

Hands Gestures Recognition in Stereo Images

Tudor I. Cerlincă, Ștefan Gh. Pentiuc

Abstract—This paper presents a fast, accurate and innovative method for hands gestures recognition in stereo images. Hands gestures recognition is a Computer Vision key problem with lots of applications in different area such as: HRI(Human-robot Interaction), HCI(Human-Computer Interaction), virtual reality, human emotions and behavior understanding etc. Although there are many high-performance hands gestures recognition methods, only few of them can be integrated in real time HRI applications. Perhaps the most important limitation is related to the maximum distance at witch the system is still able to recognize the hands gestures. Few methods support distances above 3-4 meters. The method we proposed can detect hands gestures up to five meters and was designed in such a manner that it can be easily integrated in various vision-based applications that need real-time detection and recognition of hand gestures. The final goal was the integration of our method into a more complex system in which the motion of a mobile robot can be controlled only through hands gestures.

Index Terms—HRI, hands gestures, stereo vision

I. INTRODUCERE

PROBLEMATICA recunoașterii în timp real a gesturilor mâinilor este considerată o componentă cheie în foarte multe aplicații Computer Vision: interacțiunea om-robot, interacțiunea om-calculator, realitatea virtuală etc.. Acest domeniu are aplicații și în studiu emoțiilor sau a comportamentului uman.

În acest domeniu există multe metode performante de recunoaștere a gesturilor mâinilor dar acestea nu pot fi întotdeauna integrate în aplicații din lumea reală. Cercetările efectuate au arătat că algoritmi actuali se confruntă cu următoarele probleme:

- distanța dintre sistemul de camere video și mână trebuie să fie mică; cu cât distanța este mai mare cu atât procentajul clasificărilor corecte este mai mic;
- modelarea corpului uman prin metode geometrice nu furnizează rezultatele scontate, în cazul imaginilor preluate de la mare distanță;
- schimbarea condițiilor de iluminare poate fi un inconvenient major.

Aplicațiile actuale sunt complexe și necesită sisteme de recunoaștere robuste, capabile să recunoască și să interpreteze gesturile mâinilor în aproape orice condiții. În funcție de tipul imaginilor prelucrate, sistemele de recunoaștere a gesturilor mâinilor pot fi împărțite în trei mari categorii:

- *single-cam* – prelucrează imagini provenite de la o singură cameră video;
- *multi-cam* – prelucrează în mod individual imagini provenite de la un sistem de camere video;
- *stereo-cam* – prelucrează o imagine stereo, obținută prin combinarea imaginilor provenite de la un sistem format din două sau mai multe camere video.

Sistemele *single-cam* sunt relativ ușor de implementat, dar au o serie de dezavantaje majore, cum ar fi:

- câmpul de vizibilitate este limitat;
- nu pot rezolva cazurile de ambiguitate (ex.: două persoane, aflate una în spatele celeilalte execută același gest; o persoană aflată în fața/spatele celei autorizate să execute gesturi, execută un gest valid);

Sistemele *multi-cam* au avantajul unui câmp de vizibilitate lărgit dar tot nu rezolvă cazurile de ambiguitate. În plus, aceste sisteme se confruntă cu noi cazuri de ambiguitate (ex.: pentru fiecare cameră video, sistemul de recunoaștere identifică execuția unor gesturi diferite).

Sistemele *multi-cam* sunt mult mai dificil de implementat, dar rezolvă toate problemele cu care se confruntă sistemele *single-cam* și *multi-cam*. Aceste sisteme pot fi integrate cu succes în aplicații specifice domeniului HRI. Sistemele *single-cam* pot fi integrate cu succes în aplicații HCI deoarece, în acest caz, distanța dintre operatorul uman și camera video este relativ mică iar câmpul vizual oferit de camera video este suficient de mare. În plus, situațiile de ambiguitate apar foarte rar, deoarece, de cele mai multe ori, în fața camerei video se află o singură persoană. Sistemele *multi-cam* pot fi integrate atât în aplicații specifice domeniului HRI cât și în cele specifice domeniului HCI.

Spre deosebire de HCI, HRI este un domeniu care impune o condiție, foarte greu de realizat de cele mai multe ori: toate aplicațiile trebuie să fie capabile să ruleze în medii fără constrângere. Astfel, schimbarea condițiilor de mediu nu trebuie să influențeze în nici un fel buna funcționare a algoritmilor.

Orice sistem de interacțiune gestuală va defini un dicționar de gesturi care va reprezenta limbajul de comunicare dintre operatorul uman și sistemul robotic. Interacțiunea va fi de tipul comandă-răspuns. Este binecunoscut faptul că cele mai multe sisteme de interacțiune gestuală recunosc și interpretează doar gesturile degetelor. În acest caz, distanța maximă de la care poziția degetelor mai poate fi recunoscută și interpretată este practic foarte mică, ceea ce conduce, de cele mai multe ori, la imposibilitatea realizării unui sistem performant de

interacțiune gestuală. Astfel de sisteme de interacțiune gestuală vor avea un scop pur didactic, integrarea lor în mediul industrial fiind practic imposibilă.

