# Detecţia mâinii într-o scenă încărcată de obiecte, în scopul recunoaşterii gesturilor

Delia Mitrea
Universitatea Tehnică Cluj-Napoca,
Facultatea de Automatică şi Calculatoare,
str. G. Baritiu nr.26-28, Cluj-Napoca
delia.mitrea@cs.utcluj.ro

Sergiu Nedevschi
Universitatea Tehnică Cluj-Napoca,
Facultatea de Automatică si Calculatoare,
str. G. Baritiu nr.26-28, Cluj-Napoca
sergiu.nedevschi@cs.utcluj.ro

Dorian Gorgan
Universitatea Tehnică Cluj-Napoca,
Facultatea de Automatică si Calculatoare
str. G. Baritiu nr.26-28, Cluj-Napoca
dorian.gorgan@cs.utcluj.ro

## REZUMAT

Detecţia vizuală a mâinii, într-o scenă încărcată de obiecte, este o etapă importantă în scopul recunoaşterii gesturilor, în contextul unei interfeţe mutimodale ce implica mijloace de interacţiune naturală cu utilizatorul. Problema este aceea de a distinge, în timp real, obiectul cautat faţa de fond si faţa de celelalte obiecte din scenă, indiferent de poziţia sa şi de condiţiile de iluminare, respectiv de acoperirile cu alte obiecte. Se va adopta o metodă robustă bazată pe caracteristici de textură, culoare şi formă.

## Categorii şi descriptori ai subiectelor

**H.1.2. [User/Machine Systems]** Human Factors, Human Information Processing
**I.4** [**Immage Processing and Computer Vision**]: Segmentation - *edge and feature detection*, *pixel classification*, Feature Measurment - *texture*, Scene Analysis - *shape* , *object recognition*

http://www.acm.org/class/1998/

## Termeni generali

*Documentation*, *Theory*, *Human Factors*, *Design, Algorithms, Performance*

## Cuvinte-cheie

decor multi-modal, scenă naturală, invarianţă vizuală, detecţia mâinii, gesturi

## 1. INTRODUCERE

Scopul acestei lucrări este acela de a dezbate problema detecţiei vizuale a mâinii într-o scenă încărcată de obiecte şi de a propune o metodă robustă pentru soluţionarea acesteia.

Detecţia vizuală a mâinii prezintă o deosebită importanţă în contextul recunoaşterii gesturilor, constituind prima etapă a acestui proces complex care implică, în primul rând, localizarea mâinii, urmând ca abia apoi să se facă demersurile necesare pentru identificarea unor poziţii şi expresii specifice ale mâinii, respectiv urmărirea mişcării acesteia si recunoaşterea traiectoriei parcurse – ceea ce conferă o caracterizare completă a gestului. Identificarea şi localizarea mâinii pe cale vizuală , în comparaţie cu alte modalităţi de localizare a mâinii (senzori, termo-viziune artificială) prezintă următoarele avantaje:

1. metoda este apropiată de percepţia umană, procesul de recunoaştere umană a gesturilor fiind unul pur vizual
2. costul echipamentelor hardware necesare (cameră video, PC ) este mai redus decât în celelalte cazuri.

Detecţia mâinii într-o scenă încărcată de obiecte este o instanţă a problemei generale de detecţie a obiectelor în scenă. Se vor discuta aspectele legate de această problemă, dificultăţile cu care se confruntă literatura internaţională de specialitate în a o soluţiona, apoi se va trece la cazul particular al detecţiei mâinii. Se vor introduce si defini termeni noi, specifici, întâlniţi în literatura de specialitate, cărora li se va găsi un corespondent adecvat în limba română – decor multimodal, invarianţă vizuală, scenă încărcată de obiecte. Se vor trece apoi în revistă metodele existente în literatura de specialitate pentru detecţia obiectelor, respectiv a mâinii şi se vor identifica principalele direcţii ale acestor metode. Metoda propusă va urmări găsirea acelor caracteristici (trăsături) ale obiectelor care sunt invariante la schimbarea condiţiilor precum poziţia obiectului, direcţia de iluminare, distanţa faţă de observator ; de asemenea, se va adopta un algoritm care să furnizeze rezultatul aşteptat în timp real – aceasta din urmă fiind o necesitate imperativă în cazul interfeţelor om-calculator bazate pe interacţiune naturală. Se doreşte şi obţinerea unui anumit grad de generalitate a metodei, pentru a face posibilă utilizarea acesteia atât în cazul unei aplicaţii comandată prin gesturi, cât şi în cazul unui translator al limbajului mimico-gestual.

