

Bogdan IONESCU

***ANALIZA ȘI PRELUCRAREA
SECVENȚELOR VIDEO
Indexarea automată după conținut***



București, 2009

Copyright © 2009, Bogdan IONESCU
Toate drepturile sunt rezervate autorului.

*Adresă: Editura TEHNICĂ
Str. Olari, nr. 23, sector 2
București, România
cod 024056*

www.tehnica.ro

Referenți științifici:

Prof. univ. dr. ing. Constantin VERTAN
Conf. univ. dr. ing. Mihai CIUC

Lucrare editată cu sprijinul
Consiliului Național al Cercetării Științifice
din Învățământul Superior

Descrierea CIP a Bibliotecii Naționale a României
IONESCU, BOGDAN

**Analiza și prelucrarea secvențelor video: indexarea
automată după conținut / Bogdan Ionescu. – București:**
Editura Tehnică, 2009

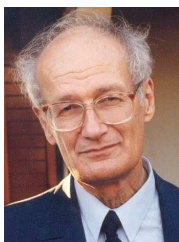
Bibliogr.

ISBN 978-973-31-2354-5

621.397.42

681.772.7

Prefață



Fostul meu student și actualul coleg dr. Bogdan Ionescu m-a rugat să-i prefătez această primă carte, ceea ce fac cu plăcere - el fiind dintr-o pleiadă de tineri care aduc cinste școlii românești. Bogdan Ionescu și-a terminat de curând (un an) un doctorat în cotutelă (România - Franța) în domeniul relativ nou și foarte important al prelucrării semnalelor multidimensionale, și a avut șansa unui subiect a cărui stringență crește pe zi ce trece: indexarea după conținut a bazelor de date video (mai pe românește: a bibliotecilor de filme - în cazul lui, de filme artistice de animație).

Ce vrea să zică asta - indexarea după conținut - cititorul va găsi în primul capitol, dar sunt tentat să zic și aici, în aceste rânduri, câteva cuvinte: problema nu e chiar nouă. Cu ceva zeci de ani în urmă am aflat că pe alte meleaguri oamenii se ocupau, pentru cuvinte, cu alcătuirea unor asemenea dicționare. Cele alfabetice, pe care le avem și noi, îți explică ce vrea să zică un cuvânt pe care îl ai dar al cărui sens nu îl știi; dar sunt și probleme de alt fel: acolo era un exemplu de întâmplare în academia spaniolă - un vorbitor nu-și aducea aminte cum se cheamă un om născut pe vapor (noi n-avem cuvânt pentru acest concept). Ne trebuie dicționare care să ne ducă de la concept la cuvânt. Despre unele popoare primitive se zice că aveau zeci de cuvinte pentru a denumi diferite tipuri de nori; noi n-avem, dar am putea eventual descrie forma lor, mișcarea lor, ca să precizăm la care ne referim când vrem să povestim o întâmplare concretă.

Într-o bibliotecă de un miliard de cărți, cu câte 500 de pagini fiecare și cu 2000 de semne pe pagină avem nevoie doar de 50 de cifre binare pentru a identifica orice literă, ceea ce mi se pare extrem de puțin - la îndemâna

umanului: le cuprindem cu ochiul dintr-o privire, pe un rând. Oare nu e posibil să avem căi/o cale de a ajunge la "obiectul" dorit dintr-o colecție vastă, cunoscându-l prin calitățile sale (făcute cumva măsurabile: da-nu, roșu-albastru-galben-verde, o valoare întreagă între 1 și 100, 17 grade de turtire a unui cerc în elipsă, etc.)? "Obiectele" de care vorbeam pot fi entități foarte complexe: o imagine, o secvență de film mut, entități "multimodale" (vorba, sunete, imagini, text, etc.). Parcă suntem tentați a zice da. Dar acum vine partea dificilă a problemei, și în același timp frumoasă prin efortul de creație pe care ni-l cere (aspectul care ne provoacă, ne desfide, englezul ar zice "challenging"): pe de o parte, în cazul concret al unei colecții de un tip dat (de pietre, de găze, de filme), care sunt atributele, cum le definim ca să caracterizăm cât mai compact și mai corect, acea colecție; pe de altă parte, în fața unui obiect din colecție, cum măsurăm *automat*, adică nu prin intervenția omului (în cazul ăsta avem nevoie de un specialist în domeniu!), aceste atribute.

Fără acest mic amănunt aici, "automat", suntem pierduți fiindcă operația manuală de adnotare cu atribute a obiectelor este consumatoare de timp în așa măsură că ne face întreprinderea lipsită de sens.

În momentul de față al scurtei noastre istorii de câteva sute de ani, suntem în pericol de a fi "înecați în informații" care pe de o parte multe ne sunt vitale și pe de alta, în ansamblul lor ne copleșesc, fără a putea ajunge la cele de care avem nevoie suntem ca însetatul din pustiu peste care năvălește marea. Indexarea automată după conținut ne poate salva.

La laboratorul nostru din Politehnica bucureșteană, aceste preocupări sunt de dată mai veche (aș menționa aici preocupările prof. Constantin Vertan în timpul unor stagii în Franța și apoi aici), dar tomul lui Bogdan Ionescu este prima carte dedicată acestui subiect, și în particular indexării video, și cred că trebuie s-o salutăm cu entuziasm fiind sosită într-un moment când e nevoie de ea. Sper că o vor urma altele și că subiectul va atrage și pe alți tineri cercetători spre binele nostru al tuturor. Felicitări autorului pentru munca asiduă depusă și calitatea lucrării rezultate.

Prof. Dr. Ing. Vasile BUZULOIU
București 17 noiembrie 2008

Cuvântul autorului

Indexarea automată după conținut a datelor este un domeniu ce câștigă din ce în ce mai mult teren, datorită necesității crescânde de exploatare a volumelor mari de date. Dacă, nu până demult, puteam vorbi de o lipsă informațională, progresul tehnologic a făcut ca în zilele noastre să ne confruntăm cu o adevărată explozie de informație.

Din acest amalgam informațional, un interes aparte îl au *informațiile multimedia*, ce sunt definite ca fiind o combinație de tipuri de conținut, printre care cele mai uzuale sunt: textul, sunetul și imaginile.

În societatea modernă, informația multimedia face parte din viața noastră cotidiană și imi este greu să-mi imaginez că vom mai putea vreodată să ne lipsim de ea. De exemplu, telefonul portabil a devenit indispensabil și ne însoțește pretutindeni, acesta fiind un adevărat centru multimedia în miniatură. Prin intermediul acestuia, putem accesa informațiile multimedia din rețeaua Internet, putem folosi mesageria electronică, putem înregistra, stoca, reda și distribui filme sau imagini în orice moment. Fiecare persoană, a devenit astfel, cu voie sau fără voie, un "consumator" de date multimedia.

Motivată în principal de un interes comercial, dezvoltarea infrastructurii de stocare și transmisie a datelor a dus la apariția unei noi probleme, și anume: *Cum accesăm informația multimedia utilă dintr-un vast amalgam de date? Cum facem să găsim aceea informație pe care o dorim?* Problema ar fi una simplă în cazul a câtorva date, dar când o astfel de colecție poate conține la ordinul a sute de mii de documente video, de exemplu, care la rândul lor conțin sute de mii de imagini, problema pare imposibil de rezolvat.

Soluția actuală existentă este dată de *sistemele de indexare după conținut* a bazelor de date. Conceptul de indexare este definit ca fiind procesul de adnotare a informației existente într-o colecție de date, prin adăugarea de

informații suplimentare despre conținutul acesteia, informații numite și *indici* de conținut. Pe baza indicilor, sistemul poate grupa datele în funcție de similaritate, în categorii, subcategorii și așa mai departe. De exemplu, dacă dispunem de o bază de documente video, ideal, în urma indexării automate, acestea pot fi regrupate în funcție de gen în: filme, muzică, desene animate, știri, etc., sau la un nivel de detaliu mai ridicat, în subcategorii precum: film de ficțiune, dramă, documentar, etc. În acest fel, căutarea informației dorite este restrânsă la căutarea într-o subcategorie din care aceasta face parte, reducând astfel timpul de căutare și totodată îmbunătățind precizia căutării.

Pe de altă parte, procesul de indexare nu este opțional, ci este strict necesar într-o colecție mare de date. În acest caz, o informație care nu a fost indexată este practic inexistentă pentru utilizator, cu toate că aceasta este prezentă în bază. Să luăm exemplul simplu al unui sistem de indexare a fișierelor, prezent în orice sistem de operare. Acesta ordonează datele în funcție de nume, tipul conținutului, data creării etc., în directoare și subdirectoare. Dacă un anumit fișier nu a fost indexat, cu toate că acesta se află fizic pe dispozitivul de stocare, acesta este transparent pentru utilizator, fiind imposibil de localizat.

Sistemele de indexare, pe parcursul evoluției, au trecut de la o abordare sintactică a procesului de adnotare la o abordare semantică, cum este cazul sistemelor actuale. Diferența dintre acestea este una semnificativă. *Adnotarea sintactică* se limitează la caracterizarea conținutului datelor cu atribute numerice de nivel scăzut, precum măsuri statistice, diverși parametri, etc. Din păcate, o astfel de abordare este implicit adresată unui public avizat în domeniu, căutarea informației necesitând cunoștințe tehnice. Pe de altă parte, *adnotarea semantică* are ca scop descrierea conținutului datelor într-un mod cât mai apropiat de modul de percepție uman. Astfel, localizarea datelor devine naturală și accesibilă publicului larg, fiind ghidată de un limbaj textual. De exemplu, căutarea filmelor în funcție de valorile vitezei medii de deplasare a obiectelor în scenă nu este evidentă, pe când o căutare în funcție de conținutul de acțiune (redus, ridicat) este pe înțelesul tuturor.

Această lucrare vine să adreseze tocmai această problemă a indexării automate după conținut a datelor multimedia, punând accentul pe secvențele de imagini, domeniu de mare actualitate în străinătate în acest moment, dar încă la începuturi în România.

Această lucrare propune un studiu bibliografic detaliat al literaturii de specialitate din acest domeniu, abordând direcțiile fundamentale de analiză și prelucrare a secvențelor de imagini în contextul sistemelor de indexare după conținut. Astfel, sunt prezentate atât aspecte teoretice (principii și metode), cât și exemple concrete (sisteme, aplicații), punând la dispoziția cititorului

o bibliografie mai mult decât generoasă (peste 280 de citări ale unor articole din reviste și conferințe internaționale de specialitate). Cartea este adresată atât începătorilor în domeniul prelucrării și analizei de imagini și video, cât și celor deja experimentați, constituind un ghid de bună practică și totodată un sistem de indexare a realizărilor semnificative din domeniu.

Ideea scrierii acestui manuscris, a apărut în urmă cu mai bine de cinci ani, odată cu demararea tezei mele de doctorat realizată în cotutelă, pe de-o parte la laboratorul LAPI - Laboratorul de Analiza și Prelucrarea Imaginilor din Universitatea "Politehnica" din București, sub îndrumarea Domnului Profesor Vasile Buzuloiu, cât și la laboratorul, la vremea respectivă, LAMII - Laboratoire d'Automatique et de Micro-Informatique Industrielle din Université de Savoie, sub îndrumarea Domnului Profesor Patrick Lambert. Tematica abordată a constat în studiul și dezvoltarea unui sistem de indexare automată după conținut a secvențelor de animație din cadrul Festivalului Internațional al Filmului de Animație de la Annecy, echivalentul în domeniul animației al festivalului de film de la Cannes. Studiul bibliografic și cercetarea detaliată realizată cu această ocazie, precum și faptul că doar o parte din acestea au putut fi valorificate în teza de doctorat (din motive obiective de spațiu), m-au condus spre ideea unei posibile redactări ulterioare a unei cărți dedicate.