În această lucrare este prezentat un sistem inovator de recunoaștere a gesturilor mâinilor în imagini stereoscopice. Sistemul a fost testat în cadrul unei aplicații software de comandă a unui robot mobil.

II. SISTEMUL DE RECUNOAȘTERE A GESTURILOR MÂINILOR

Sistemul de recunoaștere a gesturilor mâinilor nu ia în considerare poziția degetelor ci poziția întregului antebraț, deoarece acesta poate fi detectat de la distanțe mult mai mari. Gesturile pot fi identificate și interpretate chiar dacă se îndepărtează destul de mult de cazul ideal. Sistemul poate recunoaște șase gesturi diferite ce corespund unui număr de patru comenzi: LEFT, RIGHT, FORWARD și BACKWARD.

S-a avut în vedere ca sistemul de recunoaștere a gesturilor să fie unul total adaptiv și toate valorile de prag să fie determinate și ajustate în mod automat. Sistemul inteligent de recunoaștere a gesturilor este construit în totalitate pe baza unui clasificator statistic care are rolul de a estima poziția mâinilor funcție de poziția și dimensiunile capului uman.

Recunoașterea gesturilor este un proces care se desfășoară în 8 etape:

- achiziția imaginilor de la sistemul de camere video;
- detecția și predicția poziției capului uman;
- estimarea poziției mâinilor pentru fiecare gest;
- detecția pielii umane (localizarea exactă a mâinilor);
- detecția componentelor mâinilor;
- conectarea componentelor mâinilor;
- detecția mișcării componentele mâinilor. Analiza istoricului mișcărilor;
- selecția automată a gestului.

În primele două etape, sistemul de recunoaștere a gesturilor va prelua câte o imagine de la fiecare cameră video, va combina cele două imagini și apoi va încerca să detecteze capul uman. Dacă operația de detecție eșuează, atunci sistemul va încerca să estimeze poziția capului pe baza pozițiilor anterioare. Sistemul poate detecta capul uman atât în imagini ce conțin o vedere frontală a feței cât și în imagini ce conțin o vedere din profil a acesteia. Sistemul de recunoaștere a gesturilor va furniza rezultate corecte chiar dacă imaginile supuse prelucrării conțin mai multe persoane. Se poate considera că această etapă este vitală în procesul de recunoaștere a gesturilor deoarece trecerea la celelalte etape se poate realiza numai după detecția capului uman.

În cea de-a treia etapă, sistemul va încerca să estimeze poziția și dimensiunea fiecărei mâini. Pentru obținerea unui timp de răspuns cât mai mic, estimarea se realizează prin intermediul unei proceduri statistice. Această procedură se bazează în întregime pe o serie de rapoarte statistice ce există între diferitele părți ale corpului uman. Este binecunoscut faptul că dimensiunile mâinilor umane se află într-o strânsă relație cu dimensiunile capului uman, astfel că, dacă se

reușește localizarea capului (poziție și dimensiuni), atunci se pot estima: poziția și dimensiunea mâinilor, pentru fiecare din cele șase gesturi. Pentru fiecare mână, sistemul va determina în mod dinamic (funcție de poziția și dimensiunea capului), câte trei dreptunghiuri de încadrare diferite ce acoperă toată aria în care se poate mișca o mână pentru a executa un gest valid (figura 1).

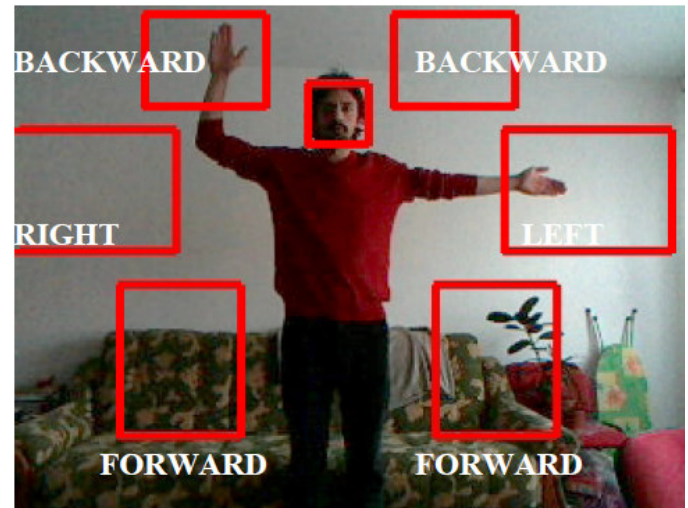


Fig. 1. Estimarea poziției mâinilor.