## 2. DEFINIRE CONCEPTE

**Fundal:** planul, sau suprafaţa generată de mai multe plane, care mărgineşte, în extremitatea opusă observatorului, spaţiul scenei.

**Decor:** obiectele din planul secundar, care diferă faţă de obiectul de interes; termenul originar este acela de *background* -care desemnează atât fundalul, cât şi decorul;

**Decor multi-modal:** termenul originar fiind acela de *multi-modal background* [5] , conceptul se referă la un decor în care se

petrec mişcări repetitive ale obiectelor, precum mişcările repetate ale unei ape sau mişcările crengilor arborilor datorate vântului;

**Scenă**: desemnează spaţiul vizual ce constă din fundal, decor şi obiectul de interes;

**Scenă naturală:** desemnează scena obişnuită a obiectelor într-o situaţie reală, fiind caracterizată prin decor multimodal, având loc atât mişcări periodice sau izolate ale elementelor prezente permanent în scenă, cât şi apariţia şi dispariţia unor obiecte din scenă, respectiv acoperiri între obiectele scenei; termenul de provenienţă este acela de *cluttered scene (background)* [2],[8] – scenă încărcată de obiecte;

**Detecţia obiectelor** implică formularea unui răspuns pozitiv sau negativ privind prezenţa într-o scenă naturală a unui obiect aparţinând unei anumite categorii. Spre deosebire de recunoaştere, care presupune identificarea obiectului căutat – faţa unei anumite persoane, mâna într-o poziţie specifică, un anumit automobil- detecţia obiectelor presupune găsirea unui obiect aparţinând unei clase mai largi – feţe umane, mâini, automobile.

**Invarianţă vizuală:** capacitatea unor elemente (textură, formă, culoare) sau părţi componente ale obiectelor de a nu-şi modifica proprietăţile vizuale în momentul modificării unor condiţii precum schimbarea orientării obiectului relativ la observator, schimbarea direcţiei de iluminare, modificarea distanţei obiectului faţă de observator sau a rezoluţiei imaginii, acoperirile cu alte obiecte; termenul originar este adjectivul *view-invariant* [6]

**Textură:** Se referă la aranjamentul regulat al elementelor vizuale dintr-o anumită regiune a scenei, aranjament ce poate fi caracterizat prin mărimi statistice.

**Texton:** microstructura (elementul vizual) fundamental prin repetarea căruia rezultă textura

# 3. DETECŢIA OBIECTELOR PRIN TEHNICI VIZUALE. PARTICULARIZARE: DETECŢIA MÂINII

## 3.1 Problema generală a detecţiei obiectelor prin tehnici vizuale

Întrucât, aşa cum rezultă din definiţia dată în paragraful anterior, detecţia obiectelor în imagini înseamnă căutarea în imagine a acelor obiecte aparţinând unor categorii precizate la un moment dat; astfel, problema detecţiei implică clasificare, recunoaştere, si, deci, un anumit grad de complexitate.