Această idee avea să se concretizeze după susținerea tezei de doctorat, când am participat la competiția de granturi de Resurse Umane, organizată de CNCSIS - Consiliului Național al Cercetării Științifice din Învățământul Superior, programul RP de stimulare a revenirii în țară a tinerilor cercetători români. Proiectul propus venea să continue natural cercetarea realizată în străinătate până în acel moment, și anume propunea extinderea studiului indexării spre baze de date generice de documente video, precum și dezvoltarea unei aplicații software de adnotare și navigare virtuală în baza de date. Obținerea grantului RP-2 mi-a permis actualizarea studiului bibliografic realizat anterior, îmbunătățirea acestuia, precum și dezvoltarea de noi direcții de studiu.

Astfel, rezultatele cercetării până în acest moment s-au concretizat în șapte capitole. În primul capitol am detaliat problematica indexării după conținut a datelor multimedia, punând accentul pe metodele de analiză și adnotare de conținut a secvențelor de imagini, ce fac subiectul acestei lucrări. De asemenea, am realizat o trecere în revistă a tehnicilor de indexare a imaginilor, sunetului, secvențelor de imagini și, respectiv, video.

Capitolul al doilea abordează o problemă de prelucrare a secvențelor de imagini ce este premergătoare adnotării propriu-zise a conținutului, dar totodată necesară, și anume segmentarea temporală a secvenței, atât în unități sintactice (plane video), cât și semantice (scene video). Segmentarea

temporală, prin detecția schimbărilor de plan, permite înțelegerea structurii temporale a secvenței, necesară în etapele ulterioare de prelucrare, indiferent că este vorba de o indexare sintactică sau de nivel semantic superior.

Capitolul al treilea propune o analiză a metodelor de caracterizare a informației fundamentale a secvențelor de imagini și anume mișcarea. Pornind de la studierea problematicii estimării mișcării la nivel de imagine, am realizat o trecere în revistă a diverselor direcții de studiu abordate de metodele de analiză și caracterizare a conținutului de mișcare din secvență.

Capitolul al patrulea abordează o altă informație reprezentativă a secvențelor de imagini, ce joacă un rol important în percepția vizuală, și anume, conținutul de culoare. Pornind de la modalitățile clasice de reprezentare a culorilor folosind spațiile de culoare, și ajungând până la o descriere perceptuală cu ajutorul teoriei culorilor, am realizat o trecere în revistă a modalităților de caracterizare a conținutului de culoare, atât static, la nivel de imagine, cât și dinamic, la nivel de secvență de imagini.

Capitolul al cincilea propune un studiu al metodelor de rezumare automată a conținutului secvențelor de imagini, atât statică (în imagini) cât și dinamică (în mișcare). Rezumarea de conținut joacă un rol important pentru indexare, deoarece permite pe de-o parte reducerea drastică a timpului vizualizării datelor dintr-o bază mare de date, cât și reducerea redundanței informaționale pentru alte etape de prelucrare.

Capitolul al șaselea face trecerea dintre nivelul sintactic de adnotare și cel semantic, prin abordarea tehnicilor de formalizare cu concepte fuzzy a datelor numerice de nivel scăzut.

În final, capitolul al șaptelea prezintă un studiu al tehnicilor de clasificare nesupervizată a datelor (automată), cât și al tehnicilor de clasificare supervizată (ce folosesc o etapă de învățare). Tehnicile de clasificare prezintă un real interes pentru procesul de indexare, deoarece pe baza atributelor de conținut, determinate în etapa de adnotare, acestea pot grupa datele în colecții de date omogene.

Pentru mai multe detalii referitor la aspecte aplicative ale prelucrării și analizei secvențelor de imagini în contextul indexării, cititorul poate consulta site-ul proiectului de indexare RP-2, și anume: <http://alpha.imag.pub.ro/VideoIndexingRP2/>.

Sper sincer ca această lucrare să constituie un ajutor și o referință pentru cei interesați de problemele prelucrării secvențelor de imagini, și că alții o să-mi urmeze exemplul și o să ducă mai departe cercetarea românească din acest domeniu.

Ș.l. Dr. Ing. Bogdan IONESCU
București 30 noiembrie 2008

Mulțumiri

Această lucrare nu s-ar fi concretizat fără suportul grantului de cercetare CNCSIS - Consiliului Național al Cercetării Științifice din Învățământul Superior, Resurse Umane, RP-2 (2007-2009), intitulat "Dezvoltarea de Metode de Indexare Semantică după Conținut a Bazelor de Documente Video: Aplicații la Navigare, Căutare și Rezumare Automată a Conținutului"¹. În acest sens, țin să mulțumesc *Domnului Președinte CNCSIS Profesor Ioan Dumitrache, Domnului Vicepreședinte Profesor Mihai Gîrțu și Domnului Director Profesor Adrian Curaj*, inițiatorii programului RP de reintegrare. De asemenea, țin să mulțumesc *Doamnei Director Adjunct Magdalena Crîngașu și Doamnei Consilier Adriana Rotar*, pentru ajutorul acordat cât și pentru informațiile prețioase oferite pe durata desfășurării proiectului.

Țin să mulțumesc laboratorului LAPI - Laboratorul de Analiza și Prelucrarea Imaginilor, din Universitatea "Politehnica" din București, și astfel *Domnului Profesor Vasile Buzuloiu*, pentru acceptarea mea în colectivul de cercetare, pentru prietenia arătată de-a lungul timpului cât și pentru încadrarea prețioasă acordată pe parcursul formării mele profesionale. Mulțumesc colegilor mei profesori, *Constantin Vertan și Mihai Ciuc*, pentru ajutorul important, pentru atmosfera cordială din cadrul laboratorului precum și pentru modelul de conduită arătat.

Mulțumesc în mod special *Domnului Profesor Adrian Badea și Domnului Profesor Corneliu Burileanu*, pentru prietenia acordată, pentru ajutorul prețios, pentru sugestiile valoroase și pentru suportul constant de-a lungul formării mele științifice.

Țin să mulțumesc *Domnului Profesor Nicolae Vasiliu* pentru ajutorul

¹vezi site-ul proiectului "<http://alpha.imag.pub.ro/VideoIndexingRP2/>".

acordat publicării acestei lucrări și pentru sfaturile prețioase. De asemenea, mulțumesc *Domnului Profesor Ilie Prisecaru* pentru suportul acestuia și pentru ajutorul acordat.

Mulțumesc de asemenea *Domnului Profesor Teodor Petrescu* și *Domnului Profesor Dan Stoichescu* pentru sprijinirea activității mele de cercetare în cadrul Facultății de Electronică, Telecomunicații și Tehnologia Informației și a Catedrei de Electronică Aplicată și Ingineria Informației, precum și pentru ajutorul acordat.

Vreau să mulțumesc laboratorului LISTIC - Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance, Annecy, Franța, și astfel *Domnului Profesor Philippe Bolon* pentru co-finanțarea tezei mele de doctorat realizată în domeniul analizei și prelucrării secvențelor de imagini. De asemenea, mulțumesc călduros *Domnului Profesor Patrick Lambert* și *Domnului Profesor Didier Coquin* pentru încadrarea mea pe parcursul tezei de doctorat, pentru sfaturile acordate cât și pentru suportul constant al acestora pe toată durata cercetării efectuate la Annecy.

De asemenea, țin să mulțumesc tuturor colaboratorilor externi ce au susținut proiectul de indexare video pe care l-am inițiat:

- *Domnul Profesor Daniel Bouillot*, IMUS - Institut de Management de l'Université de Savoie și CITIA - Cité de l'Image en Mouvement,
- *Domnul Profesor Patrick Lambert* și *Domnul Profesor Philippe Bolon* - LISTIC, Polytech'Savoie, Annecy-Franța,
- *Domnul Profesor Robert Laganière* - VIVA - The Video, Image, Vision and Autonomous Systems Research Laboratory, Ottawa-Canada,
- *Domnul Emmanuel Quillet* și *Domnul Director Hervé Lièvre* CERIMES - Centre de Ressources et d'Information sur les Multimédias pour L'Enseignement Supérieur din Ministère Enseignement Supérieur et Recherche Français.

Mulțumesc în mod special prietenei mele Monica care m-a sprijinit în tot ce am întreprins până în prezent și care a avut răbdarea să corecteze acest manuscris.

Cu această ocazie, țin să mulțumesc călduros *Doamnei Eugenia Burcea* pentru tot sprijinul prețios acordat și *Doamnei Director Diana Cocârță* ce mi-a sugerat redactarea acestei cărți.

Nu în ultimul rând, vreau să mulțumesc călduros Editurii Tehnice și astfel *Domnului Director Roman Chirilă* pentru acceptarea publicării acestei cărți, pentru finanțarea a o parte din costurile de publicare, precum și pentru munca deloc neglijabilă depusă pentru aducerea la viață a manuscrisului acestei cărți.

Remerciements

Ces travaux ne se seraient véritablement concrétisés sans le soutien assuré par le Grant de recherche du CNCSIS, le Conseil National de la Recherche Scientifique de l'Enseignement Supérieure de la Roumanie, Ressources Humaines, RP-2 (2007-2009), intitulé "Le Développement de Méthodes d'Indexation Sémantique du Contenu des Documents Vidéo: Application à la Navigation, Recherche et Résumé Automatique du Contenu"². A ce titre là, je tiens à remercier *M. Ioan Dumitrache*, *M. Mihai Gîrțu* et *M. Adrian Curaj*, les promoteurs du programme RP. Je tiens également à remercier *Mme Adriana Rotar* et *Mme Magdalena Crîngașu* pour leur aide et pour les informations précieuses fournies pendant le déroulement du projet.

Je tiens à remercier le laboratoire LAPI de l'Université "Politehnica" de Bucarest, Laboratoire d'Analyse et Traitement d'Images, et particulièrement *M. Vasile Buzuloiu*, pour m'avoir accueilli au sein de son équipe de recherche, pour son amitié et son encadrement précieux pendant ma formation professionnelle. Je remercie également mes collègues professeurs, *M. Constantin Vertan* et *M. Mihai Ciuc*, pour leur aide précieuse, pour la bonne ambiance qu'ils ont su créer au sein du laboratoire et pour l'exemple de conduite qu'ils ont été.

Je remercie particulièrement à *M. Adrian Badea* et *M. Corneliu Burileanu*, pour leur amitié, leur aide précieuse, leurs nombreux conseils et leur soutien constant tout au long de ma formation scientifique.

Je remercie *M. Nicolae Vasiliu* pour m'avoir aidé dans la publication de ce livre et pour ses précieux conseils. Egalement, je remercie *M. Ilie Prisecaru* pour son soutien et son aide.

²voir la page web du projet <http://alpha.imag.pub.ro/VideoIndexingRP2>.

Je veux également adresser tous mes remerciements à *M. Teodor Petrescu* et *M. Dan Stoichescu* pour leur aide et pour avoir encouragé et soutenu mon activité de recherche à la Faculté d'Electronique, Télécommunications et Technologie de l'Information et au département d'Electronique Appliquée et d'Ingénierie de l'Information.

Je tiens à remercier le laboratoire LISTIC, Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance d'Annecy, ainsi que *M. Philippe Bolon*, directeur du LISTIC, pour le cofinancement de ma thèse de doctorat portant sur l'analyse et le traitement de séquences d'image. Je remercie également chaleureusement *M. Patrick Lambert* et *M. Didier Coquin* pour leur encadrement pendant ma thèse, leurs nombreux conseils et leur soutien constant tout au long de mes stages de recherche à Annecy.

