Computer Vision

Cristian Dima — mailto:cdima+@cs.cmu.edu
Mihai Budiu — mailto:mihaib+@cs.cmu.edu

mai 2000

Subiect: O introducere în Computer Vision

Cunoştinţe necesare: nici una

Cuvinte cheie: computer vision, imagine, segmentare, urmărire (tracking), eşantionare, vector

Cuprins

În textul nostru despre "cercetarea în robotică", din numărul curent al PC Report, arătăm că *percepţia* este unul din elementele cheie ale roboticii. În acest text explorăm o ramură a percepţiei robotice, şi anume "vederea automată", în engleză *computer vision* (vom abrevia de aici înainte cu CV această sintagmă).

1 Introducere

Dintre toate domeniile în care un robot, sau un calculator, poate fi comparat cu performanţele umane, percepţia (şi componenta sa cea mai importantă, văzul) este probabil mai puţin satisfăcătoare prin calitatea rezultatelor. De ce stau lucrurile aşa? Pentru că, chiar dacă suntem în stare să construim senzori mult mai sensibili decît organele omeneşti de simţ, nu putem egala resursele computaţionale dedicate de creier acestei activităţi. În mod paradoxal, progresele din CV nu fac decît să sporească admiraţia pentru modul în care creierul uman procesează imaginile, de la fizica formării imaginii, pînă la sistemele de prelucrare şi analiză, precum şi pentru adaptabilitatea remarcabilă, la condiţii de mediu extrem de schimbătoare. Din păcate, foarte multe lucruri ne sunt necunoscute despre felul în care creierul procesează informaţiile vizuale, aşa că nu putem "copia" soluţia sa.

Datorită vastităţii domeniului, chiar şi o simplă încercare de a enumera temele de cercetare ar fi sortită eşecului. Importanţa percepţiei şi dificultăţile legate de obţinerea unor rezultate semnificative fac ca în fiecare subdomeniu să existe o activitate febrilă ce cercetare şi sumedenie de rezultate. Fiecare metodă propusă în literatură oferă adesea rezultate bune, dar pe o problemă precis delimitată. Limitările provin uneori din modele matematice simplificate al fenomenelor; pentru fiecare metodă putem imagina un context în care nu funcţionează satisfăcător.

Nici măcar cărţi de sute de pagini dedicate în mod special CV nu reuşesc sa trateze decît anumite zone ale domeniului; nu e deci surprinzător că scurta noastră introducere va fi incompletă şi relativ superficială: vom prezenta doar *genul* de probleme existente şi vom încerca să rezumăm toată cercetarea din domeniu în cîteva pagini. Vom încerca să prezentăm acele teme din Computer Vision care sunt legate strîns de aplicaţiile robotice. Rugăm cititorul să reţină că acestea sunt adeseori interdependente, sau dependente de alte domenii, ce nu vor fi menţionate.

2 Teme ale CV în robotică

Robotica este unul din domeniile în care sistemele de viziune sunt foarte des folosite. Un robot care asamblează piese trebuie să le poată identifica; are nevoie deci de mecanisme de *percepţie* şi de un model despre modul în care îşi *formează imaginile*, pentru a înţelege ce sunt informaţiile primite.

Un robot ce navighează pe un teren unde există iarbă şi stînci trebuie să fie capabil sa distingă între zonele cu iarbă şi obstacole. Tehnica folosită pentru a distinge obiecte se numeşte *segmentare de imagine*.

Roboţii se mişcă într-un spaţiu tridimensional; este deci nevoie de a extrage informaţie tridimensională; în cazul acesta se poate apela la *viziunea stereoscopică*, care utilizează imagini ale aceeaşi scene luate din puncte diferite.

Foarte frecvent roboţii combină mai mulţi senzori pentru a obţine informaţie despre mediul înconjurător, sau "baleiază" mediul cu senzorii, culegînd informaţii pe suprafeţe mari. În acest caz, este nevoie ca datele provenite de la dispozitive diferite (fie de la roboţi diferiţi, fie de la acelaşi senzor la momente de timp diferite) să fie aliniate între ele, *reconstituind din mozaicul de imagini* un model unic al realităţii.

Identificarea obiectelor de interes şi clasificarea lor se poate face cu tehnici folosite şi pentru, manipularea şi căutarea de informaţii în *baze de date multimedia*; vom vedea unele tehnologii din acest domeniu la lucru în acest text.

Adesea nu este suficient ca un obiect să poată fi identificat în mod static; robotul trebuie să fie capabil să urmărească obiectul în mişcare (motion tracking).

În fine, unii roboţi trebuie să fie educaţi pentru a învăţa ce anume prezintă interes din mediu; putem deci vorbi şi despre *învăţare* în contextul CV.

Este de reţinut faptul că diferitele domenii ale CV sunt întrepătrunse. Un simplu exemplu: pentru a recunoaşte obiecte, este uneori nevoie de a separa obiectul de interes de fundalul imaginii. Dar acest procedeu cade în sfera segmentării imaginilor, şi deci toate problemele şi dificultăţile existente în domeniul segmentării apar în mod indirect în anumite metode de recunoaştere. Deşi nu legate în mod direct de vederea utilizată în domeniul roboticii, aproape toate domeniile de cercetare din CV se dovedesc astfel utile în anumite aplicaţii robotice.

3 Formarea imaginilor

În această diviziune a CV intră tot ceea ce are legatură cu procedeele de obţinere a imaginilor. Există trei aspecte principale ale formării imaginilor: geometria, fotometria şi captarea a imaginii.