Demersul necesar detecţiei obiectelor poate fi împărţit în două etape:

- separarea regiunilor distincte din scenă (segmentarea secenei)
- localizarea obiectului căutat

Prima etapă, de segmentare a scenei, implică sesizarea graniţelor dintre regiunile distincte ale imaginii care pot să corespundă obiectelor distincte din scenă. Aici intervin, de obicei, metodele bazate pe detecţie de muchii, respectiv de contururi. Se vor pune în evidenţă, aşadar, trecerile de la regiuni cu anumite caracteristici, la regiuni cu caracteristici diferite. Complexitatea scenelor din lumea reală şi varietatea tipurilor de obiecte impun metode sofisticate de detecţie a graniţelor între obiecte - simple treceri de la un nivel de intensitate la un alt nivel de intensitate, treceri de la o textură la o alta textură, contururi iluzorii.

Cea de-a doua etapă implică localizarea obiectului căutat, despre care se stie că are anumite caracteristici: culoare, formă, textură, sau grupare de regiuni cu caracteristici bine precizate. De obicei, această etapă este precedată de una de învăţare, în care se reţin caracteristicile definitorii ale obiectului căutat, dintr-un set de imagini de antrenament.

Complexitatea procesului de separare rezidă în discontinuitatea contururilor, datorată acoperirilor dintre obiecte, schimbarea aparenţei obiectului căutat la modificarea poziţiei, a condiţiilor de iluminare, a distanţei faţă de observator, dinamicii scenelor naturale, decorurilor multimodale. Percepţia umană se bazează pe combinarea caracteristicilor de textură, formă, culoare, pe memorie, capacitate deductivă. Această combinaţie de factori trebuie integrată într-un sistem computaţional, astfel încât acesta să funcţioneze în timp real. Aşadar, se vor căuta acele caracteristici ale obiectelor care sunt invariante la schimbarea conditiilor externe, iar pe de altă parte, se vor implementa metode şi algoritmi performanţi, cât mai rapizi, în scopul recunoaşterii obiectului in condiţii variate.

Metodele generale de detecţie a obiectelor indiferent de clasa de apartenenţă, implică, până în momentul actual, o complexitate computaţională deosebit de ridicată si o dimensiune foarte mare a setului de antrenament. De aceea, în cazul metodelor robuste, se preferă căutarea doar a unui anumit obiect specific, după un număr cât mai mic, dar suficient, de carcateristici, invariante la modificarea condiţiilor externe.

În ceea ce priveşte detecţia mâinii, se pot remarca următoarele caracteristici invariante: textura şi culoarea pielii, forma 3D specifică, dependentă de câteva poziţii de bază ale acesteia. Forma, mărimea, precum şi uniformitatea culorii şi a texturii, lipsa altor elemente componente precum ochi, ochelari, nas, gura, mustaţă, păr, contribuie la distincţia mâinii faţă de alte obiecte având aceeaşi textură şi culoare – cea a pielii , precum faţa umană.

## 3.2 Metode existente in literatura de specialitate

Pentru detecţia muchiilor, contururilor şi regiunilor distincte din imagine, metodele existente variază, de la detecţia de muchii pe baza nivelurilor de intensitate - metoda Gradientului sau a Laplacianului [13], la metode mai complexe, precum prăguirea (*thresholding*) pe baza nivelelor de intensitate ale pixelilor [9], extragerea fondului [9], metode bazate pe textură [3], iar metodele pentru extragerea conturului, de la transformata Hough [13], la modele active [7].

În ceea ce priveşte localizarea obiectelor aparţinând unor categorii specificate, există următoarele modalităţi de soluţionare a problemei: detecţie de contur, potrivirea formelor 2D ale acestor contururi [1], metode robuste, de timp real, bazate pe determinarea distanţei dintre semnături ce surprind caracteristici de textură, culoare si formă a obiectului cautat [14], metode ce considera obiectul ca o constelatie de părţi şi utilizează clasificatori bazaţi pe reguli structurale, precum şi reprezentări probabilistice care ţin cont de scală, formă şi ocluziile dintre obiecte [4], detecţia, independentă de orientare si scală, a unor obiecte aparţinând unor clase multiple [14]. Multe metode