J'adresse tous mes remerciements aux collaborateurs étrangers qui ont soutenu le projet d'indexation vidéo que j'ai monté à Bucarest:

- *M. Daniel Bouillot* de l'IMUS - Institut de Management de l'Université de Savoie et CITIA - Cité de l'Image en Mouvement,
- *M. Patrick Lambert* et *M. Philippe Bolon* du LISTIC, Polytech'Savoie, Annecy-France,
- *M. Robert Laganière* de VIVA - The Video, Image, Vision and Autonomous Systems Research Laboratory, Ottawa-Canada,
- *M. Emmanuel Quillet* et *M. Hervé Lièvre* du CERIMES - Centre de Ressources et d'Information sur les Multimédias pour l'Enseignement Supérieur du Ministère Enseignement Supérieur et Recherche Français.

Je remercie tout particulièrement mon amie Monica pour son soutien, pour la patience dont elle a fait preuve pour corriger ce manuscrit et pour avoir été constamment proche de moi.

A cette occasion, je remercie chaleureusement *Mme Eugenia Burcea* pour son aide très précieuse et *Mme Diana Cocârță* qui m'a suggéré de réaliser ce livre.

Enfin, je tiens à remercier chaleureusement la Maison d'Edition "Editura Tehnică" ainsi que *M. Roman Chirilă* pour avoir accepté la publication de ces travaux, pour le financement d'une partie des frais et aussi pour l'aide apportée pour donner vie à ce manuscrit.

Cuprins

1	Conceptul de indexare după conținut	1
1.1	Definirea conceptului de indexare	2
1.2	Sistemele de indexare de imagini	5
1.3	Sistemele de indexare a sunetului	7
1.4	Sistemele de indexare a secvențelor de imagini	8
1.4.1	Principiul adnotării de conținut	10
1.4.2	Adnotarea semantică a conținutului	14
1.4.3	Sistemul de navigare în baza de date	20
1.4.4	Sistemul de căutare în baza de date	22
1.5	Sistemele de indexare video	26
1.6	Concluzii	28
2	Segmentarea temporală	29
2.1	Structura temporală a unei secvențe	30
2.2	Descompunerea în plane video	33
2.2.1	Detecția de "cuts"	33
2.2.2	Detecția de "fades"	45
2.2.3	Detecția de "dissolves"	50
2.2.4	Evaluarea detecției tranzițiilor video	56
2.2.5	Constituirea planelor video	59
2.3	Detecția scenelor video	61
2.3.1	Tehnici de clasare automată a scenelor	63
2.3.2	Tehnici de descompunere în scene	65
2.3.3	Aplicații ale analizei scenelor video	68
2.4	Concluzii	71

3	Analiza mișcării	73
3.1	Estimarea mișcării	76
3.1.1	Metodele diferențiale	80
3.1.2	Metodele parametrice	83
3.1.3	Metodele stohastice	85
3.1.4	Metodele de estimare pe blocuri de pixeli	87
3.1.5	Fluxul video MPEG	97
3.2	Analiza mișcării camerei video	99
3.2.1	Analiza mișcării camerei în domeniul comprimat	101
3.2.2	Analiza mișcării în domeniul spațio-temporal	103
3.3	Concluzii	105
4	Analiza de culoare	107
4.1	Spațiile de culoare	109
4.1.1	Sisteme de culori primare	110
4.1.2	Sisteme pe bază de luminanță-crominanță	115
4.1.3	Sisteme perceptuale	117
4.1.4	Sisteme de axe independente	123
4.2	Conținutul de culoare la nivel de imagine	124
4.2.1	Analiza pe bază de histogramă	125
4.2.2	Analiza pe baza denumirii culorilor	129
4.2.3	Analiza senzației induse de culoare	133
4.3	Conținutul de culoare în secvențele de imagini	138
4.4	Concluzii	141
5	Rezumarea automată de conținut	143
5.1	Construcția rezumatelor statice	146
5.1.1	Clasificarea metodelor existente	147
5.1.2	Mecanismul de extragere a imaginilor cheie	153
5.2	Construcția rezumatelor dinamice	162
5.2.1	Informația conservată de rezumat	164
5.2.2	Procesul de generare a rezumatului dinamic	167
5.3	Metodele de evaluare a rezumatelor	170
5.3.1	Analiza descriptivă a rezultatului	171
5.3.2	Utilizarea unei măsuri matematice	171
5.3.3	Testele de evaluare	173
5.4	Concluzii	175
6	Formalizarea fuzzy	177
6.1	Introducerea conceptului de incertitudine	178
6.2	Logica booleană și logica fuzzy	181

6.3	Formalizarea pe baza regulilor fuzzy	184
6.3.1	Variabilele fuzzy	185
6.3.2	Principiul inferenței fuzzy	187
6.4	Avantajele reprezentării fuzzy	194
6.5	Aplicabilitatea sistemelor fuzzy	195
6.6	Concluzii	199
7	Clasificarea după conținut a datelor	201
7.1	Clasificarea nesupervizată a datelor	204
7.1.1	Etapele clasificării nesupervizate	205
7.1.2	Metodele existente de clasificare nesupervizată	206
7.2	Clasificarea supervizată	217
7.2.1	Etapele clasificării supervizate	218
7.2.2	Metodele existente de clasificare supervizată	220
7.3	Concluzii	233
	Bibliografie	235

tomată a caracterelor sau OCR ("Optical Character Recognition"⁷). Un exemplu este sistemul propus în [Kim 00b] unde regiunile din imagine ce conțin text încrustat sunt mai întâi izolate folosind o clasificare pe bază de rețele neuronale, iar mai departe literele sunt segmentate și identificate.

1.4.2 Adnotarea semantică a conținutului

Metodele existente de adnotare a conținutului secvențelor de imagini se împart în două mari categorii:

- **metode de adnotare sintactică**, ce sunt utilizate de prima generație de sisteme de indexare, precum cele enumerate în paragrafele anterioare,
- **metode de adnotare semantică**, ce reprezintă noua direcție de analiză folosită de marea parte a sistemelor de indexare actuale.

Adnotarea sintactică este definită generic ca fiind adnotarea ce se referă la *relațiile dintre unitățile de nivel scăzut constituente ale secvenței și modul de constituire a structurii acesteia*. Aceasta se poate realiza pe baza atributelor de nivel scăzut extrase din secvență, precum parametri statistici calculați la nivel de pixel sau regiuni de pixeli, proprietăți geometrice ale obiectelor, structura temporală a secvenței sau vectori de mișcare. De regulă, indicii obținuți în urma procesului de adnotare sunt valori numerice ce descriu atributele enumerate mai sus dar și relațiile sintactice ce pot exista între acestea. Extrași la acest nivel de percepție, indicii sintactici sunt accesibili doar pentru utilizatorul avizat. De exemplu, căutarea unei secvențe de imagini care să conțină 30% mișcare de translație și 20% mișcare de rotație, nu constituie o formulare prea relevantă pentru utilizator.

În contrast cu adnotarea sintactică, adnotarea semantică a conținutului propune o descriere perceptuală la un nivel similar cu nivelul de percepție uman. Informațiile de nivel scăzut obținute în urma analizei sintactice pot fi convertite în concepte lingvistice folosind informații "a priori" despre conținutul secvenței. Totuși, obținerea unei descrieri semantice de conținut necesită înțelegerea completă a conținutului secvenței, astfel că pentru aceasta se preferă o abordare multimodală (image-sunet-text).

Un sistem semantic este definit generic ca fiind *orice sistem ce implică o colecție de simboluri (vocabularul sistemului), reguli ce permit constituirea de propoziții, reguli de desemnare și reguli de validare*. În cazul sistemelor

⁷recunoașterea automată a caracterelor reprezintă procesul mecanic sau electronic de traducere a imaginilor ce conțin scris de mână, scris de mașină sau text imprimat (de regulă rezultate în urma procesului de scanare) în text editabil de către calculator.

de indexare, termenul de "semantic" își conservă acest sens. Acesta se traduce prin *codarea interpretării datelor pentru a servi unei aplicații specifice* [Smeulders 00]. Astfel, sistemele de indexare semantică implică existența unui *set de simboluri și reguli* ce permit interpretarea lingvistică a anumitor evenimente sau proprietăți ale secvențelor de imagini.

Adnotarea semantică a conținutului a fost abordată pentru prima oară în sistemele de indexare a imaginilor, dar aceasta era greu de realizat deoarece proprietățile semantice ale scenei sunt dificil de extras dintr-o simplă imagine statică. Datorită informațiilor suplimentare furnizate de secvențele de imagini (informația spațio-temporală și de mișcare), analiza semantică devine mai naturală în acest caz. De exemplu, dacă luăm cazul unei imagini ce surprinde un jucător de fotbal, singurele caracteristici ce reies din analiza imaginii sunt fizionomia acestuia și prezența sa în scenă. Pe de altă parte, dacă dispunem de secvența ce îl surprinde pe jucător, putem determina dacă acesta va marca golul, modul în care acesta joacă, despre ce meci este vorba, etc., informații semantice esențiale pentru înțelegerea conținutului secvenței.

Pentru a înțelege mai bine diferența dintre cele două categorii de adnotări, în Figura 1.3 am ilustrat un exemplu concret de adnotare sintactică și respectiv semantică în cazul unei secvențe de fotbal (axa orizontală reprezintă axa temporală, secvența este rezumată în doar câteva imagini reprezentative).

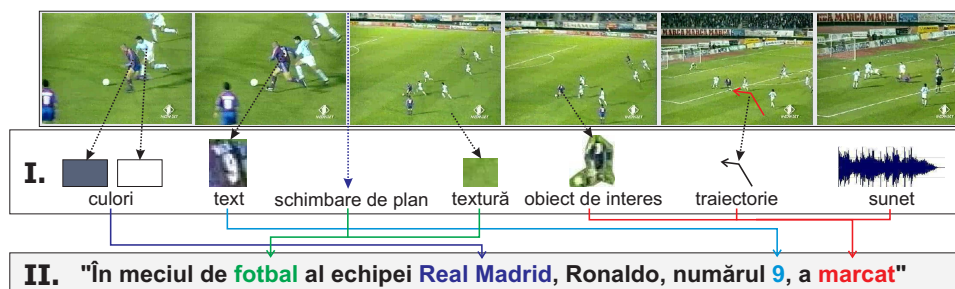


Figura 1.3: Diferența dintre adnotarea sintactică (punctul I.) și semantică (punctul II.). Săgețile colorate indică gradul de implicare al parametrilor de nivel scăzut în construirea descrierii semantice.

Astfel, în acest caz, adnotarea sintactică ne va furniza doar informații relative la scenă și la proprietățile acesteia, precum culoare, prezență text, textură, traiectoria obiectelor în mișcare, ritmul de desfășurare al acțiunii, etc. Pe de altă parte, adnotarea semantică va da sens acestor informații: în mod ideal textura verde va indica că este vorba de un meci de fotbal, culorile jucătorilor (obiecte în mișcare) vor dezvălui echipele, recunoașterea

numerelor de pe tricou va identifica jucătorii, segmentarea obiectului de interes, urmărirea acestuia și prezența zgomotului specific vor indica marcarea golului. Punând cap la cap toate informațiile, sistemul va "înțelege" că este vorba despre un meci de fotbal al echipei Real Madrid în care jucătorul cu numărul 9, Ronaldo, marchează.

