[SAB 95] Sabourin R., Genest G., "Definition et evaluation d'une famille de representations pour la verification hors-ligne des signatures", Traitement du signal, vol.12, No 4-6, Avril 1995.
- 172.[SAB 97] Sabourin R., Genest G., Préteux F.J., "Off-line Signature Verification By Local Granulometric Size Distributions", IEEE Trans. on PAMI, vol.19, No.9, September 1997.
- 173.[SAH 93] Saha P.K., Khanda B., Majungder D.D., "A Single Scan Boundary Removal Thining Algorithm for 2-D Binary Object", Pattern Recognition Letters, North Holland, vol.14, 1993, pp.173-179.
- 174.[SAL 94] Saleh A., "A Method Of Coding Handwritten Arabic Characters And Its Application To Context-Free Grammar", Pattern Recognition Letters, North Holland, vol.15, No.12, December 1994, pp.1265-1271.
- 175.[SAN 97] Sanchez A., Benedi J.M., "Consistency of Stochastic Context-Free Grammars from Probabilistic Estimation Based on Growth", IEEE Trans. on PAMI, vol.19, No.9, September 1997.
- 176.[SAR 98a] Sarkar P., Nagy G., Zhou J., Lopresti D., "Spatial Sampling of Printed Patterns", IEEE Trans. on PAMI, vol.20, No.3, March 1998.
- 177.[SAR 98b] Sarukkai R.R., Ballard D.H., "Phonetic Set Indexing for Fast Lexical Access", IEEE Trans. on PAMI, vol.20, No.1, January 1998.
- 178.[SCH 92] Schalkoff R.J., "Pattern Recognition: Statistical, Structural and Neural Approaches", Wiley, 1992.
- 179.[SEN 94] Senior W.A., "Off-line Handwriting Recognition Using Neural Networks" Ph. D. Thesis, University of Cambridge, 1994.
- 180.[SEN 96] Seni G., Shrihari R.K., Nasrabadi N., "Large Vocabulary Recognition Of On-line Handwritten Cursive Words", IEEE Transactions on PAMI, vol.18, No.7, July 1996, pp.757-762.
- 181.[SEN 98] Senior A.W., Robinson A.J., "An Off-line Cursive Handwriting Recognition System ", IEEE Trans. on PAMI, Vol. 20, No. 3, March 1998.
- 182.[SHA 93] Shapiro V., Gluchev G., Sgurev V., "Handwritten Document Image Segmentation and Analysis", Pattern Recognition Letters, North Holland, vol.14, No.1, January 1993, pp.71-78.
- 183.[SHI 96] Shi Z., Govindaraju V., "Character Image Enhancement By Selective Region Growing", Pattern Recognition Letters, North Holland, vol.17, No.5, May 1996, pp.523-527.
- 184.[SHU 96] Shustorovich A., Thraser C.W., "Neural Network Positioning And Classification Of Handwritten Characters", IEEE Transactions on PAMI, vol.18, No.6, June 1996, pp.685-693.
- 185.[SIN 97] Sin B.K., Kim J.H., "Ligature Modeling For On-line Cursive Script Recognition", IEEE Trans. on PAMI, Vol.19, No.5, May 1997.
- 186.[SON 93] Sonka M., Hlavac V. & Boyle R., "Image Processing, Analysis And Machine Vision", Chapman & Hall, Glasgow, 1993.

- 187.[SPI 97] Spitz A.L., "Determination Of The Script And Language Content Of Document Images", IEEE Trans. on PAMI, vol.19, No.3, March 1997.
- 188.[SRI 93] Shrihari S.N., "Recognition Of Handwritten And Machine-Printed Text For Postal Address Interpretation", Pattern Recognition Letters, North Holland, vol.14, No.4, April 1993, pp.291-302.
- 189.[SRI 94] Setlur S., Govindaraju V., "Generating Manifold Samples From Handwritten Word", Pattern Recognition Letters, North Holland, vol.15, No.9, September 1994, pp.901-905.
- 190.[SRI 96] Srihari S.N., Shin Y.C., Ramanaprasad V., "A System To Read Names And Addresses On Tax Forms", Proceedings of the IEEE vol.84, July 1996, pp.1034-1049.
- 191.[SUE 93] Suen C.Y., Legault R., "Building A New Generation of Handwriting Recognition System", Pattern Recognition Letters, North Holland, vol.14, No.4, April 1993, pp.303-315.
- 192.[TAN 97] Tang Y.Y., Ma H., Liu J., Li B.F., Xi D., "Multiresolution Analysis In Extraction Of Reference Lines From Documents With Gray Level Background", IEEE Trans. on PAMI, vol.19, No.8, August 1997.
- 193.[TAN 98] Tan T.N., "Rotation Invariant Texture Features And Their Use In Automatic Script Identification", IEEE Trans. on PAMI, vol.20, No.7, July 1998.
- 194.[TAY 83] Taylor I., Taylor M.M., "Psychology Of Reading", Academic Press NY 1983.
- 195.[TOD 94] Todorean G., Coșteiu M., Giurgiu M., "Rețele neuronale", ed. Microinformatica Cluj, 1994.
- 196.[TOF 91] Toffler A. "Șocul viitorului", Ed. Politică, București, 1982.
- 197.[TOM 91] Toma I. C., Oteșteanu M., Gui V., Vasiu R., "Televiziune, înregistrarea magnetică și prelucrarea numerică a imaginilor" Lit. IPTVT, Timișoara, 1991.
- 198.[TOR 96] Toraichi K., Kurmamoto T., Yamamoto K., Yamada H., "Feature Analysis Of Hand-printed Chinese Characters", Pattern Recognition Letters, North Holland, vol.17, No.7, June 1996, pp.795-800.
- 199.[TRI 95] Trier O., Taxt T., "Improvement Of Integrated Function Algorithm For Binarization Of Document Images", Pattern Recognition Letters, North Holland, vol.16, No.3, March 1995, pp.277-283.
- 200.[TRI 97] Trier O., Taxt T., Jain A.K., "Recognition Of Digits In Hydrographic Maps: Binary Versus Topographic Analysis", IEEE Trans. on PAMI, vol.19, No.4, April 1997.
- 201.[USP 98] United States Postal Service, "Future Perspectives" www.usps.gov.
- 202.[VER 92] Verikas A.A., Bachauskene M.I., "Adaptive Character Recognition System", Pattern Recognition Letters, North Holland, vol.13, 1992, pp.207-212.
- 203.[VID 95] Vidal E., Marzal A., Aibar P., "Fast Computation Of Normalized Edit Distances", IEEE Trans. on PAMI, vol.17, No.9, September 1995.
- 204.[VIN 92] Vincent J.M., Myers D.J., "Image Feature Location In Multi-Resolution Images Using A Hierarchy Of Multilayer Perceptron", in "Neural Network For Vision Speech And Natural Language", editor Linggard R., Cambridge, 1992.
- 205.[WAA 94] Waard W.P., "Neural Technique And Postal Code Detection", Pattern Recognition Letters, North Holland, vol.15, No.2, February 1994, pp.199-206.