3.1 Geometria formării imaginii

Geometria formării imaginii descrie modelul procedeului de formare a imaginii în dispozitivul optic utilizat.

Modelul cel mai frecvent utilizat este cel al camerei obscure, în care imaginea se formează prin trecerea razelor luminoase printr-un orificiu infinitezimal ("centrul de proiecţie") şi proiectarea lor pe un plan. Într-un aparat tradiţional, acest plan este un material fotosensibil; în camerele digitale de astăzi planul este compus din elemente electronice sensibile la radiaţia din spectrul de interes (vizibil, infraroşu, etc.). Transformarea geometrică în modelul camerei obscure este intuitivă şi uşor de exprimat matematic; tehnologia este cunoscută din epoca renascentistă şi poartă numele de *proiecţia perspectivă*. Marea problemă a acestei transformări este însă faptul că relaţia dintre coordonatele punctelor din imagine şi coordonatele tridimensionale ale lumii fotografiate este inerent neliniară[1].

Din această cauză, adesea se foloseşte un alt tip de proiecţie, cea *ortografică*, în care un obiect în spaţiul 3D este întîi proiectat perpendicular pe planul imaginii, ignorînd *profunzimea* (adică un punct (x, y, z) în 3 dimensiuni devine un punct (x, y) în planul imaginii). Apoi această proiecţie este redusă la scară, în funcţie de distanţa medie a obiectului faţă de cameră.

[1]Un punct cu coordonatele (X, Y, Z) este proiectat în punctul (x, y), unde $x = f \cdot X/Z$, unde f este distanţa focală; această relaţie nu este o funcţie lineară în Z.

Modelul acesta, care foloseşte o proiecţie ortografică urmată de o reducere la scară este o aproximare acceptabilă a modelului din perspectivă doar în cazurile în care obiectul este aproape de axa optică a camerei şi relativ mic, în comparaţie cu distanţa dintre obiect şi aparatul optic.

O altă transformare lineară, mai puţin "artificială" decît cea ortografică, este *proiecţia paralelă*. Şi în acest caz obiectul este proiectat pe planul imaginii şi apoi redus la scară, însă, de data aceasta, proiecţia este realizată prin raze paralele orientate în direcţia "medie" a razelor ce ar alcătui proiecţia prin perspectivă (dacă preferaţi, de-a lungul unei raze din centrul camerei în centrul obiectului). Ca atare, această metodă dă rezultate bune chiar dacă obiectul nu este aproape de axa optică. În cazul în care obiectul este chiar pe axa optică, proiecţia paralelă şi cea ortografică sunt echivalente. Cele trei proiecţii sunt ilustrate în figura 1.

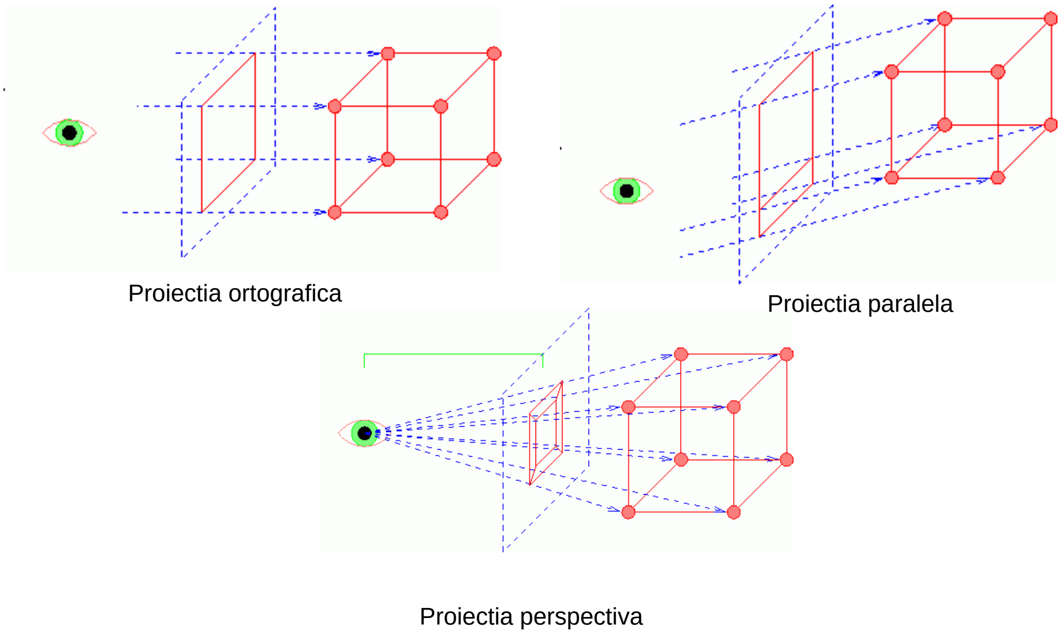

Figura 1: Cele trei tipuri de proiecţie folosite în modelarea aparatelor care captează imaginea.