generice, ce detectează, de obicei, mai multe clase de obiecte, utilizează clasificatori multipli de tip ada-boost, ce determină trăsăturile cele mai relevante pentru clasificare[14]. O altă categorie de metode, ce urmăresc detecţia obiectului indiferent de poziţia sa, sau – în cazul feţelor umane, de expresia feţei, utilizează aşanumitele modele de aparenţă, ce surprind variaţiile posibile ale acestor obiecte şi utilizează clasificatori de tipul reţelelor neuronale, lanţurilor Markov ascunse, clasificatorul Bayesian. Aceste metode implică o complexitate computaţională deosebit de ridicată.

Un interes deosebit îl prezintă metodele robuste de localizare a obiectelor în mişcare, precum metoda [5], ce utilizează împărţirea imaginii în blocuri şi compararea blocurilor corespondente din cadre consecutive ale unei secvenţe video, pe baza unei trăsături statistice, ce caracterizează textura acelui bloc, anume parametrul LBP (*Local Binary Pattern*):

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)^2 \, p \quad (1)$$

De un interes aparte este şi metoda dezvoltată de renumitul sistem „*Pfinder"* – de urmărire în timp real a persoanelor, care defineşte regiuni cu caracteristici eterogene numite „*blobs" şi* calculează probabilitatea de apariţie a unei astfel de regiuni prin formula gaussiană de probabilitate. [9]

În ceea ce priveşte metodele specifice de detecţie a mâinii, se preferă metode robuste, de tipul celor ce utilizează trăsături spectrale, extrase dupa calculul transformatelor Fourier, Gabor sau Wavelet ale imaginii [8]; metode bazate pe histograme de culori sau statistici ce caracterizează textura pielii [2], metode ce consideră mâna ca o constelaţie de părţi, metode asemănătoare cu cele potrivite pentru detecţia feţelor: bazate pe trăsături, pe potrivirea şabloanelor, pe modele de aparenţă, pe culori, pe detecţia mişcării [6].

## 3.3 Metoda propusă pentru detecţia mâinii

Metoda propusă urmăreşte localizarea mâinii în imagini reprezentând scene naturale de tipuri variate, cu decor multi-modal, adecvate situaţiilor specifice pentru recunoaşterea gesturilor, care conţin mâna umană ca element esenţial, dar şi alte obiecte din decorul care completează imaginea – cel mai frecvent întâlnite fiind elementele din partea superioară a corpului uman.

Elementul vizual de bază care se ia în considerare în segmentare este textura, care, din punct de vedere tehnic, reprezintă aranjamentul regulat al nivelurilor de intensitate ale pixelilor dintr-o anumită regiune. Dincolo de această definiţie clasică, textura 3D poate fi privită şi ca o configuraţie bine definită de microstructuri numite textoni, acestea reprezentând elementele fundamentale ale texturii. Aceste microstructuri sunt de următoarele tipuri: muchie (*edge*), creastă (*ridge*), pată (*spot*), undă (*wave*), undişoară (*ripple*). Conform acestei viziuni, caracterizarea texturii se va face prin histograma textonilor; procesul de construire a histogramei textonilor implică următorii paşi [11]:

- calculul unui vector de trăsături pentru fiecare pixel, ce va conţine rezultatele aplicării nucleelor de convoluţie Laws [3] respectiv nucleul de convoluţie Laplacian al Gaussianului [13]

- formarea textonilor, prin gruparea pixelilor cu caracteristici similare, folosind metoda numita *k-means clustering* [13]

- marcarea fiecărui pixel cu eticheta textonului căruia îi aparţine

- construirea histogramei textonilor, care memorează numărul de etichete aparţinând fiecărei clase din regiunea pentru care se face caracterizarea texturii

Pentru ca analiza să fie independentă de orientarea texturii, respectiv de condiţiile de iluminare, se va considera aceeaşi regiune de material în imagini diferite, reprezentând condiţii diferite de orientare, respectiv de iluminare. Vectorul de trăsături pentru un pixel va rezulta din concatenarea vectorilor de trăsături ai pixelului respectiv calculat în fiecare astfel de imagine. Se va obţine, în acest mod, o caracterizare a texturii general valabilă, independendentă de orientarea, respectiv direcţia de iluminare a materialului.