În cea de-a patra etapă, sistemul va încerca să determine poziția exactă a mâinilor. Căutarea nu se realizează în întreaga imagine ci doar în cadrul celor șase dreptunghiuri de încadrare determinate anterior, ceea ce conduce în final la obținerea unor timpi de răspuns foarte mici. Localizarea mâinilor se realizează prin intermediul modulului de detecție a pielii umane. În această etapă se realizează doar o analiză cantitativă a informațiilor conținute în cele șase dreptunghiuri de încadrare.

În următoarele trei etape se realizează o analiză calitativă a informațiilor cu scopul de a identifica obiectele de culoare apropiată pielii, în fiecare din cele șase dreptunghiuri de încadrare. În cazul ideal, modulul de detecție a mâinilor va detecta întreaga mână. Experimentele efectuate au arătat însă că în cazul imaginilor preluate de la o distanță mai mare de 2.5m, modul de detecție a mâinilor nu va detecta întotdeauna întreaga mână ci o serie de zone corespunzătoare mâinii umane. Rezultatul acestei analize este reprezentat de o mulțime de obiecte care aparțin mâinii umane și pot fi conectate între ele.

Totuși, sistemul de recunoaștere a gesturilor poate furniza rezultate eronate în cazul în care, dreptunghiurile de încadrare conțin obiecte (cărămizi, cutii de carton etc.) de culoare apropiată pielii și arie comparabilă cu cea a mâinii. În cea de-a șaptea etapă sistemul va realiza o analiză cantitativă și una calitativă a istoricului mișcărilor (MHI – Motion History Image) din cele șase dreptunghiuri de încadrare, cu scopul de a reduce din punct de vedere numeric cazurile în care sistemul furnizează rezultate eronate.

Se va considera că utilizatorul a executat un gest valid doar atunci când:

- în dreptunghiurile de încadrare a avut loc o mișcare uniformă a obiectelor ce intră în componența mâinilor;
- diferența de profunzime dintre capul și mâinile umane este foarte mică.

În ultima etapă, pe baza dicționarului de gesturi, sistemul va determina gestul executat de către operatorul uman.

III. DETECȚIA CAPULUI UMAN

Se poate considera că această etapă este vitală în procesul de recunoaștere a gesturilor deoarece trecerea la celelalte etape se poate realiza numai după detecția sau estimarea poziției capului uman. Operația de detecție a capului uman a fost realizată prin implementarea unui sub-sistem cu “învățare supravegheată”. Acest sub-sistem are la bază caracteristici Haar simple [1] (denumite astfel pentru că sunt calculate într-un mod similar cu coeficienții transformatei wavelet Haar) și un set de clasificatori AdaBoost conectați în cascadă, fiecare clasificator din cascadă având un grad de complexitate mult mai ridicat față de predecesorul său. Algoritmul implementat poate detecta capul uman în imagini care conțin o vedere frontală sau din profil a capului uman. Algoritmul de detecție a capului uman este prezentat în detaliu în [2].

Experimentele efectuate au arătat că în unele cazuri algoritmul de detecție nu va reuși să detecteze capul uman, chiar dacă imaginea supusă prelucrării conține o vedere frontală sau din profil a acestuia. În acest caz se încearcă estimarea poziției capului, pe baza pozițiilor anterioare. Estimarea se realizează prin intermediul unei proceduri probabilistice. Operația de estimare se realizează doar pentru imaginile care satisfac

următoarea condiție: $t_{curr_img} - t_{head} < Th_{MaxDiffTime}$

unde:

t_{curr_img} - data când a fost preluată imaginea curentă;

t_{head} - data când a fost înregistrată ultima operație de detecție a capului uman;

$Th_{MaxDiffTime} \in [0_{ms}, 200_{ms}]$ - intervalul de timp maxim pentru care se realizează estimarea.

Probabilitatea ca imaginea curentă să conțină o vedere (frontală sau din profil) a capului uman este cu atât mai mare

cu cât valoarea $t_{curr_img} - t_{head}$ este mai mică. Experimentele efectuate au arătat că probabilitatea ca operatorul uman să-și miște capul uman în intervalul $Th_{MaxDiffTime}$ este foarte mică.

Se poate considera că o imagine conține o vedere a capului uman atunci când:

- $t_{curr_img} - t_{head} < Th_{MaxDiffTime}$;
- ultimele Th_{N_Head} imagini conțin o vedere a capului uman;
- distanța dintre pozițiile centrelor de greutate ale capului uman este foarte mică.

Figurile 2 și 3 prezintă rezultatele algoritmului de detecție a

capului uman.

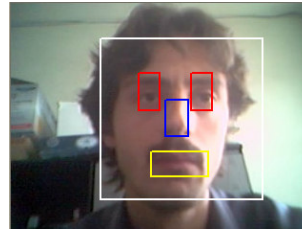


Fig. 2. Detecția capului. O singură persoană.