O întrebare legitimă este de ce modelul camerei obscure e aproape universal acceptat, dar toate camerele de filmat/fotografiat în uz folosesc de fapt un sistem cu lentile. Iată de ce: pentru a obţine o imagine cît mai clară cu o cameră obscură, orificiul prin care razele de lumină sunt proiectate ar trebui să fie cît mai mic. Pe masură ce orificiul devine mai mic, fenomenul de difracţie devine din ce în ce mai semnificativ, rezultînd într-o imagine neclară. (Pe lîngă asta, energia semnalului optic care trece printr-o găurică mică este foarte redusă.) Scopul lentilelor este de a menţine geometria camerei obscure, în timp ce dimensiunile orificiului sunt suficient de mari ca difracţia să fie nesemnificativă. Lentilele prezintă şi ele o serie întreagă de alte probleme: aberaţii cromatice, radiale, distorsiuni şi alte efecte nedorite, însă constituie un substanţial avantaj faţă de camera obscură.

3.2 Fotometria

Fotometria se ocupă cu măsurătoarea luminii. Fotometria este foarte importantă pentru CV, datorită dificultăţii stabilirii unei relaţii cantitative între caracteristicile fizice ale unui obiect, condiţiile de iluminare şi imaginea obţinută. Intensitatea unui pixel dintr-o imagine depinde de cantitatea de lumină reflectată în direcţia elementului fotosensibil; cantitatea de lumină reflectată însă depinde atît de natura suprafeţei reflectante cît şi de direcţia şi intensitatea surselor de iluminaţie, cum ilustrează şi figura 2 (în măsura în care tipografia poate reproduce diferenţele).

Figura 2: Acelaşi obiect fotografiat cu trei lumini ambiante diferite. Care e culoarea adevărată?

Să presupunem că dorim să implementăm un sistem care numără bilele roşii din cîmpul vizual al unei camere de luat vederi. Cum putem deosebi o bilă roşie sub lumină albă de o bilă albă luminată cu un bec roşu? Dacă nu ştim nimic despre culoarea sursei de lumină sau a suprafeţei, nu le putem deosebi! Acesta este un exemplu de *indeterminare*: putem afla culoarea sursei de lumină dacă ştim culoarea bilei, şi putem afla culoarea bilei dacă ştim culoarea sursei. E însă imposibil să extragem informaţii despre ambele dintr-o singură imagine.

Numărătorul de bile roşii este un proiect foarte simplu, şi totuşi o soluţie a problemei nu este evidentă. Ce putem face în cazuri mult mai complicate, cum ar fi recunoaşterea feţelor umane, detectarea obstacolelor în aer liber — caz în care un nor poate schimba drastic condiţiile de iluminare — sau în alte aplicaţii în care sursele de lumină sunt incontrolabile sau în schimbare? Problema inconsistenţei culorilor în timp este extrem de dificilă. Deşi au existat, şi există în continuare, cercetători care studiază modalităţi de creştere a robusteţii sistemelor bazate pe analiza culorilor, se pare că ideea de a utiliza culoarea pentru a obţine mai multă informaţie decît cea din imagini alb-negru este nepopulară.

3.3 Captarea imaginii

Captarea imaginii nu este probabil un domeniu de cercetare academică, însă face parte din cunoştinţele obligatorii pentru cei care lucrează în CV.

Putem privi o imagine ca fiind o funcţie, definită în două dimensiuni (cele două axe ale imaginii) a cărei valoare este culoarea (sau intensitatea luminii, în cazul imaginilor alb-negru).

În această accepţiune, atît domeniul cît şi codomeniul funcţiei sunt continue. Pentru că fiecare element fotosensibil are dimensiuni finite, în procesul de captare a imaginii domeniul este discretizat într-o grilă; fiecare celulă a grilei corespunde unui element fotosensibil (pixel). Această transformare din spaţiul continuu în cel discret se numeşte *eşantionare* (sampling).

În plus, valoarea culorii sau a intensităţii luminii este de asemenea într-un spaţiu continuu; aparatele vor discerne însă un număr finit de valori diferite. Aceasta este *cuantizarea* valorilor. De exemplu aparatele de filmat alb-negru se limitează adesea la 256 nivele de gri. O bună înţelegere a limitărilor fizice şi teoretice ale senzorilor este un element cheie în adaptarea algoritmilor pentru diferitele aplicaţii.

4 Comparaţia imaginilor

Una dintre cele mai elementare operaţii pe care le putem face cu imagini este de a întreba "cît de asemănătoare sunt aceste două imagini"? Dacă putem răspunde în mod rezonabil la o astfel de întrebare, deschidem posibilitatea foarte multor aplicaţii: căutare de imagini pe Internet sau în biblioteci digitale sau recunoaşterea obiectelor (prin compararea cu o imagine-prototip), care la rîndul lor permit apoi aplicaţii mult mai sofisticate.

Cu toate că întrebarea asta pare simplă, chiar şi pentru un om răspunsul nu este întotdeauna evident. De fapt, ca să putem răspunde, trebuie să clarificăm care este măsura noastră de "apropiere" sau "distanţă" între două imagini. În plus, aplicaţii diferite vor avea probabil noţiuni diferite de "aproape": de exemplu un botanist va considera două imagini care conţin o tufă cu gogoşari asemănătoare, chiar dacă unul din gogoşari este într-un ghiveci în interior iar celălalt undeva pe cîmp, pe cînd un specialist în decoraţiuni interioare va socoti cele două imagini ca fiind foarte diferite (şi probabil va fi oripilat de ideea de a avea pe post de ornament o tufă cu gogoşari).

Din fericire algebra ne pune la dispoziţie nişte unelte foarte simple cu care putem rezolva în mod foarte eficace măcar o parte din problemă. Iată soluţia folosită în practică, ilustrată în figura 3:

- Pentru fiecare imagine "măsurăm" într-un anume fel o serie de parametri care ni se par importanţi pentru clasificare. Pentru fiecare astfel de parametru obţinem o valoare.