Paradigma senzorială enunțată în cazul sistemelor de indexare a imaginilor este mai puțin pronunțată în cazul sistemelor de indexare a secvențelor de imagini, acest lucru datorându-se în principal informațiilor suplimentare ce facilitează înțelegerea conținutului secvenței. Cu toate acestea, paradigma semantică, de asemenea prezentă în sistemele de indexare a imaginilor, ia amploare în cazul secvențelor de imagini datorită *lipsei de corelație dintre informația pe care o putem recupera din conținutul datelor și interpretarea care i-o atribuim* [Smeulders 00].

Astfel că, un sistem de indexare semantică eficient, trebuie să reunească următoarele trăsături importante [Naphade 02]:

- în primul rând este *capacitatea de analiză semantică* pe baza cererilor de căutare formulate de utilizator (vezi Secțiunea 1.4.4),
- un sistem eficient trebuie să fie *multimodal*, reunind și armonizând metode de analiză ce folosesc diversele modalități ale secvenței: imagine, text, sunet, etc.,
- relațiile existente între atributele de nivel scăzut și percepția lor semantică trebuie *rezumate în mod eficient* pentru ca sistemul să fie capabil să ofere utilizatorului o descriere semantică coerentă.

Tendința actuală a sistemelor de indexare a secvențelor de imagini către analiza semantică a fost motivată și de atenția acordată relativ noului standard de compresie video și anume standardul MPEG-7⁸ [Wang 00]. Noul standard video încearcă să introducă în procesul de codare, direct în fluxul de date, informații semantice referitoare la conținutul secvenței. Astfel, în momentul indexării, acestea vor putea fi recuperate direct din fluxul MPEG, eliminând astfel procesul de adnotare.

Pentru o descriere mai amănunțită a sistemelor de indexare semantică, cititorul se poate raporta la studiile prezentate în [Naphade 02] și [Snoek 05b]. În cele ce urmează vom prezenta obiectivele sistemelor de indexare semantică precum și dificultățile impuse de analiza semantică a conținutului.

⁸standardul MPEG-7 este un standard de descriere a conținutului multimedia. Acesta folosește descrieri suplimentare atașate conținutului video clasic MPEG, pentru a facilita indexarea automată după conținut. Standardul MPEG-7 este denumit formal și Interfață de Descriere a Conținutului Multimedia.

numit în literatura de specialitate și "final cut". În acest sens, segmentarea temporală poate fi percepută ca fiind procesul invers editării secvenței ce are loc în studio.

2.1 Structura temporală a unei secvențe

Din punct de vedere al structurii temporale, o secvență de imagini poate fi reprezentată pe mai multe niveluri ierarhice. Acestea sunt ilustrate în Figura 2.1, astfel:

- **nivelul imagine:** reprezintă nivelul structural cu gradul de granularitate cel mai mare (cel mai detaliat nivel) și este reprezentat de toate imaginile conținute în secvență.
- **nivelul planelor video:** corespunde imaginilor secvenței ce au fost filmate între două porniri consecutive ale camerei video. Secvența de imagini astfel obținută are proprietatea de continuitate vizuală (vezi [Corridoni 95]).

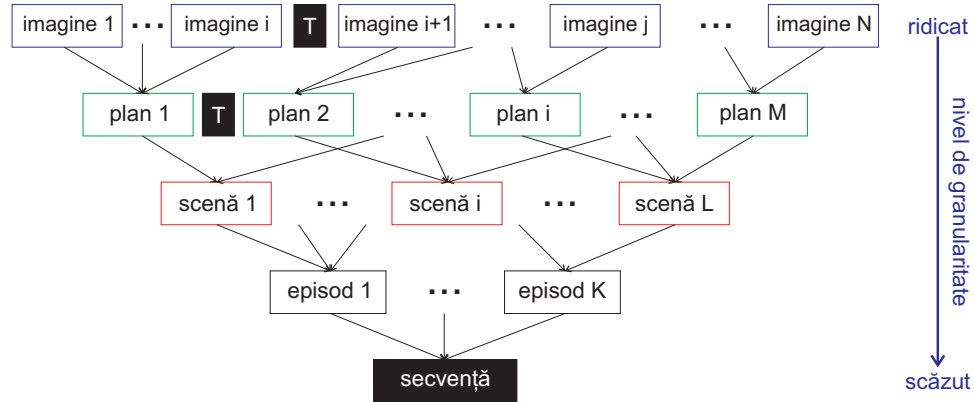


Figura 2.1: Structura ierarhică a unei secvențe de imagini (T reprezintă o tranziție video).

- **nivelul scenelor:** corespunde grupurilor de plane video ce sunt corelate din punct de vedere al conținutului semantic. Acestea trebuie să respecte regula celor trei unități: unitate de loc, unitate de timp și unitate de acțiune [Corridoni 95].

- **nivelul episoadelor:** corespunde grupurilor de scene ce sunt similare din punct de vedere al acțiunii globale (de exemplu, episoadele unei serii TV) [Bimbo 99].
- **nivelul secvenței:** este nivelul structural cu gradul de granularitate cel mai mic și este reprezentat de secvența însăși.

Marea parte a metodelor de analiză a secvențelor de imagini prelucrează secvența la nivel de plan video. Celelalte niveluri ierarhice, precum scenele sau episoadele, sunt folosite cu predilecție de sistemele de indexare semantică, deoarece detecția acestora presupune o analiză perceptuală de conținut.

Într-o secvență, planele video sunt concatenate pe baza **tranzițiilor video** (vezi Figura 2.1). O tranziție video este un efect vizual folosit pentru a lega imaginea de sfârșit a unui plan, de imaginea de început a planului următor. În funcție de tipul transformărilor 2D aplicate imaginilor, tranzițiile video existente se împart în cinci clase:

- **clasa de identitate:** tranzițiile din această categorie nu modifică imaginile planelor video și nici nu adaugă imagini suplimentare (vezi [Lienhart 01b]). În această categorie se află doar tranzițiile de tip "cut", numite și tranziții abrupte. Un "cut" produce o discontinuitate vizuală în secvență, deoarece planele vecine sunt alipite în mod direct (vezi Figura 2.2).
- **clasa spațială:** din această categorie fac parte tranzițiile ce aplică imaginilor planelor transformări spațiale [Lienhart 01b] (vezi Figura 2.2). Ca exemple putem menționa efectele de tip "wipes" în care o imagine este înlocuită progresiv de o alta folosind o margine de o anumită formă, efectele de tip "mattes" care de regulă sunt folosite pentru a combina imaginea din planul principal cu imaginea de fundal sau efectele de tip "page turns" în care noua imagine este descoperită simulând răsfoirea paginii unei cărți.
- **clasa cromatică:** în acest caz, imaginile planelor video sunt modificate prin transformări de culoare [Lienhart 01b]. Ca exemple putem menționa tranzițiile de tip "fade" și "dissolve" (vezi Figura 2.2). Un "fade" este o tranziție ce permite, fie dizolvarea progresivă a unei anumite imagini într-o imagine constantă, de regulă neagră, ceea ce numim "fade-out", fie apariția progresivă a unei imagini dintr-o imagine constantă, proces numit "fade-in". O tranziție de tip "dissolve" este în general definită de superpoziția unui efect "fade-out" peste un efect "fade-in", suprapunere ce are ca efect vizual dizolvarea unei imagini în alta.

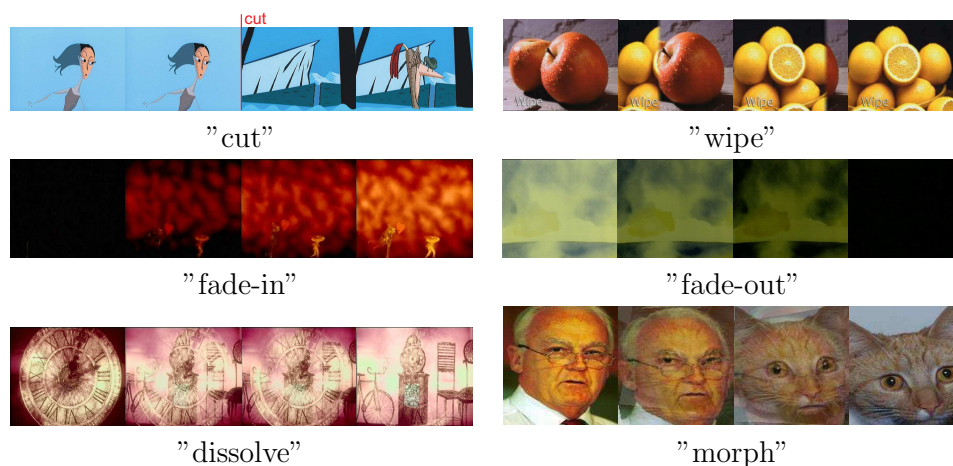


Figura 2.2: Exemple de tranziții video (pentru fiecare tranziție au fost prezentate doar câteva imagini reprezentative, axa orizontală reprezentând axa temporală, sursă imagini [Folimage 06] [Wikipedia 08] [Morphing 08]).

- **clasa spațio-cromatică:** tranzițiile video din această categorie sunt o combinație a clasei spațiale și cromatice, imaginile planelor fiind modificate atât prin transformări spațiale, cât și cromatice [Lienhart 01b]. În această categorie se regăsesc toate efectele de tip "morphing"¹ (vezi Figura 2.2). Cu toate acestea, anumite transformări din clasa cromatică pot fi încadrate și în această categorie, un exemplu fiind transformările de tip "dissolve" ce înglobează mișcări ale camerei video.
- **clasa temporală:** reprezintă o categorie aparte de tranziții video. În anumite situații, tranziția de la un plan video la altul se face temporal folosind o mișcare 3D a camerei video (vezi Secțiunea 3.2). De exemplu, camera video filmează un obiect de interes, iar apoi se translatează și se focalizează pe un punct de interes îndepărtat, din fundalul imaginii. Astfel, anumite mișcări 3D ale camerei video, cu toate că nu sunt tranziții video propriu-zise în sensul definiției enunțate anterior, au rolul de a face legătura între două momente distincte ale secvenței (două plane diferite), putând astfel fi considerate drept tranziții.

Din punct de vedere al duratei, tranzițiile video se împart în două categorii, astfel întâlnim *tranziții abrupte* sau "cuts" și *tranziții graduale*, precum

¹"morphing" este un efect special ce presupune metamorfozarea unei imagini în alta prin tranziții uniforme și constante.

tuia, "principala ambiție a lui MPEG-7 este de a face conținutul informațiilor multimedia să fie la fel de ușor de accesat pe Internet precum informațiile textuale". În ceea ce privește informația de mișcare, standardul MPEG-7 selectează și integrează *unele dintre cele mai performante metode existente* de analiză a mișcării. Astfel, tehnicile existente se grupează în două categorii principale [Jeannin 01]:

- pe de-o parte sunt metodele de analiză globală, bazate pe **analiza mișcării globale** a camerei video. În acest caz, analiza mișcării este realizată la nivel de segment video (pasaj al secvenței). Dintre aplicațiile analizei de mișcare globală putem enumera: recunoașterea mișcării camerei video, detecția activității de mișcare sau generarea imaginilor de tip "mozaic".
- pe de altă parte sunt metodele de analiză locală, ce sunt bazate pe **analiza mișcării obiectelor** din scenă. Acestea analizează mișcarea la nivel de regiuni spațiale de pixeli din imagine. De regulă, analiza locală este folosită pentru segmentarea și urmărirea temporală a obiectelor în mișcare.

Aceste două direcții de studiu sunt sintetizate în Figura 3.1. În cele ce urmează vom face o trecere în revistă a tehnicilor folosite de fiecare dintre acestea.