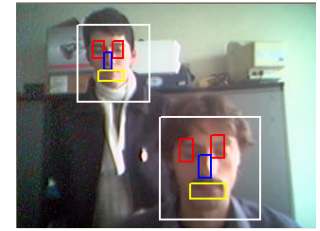


Fig. 3. Detecția capului. Două persoane.

IV. STEREOSCOPIA

Stereoscopia este o ramură a opticii care se ocupă cu procedeele de înregistrare și redare a imaginilor în relief și cu studiul proprietăților opto-mecanice ale instrumentelor folosite în acest scop. Principiul care stă la baza stereoscopiei este cel al percepției umane: 2 ochi = 2 vederi. Astfel, fiecare ochi captează o imagine proprie care este trimisă creierului uman pentru a fi procesată. Când creierul primește simultan două imagini, acestea vor fi combinate în una singură. Procesul de combinare a imaginilor constă în regăsirea zonelor similare și adăugarea micilor diferențe. Micile diferențe dintre cele două imagini conduc la diferențe mari în imaginea finală. Imaginea obținută este mai mult decât o sumă a celor două imagini, este chiar o imagine stereo tridimensională.

În domeniul Computer Vision ochiul uman este înlocuit de către o cameră video iar imaginile ce vor fi combinate provin de la două sau mai multe camere video diferite. Sistemul de recunoaștere a gesturilor nu are drept scop obținerea unei imagini stereoscopice ci determinarea profunzimii capului și mâinilor umane.

Pentru cazul camerelor video paralele care sunt îndreptate în aceeași direcție (figura 4), profunzimea unui obiect (distanța dintre sistemul de camere video și obiectul în cauză) poate fi calculată cu următoarea formulă [3]:

$$Z = \frac{f * B}{d} \quad (1)$$

unde:

- f este distanța focală a celor două camere video;
- B este distanța dintre obiectivele celor două camere video;
- d este disparitatea stereo (distanța dintre un obiect din prima imagine și același obiect din cea de-a doua imagine);
- Z este distanța până la obiectul de interes.

În acest caz, disparitatea este invers proporțională cu distanța până la obiectul de interes. Calculul disparității presupune identificarea perechilor de obiecte identice din cele două imagini. Cele mai cunoscute metode de calcul a disparității sunt: metoda corelației, a corespondenței și a diferenței de fază. În cazul sistemului de recunoaștere a gesturilor, disparitatea se poate calcula foarte ușor, după identificarea capului și mâinilor, în fiecare din cele două imagini.

În continuare, vom considera cazul a două imagini care prezintă aceeași scenă (figurile 5 și 6) dar au fost preluate de la două camere video diferite. Figura 7 prezintă rezultatul suprapunerii celor două hărți de contururi și distanța dintre sistemul de camere video și capul uman (obiectul de interes).

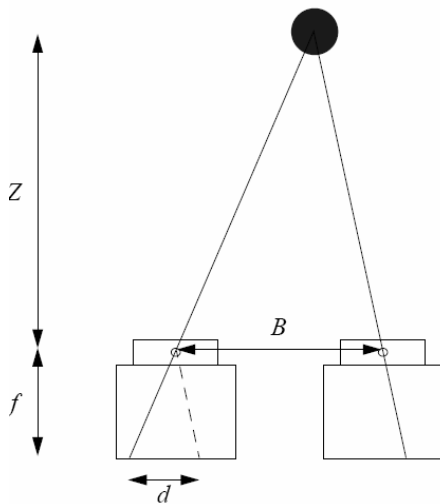


Fig. 4. Disparitatea și distanța până la obiectele de interes. Cazul camerelor video paralele [2].



Fig. 5. Imagine preluată cu prima cameră video.



Fig. 6. Imagine preluată cu cea de-a doua cameră video.



Fig. 7. Hărțile de contururi și distanța până la obiectul de interes.

V. ESTIMAREA POZIȚIEI MĂINILOR

În această etapă sistemul va încerca să estimeze, pentru fiecare gest în parte, poziția și dimensiunea mâinilor. Pentru obținerea unui timp de răspuns cât mai mic, estimarea se realizează prin intermediul unei proceduri statistice. Această procedură se bazează în întregime pe o serie de rapoarte statistice ce există între diferitele părți ale corpului uman.