  Iată cîteva exemple de parametri:

  - mărimea imaginii în pixeli;
  - numărul de culori sau nuanţe de gri diferite;
  - dimensiunea celei mai mari "pete" de culoare mată din imagine;
  - numărul de puncte dintr-o anumită culoare;
  - numărul mediu de puncte consecutive dintr-o anumită culoare pe o direcţie dată;
  - media culorilor din imagine;
  - intensitatea maximă şi minimă a punctelor din imagine.

  Asamblăm pentru fiecare imagine colecţia aceasta de valori într-un *vector*.

  Dacă avem n valori diferite pentru o imagine, obţinem un vector într-un spaţiu n-dimensional. În caz extrem, putem vedea întreaga imagine ca pe un vector într-un spaţiu $x \times y$-dimensional, unde x şi y sunt dimensiunile în pixeli, iar valoarea în fiecare punct din imagine este intensitatea culorii din acel punct.

  În general această soluţie extremă nu este practică, din două motive:

  - Vrem ca toate imaginile cu care operăm să fie reprezentate de vectori cu acelaşi număr de dimensiuni, pentru că aceasta este crucial pentru a le compara[2]

[2] Această problemă poate fi rezolvată re-eşantionînd toate imaginile la o aceeaşi rezoluţie.

  - Numărul de pixeli dintr-o imagine este de obicei mult prea mare, comparat cu numărul de trăsături care ne interesează. Adesea vrem să *comprimăm* descrierea imaginii într-un număr relativ mic de trăsături (de la cîteva la cîteva sute).

- Odată ce avem o descriere a unei imagini printr-un vector, putem folosi două metrici simple pentru a compara vectorii:

  - Putem calcula distanţa dintre doi vectori;
  - Putem calcula unghiul plan dintre doi vectori.

  A doua metodă este recomandabilă, pentru că este insenzitivă la scalarea vectorului (de exemplu, pentru parametrii exemplificaţi mai sus, unghiul dintre doi vectori nu se schimbă dacă dublăm dimensiunea uneia dintre ele). Ambele metrici se pot calcula foarte uşor: prima necesită n scăderi, iar a două n înmulţiri şi adunări plus un radical (lăsăm detaliile pe seama unui curs introductiv de algebră lineară).

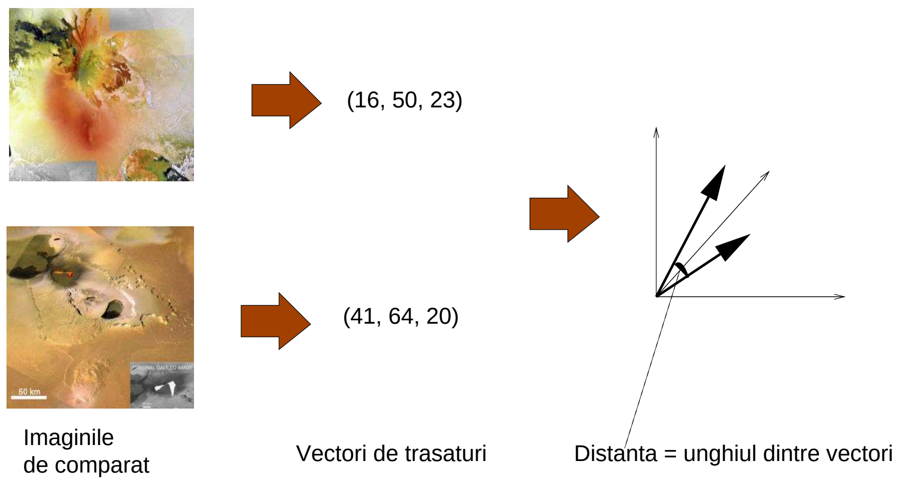

Figura 3: Pentru a compara două imagini măsurăm unele trăsături ale lor; valorile trăsăturilor formează în vectori într-un spaţiu n-dimensional (3-dimensional în această figură). “Distanţa” dintre două imagini este dată de unghiul dintre vectori. Aceste imagini înfăţişează satelitul lui Jupiter, Io, din misiunea spaţială a sondei spaţiale Galileo, de la NASA.

Desigur, într-un fel am ascuns praful sub covor: am transformat problema comparării de imagini în problema calculului de trăsături “interesante”. Dar acesta este un progres substanţial; în mod surprinzător, foarte multe din trăsăturile relativ simple din exemplele de mai sus se dovedesc foarte eficace în practică pentru a compara imagini.

5 Baze de date multimedia

Deşi nu sunt un subiect central în CV în robotică, vom trece în revistă tehnicile de căutare a imaginilor în bazele de date multi-media.

Distincţia între între procesul de căutare a unei imagini după conţinut şi cel de recunoaştere a imaginii este destul de fină. Numeroase sisteme de recunoaştere pot fi folosite pentru a căuta imagini într-o bază de date. Distincţia constă în faptul că, în cazul bazelor de date, imaginile sunt disponibile *înainte* de a începe căutarea deci ele pot fi pre-procesate. Putem astfel crea structuri de date off-line cu algoritmi prea costisitori pentru recunoaşterea interactivă a obiectelor.

Să luăm pentru exemplu un sistem care sumarizează filme video; dorim să folosim acest sistem pentru a vedea numai secvenţele în care apare un anumit actor. Dacă sistemul poate înregistra filmul pentru prelucrare, tehnicile folosite sunt foarte diferite decît în cazul în care trebuie să recunoască insul atunci cînd apare la televizor.