În scopul segmentării scenei în obiecte, se va urmări determinarea graniţelor dintre texturile care compun scena în felul următor:

- se va împărţi imaginea în blocuri de dimensiuni mai mici decât 10 pixeli;

- se va calcula histograma textonilor pentru fiecare bloc

- se vor calcula distanţele dintre histogramele blocurilor

$$\chi^2(h_1, h_2) = \frac{1}{2} \sum_{n=1}^{\#bins} \frac{(h_1(n) - h_2(n))^2}{h_1(n) + h_2(n)} \quad (2)$$

învecinate, folosind distanţa $\chi^2$:

- se va considera că există o trecere de la o textură la o alta textură, respectiv de la fundal la obiectele scenei şi invers doar dacă distanţa $\chi^2$ este mai mare decât un prag stabilit în felul următor:

$$\text{Prag} = (\chi^2_{min} + \chi^2_{max})/2 + \sigma^2_{\chi} \quad (3)$$

unde $\chi^2_{min}$ si $\chi^2_{max}$ reprezintă valorile minimă, respectiv maximă a distanţelor calculate între toate blocurile adiacente ale imaginii, de la stânga spre dreapta, $\sigma^2_{\chi}$ reprezintă abaterea medie pătratică a acestor distanţe.

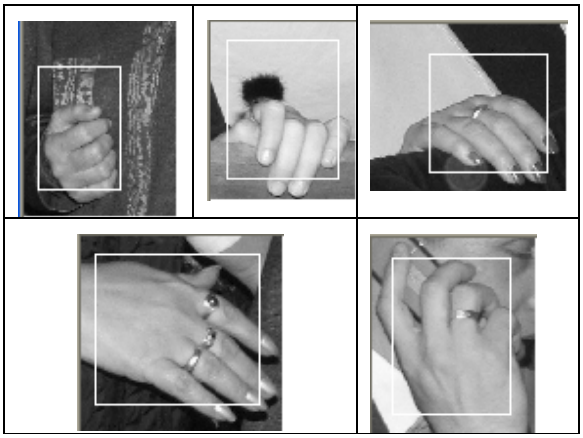
Se vor marca cu 0 regiunile aparţinând fundalului şi cu 1 regiunile din interiorul obiectelor.

Se vor căuta doar acele obiecte pentru care textura din interiorul blocurilor marcate corespunde cu textura din setul de antrenament , în cazul de faţă textura mâinii. Recunoaşterea texturii se va face tot prin intermediul metodei bazate pe textoni, în varianta independentă de orientare şi condiţii de iluminare, utilizând ca metrică distanţa dintre histograme.

Pentru a perfecţiona rata de recunoaştere, se vor lua în considerare şi caracteristici de formă şi culoare – histograma de culori sau semnături RGB ale imaginii, pentru caracterizarea culorii, respectiv spectrul de formă ISS (Image Shape Spectrum) [12], a cărui valoare într-un punct caracterizează curbura locală a suprafeţei de intensitate [12]. Astfel, setul de antrenament va conţine imagini reprezentând mâna în câteva poziţii de bază ale acesteia .

## 4. REZULTATE EXPERIMENTALE

**Tabelul 1: Cazuri de detecţie a mâinii**



Tabelul de mai sus ilustrează câteva rezultate experimentale obţinute în urma aplicării metodei descrise in secţiunea 3.3 – detecţia mâinii pe baza informaţiilor de textură. Mâna este surpinsă în diverse poziţii ale sale, prin intermediul unui dreptunghi. Decorul este compus din elemente de îmbrăcăminte umană sau din accesorii precum inelele, alte părţi ale corpului precum faţa, obiecte ţinute în mână (telefon). S-au utilizat imagini cu niveluri de gri reprezentate în formatul 8 biţi/pixel, preluate la o rezoluţie de 5 MP, fiind împarţite în blocuri de dimensiuni 10, 7, 5, respectiv 3 pixeli.