- 206.[WAD 93] Wada M., "Proposal For Fully Automated Mail Processing System For The 21st Century", Pattern Recognition Letters, North Holland, vol.14, No.4, April 1993, pp.281-290.
- 207.[WAD 98] Wada K, Mori K., Toraichi K., "PaRM: A Parallel Relaxation Machine For Handwritten Character Recognition", Pattern Recognition Letters, North Holland, vol.19, No.5-6, April 1998, pp.475-481.
- 208.[WAK 93] Wakahara T., "Toward Robust Handwritten Character Recognition", Pattern Recognition Letters, North Holland, vol.14, No.4, April 1993, pp.345-354.
- 209.[WAK 97] Wakahara T., Odaka K., "On-line Cursive Kanji Character Recognition Using Stroke-Based Affine Transformation", IEEE Trans. on PAMI, vol.19, No.12, December 1997.
- 210.[WAN 93] Wang L., Pavlidis T., "Detection Of Curved And Straight Segments From Gray Scale Topography", CVGIP: Graphical Models And Image Processing, vol.58, No.11, November 1993, pp.352-365.
- 211.[WEN 96] Wen C.J., Yau H.F., Jong B.S., "Optimal Segmentation Of Handwritten Chinese Signature Using Wavelet Transforms", Optical Engineering, vol.35, September 1996, pp.2721-2729.
- 212.[WEN 98] Wenyin L., Dori D., "Incremental Arc Segmentation Algorithm And Its Evaluation", IEEE Trans. on PAMI, vol.20, No.4, April 1998.
- 213.[WIL 96] Wilfong G., Sinden F., Ruedisueli L., "On-line Recognition Of Handwritten Symbols" IEEE Transactions on PAMI, vol.18, No.9, September 1996, pp.935-940.
- 214.[WIL 97] Williams C.K.I., Revow M., Hinton G.E., "Instantiating Deformable Models With A Neural Net" Computer Vision and Image Understanding, vol.68, No.1, October 1997, pp.120-126.
- 215.[WON 95] Wong W.H., Siu W.C., Lam K.M., "Generation Of Moment Invariants And Their Uses For Character Recognition", Pattern Recognition Letters, North Holland, vol. 16, No. 2, January 1995, pp.115-123.
- 216.[WU 97] Wu Q.W., Lee S.Y., Jou I.C., "On-line Signature Verification Based On Split-And-Merge Matching Mechanism", Pattern Recognition Letters, North Holland, vol.18, No.7, July 1997, pp.665-673.
- 217.[YAN 94a] Yan H., "Handwritten Digit Recognition Using An Optimal Nearest Neighbour Classifier", Pattern Recognition Letters, North Holland, vol.15, No.2, February 1994, pp.207-211.
- 218.[YAN 94b] Yan H., Wu J., "Character and Line Extraction From Colour Map Images Using A Multi-Layer Neural Network", Pattern Recognition Letters, North Holland, vol.15, No.1, January 1994, pp.97-103.
- 219.[YON 93] Yong J.D., Prasad R., "On-line Recognition of Handwritten Characters Using Differential angles and Structural Descriptors", Pattern Recognition Letters, North Holland, vol.14, No.12, December 1994, pp.141-150.

- 220.[YON 94] Yong J.D., Chen Y.S. & Hsu W.H., "Adaptive Thresholding Algorithm and Its Hardware Implementation", Pattern Recognition Letters, North Holland, vol.15, No.2, February 1994, pp.141-150.
- 221.[YU 96] Yu B., Jain A.K., "A Generic System for Form Dropout", IEEE Trans. on PAMI, Vol.18, No.11, November 1996.
- 222.[YUE 97] Yuen I., Hanzo L., "Block-Run Length Coding Of Handwriting And Bilevel Graphics Based On Quadtree Segmentation", Pattern Recognition Letters, North Holland, vol. 18, No. 1, February, 1997, pp.187-191.
- 223.[ZHO 94] Zhou J., Pavlidis T., "Discrimination Of Characters By A Multi-stage Recognition Process", Pattern Recognition, vol. 27, No. 11, November, 1994, pp.1539-1549.
- 224.[ZLA 94] Zlatopolsky I., "Automated Document Segmentation", Pattern Recognition Letters, North Holland, vol. 15, No. 7, July, 1994, pp.699-704.