Există două tipuri de căutări: cele care cer de la utilizator o imagine-exemplu, şi cele care pornesc de la o interogare sub formă de cuvinte. Asemenea tipuri de sisteme, deşi departe de a fi perfecte, există deja în operaţie pe Internet. Pentru prima categorie găsim ilustrativ sistemul dezvoltat la INRIA (Institut National de Recherche en Informatique et Automatique), în Franţa: `http://www-rocq.inria.fr/cgi-bin/imedia/surfimage.cgi`.

Sistemele din a doua categorie sunt mai puţin robuste; de exemplu cel dezvoltat la C&C Research Laboratories, NEC USA, `http://www.ccrl.com/amore/`. Rezultatul unei căutări executate de noi este prezentat în figura 4.

5.1 Recunoaşterea obiectelor

Pentru a recunoaşte obiecte, se porneşte de la un set de modele ale obiectelor; cînd o imagine este prezentată sistemului pentru analiză, modelele disponibile sunt folosite pentru a determina care sunt obiectele din imagine. Problemele de rezolvat sunt însă foarte dificile: pentru a recunoaşte un obiect într-o imagine, ar trebui să ştim cum arată văzut din orice poziţie, şi cu orice sursă de iluminare. În practică această cerinţă este imposibil de satisfăcut. În cazul în care avem un model complet al obiectului, algoritmii de recunoaştere încearcă să determine care este poziţia şi iluminarea cea mai plauzibilă pentru a "explica" imaginea.

6 Segmentare

O alta problemă fundamentală în CV este cea de a separa într-o imagine diferitele obiecte care apar. Dificultatea acestei probleme variază enorm în funcţie de imaginea cu pricina; cu cît iluminaţia este mai uniformă, obiectele au culori mai contrastante şi imaginea conţine mai puţin zgomot, cu atît problema segmentării este mai simplă.

Există foarte multe metode pentru a segmenta imagini. Vom indica aici schematic una dintre ele, care "creşte" regiunile în jurul unor nuclee:

- Printr-o metodă oarecare, marcăm "centrele" obiectelor principale din imagine. Putem face asta fie cu informaţii de la utilizator, fie alegînd punctele de contrast maxim, fie alte informaţii, depinzînd de aplicaţie.

- Calculăm apoi "distanţa" între fiecare două puncte vecine în imagine (putem folosi fie patru fie opt vecini) astfel: cu cît culorile sunt mai diferite, cu atît distanţa este mai mare.

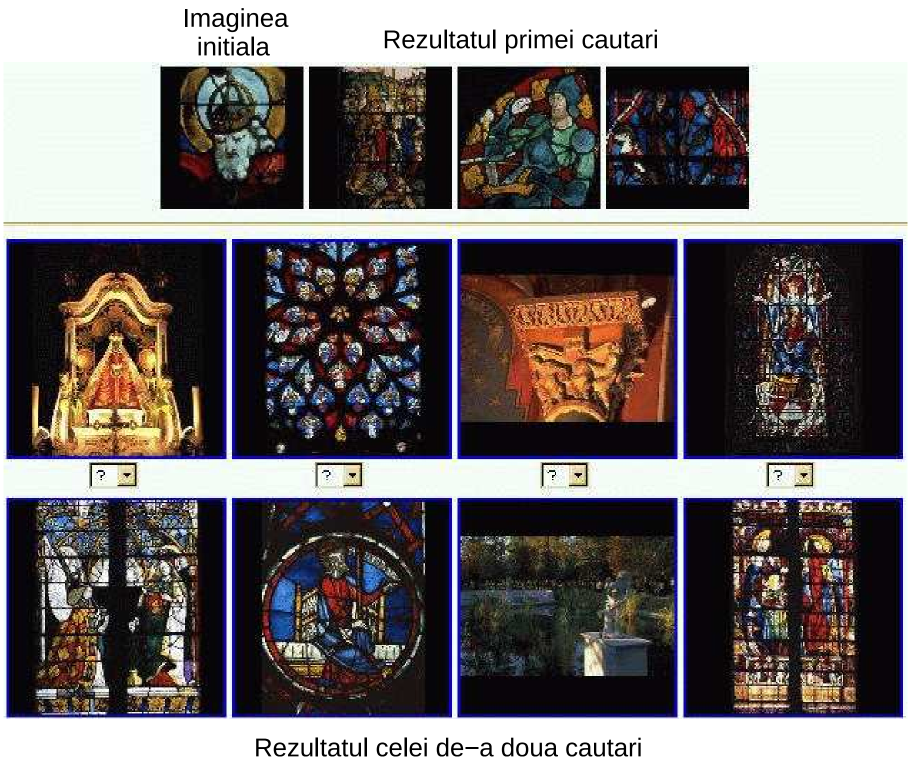

Figura 4: Căutarea în baze de date de imagini. Acestea sunt rezultatele obţinute pornind de la imaginea din stînga-sus, folosind sistemul dezvoltat la INRIA. Pornind de la o imagine, sistemul returnează o serie imagini pe care le "crede" similare. Utilizatorul le selecţionează pe acelea care sunt reprezentative. După prima iteraţie, algoritmul a găsit cele trei imagini aflate la dreapta celei iniţiale şi alte cinci imagini "incorecte", înfăţişînd diferite picturi. Cum eram în căutare de vitralii, le-am selecţionat pe cele trei şi am lansat o nouă căutare. De data aceasta, algoritmul a găsit cinci noi imagini reprezentînd vitralii.