Este binecunoscut faptul că dimensiunile mâinilor umane se află într-o strânsă legătură cu dimensiunile capului uman. Dacă se reușește localizarea capului (poziție și dimensiuni), atunci se pot estima, poziția și dimensiunea mâinilor, pentru fiecare din cele șase gesturi. Pentru fiecare mână, sistemul va determina în mod dinamic (funcție de poziția și dimensiunea capului), câte trei dreptunghiuri de încadrare diferite ce

acoperă toată aria în care se poate mișca o mână pentru a executa un gest valid. Într-o primă fază, un gest va fi considerat valid atunci când mâinile pot fi localizate în interiorul dreptunghiurilor de încadrare corespunzătoare iar combinația acestor dreptunghiurilor aparține dicționarului de gesturi. Dimensiunea dreptunghiurilor de încadrare nu este constantă ci determinată, la fiecare pas, în funcție de dimensiunea actuală a capului uman. Trebuie precizat faptul că scopul acestei etape nu este de a determina poziția exactă a mâinilor ci doar de a determina ariile rectangulare în care trebuie să se realizeze căutarea.

Procedura de determinare a dreptunghiurilor de încadrare are la bază studiul lui Leonardo da Vinci despre proporțiile corpului uman [4]. În cadrul acestui studiu se consideră că dimensiunile unor părți ale corpului uman pot fi determinate în funcție de dimensiunea capului. Conform acestui studiu, se poate considera că:

- lungimea gâtului este aproximativ egală cu 1/4 din lungimea capului;
- lungimea unei palme este aproximativ egală cu 3/4 din lungimea capului;
- lungimea antebrațului este aproximativ egală cu 5/4 din lungimea capului;
- lungimea brațului este aproximativ egală cu 3/2 din lungimea capului.

După cum se poate observa și din figura 1 dimensiunea dreptunghiurilor de încadrare este mai mare decât cea a mâinilor. S-a optat pentru această soluție deoarece:

- este foarte dificil de determinat poziția și dimensiunea exactă a mâinilor;
- gestul poate fi recunoscut chiar dacă se îndepărtează destul de mult de cazul ideal.

VI. DETECȚIA MĂINILOR

În cea această etapă, sistemul va încerca să determine poziția exactă a mâinilor. Căutarea nu se realizează în întreaga imagine ci doar în cadrul celor șase dreptunghiuri de încadrare, ceea ce conduce în final, la obținerea unor timpi de răspuns foarte mici. Localizarea mâinilor se realizează prin intermediul unui modul de detecție a pielii umane. Detecția pielii umane este o componentă cheie în foarte multe aplicații Computer Vision cum ar fi: detecția feței, recunoaștere gesturilor, realitate virtuală [5] etc.

În general, operația de detecție a pielii este realizată prin intermediul unui filtru de piele. Literatura de specialitate marchează existența mai multor metode de implementare a filtrului de piele: clasificatori Bayes, mixturi Gaussiene în spațiul de culoare RGB, SOM (Self-Organizing Maps) etc. De asemenea se pot utiliza diferite spații de culoare: RGB, RGB normalizat, HSV, YcrCb etc.

Filtrul de piele a fost implementat în spațiul de culoare HSV (Hue – Saturation – Value) datorită similitudinilor dintre acesta și modul în care ochiul uman percepe culorile. Într-o primă fază imaginile sunt convertite din format RGB în HSV. Transformarea din spațiul de culoare RGB în cel HSV este

prezentată în 2.

$$MAX = \max(R, G, B)$$

$$H = \begin{cases} \text{undefined}, MAX = MIN \\ 60^\circ \times \frac{G - B}{MAX - MIN}, MAX = R, G \geq B \\ 60^\circ \times \frac{G - B}{MAX - MIN} + 360^\circ, MAX = R, G < B \\ 60^\circ \times \frac{B - R}{MAX - MIN} + 120^\circ, MAX = G \\ 60^\circ \times \frac{R - G}{MAX - MIN} + 240^\circ, MAX = B \end{cases} \quad (2)$$

$$S = \begin{cases} 0, MAX = 0 \\ 1 - \frac{MIN}{MAX}, \text{altfel} \end{cases}$$

$$V = MAX$$

În general, se poate considera că un punct aparține pielii umane dacă sunt îndeplinite simultan următoarele condiții:

- $H > h_{\min}$
- $H < h_{\max}$
- $S > s_{\min}$
- $S > s_{\max}$

unde: h_{\min} , h_{\max} , s_{\min} și s_{\max} reprezintă valorile de prag utilizate în cadrul detecției.

Valorile h_{\min} , h_{\max} , s_{\min} și s_{\max} nu sunt constante și nici nu vor fi stabilite de către operatorul uman ci vor fi determinate în mod automat, pentru fiecare imagine în parte. Operația de determinare a acestor valori se bazează pe rezultatele etapei de detecție a capului uman și pornește de la presupunerea că marea majoritate a punctelor din dreptunghiul de încadrare corespunzător capului uman vor fi puncte care aparțin feței umane. Dacă se realizează histograma H-S pentru dreptunghiul de încadrare corespunzător capului uman, atunci valorile maxime vor fi asociate punctelor corespunzătoare pielii din zona feței umane.