## 5. CONCLUZII

Metoda este potrivită pentru detecţia mâinii în scene naturale, fiind capabilă de a localiza obiectul de interes într-un decor variat. Pentru perfecţionarea ratei de succes, ne propunem utilizarea şi a unor alte informaţii precum cele de formă şi culoare, precum şi detecţia exactă a conturului mâinii prin intermediul unor modele active. În ceea ce priveşte viteza de execuţie, metoda va fi comparată cu cea descrisă în [5], special concepută pentru detecţia, pe baza texturii, a obiectelor aflate în mişcare, în timp real, dar care are drept inconvenient lipsa capacităţii de a detecta obiectele ce nu-şi modifică poziţia între două cadre succesive ale unei secvenţe video.

## 6. REFERINŢE


[1] Belongie, S., Malik, J. and Puzicha, J., *Shape matching and object recognition using shape contexts*, IEEE PAMI, vol. 24, no. 24, April 2002

[2] de Campos, Teofilo Emidio, University of Oxford, *3D Hand and Object Tracking for Intention Recognition*, Transfer Report, 2003

[3] Davis, Larry, Department of Computer Sciences, University of Texas at Austin, Austin, *Image Texture AnalysisTechniques – A Survey*

[4] Fergus, R; Perona, P., Zisserman, A., *Object class recognition by unsupervised scale-invariant learning*, CVPR, 2003

[5] Heikkila, M., Pietikainen M., J. Heikkila, Department of Electrical and Information Engineering, University of Oulu, Finland, *A Texture-based Method for Detecting Moving Objects*, 2004

[6] Hsuan, Ming Yang, Honda Research Institute, Mountain View, California, USA, *Advances in Face Processing: Detection*, ICPR 2004

[7] Kass, M., Witkin, A., Terzopoulos, D., *Snakes: Active contour models*, International Journal of Computer Vision. v. 1, n. 4, pp. 321-331, 1987

[8] Kolsch, M., Turk, M., Department of Computer Science, University of California, Santa Barbara *Robust Hand Detection,* 2003

[9] Lombardi, Paolo, Dipartimento di Informatica e Sistemistica, Università di Pavia, Italy, *A survey on pedestrian detection for autonomous driving systems*, 2001

[10] Mitrea, Delia, Gorgan, Dorian, *Comportament Adaptiv pentru Interfete Utilizator Avansate prin Intermediul Gesturilor*, Volumul de lucrari al Primei Conferinte Nationale de Interactiune Om-Calculator –RoCHI 2004, Universitatea Politehnica Bucuresti 23-24 Septembrie 2004, Editura Printech, Bucuresti, 2004, ISBN 973-718-053-4, pg. 33-42

[11] Mitrea Delia, Nedevschi Sergiu, Technical University of Cluj-Napoca, *Road Quality Evaluation and Road Material Recognition using 3D textons*, 8-th IEEE International Conference of Intelligent Engineering Systems, INES 2004, Cluj-Napoca, Romania, septembrie 2004 Proceedings, Editura U.T.Press Cuj-Napoca, pg.236-241

[12] Nastar, Chahab, Mitschke, Matthias, INRIA, France, *Real-Time Face Recognition using Feature Combination*, Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, 14-16 April 1998, Nara, Japan

[13] Nedevschi, Sergiu, Editura Albastra, Cluj-Napoca, 1998: *Prelucrarea imaginilor si Recunoasterea Formelor*

[14] Torralba, A., Murphy, K.P., and Freeman,W.T., *Sharing visual features for multiclass and multiview object detection*, CVPR'2004