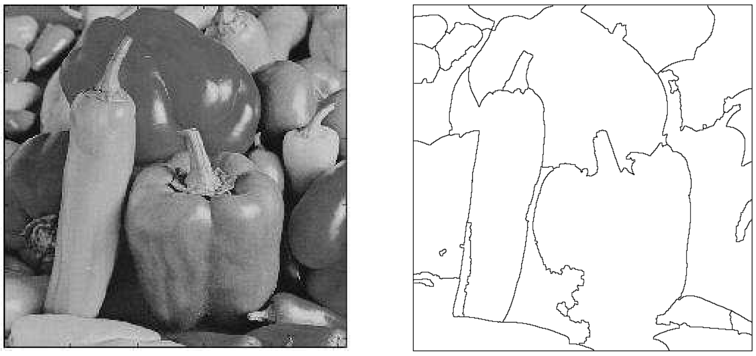

Figura 5: *Segmentarea de imagini* încearcă să identifice obiectele care apar într-o imagine (adică să atribuie fiecare pixel unui obiect). Acest exemplu arată un original şi rezultatul segmentării, cu conturul trasat în jurul diferitelor obiecte identificate.

- Rulăm apoi un proces iterativ de calcul în care fiecare punct este atribuit în aceeaşi regiune cu cel mai apropiat vecin. Acest proces se poate efectua foarte eficace pe o maşină cu mai multe procesoare.

- Algoritmul se termină cînd fiecare punct este atribuit uneia dintre regiunile iniţiale.

Imaginea 5 arată un exemplu de imagine segmentată automat pornind de la cîteva puncte etichetate manual.

7 Mozaicuri şi reconstrucţie

Problema mozaicurilor este de a îmbina a mai multe imagini parţiale ale unei scene într-o singură imagine cît mai cuprinzătoare. Una dintre aplicaţiile acestei tehnici este alcătuirea unor hărţi detaliate din imagini aeriene. În acest caz "lipirea" imaginilor este relativ simplă, întrucît mişcarea este rectilinie. Mult mai spectaculos şi mai dificil este să folosim o secvenţă video pentru a construi o imagine de ansamblu, ca în figura 6.

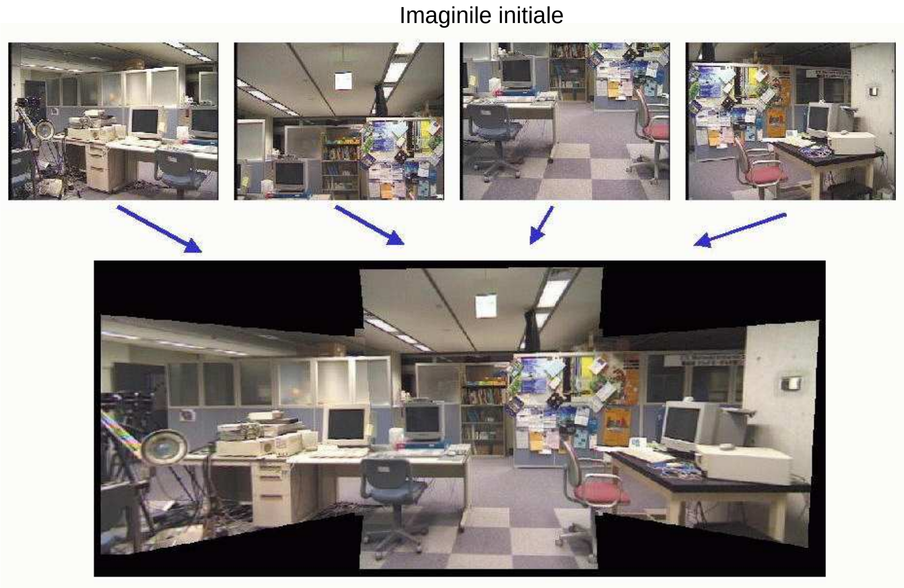

Figura 6: Reconstrucţia unei scene din imagini parţiale. Imagine obţinută la Waseda University, School of Science and Engineering, Tokyo, Japonia.

Dificultatea majoră a alcătuirii mozaicurilor constă în selecţionarea regiunilor din fiecare imagine care trebuie folosite în imaginea panoramică, şi mai ales în calculul deformărilor aplicate fiecărei regiuni pentru ca ele să se îmbine corect. Gradul de dificultate depinde de mişcarea pe care camera de luat vederi o întreprinde. În cazul din figura 6, camera a efectuat două rotaţii, în plan orizontal şi vertical, de unde imaginea panoramică în forma de cruce. În acest caz algoritmii urmăresc "mişcarea" pixelilor de la o imagine la alta, pentru a calcula poziţia fiecăreia dintre imagini în cea finală.

Mai recent s-au elaborat algoritmi eficienţi pentru cazuri mult mai dificile, în care secvenţele video conţin efecte “zoom” sau mişcare înainte–înapoi. Algoritmii fac o proiecţie pe o “conductă” virtuală, care permite alipirea imaginilor cu rezultate spectaculoase.