Valorile de prag utilizate în cadrul detecției vor fi calculate astfel:

- $h_{\min} = H_Histo_{\max} - H_{th_exp}$
- $h_{\max} = H_Histo_{\max} + H_{th_exp}$
- $s_{\min} = S_Histo_{\max} - S_{th_exp}$
- $s_{\max} = S_Histo_{\max} + S_{th_exp}$

unde: H_{th_exp} și S_{th_exp} sunt valori determinate în mod experimental. Aceste valori de prag pot fi utilizate și pentru detecția pielii din zona mâinilor, deoarece în spațiul de culoare HSV, zonele corespunzătoare pielii au aceeași distribuție a

componentelor H, S și V atât pentru pielea de pe față cât și pentru cea de pe mână (figura 8). Pentru localizarea unei mâini se aplică filtrul de piele în fiecare din cele 3 dreptunghiuri de încadrare corespunzătoare. Pentru fiecare dreptunghi în parte, se determină numărul de puncte care aparțin pielii. Mâna va putea fi localizată în dreptunghiul pentru care numărul de puncte corespunzătoare pielii este maxim, doar dacă acest

maxim este mai mare decât o valoare de prag Th_{piele} . Această valoare de prag nu este o constantă ci este determinată în mod automat în funcție dimensiunea estimată a mâinilor.

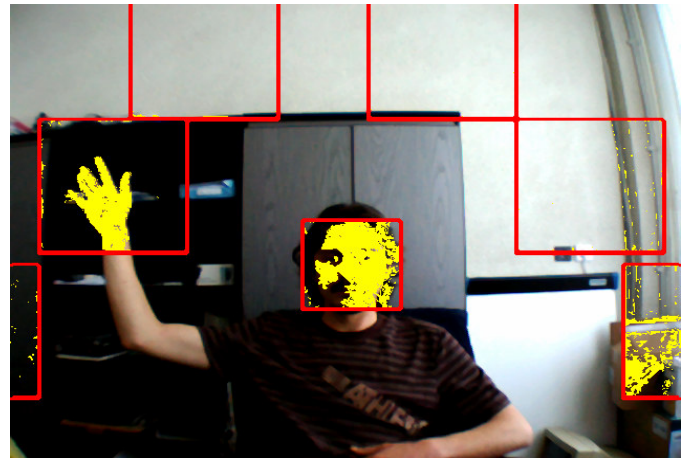


Fig. 8. Detecția pielii umane.

Un sistem software de recunoaștere a gesturilor mâinilor realizat doar pe baza primelor patru etape va prezenta următoarele avantaje:

- localizarea rapidă a mâinilor;
- identificarea rapidă a gesturilor mâinilor;
- adaptabilitatea la schimbarea condițiilor de mediu.

Totuși, rezultatele experimentale au arătat că sistemul poate furniza rezultate eronate atunci când zonele corespunzătoare dreptunghiurilor de încadrare conțin obiecte statice de culoare apropiată pielii. Din punctul de vedere al percepției umane, culoarea pielii diferă destul de mult de cea a unor cutii de carton sau a unor cărămizi. În spațiul de culoare în care se realizează (HSV), distribuția componentelor de culoare este aproximativ aceeași atât pentru mâna umană cât și pentru cutiile de carton sau pentru cărămizi.

Rezultatele eronate pot fi cauzate de un grup de obiecte (fiecare obiect individual având aria mai mică decât cea a mâinii) sau de un singur obiect de arie comparabilă cu cea a mâinii.

Algoritmii implementați în cadrul următoarelor trei etape (detecția componentelor mâinilor, conectarea componentelor mâinilor și detecția mișcării pentru componentele mâinilor) au condus la o creștere substanțială a performanțelor sistemului de recunoaștere a gesturilor mâinilor.

VII. DETECȚIA COMPONENTELOR MÂINILOR

Scopul principal al acestei etape este de a identifica obiectele de culoare apropiată pielii, în fiecare din cele șase

dreptunghiuri de încadrare. Identificarea obiectelor are drept scop realizarea unei analize calitative a informațiilor conținute în dreptunghiurile de încadrare. Operația de detecție a obiectelor a fost realizată cu ajutorul unui algoritm de etichetare iterativ ce acționează asupra imaginii binarizate. Mulțimea obiectelor a fost supusă unei operații de filtrare cu scopul de a elimina obiectele care nu pot fi considerate ca făcând parte din mâna umană. Aceste obiecte vor avea aria mai mică decât o valoare de prag **ThMinArea** sau mai mare decât aria estimată a mâinii. Valoarea **ThMinArea** este determinată în mod automat în funcție de aria estimată a mâinii.

Pentru umplerea golurilor și a micilor discontinuități, operația de detecție a obiectelor este urmată de una de închidere a imaginii. Închiderea unei imagini este un proces din doi pași: dilatare și eroziune. Ordinea operațiilor este foarte importantă, astfel că dacă cele două operații sunt realizate în ordine inversă se obține o deschidere a imaginii și nu o închidere.