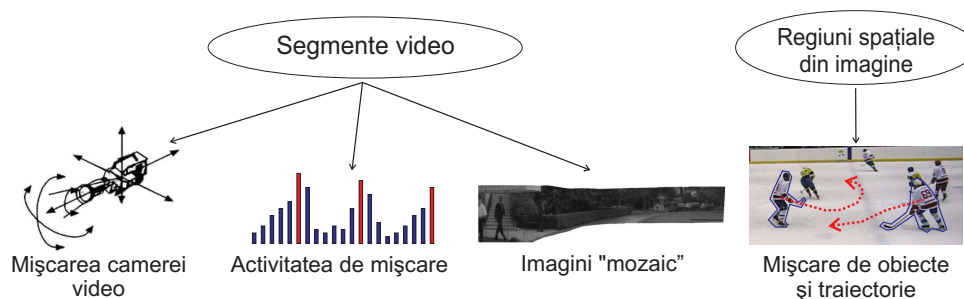


Figura 3.1: Principalele direcții de analiză a mișcării în secvențele de imagini: nivel global (segment) și nivel local (regiune) (sursă standard MPEG-7 [Jeannin 01]).

Mișcarea globală. Analiza mișcării globale a scenei este realizată la nivel de segment video sau de grup de imagini. O primă informație extrasă din secvență este *tipul mișcării camerei video*, ca de exemplu: mișcare translatională, mișcare de rotație, mișcare de apropiere etc. (vezi Secțiunea 3.2).

Informațiile reținute în acest caz pentru o anumită categorie de mișcare sunt de regulă amplitudinea mișcării, durata mișcării precum și localizarea acesteia în secvență. Analiza mișcării camerei video este importantă deoarece permite în anumite situații înțelegerea conținutului secvenței prin identificarea anumitor pasaje de interes din aceasta. De exemplu, focalizarea asupra unui anumit personaj se traduce printr-o mișcare a camerei de tip ”zoom-in” (mărire), sau, creșterea suspansului acțiunii poate fi marcată de o mișcare de translație foarte rapidă.

O altă informație exploatată este *activitatea de mișcare*. Aceasta este o măsură a percepției vizuale pe care o avem asupra mișcării conținute în secvență. Activitatea de mișcare este determinată pe baza clasificării mișcării globale în funcție de o serie de parametri de nivel scăzut (de exemplu, dispersia amplitudinii vectorilor de mișcare). Clasificarea este realizată pe mai multe niveluri de activitate, în funcție de intensitatea acțiunii. La clasificare este luată în calcul și situația în care acțiunea este absentă, aceasta reprezentând nivelul minim de activitate. Un nivel de activitate intens corespunde evenimentelor dinamice, ca de exemplu scenele de gol din secvențele de fotbal sau scenele de urmărire de mașini din secvențele de știri. Pe de altă parte, un nivel de activitate redus corespunde scenelor cu un conținut ”sărac” în mișcare, ca de exemplu scenele de dialog dintre personaje sau scenele de interviu din secvențele de știri sau documentare.

Tot pe baza analizei mișcării globale este și *construcția imaginilor de tip ”mozaic”* [Aner 01]. O imagine de tip ”mozaic” este o imagine statică ce rezumă conținutul de mișcare global al unui pasaj al secvenței (de regulă un plan video). Aceasta este realizată prin regruparea și suprapunerea diverselor imagini ale segmentului, după recalarea geometrică în funcție de deplasarea globală a scenei (vezi Figura 5.2 de la pagina 150). Imaginile de tip ”mozaic” sunt folosite drept rezumate compacte ale diverselor pasaje ale secvenței și de regulă au o complexitate de calcul ridicată. Totuși, aceasta poate fi redusă prin folosirea parametrilor de deformare furnizați de standardul MPEG-7.

Mișcarea locală. Analiza mișcării locale sau a deplasării obiectelor, este efectuată la nivel de regiuni de pixeli. Dacă pentru caracterizarea globală a mișcării, vectorii de mișcare puteau fi estimați la un nivel de detaliu mai redus (de exemplu, la nivel de blocuri de pixeli), furnizând astfel o aproximație grosieră a fluxului optic, în cazul analizei mișcării locale a obiectelor, vectorii de mișcare sunt estimați de regulă la nivel de pixel pentru obținerea unui nivel de detaliu ridicat. Pentru analiză, metodele existente folosesc de regulă o modelare parametrică a mișcării. Aceasta permite localizarea în secvență a obiectelor cu deplasări similare, în ciuda diverselor deformări geometrice suportate de acestea. În general, rezultatul analizei mișcării obiectelor este

cuantificat prin furnizarea traiectoriei acestora sub formă de evoluție temporală a anumitor puncte de interes, precum centrul de greutate sau anumite puncte de contur.

În contextul indexării după conținut a secvențelor de imagini, metodele de analiză a traiectoriei obiectelor ("object tracking") sunt cu mult mai studiate decât metodele de analiză globală a mișcării camerei video. Acest lucru se datorează în principal faptului că într-o secvență marea majoritate a evenimentelor de interes implică, și sunt legate, de mișcarea obiectelor. De exemplu, într-o secvență sportivă, va fi mult mai interesant și totodată reprezentativ pentru analiză să dispunem de traiectoria unui anumit jucător care este într-o acțiune de atac, decât să caracterizăm mișcarea globală a camerei video ce urmărește jucătorul. Pentru un studiu biliografic complet al tehnicilor de analiză a mișcării obiectelor, cititorul se poate raporta la studiile [Dagtas 00], [Fablet 02] sau [Smith 04].

În concluzie, toate metodele existente de analiză a mișcării, fie că este vorba de mișcare globală sau locală, folosesc ca punct de plecare *estimarea mișcării*. Aceasta, pe baza măsurării deplasării pixelilor, sau a regiunilor de pixeli, de la un cadru la altul, furnizează un câmp vectorial de mișcare. În cele ce urmează vom face o trecere în revistă a tehnicilor de estimare a mișcării existente.

3.1 Estimarea mișcării

Principiul estimării de mișcare constă în determinarea deplasării unui pixel, sau a unui bloc de pixeli, între două imagini succesive ale secvenței, pe baza minimizării variației intensității acestuia, numită și DFD sau "Displaced Frame Difference". Această variație poate fi reprezentată sub forma următoare:

$$DFD(\vec{r}, \vec{d}, \Delta t) = I(\vec{r} + \vec{d}, t + \Delta t) - I(\vec{r}, t) \quad (3.1)$$

unde $\vec{r}$ reprezintă poziția pixelului sau a blocului de pixeli în imaginea analizată, $\vec{d}$ reprezintă vectorul de deplasare între momentele t și $t + \Delta t$ exprimat în funcție de deplasarea pe cele două axe, oX și respectiv oY , $\vec{d} = (dx, dy)$, iar $I(t)$ reprezintă imaginea la momentul t .

Acest principiu de estimare se bazează pe ipoteza conform căreia intensitatea pixelilor nu variază semnificativ de la o imagine la alta. Un exemplu de vectori de mișcare obținuți la nivel de blocuri de pixeli este prezentat în Figura 3.2. Secvența folosită, pentru care am ilustrat câteva imagini reprezentative, conține o deplasare a camerei video către dreapta, de aici

ce numim o "sferă de culoare". De-a lungul istoriei, au fost propuse mai multe astfel de sisteme de reprezentare perceptuală a culorilor, dintre acestea menționăm pe cele mai cunoscute ce poartă și numele creatorilor lor, și anume: sfera de culoare a lui Runge, roata de culoare a lui Chevreul, spațiul culorilor opuse a lui Hering, solidul de culoare a lui Munsell, roata de culoare a lui Itten, etc. (vezi Figura 4.11).

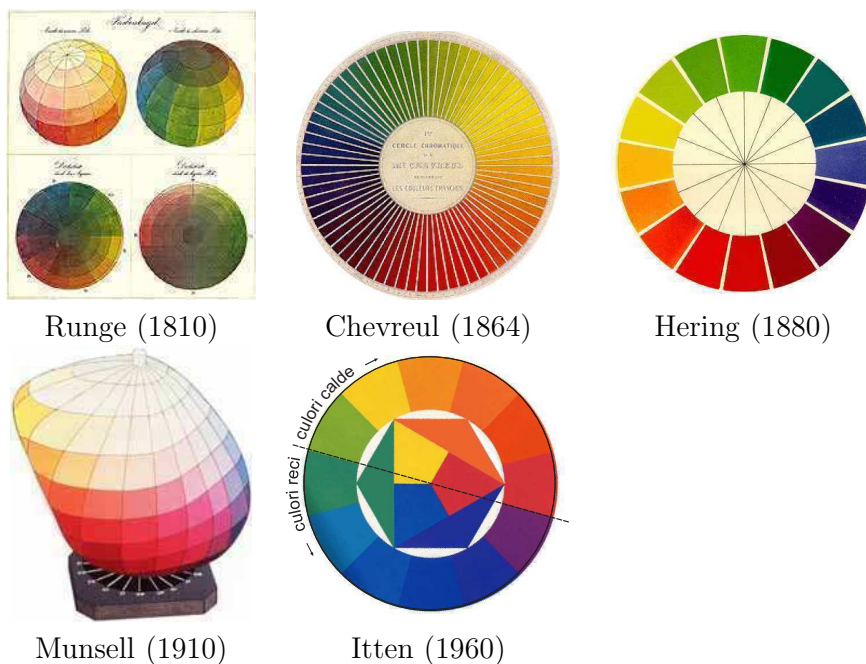


Figura 4.11: Reprezentarea culorilor sub formă perceptuală pe baza roților și a sferelor de culoare.

O astfel de reprezentare a culorilor este foarte utilă în cazul în care se dorește studierea relațiilor perceptuale dintre culori. În domeniul artei, conceptul de relație între culori, unde culorile sunt combinate pe baza unei roți de culoare, este un studiu esențial (vezi Josef Albers, Faber Birren, Johannes Itten, etc.).

Dacă luăm ca exemplu roata de culoare a lui Itten, care este și una dintre cele mai cunoscute roți de culoare, în aceasta culorile sunt aranjate cu un anumit scop, astfel: culorile considerate ca fiind calde se găsesc în prima jumătate a roții, începând cu culoarea "spring", continuând cu galben și finalizând cu magenta, în timp ce culorile considerate ca fiind reci se găsesc în cealaltă jumătate, pornind de la violet, continuând cu albastru și finalizând

cu verde (vezi Figura 4.11). Mai mult, culorile ce sunt opuse din punct de vedere al percepției se găsesc poziționate diametral opus (de exemplu albastru și galben) în timp ce culorile considerate analoage sunt culori vecine pe roata de culoare (de exemplu galben și portocaliu).

Astfel, folosind teoria de culoare dezvoltată de Itten, precum și reprezentarea perceptuală a culorilor pe baza roților de culoare, putem caracteriza conținutul vizual în termeni de *contrast* și respectiv *concordanță* de culoare. Itten [Itten 61] definește percepția de culoare pe baza a *șapte contraste de culoare* ce sunt exemplificate în Figura 4.12, astfel:

- *contrastul de nuanță*: acest contrast vizual este realizat prin juxtapunerea de diverse nuanțe de culoare. Cu cât acestea sunt mai diferite din punct de vedere perceptual, cu atât mai puternic este contrastul obținut (distanța dintre culori este evaluată folosind o roată de culoare). Un exemplu este ilustrat în Figura 4.12.a.