O aplicaţie originală a imaginilor-mozaic este utilizarea lor pentru ghidarea vehiculelor robotice în medii controlate. Un proiect al Institutul de Robotica al Universităţii Carnegi Mellon utilizează mozaicuri imense, formate din imagini ale podelei unei hale, pentru a permite localizarea în interior a unei echipe de elevatoare robotizate. Un vehicul de recunoaştere trece în prealabil deasupra întregii suprafeţe a halei, înregistrînd imagini şi “lipindu-le”. În timpul funcţionării, celelalte elevatoare pleacă de la puncte fixe, şi compară încontinuu imaginea de la propria cameră cu imaginea panoramică obţinută în prealabil. În felul acesta se poate obţine o localizare cu o precizie de cîţiva milimetri.

8 Mişcare, urmărire şi analiza secvenţelor de imagini

În numeroase situaţii informaţia vizuală este captată în mod continuu, în intenţia de a detecta schimbările care se petrec. Un exemplu clasic provine din aplicaţiile de supraveghere (surveillance). Dacă avem o incintă cu un sistem de camere video a căror orientare poate fi controlată de la distanţă, vrem ca în cazul în care un intrus este descoperit, camerele să-şi schimbe orientarea astfel ca intrusul să rămînă mereu în cîmpul vizual. În acest caz va trebui să calculăm mişcarea siluetei între fiecare două imagini consecutive.

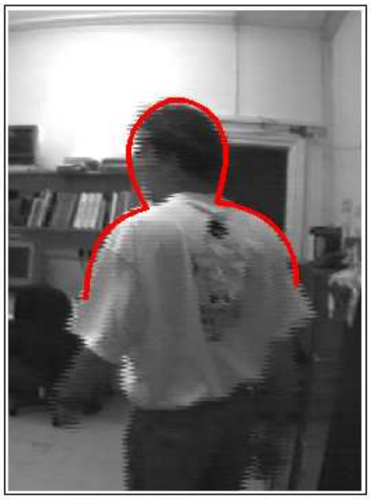

Figura 7: Urmărire automată a unui intrus. Sistemul va detecta şi urmări silueta, astfel încît să rămînă mereu în centrul imaginii.

O altă posibilă aplicaţie este navigarea automobilelor robotizate (vedeţi exemplul NAVLAB în articolul despre cercetarea în robotică), sau monitorizarea automată a traficului. În cadrul unui proiect de cercetare al universităţii Stanford din California, s-au dezvoltat sisteme care permit detectarea şi urmărirea siluetelor automobilelor pe străzi. Urmărind viteza de mişcare a automobilelor în imagine se poate evalua viteza lor reală; se pot detecta în felul acesta fie excesele de viteza, fie blocajele de circulaţie. În cazul automobilelor robotizate, urmărirea contururilor celorlalte maşini din cîmpul vizual este folosită pentru a evita coliziunile.

Exista numeroşi algoritmi pentru analiza mişcării şi urmărirea contururilor. Poate cei mai simpli şi mai răspîndiţi sunt cei bazaţi pe analizarea *fluxului optic* (optical flow). O metodă de a determina deplasarea fiecărui pixel între două imagini succesive, pornind de la presupunerea că în imagine culorile variază gradual. Să presupunem că imaginea este un gri din ce în ce mai închis, de la stînga la dreapta. Pentru a determina mişcarea, comparăm culoarea unui punct din imaginea originală cu a unui punct din cea finală; dacă punctul în final e mai închis la culoare, obiectul s-a deplasat spre stînga. Pentru fiecare punct din imagine se obţine astfel un vector care indică direcţia şi deplasarea punctului pentru a ajunge la poziţia sa din imaginea succesivă. În felul acesta, pentru a urmări un contur de-a lungul unei secvenţe video, putem aplica aceiaşi vectori de deplasare fiecărui punct de pe contur.

9 Stereoscopie

Dacă există o tehnologie consacrată în CV, atunci e vorba de stereoscopie. Stereoscopia este procedeul prin care se poate obţine informaţie tridimensională pornind de la două sau mai multe imagini. Principiul este cel pe care se bazează percepţia umană. Faceţi următorul experiment: ţineţi pentru cîteva secunde un deget la cîţiva centimetri de ochi. Închideţi şi deschideţi ochii pe rînd: veţi observa cum degetul "sare", datorită poziţiei diferite faţă de "dispozitivul" folosit pentru obţinerea imaginii. Depărtaţi apoi degetul la distanţa maximă posibilă şi încercaţi din nou; de data aceasta, deplasarea va fi considerabil mai mică. Distanţa dintre poziţiile degetului în cele doua imagini (cea de la ochiul stîng şi cea de la cel drept) se numeşte *disparitate*. Ea este proporţională cu distanţa dintre ochi şi invers proporţională cu distanţa între deget şi ochi.

Figura 8 explică procedeul folosit de algoritmii pentru procesarea imaginilor stereoscopice. Dacă reuşim să identificăm punctele corespunzătoare din cele două imagini putem calcula disparitatea d. Din Formula $d = f \cdot B/z$ deducem distanţa z pînă la obiect (f fiind distanţa focală, şi B distanţa dintre cele două obiective). Dacă putem calcula z pentru fiecare punct din imagine, înseamnă că putem afla distanţele pînă la toate punctele vizibile în ambele imagini.