Figurile 9, 10, 11 și 12 prezintă efectul închiderii pentru cazul unor imagini preluate de la sistemul de camere video. Se poate observa că:

- au fost umplute golurile mici;
- au fost conectate obiectele disjuncte.



Fig. 9. Imaginea originală.



Fig. 10. Imaginea după închidere.

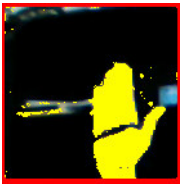


Fig. 11. Imaginea originală.

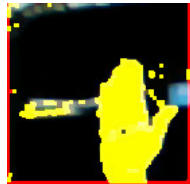


Fig. 12. Imaginea după închidere.

VIII. CONECTAREA COMPONENTELOR MÂINILOR

Cea mai importantă etapă în procesul de analiză calitativă a informațiilor conținute în cele șase dreptunghiuri de încadrare este cea de conectare a obiectelor care pot fi considerate ca aparținând mâinii umane. Rezultatul operațiilor de detecție a obiectelor și închidere a imaginii este o mulțime **M** de obiecte disjuncte.

Când cardinalul mulțimii **M** este mai mare decât o valoare maximă **ThMaxNumObiecte**, se consideră că obiectele nu pot fi conectate, ele fiind rezultatul unei detecții defectuoase.

Două obiecte **O1** și **O2** pot fi conectate doar dacă distanța dintre centre lor de greutate este mai mică decât o valoare **ThMinCG**. Această valoare nu este o constantă ci este determinată în mod automat în funcție de aria estimată a mâinii. Rezultatul operației de conectare este o mulțime $M_1 \subseteq M$ de obiecte ce pot fi conectate între ele.

Dacă aria totală a obiectelor din mulțimea M_1 este mult mai mare decât aria estimată a mâinii atunci se renunță în mod succesiv la obiectele cu cea mai mică arie, până în momentul în care aria totală este aproximativ egală cu cea a mâinii.

Rezultatul acestei etape este o mulțime de obiecte $M_2 \subseteq M_1$ pentru care:

- $\forall O \in M_2, \text{ThMinArea} \leq \text{Aria}(O) \leq \text{AriaEstimataMana}$
- $\forall O_i, O_j \in M_2, \text{Dist}(\text{CG}(O_i), \text{CG}(O_j)) < \text{ThMinCG}$
- $\text{Card}(M_2) \leq \text{ThMaxNumObiecte}$
- $\sum_{i=1}^{\text{Card}(M_2)} \text{Aria}(O_i) \cong \text{AriaEstimataMana}$

Figurile 13, 14, 15 și 16 prezintă rezultatele etapelor de detecție și conectare a componentelor mâinilor. Obiectele individuale sunt colorate utilizând culori diferite.



Fig. 13. Imaginea originală.



Fig. 14. Imaginea după detecția și conectarea obiectelor.



Fig. 15. Imaginea originală.



Fig. 16. Imaginea după detecția și conectarea obiectelor.

IX. DETECȚIA MIȘCĂRII COMPONENTELOR MÂINILOR. ANALIZA ISTORICULUI MIȘCĂRIILOR

Rolul acestei etape este de realiza o analiză cantitativă și una calitativă a istoricului mișcărilor (MHI – Motion History Image) din cele șase dreptunghiuri de încadrare, cu scopul de a reduce din punct de vedere numeric cazurile în care sistemul furnizează rezultate eronate (ex.: existența unor obiecte statice de culoare apropiată pielii și arie comparabilă cu cea a mâinii). Se va considera că utilizatorul a executat un gest valid doar dacă în dreptunghiurile de încadrare a avut loc o mișcare uniformă a obiectelor.

Analiza mișcării este un proces din șase pași:

- detecția mișcării;
- extragerea conturului tuturor obiectelor aflate în mișcare;
- determinarea direcției de mișcare;
- actualizarea istoricului de mișcare, pentru fiecare dreptunghi de încadrare;
- analiza cantitativă a istoricului;
- analiza calitativă a istoricului.

Operația de detecție a mișcării a fost realizată cu ajutorul unui algoritm de “background subtraction” adaptiv [6]. Rezultatul acestui algoritm este reprezentat de o mulțime de obiecte aflate în mișcare. Fiecărui obiect aflat în mișcare i-a fost extras conturul și apoi adăugat într-un vector de mișcare (istoric de mișcare). Operația de construcție a vectorului de mișcare durează o secundă și jumătate. Operația de determinare a direcției de mișcare se realizează în același timp cu cea de construcție a vectorului de mișcare. La fiecare pas se determină noua direcție de mișcare prin compararea punctelor de pe contururile obiectelor din ultimele două imagini.