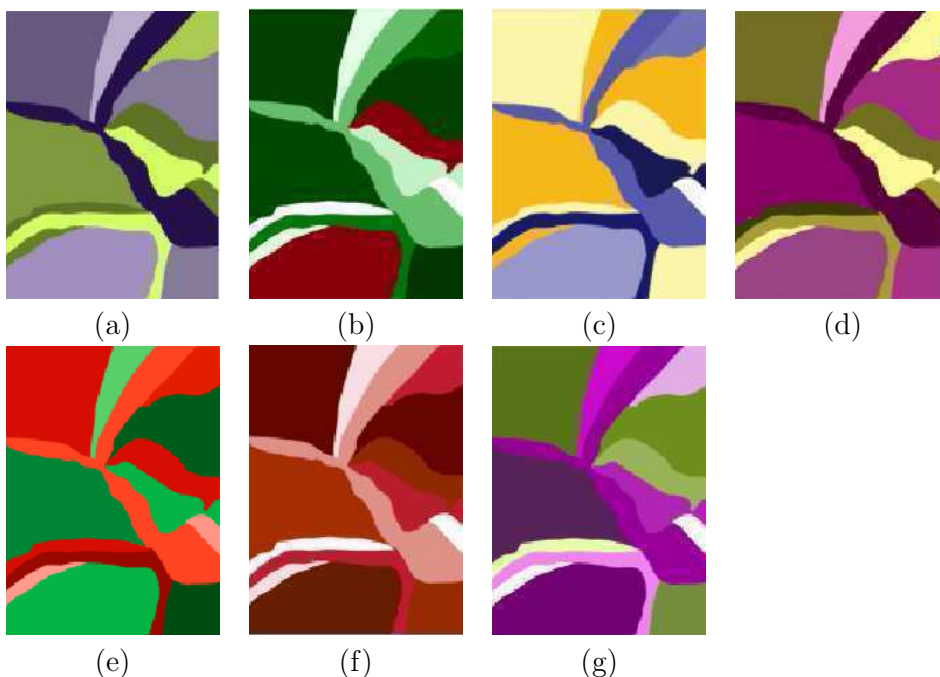


Figura 4.12: Cele șapte contraste ale lui Itten: (a) Contrastul de nuanță, (b) Contrastul închis-deschis, (c) Contrastul cald-rece, (d) Contrastul de complementaritate, (e) Contrastul de simultaneitate, (f) Contrastul de saturație, (g) Contrastul de extensie (sursă imagini ”<http://www.worqx.com/color/itten.htm>”).

- *contrastul închis-deschis*: acest contrast este legat de gradul de percepție al intensității luminoase. La extreme se găsesc negrul (absența luminii) și albul (intensitatea maximă), iar între acestea sunt nivelele de gri și nuanțele cromatice. Contrastul este realizat prin juxtapunerea atât a culorilor deschise cât și închise (vezi Figura 4.12.b).
- *contrastul cald-rece*: acest contrast corespunde senzației de căldură transmisă de anumite culori. În domeniul artei, culorile prezintă o anumită temperatură sau căldură. Astfel, galben, galben-portocaliu, portocaliu, roșu-portocaliu, roșu și roșu-violet sunt considerate ca nuanțe calde, pe când galben-verde, verde, albastru-verde, albastru, albastru-violet și violet sunt considerate ca nuanțe reci. Contrastul de culoare este realizat prin juxtapunerea atât a culorilor calde cât și reci (vezi Figura 4.12.c).
- *contrastul de complementaritate*: acest contrast corespunde relațiilor de complementaritate existente între culori. În practică, pe o roată de culoare (de exemplu roata de culoare a lui Itten, vezi Figura 4.11) perechile de culori complementare (opuse ca percepție) sunt determinate de linia dreaptă ce trece prin centrul roții și care leagă două culori diametral opuse. Contrastul de complementaritate este astfel realizat prin folosirea de culori opuse din punct de vedere al percepției, obținându-se astfel o anumită simetrie vizuală (vezi Figura 4.12.d).
- *contrastul de simultaneitate*: acest contrast se folosește de răspunsul asimetric al percepției umane la fenomenul culorilor opuse. Contrastul este realizat prin "vibrarea" percepției frontierelor dintre culori. Cu acest contrast se pot realiza o serie de iluzii optice interesante. Un exemplu este ilustrat în Figura 4.12.e.
- *contrastul de saturație*: acest contrast este realizat prin juxtapunerea de nuanțe pure saturate cu nuanțe diluate de saturație redusă. Acest contrast se dovedește însă a fi relativ, deoarece anumite culori pot apărea ca fiind mai saturate prin contrast dacă sunt alăturate unei culori mai puțin saturate, și vice-versa. Un exemplu este ilustrat în Figura 4.12.f.
- *contrastul de extensie*: acest contrast este legat de proporția în care sunt folosite culorile în imagine. Percepția vizuală a unei culori este direct influențată de gradul de luminanță folosit, precum și de suprafața spațială ocupată de culoare în imagine. Contrastul de extensie este astfel realizat prin asocierea de culori regiunilor fizice din imagine ce au o suprafață proporțională cu ponderea perceptuală vizuală a culorii (vezi Figura 4.12.g).

în cele mai multe cazuri, are nevoie să vizualizeze conținutul secvențelor. Problema nu este una dificilă în cazul în care ar fi vorba de doar câteva secvențe, dar vizualizarea a mii de secvențe este un lucru aproape imposibil de realizat. Una dintre soluțiile adoptate constă în folosirea *rezumatelor de conținut*.

Un rezumat al unei secvențe de imagini poate fi definit în linii mari ca fiind *o reprezentare compactă a conținutului acesteia* [Li 01]. Mai riguros, rezumatul unei secvențe de imagini reprezintă *o colecție, de dimensiuni reduse, de imagini fixe (colecție de imagini) sau în mișcare (colecție de segmente), ce redă conținutul secvenței în așa fel încât partea esențială a acestuia să fie conservată iar utilizatorul să fie informat rapid și concis* [Pfeiffer 96].

Interesul în a dispune de o reprezentare compactă a secvenței nu se rezumă doar la reducerea timpului necesar căutării și navigării în conținutul bazei de date. Rezumatul de conținut poate fi folosit și pentru a reduce timpul de calcul în anumite metode de analiză și prelucrare a secvențelor de imagini, prin reducerea volumului de date ce trebuiesc prelucrate. Mare majoritate a metodelor existente nu folosesc integral conținutul secvenței, ci numai informația furnizată de un anumit număr sau grup de imagini reprezentative ("imagini cheie"). Acestea sunt alese astfel încât, pentru prelucrarea vizată, informația necesară din secvență să nu fie alterată. De exemplu, pentru a calcula și analiza distribuția globală de culoare a unei secvențe de imagini, folosirea unei singure imagini pentru fiecare plan video oferă o precizie similară folosirii tuturor imaginilor din secvență [Ionescu 05c].

Din punct de vedere al procesului de generare, un rezumat poate fi construit *manual*, *semi-automat* (intervenția umană este parțială, de regulă folosită ca validare) sau în mod *automat*. Având în vedere volumul mare de date conținute chiar la nivelul unei singure secvențe, metodele de generare manuală, și chiar semi-automată, sunt evitate datorită implicării unui număr important de resurse umane în procesul de selecție a conținutului reprezentativ al secvenței. Tendința actuală este de automatizare completă a procesului de rezumare pentru a putea fi astfel folosit în timp real la indexarea conținutului bazelor de secvențe de imagini.

După cum reiese din însăși definiția conceptului de rezumat, este posibilă generarea a două categorii distincte de rezumat, și anume:

- pe de-o parte sunt **rezumatele în imagini** sau *rezumatele statice*: acestea reprezintă un fel de "storyboard"² simplificat al secvenței și sunt

²un "storyboard" reprezintă o modalitate de organizare grafică sub forma unei serii de ilustrații sau imagini, prezentate similar unei benzi desenate ("comics"), a momentelor importante din conținutului unui document video, film, etc. Acesta are ca scop previzualizarea conținutului și precede procesul de creare propriu-zisă.

la bază o colecție de imagini reprezentative pentru conținutul secvenței. În literatura de specialitate acestea sunt cunoscute sub numele de "video summaries".

- pe de altă parte sunt **rezumatele în mișcare** sau *rezumatele dinamice*: acestea reprezintă o colecție de segmente ale secvenței, fiind ele însele niște secvențe de imagini, dar de o durată mult inferioară secvenței inițiale. Uzual, dacă este vorba de rezumarea unui document video, în rezumatul dinamic este prezentă și informația audio. Rezumatele dinamice sunt cunoscute în literatura de specialitate sub numele de "video skims".

Aceste două modalități de rezumare a conținutului unei secvențe de imagini prezintă fiecare o serie de avantaje și dezavantaje. Astfel, în ceea ce privește rezumatele în imagini, principalele avantaje ale acestora pot fi sintetizate cu următoarele:

- pot fi *generate rapid* deoarece nu iau în calcul decât informația vizuală (sunetul și textul nu sunt prezente),
- pot fi *vizualizate foarte ușor*, fiind doar un ansamblu de imagini ce nu necesită sincronizarea sau temporizarea datelor (de exemplu, sincronizarea sunetului și a imaginii),
- pot facilita construirea *imaginilor de tip "mozaic"*³ [Aner 01],
- pot fi ușor de imprimat pe un suport fizic pentru a ține loc, de exemplu, de "storyboard" al secvenței,
- permit *reducerea complexității de calcul* pentru anumite metode de analiză ce pot fi aplicate direct acestora.

Pe de altă parte, rezumatele dinamice prezintă și ele o serie de avantaje fundamentale, astfel:

- acestea *au mai mult sens* decât rezumatele statice deoarece conțin informație temporală de mișcare, informație ce este pierdută în rezumatele statice,
- un rezumat dinamic este *mai bogat în informație*, acesta putând conține și alte informații precum sunetul,

³vezi explicația de la pagina 150.

- vizualizarea rezumatelor dinamice este mult mai *naturală și atractivă*: este mult mai interesant pentru utilizator să vizualizeze, de exemplu, reclama unui nou film în curs de apariție ("trailer"), decât să vizualizeze o succesiune sacadată de imagini statice din acesta [Li 01].

În ciuda faptului că rezumatele dinamice oferă un conținut informațional mult mai bogat decât rezumatele statice, acestea implică în cele mai multe situații o complexitate de calcul mult mai importantă, precum și un proces de generare mult mai laborios (sincronizare imagine și sunet, respectarea continuității și a coerenței vizuale, etc.).

În acest punct, analizând avantajele și dezavantajele celor două tipuri de rezumate, am fi tentați să alegem doar unul dintre ele ca fiind soluția optimală a problemei de rezumare de conținut. În practică, *ambele tipuri de rezumate sunt necesare* într-o aplicație de indexare după conținut, deoarece fiecare dintre acestea este adaptat unei cerințe diferite. Astfel, rezumatul în imagini permite o reprezentare rapidă și concisă, în doar câteva imagini, a conținutului vizual, ideală în cazul în care utilizatorul dorește doar să "răsfoiască" conținutul bazei de date, în timp ce rezumatul dinamic permite o reprezentare rapidă și concisă a conținutului dinamic al secvenței, permițând utilizatorului ca într-un timp relativ scurt să-și facă o idee asupra acțiunii secvenței.

Mai mult, cu toate că nu este cea mai bună strategie, fiecare tip de rezumat poate fi generat pornind de la celălalt. De exemplu, un rezumat dinamic poate fi construit pe baza unui rezumat static prin concatenarea unor segmente de o anumită durată ce conțin imaginile rezumatului static. Similar, un rezumat static poate fi generat dintr-un rezumat dinamic prin păstrarea doar a anumitor imagini reprezentative sau chiar prin sub-eșantionarea temporală a acestuia.