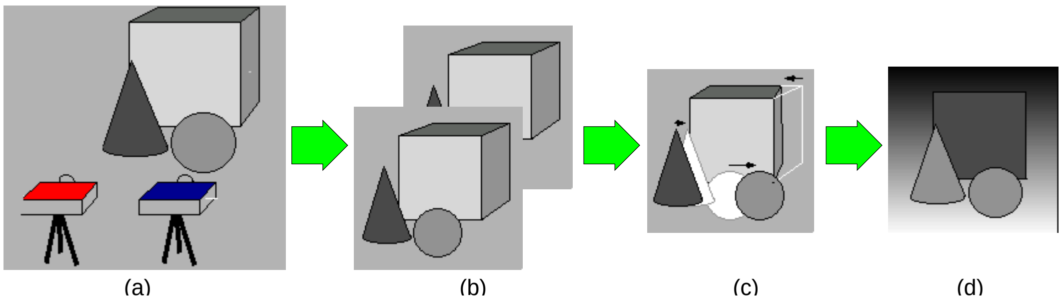

Figura 8: O scenă este înregistrată din poziţii diferite (a). Rezultatul constă în două imagini (b). Algoritmul stabileşte corespondenţe între pixelii celor două imagini care provin din acelaşi punct de pe obiect (c). "Harta de profunzime" (d) prezintă în mod sintetic distanţele calculate: albul corespunde obiectelor apropiate de cameră, negrul celor îndepărtate.

Aplicaţiile stereoscopiei sunt nenumărate: practic orice sistem care trebuie să măsoare distanţe rapid, fără prea mare precizie, poate folosi stereoscopia. Calcularea precisă a unui model 3D pornind de la un sistem stereoscopic este însă dificilă, cerînd o calibrare (calculul parametrilor sistemului, f şi B) foarte precisă. Preţul scăzut al sistemelor stereoscopice şi relativa simplitate a algoritmilor le face foarte atractive. Figura 9 prezintă stereoscopia aplicată folosind imagini aeriene.

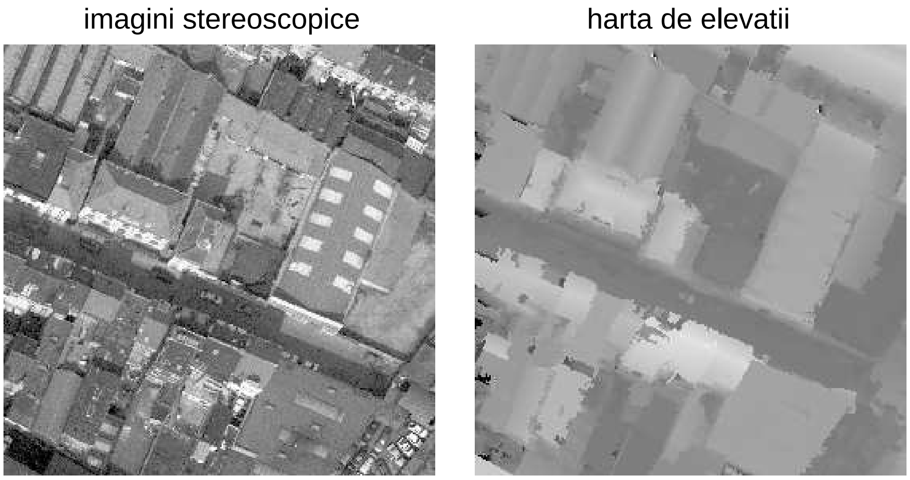

Figura 9: Pornind de la două camere de luat vederi montate pe un avion şi un sistem stereoscopic precis calibrat, se pot obţine hărţi de elevaţie; din acesta se poate apoi genera un model tridimensional aproximativ.

10 Învăţare în Computer Vision

Termenul de "învăţare" este folosit în CV cu multă generozitate. Practic orice algoritm care determină parametri de funcţionare poate fi atribuit acestei categorii. Un exemplu este "antrenarea" unui sistem pentru segmentare de imagini. Unul din autori lucrează la implementarea unui astfel de sistem. Acesta foloseşte o reţea neurală şi un clasificator bayesian. Sistemul este antrenat să separe în imaginile unei livezi regiunile de iarbă, copaci, cer şi alte obiecte. Sistemul va fi folosit de un tractor robotizat pentru detecţia obstacolelor.

Deşi foloseşte aceleaşi date, antrenarea este diferită pentru cele două sisteme. Se foloseşte un set de mici regiuni din imagini "tipice" din livadă, segmentate şi clasificate manual. Reţelei neurale i se prezintă apoi fiecare regiune, şi în funcţie de clasificare (corectă sau incorectă) se ajustează parametrii reţelei prin algoritmul backpropagation. În cazul clasificatorului bayesian, datele de antrenament sunt folosite pentru a extrage informaţii statistice despre distribuţia culorilor în fiecare clasă de obiecte (iarbă, pomi, etc). Mai precis, algoritmul utilizează datele de antrenament pentru a calcula culorile medii şi covarianţa fiecărei clase.

11 Încheiere

Încheiem aici acest mini-tur al Computer Vision. Vrem încă o dată să subliniem vastitatea domeniului, şi faptul că numeroase aspecte importante nu au fost atinse. Nu am discutat, de pildă, despre "alinierea" informaţiei provenind de la diferiţi senzori, despre calibrarea sistemelor stereoscopice, despre extragerea informaţiei 3D dintr-o secvenţă video, despre stabilizarea imaginilor şi multe alte domenii.

Sperînd ca am reuşit să vă trezim interesul, vă recomandăm o excelentă sursă de informaţii în format electronic. Proiectul "CV-online" a fost iniţiat de Catedra de Informatică a Universităţii din Edinburgh; `http://www.dai.ed.ac.uk/CVonline/`. Colaboratorii sunt cercetători de renume, care au contribuit şi contribuie în continuare benevol la cel mai mare depozit de informaţii despre CV disponibil pe Internet.