Pentru un studiu bibliografic complet al literaturii de specialitate din această direcție de studiu, cititorul se poate raporta la lucrările [Li 01] sau [Truong 07]. În cele ce urmează vom face o trecere în revistă a particularităților metodelor folosite de fiecare dintre cele două categorii de rezumat.

5.1 Construcția rezumatelor statice

După cum am menționat în partea introductivă a acestui capitol, un rezumat static este la bază o colecție de imagini fixe ce sunt considerate ca fiind reprezentative pentru conținutul secvenței. Acestea sunt numite și "imagini cheie". Din punct de vedere formal, rezumatul static al secvenței S , notat

în primul rând, mult mai mult sens, acestea devenind mai ușor accesibile publicului larg. De exemplu, descrierea distribuției de culoare a unei imagini cu parametri de nivel scăzut, precum $I = \{P_{(255,0,0)} = 50\%, P_{(0,0,255)} = 50\%\}$ (unde $P_{(c)}$ reprezintă procentul de apariție al unei culori c), va fi accesibilă doar unei persoane avizate în domeniu, în timp ce descrierea la un nivel semantic perceptual, precum ”imaginea contrastează culoarea roșie și albastră”, va permite oricărui dintre noi să-și creeze o imagine mentală a conținutului de culoare al imaginii în cauză.

În sistemele de indexare după conținut a secvențelor de imagini, descrierile semantice își găsesc un spectru de aplicabilitate mai larg [Ionescu 07a], astfel:

- permit *simplificarea navigării în baza de date*: acestea vin cu informații suplimentare despre conținutul secvențelor, facilitând astfel înțelegerea rapidă a acestuia de către utilizator. De exemplu, pe lângă rezumatele automate de conținut, o secvență poate fi acompaniată și de informații textuale, cum ar fi: genul acesteia (acțiune, dramă, documentar), tipul conținutului (natură, oraș, studio), etc.
- permit *simplificarea căutării în baza*: fiind exprimate sub formă textuală, pot fi folosite ca indici de căutare în baza de date. Astfel, utilizatorul își poate formula cererea de căutare într-un limbaj natural, apropiat de limbajul uman, de exemplu ”caută secvențele de gol din meciul echipelor X și Y” sau ”caută filmele de ficțiune”.
- constituie *un ajutor pentru specialiști*: descrierile semantice de conținut pot însoți descrierile de nivel scăzut pentru a furniza un pachet de informații complet cu privire la tehnicile folosite în secvență: structura temporală, conținut de mișcare, conținut de culoare, etc.

După cum se poate observa din cele enunțate anterior, modalitatea cea mai expresivă de exprimare a sensului semantic al datelor constă în *reprezentarea textuală a acestuia*. Dintre metodele de asociere de descrieri textuale de conținut datelor numerice, o pondere importantă o au metodele ce se folosesc de conceptul de incertitudine. Una dintre acestea o reprezintă *formalizarea fuzzy* a conceptelor semantice textuale.

6.1 Introducerea conceptului de incertitudine

Printre schimbările spectaculoase ale paradigmelor existente în diversele domenii științifice, una dintre cele mai importante o constituie fundamentarea și dezvoltarea *conceptului de incertitudine* a datelor.

Această schimbare majoră a modului de percepție științific s-a materializat prin tranziția modului tradițional de gândire, care insistă asupra faptului că noțiunea de incertitudine este o proprietate a datelor ce nu se dorește să apară și care trebuia evitată pe cât posibil, spre o viziune alternativă ce tolera tocmai această noțiune, fiind considerată de această dată inevitabilă și foarte utilă pentru analiză.

În general, în momentul proiectării unui anumit sistem de analiză, întrebarea care se pune este: "*Cum trebuie gerat sistemul și problemele asociate acestuia în cazul în care complexitatea proceselor ce trebuiesc modelizate depășește cu mult posibilitățile noastre de prelucrare?*". Cu alte cuvinte, volumul informațional disponibil este foarte ridicat pentru a putea fi controlat în totalitate iar înțelegerea proceselor este limitată. Soluția în acest caz constă tocmai în introducerea noțiunii de incertitudine pentru situațiile în care soluția ce trebuie adoptată nu este deloc evidentă, ci mai degrabă incertă.

În momentul construcției unui anumit model, se încearcă întotdeauna să se maximizeze utilitatea acestuia. Acest obiectiv este strâns legat de relațiile ce pot exista între cele trei categorii de caracteristici cheie ale unui model, și anume: *complexitatea*, *credibilitatea* și *incertitudinea* acestuia. Aceste relații nu sunt întotdeauna înțelese în totalitate. Știm doar că incertitudinea, fie că este predictivă, prescriptivă, etc., joacă rolul esențial pentru efortul de maximizare a utilității sistemului.

Totuși în cele mai multe situații, dar nu întotdeauna, incertitudinea nu reprezintă un punct forte dacă aceasta este considerată independent de alți parametri. Incertitudinea devine o informație prețioasă a sistemului dacă este analizată în raport cu alte caracteristici ale acestuia. În general, cu toate că aparent este paradoxal, cu cât se adaugă mai multă incertitudine în modelarea sistemului, cu atât complexitatea acestuia este redusă și în efect credibilitatea modelului crește [Klir 95].

În concluzie, conceptul de incertitudine este un instrument important pentru modelarea unui anumit sistem sau pentru soluționarea unei anumite probleme. Acesta permite obținerea de caracteristici "avantajoase" pentru modelul vizat, caracteristici ce vor conduce ulterior la maximizarea utilității acestuia relativ la scopul pentru care a fost creat.

Conceptul de incertitudine a fost materializat pentru prima dată în lucrările publicate de Lotfi A. Zadeh [Zadeh 65] (anticipat de filozoful Max Black în 1937). Acesta propunea o nouă teorie bazată pe reprezentarea datelor cu mulțimi fuzzy. Mulțimile fuzzy sunt mulțimi pentru care frontierele dintre date nu sunt exacte. Apartenența datelor, în acest caz, la o astfel de mulțime, nu mai este o problemă de confirmare sau negare, ci o problemă de *grad de apartenență*.

Dacă teoria probabilităților este fondată pe definirea a două valori logice de adevăr, și anume *Adevărat* (1) și *Fals* (0), în logica fuzzy, gradul de adevăr este formulat în felul următor:

dacă A reprezintă o mulțime fuzzy iar x este un obiect de interes, atunci propoziția " x este inclus în A " nu este obligatoriu să fie Adevărată sau Falsă, lucru impus de logica booleană, ci aceasta poate fi adevărată într-un anumit grad.

Acest grad de adevăr este exprimat de regulă ca o valoare cuprinsă în intervalul $[0; 1]$, unde limitele acestuia reprezintă negația totală (limita inferioară, valoare de adevăr 0) și respectiv, afirmația totală (limita superioară, valoare de adevăr 1).

Capacitatea mulțimilor fuzzy de a exprima tranziția graduală între apartenența totală și non apartenență, și vice-versa, își găsește o vastă utilitate în marea majoritate a domeniilor existente. Mulțimile fuzzy nu numai că propun o reprezentare discriminantă și plină de sens a conceptului de incertitudine, ci și o reprezentare pertinentă a conceptelor vagi ce sunt exprimate într-un *limbaj natural*.

Pentru a înțelege avantajul folosirii mulțimilor fuzzy la descrierea proprietăților anumitor procese, vom considera exemplul următor [Klir 95]: în loc să descriem prognoza meteo a zilei curente specificând procentajul exact de acoperire al cerului cu nori, P_{nori} , putem adopta o soluție mai eficientă spunând că ziua va fi, fie "însorită", fie "cu un cer acoperit".

Această descriere este o descriere vagă și puțin exactă, dar în cele mai multe cazuri este mult mai utilă decât prima modalitate de descriere. Sensul termenului de "însorită" nu este în totalitate arbitrar. O acoperire cu nori în procent de $P_{nori} = 100\%$ indică faptul că ziua nu este însorită, dar același lucru este valabil și pentru $P_{nori} = 80\%$. Astfel, pentru a desemna o "zi însorită" vom considera o serie de valori intermediare pentru P_{nori} , de exemplu: $P_{nori} \in [10\%; 20\%]$. Problema care apare este cum alegem aceste frontiere? Dacă considerăm că un procent de acoperire de mai puțin de 25% corespunde unei "zi însorite", atunci o acoperire de 26% corespunde sau nu aceluiași caz? Este evident inacceptabil în această modalitate de descriere binară ca o singură valoare a lui P_{nori} să facă diferența între două concepte opuse: "zi însorită" și respectiv "zi cu cer acoperit".

Pentru a soluționa acest conflict, termenul de "zi însorită" necesită un anumit grad de incertitudine ce va fi obținut prin introducerea unei tranziții graduale între valorile lui P_{nori} , folosite pe de-o parte pentru a desemna noțiunea de "zi însorită" și respectiv pentru conceptul opus. Acesta constituie exact principiul de bază al logicii fuzzy ce reprezintă o generalizare a logicii booleene.

laritate între vectorii de caracteristici asociați obiectelor, conceput pe baza căruia obiectele așa zise similare vor fi atribuite aceleiași clase. Principiul clasificării este sintetizat în Figura 7.1.

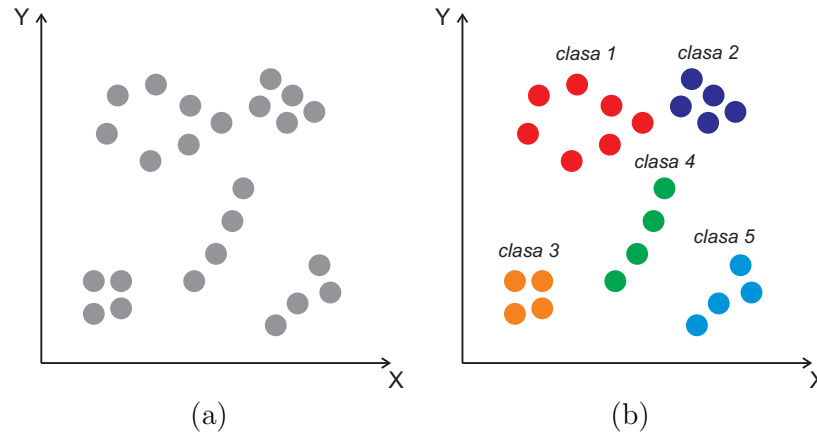


Figura 7.1: Principiul clasificării datelor: (a) datele de intrare reprezentate în spațiul de caracteristici, (b) repartiziția în clase obținută în urma clasificării (obiectele din aceeași clasă sunt reprezentate cu aceeași culoare).

Tehnicile de clasificare existente sunt utilizate într-o gamă foarte largă de aplicații. Acestea deservesc diverse obiective, dintre care putem menționa drept cele mai importante următoarele:

- **reducerea volumului informațional:** tehnicile de clasificare a datelor permit regruparea unui ansamblu de date în grupuri omogene, lucru ce facilitează reducerea volumului informațional disponibil. De exemplu, fiecare grup (clasă) de date poate fi reprezentat, în etapele de prelucrare ulterioare, doar de informația cea mai reprezentativă a grupului. La un alt nivel, clasificarea datelor permite eliminarea redundanței informaționale prin reducerea spațiului de caracteristici.
- **punerea în evidență a "cunoașterii":** tehnicile de clasificare permit localizarea într-un volum mare de date a unor grupuri de informații ce prezintă anumite caracteristici de interes. Localizarea acestora furnizează utilizatorului o cunoaștere nouă a relațiilor existente între date, cunoaștere ce nu era disponibilă anterior căutării. Acest proces este cunoscut în literatura de specialitate și sub numele de "data mining".

- **punerea în evidență a relevanței claselor:** tehnicile de clasificare permit localizarea anumitor grupuri de date ce sunt reprezentative pentru ansamblul datelor analizate,
- **punerea în evidență a datelor atipice:** tehnicile de clasificare permit de asemenea localizarea datelor ce nu corespund niciunui criteriu de similaritate, date ce sunt considerate ca fiind atipice pentru criteriile considerate. Acestea sunt importante, deoarece sunt un caz particular și trebuie analizate separat. Un astfel de exemplu sunt datele ce se găsesc pe frontiera dintre două clase diferite, date ce pot fi considerate ca aparținând ambelor clase cât și ca o clasă independentă.

Din punct de vedere al problematicii indexării datelor, subiect ce face obiectul acestei cărți, tehnicile de clasificare sunt indispensabile unui sistem de indexare după conținut. Clasificarea datelor intervine în general în însuși procesul de căutare al informației. Utilizatorul, prin formularea cererii de căutare va defini spațiul de caracteristici ce va fi folosit pentru localizarea datelor dorite. Pe baza acestuia, datele din baza de date pot fi grupate în funcție de similaritate sau cu alte cuvinte în funcție de asemănarea dintre vectorii de caracteristici asociați. Astfel, grupul sau grupurile de date ce sunt suficient de similare vectorului de caracteristici asociat cererii de căutare vor fi furnizate utilizatorului drept rezultat.

Metodele de clasificare existente se împart în două mari categorii. Prima categorie de metode o constituie *metodele probabilistice* sau de *clasificare supervizată*. Clasificarea supervizată implică clasarea datelor pe baza unor modele predefinite de clase sau "date de antrenament". Acestea reprezintă de regulă o clasificare de referință ce corespunde realității (similară unei "realități de teren" sau "groundtruth"¹), folosită inițial pentru antrenarea sistemului înaintea clasificării propriu-zise a datelor. În literatura de specialitate, termenul asociat metodelor din această categorie este de "metode de clasificare" sau "classification methods".

O a doua categorie de metode de clasificare sunt *metodele de clasificare nesupervizată sau automată*, desemnate în literatura de specialitate prin termenul de "clustering"². Clasificarea nesupervizată, spre deosebire de clasificarea supervizată, propune o partiție optimală a spațiului de caracteristici din punct de vedere al unui anumit criteriu matematic, fără a folosi informații "a priori" (de exemplu, o partiție de referință). Avantajul acestor metode este dat de faptul că sunt complet automate (nu necesită intervenția utilizatorului) și pot fi folosite pentru clasarea datelor despre care nu dispunem de

¹vezi explicația de la pagina 170.

²de notat este faptul că în limba română, termenul de clasificare este folosit generic pentru a desemna, în funcție de context, atât o clasificare supervizată cât și nesupervizată.

informații relative la conținutul acestora (număr de clase, prototipul clasei, etc.). Pe de altă parte, fiind un proces automat, relevanța claselor tinde să fie mai redusă decât în cazul clasificării supervizate, aceasta fiind dependentă de metoda folosită cât și de puterea discriminatorie a spațiului de caracteristici folosit.

În cele ce urmează, vom face o trecere în revistă a tehnicilor de clasificare supervizată și nesupervizată existente punând în evidență avantajele cât și dezavantajele fiecărei abordări.

7.1 Clasificarea nesupervizată a datelor

O catalogare interesantă a metodelor de clasificare nesupervizată existente în funcție de proprietățile contrastante ale acestora este propusă în [Jain 99], astfel, metodele de clasificare nesupervizată sunt:

- **acumulative sau partiționale:** această proprietate este legată de modul de structurare al algoritmului folosit. Metodele acumulative pornesc clasificarea de la o anumită partiție în clase, clase care pe parcursul algoritmului sunt fuzionate iterativ până când este satisfăcut un anumit criteriu de convergență. Pe de altă parte, metodele partiționale pornesc de la o singură clasă ce este divizată iterativ până când criteriul de convergență considerat este satisfăcut.
- **monotetice sau politetice**³: aceste proprietăți sunt legate de modul de utilizare a vectorilor de caracteristici în procesul de clasificare, care poate fi secvențial sau simultan. Marea parte a metodelor existente sunt politetice, astfel că pentru estimarea distanței dintre obiecte sunt folosiți toți parametrii disponibili ("features"). De asemenea, deciziile de clasare sunt luate pe baza acestei măsuri de distanță. Pe de altă parte, metodele monotetice folosesc parametrii în mod secvențial pentru a constitui progresiv clasele, de exemplu, parametrul x_1 este folosit pentru a diviza datele în două clase, mai departe, parametrul x_2 este folosit pentru divizarea claselor anterioare, și așa mai departe.
- **nete sau fuzzy.** O clasificare netă presupune alocarea fiecărui obiect unei singure clase, astfel apartenența fiind sigură. Pe de altă parte, o clasificare fuzzy asociază fiecărui obiect un grad de apartenență la una sau mai multe clase, apartenența la clase fiind de această dată incertă.

³termenul de monotetic desemnează o anumită clasă ai cărei membrii sunt identici din punct de vedere al tuturor caracteristicilor acestora, în timp ce termenul de politetic desemnează o anumită clasă ai cărei membrii sunt similari dar nu identici.

- [Lienhart 99a] R. Lienhart. *Comparison of Automatic Shot Boundary Detection Algorithms*. SPIE Storage and Retrieval for Still Image and Video Databases VII, vol. 3656, pag. 290–301, 1999.
- [Lienhart 99b] R. Lienhart, S. Pfeiffer & W. Effelsberg. *Scene Determination Based on Video and Audio Features*. IEEE International Conference on Multimedia, Computing and Systems, vol. 1, pag. 685–690, iunie, Florence-Italy 1999.
- [Lienhart 00] R. Lienhart. *Dynamic Video Summarization of Home Video*. SPIE Storage and Retrieval for Media Databases, vol. 3972, pag. 378–389, ianuarie 2000.
- [Lienhart 01a] R. Lienhart. *Reliable Dissolve Detection*. SPIE Storage and Retrieval for Media Databases, vol. 4315, pag. 219–230, ianuarie 2001.
- [Lienhart 01b] R. Lienhart. *Reliable Transition Detection in Videos: A Survey and Practitioner's Guide*. MRL, Intel Corporation, http://www.lienhart.de/Publications/IJIG_AUG2001.pdf, august, Santa Clara, USA 2001.
- [Lim 01] S.H. Lim & A. El Gamal. *Optical Flow Estimation Using High Frame Rate Sequences*. IEEE International Conference on Image Processing, vol. 2, pag. 925–928, octombrie 2001.
- [Lin 98] C.-W. Lin, Y.-J. Chang & Y.-C. Chen. *Hierarchical Motion Estimation Algorithm Based on Pyramidal Successive Elimination*. International Computer Symposium, octombrie 1998.
- [Lin 02] W.-H. Lin & A.G. Hauptmann. *News Video Classification Using SVM-Based Multimodal Classifiers and Combination Strategies*. ACM Multimedia, pag. 323–326, Juan-les-Pins, France 2002.
- [Liu 02a] C.-C. Liu & A.L.P. Chen. *3D-List: A Data Structure for Efficient Video Query Processing*. IEEE Transactions on Knowledge and Data Engineering, vol. 14, nr. 1, pag. 106–122, ianuarie-februarie 2002.

- [Liu 02b] T. Liu & J.R. Kender. *An Efficient Error-Minimizing Algorithm for Variable-Rate Temporal Video Sampling*. IEEE International Conference on Multimedia and Expo, vol. 1, pag. 413–416, 2002.
- [Liu 02c] T. Liu & J.R. Kender. *Optimization Algorithms for the Selection of Key Frames Sequences of Variable Length*. European Conference on Computer Vision, vol. 2353, pag. 403–417, London, UK 2002.
- [Liu 03] T. Liu, H.-J. Zhang & F. Qi. *A Novel Video Key-Frame Extraction Algorithm Based on Perceived Motion Energy Model*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, nr. 10, pag. 1006–1013, ottobre 2003.
- [Liu 04] T. Liu, X. Zhang, J. Freg & K. Lo. *Shot Reconstruction Degree: A Novel Criterion for Keyframe Selection*. Pattern Recognition Letter, vol. 25, nr. 12, pag. 1451–1457, settembre 2004.
- [Lu 03] S. Lu, I. King & M. Lyu. *Video Summarization Using Greedy Method in a Constraint Satisfaction Framework*. 9th International Conference on Distributed Multimedia Systems, pag. 456–461, Miami, Florida, USA 2003.
- [Lundmark 01] A. Lundmark. *Non-Redundant Search Patterns in Log-Search Motion Estimation*. Swedish Society for Automated Image Analysis Symposium - SSAB, 2001.
- [Lupatini 98] G. Lupatini, C. Saraceno & R. Leonardi. *Scene Break Detection: A Comparison*. Research Issues in Data Engineering, Workshop on Continuous Media Databases and Applications, pag. 34–41, Orlando, FL, USA 1998.
- [Ma 01] Y.F. Ma, J. Sheng, Y. Chen & H.J. Zhang. *MSR-Asia at TREC-10 Video Track: Shot Boundary Detection Task*. 10th Text Retrieval Conference, pag. 371, 2001.
- [MacQueen 67] J.B. MacQueen. *Some Methods for Classification and Analysis of Multivariate Observations*. 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, vol. 1, pag. 281–297, 1967.

- [Maillet 03] S.M. Maillet. *Content-Based Video Retrieval: An Overview*. <http://viper.unige.ch/~marchand/CBVR/>, 2003.
- [Marichal 98] X. Marichal. *Motion Estimation and Compensation for Very Low Bitrate Video Coding*. Teză de doctorat, UCL - Université Catholique de Louvain, Laboratoire de Télécommunications et Télédétection, Louvain-la-Neuve, Belgique 1998.
- [Mazière 00] M. Mazière, F. Chassaing, L. Garrido & P. Salembier. *Segmentation and Tracking of Video Objects for a Content-Based Video Indexing Context*. IEEE International Conference on Multimedia Computing and Systems, vol. 2, pag. 1191–1194, New York, USA 2000.
- [Mehtre 97] B.M. Mehtre, M.S. Kankanhalli & W.F. Lee. *Shape Measures for Content Based Image Retrieval: A Comparison*. Information Processing and Management, vol. 33, nr. 3, pag. 319–337, mai, 1997.
- [Meng 95] J. Meng, Y. Juan & S.F. Chang. *Scene Change Detection in a MPEG Compressed Video Sequence*. SPIE Symposium, vol. 2419, pag. 14–25, februarie 1995.
- [Miene 01] A. Miene, A. Dammeyer, T. Hermes & O. Herzog. *Advanced and Adaptive Shot Boundary Detection*. ECDL WS Generalized Documents, pag. 39–43, 2001.
- [Miura 03] K. Miura, R. Hamada, I. Ide, S. Sakai & H. Tanaka. *Motion Based Automatic Abstraction of Cooking Videos*. IPSJ Transactions on Computer Vision and Image Media, vol. 44, 2003.
- [Mojsilovic 00] A. Mojsilovic, J. Kovacevic, R.J. Safranek J. Hu & S.K. Ganapathy. *Matching and Retrieval Based on the Vocabulary and Grammar of Color Patterns*. IEEE Transactions on Image Processing, vol. 9, nr. 1, pag. 38–54, 2000.
- [Morphing 08] Morphing. *Introduction to Media Computation*. <http://coweb.cc.gatech.edu/mediaComp-plan/65>, 2008.