În etapa de analiză cantitativă se verifică dacă istoricul mișcărilor din dreptunghiurile de încadrare este unul semnificativ ce poate fi asociat mișcării mâinii. În etapa de analiză calitativă se verifică uniformitatea gradientului mișcării. Un gradient neuniform poate fi rezultatul mișcării unor obiecte de culoare apropiată pielii.

Implementarea algoritmilor descriși în această etapă a condus la o creștere substanțială a performanțelor sistemului de recunoaștere a gesturilor, prin reducerea din punct de vedere numeric a cazurilor în care sistemul raportează în mod eronat prezența unei comenzi.

X. REZULTATE EXPERIMENTALE.

Figurile 17, 18, 19 și 20 prezintă rezultatele experimentale ale sistemului de recunoaștere a gesturilor. Sistemul a fost testat, pentru fiecare comandă în parte, în patru situații diferite:

- **MEDIU STATIC** – testarea a fost realizată cu imagini statice, în care condițiile de iluminare și distanță a operatorului uman față sistemul de camere video sunt aproximativ constante. În acest caz, setul de testare a fost construit pe baza unui număr de 1800 imagini diferite (câte trei sute de imagini pentru fiecare gest).
- **MEDIU SEMI-DINAMIC** – testarea a fost realizată cu imagini preluate în timp real de la sistemul de camere video. S-a avut în vedere

testarea algoritmului de recunoaștere a gesturilor în diferite condiții de iluminare și distanță a operatorului uman față de sistemul de camere video. Acest mediu a fost denumit semi-dinamic deoarece între operatorul uman și aplicația de recunoaștere există o interacțiune vizuală, în sensul că operatorul uman are posibilitatea de a vizualiza atât imaginile preluate de la sistemul de camere video cât și cele șase dreptunghiuri de încadrare. În acest mod, utilizatorul va ști întotdeauna unde trebuie să poziționeze mâinile pentru a executa un gest valid.

- **MEDIU DINAMIC** – testarea a fost realizată în aceleași condiții ca și la mediul semi-dinamic cu deosebirea că în acest caz nu mai există nici o interacțiune vizuală între operatorul uman și aplicația de recunoaștere a gesturilor.
- **ROBOT MOBIL** – testarea a fost realizată cu imagini preluate în mod dinamic, de pe robotul aflat în mișcare.



Fig. 18. Gestul corespunzător comenzii FORWARD.



Fig. 17. Gestul corespunzător comenzii BACKWARD.



Fig. 19. Gestul corespunzător comenzii RIGHT.



Fig. 20. Gestul corespunzător comenzii LEFT.

XI. CONCLUZII.

În această lucrare a fost prezentat un sistem robust de recunoaștere în timp real a gesturilor mâinilor. Sistemul recunoaște șase gesturi diferite dar poate fi adaptat foarte ușor pentru a recunoaște și alte gesturi. Sistemul implementat prezintă următoarele avantaje principale:

- este total adaptiv (nu necesită intervenția operatorului uman iar toate valorile de prag sunt determinate în mod automat);
- poate să recunoască gesturile mâinilor chiar și de la distanțe mari;
- rezolvă toate cazurile de ambiguitate;
- poate să recunoască gesturile mâinilor chiar și atunci când acestea se îndepărtează destul de mult de cazul ideal;
- recunoașterea gesturilor se realizează într-un timp foarte scurt;
- operația de recunoaștere se realizează fără existența

vreunei interacțiuni vizuale cu operatorul uman.

Sistemul de recunoaștere a gesturilor mâinilor a fost integrat cu succes într-o aplicație software complexă de conducere a unui robot mobil. Comunicația dintre sistemul de recunoaștere a gesturilor și robotul mobil a fost realizată prin intermediul interfeței seriale RS232. În momentul identificării unui gest valid, sistemul de recunoaștere a gesturilor va trimite o comandă specifică către robotul mobil. La nivelul sistemului robotic, comanda va fi recepționată de către un microcontroler care va controla motoarele robotului în așa fel încât mișcarea acestuia să corespundă gestului executat de către operatorul uman.

REFERINȚE

- [1] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", Conference on Computer Vision and Pattern Recognition 2001.
- [2] T. Cerlinca and G. Pentiuc, "Detectia și urmărirea mișcărilor capului în timp real în secvențe video folosind o singură cameră", Seminarul Sisteme distribuite, 9 Decembrie 2006, Suceava, ISSN/ISBN 1842 – 6808.
- [3] D. Rosselot and E. Hall, "Processing real-time stereo video for an autonomous robot using disparity maps and sensor fusion", Center for Robotics Research, 2004.
- [4] http://en.wikipedia.org/wiki/Vitruvian_Man.
- [5] G. Pentiuc, R. Vatavu, T. Cerlinca and O. Ungureanu, "Methods and Algorithms for Gestures Recognition and Understanding", UkrObraz, 2006, Kyiv, Ukraine.
- [6] T. Cerlinca, "A DISTRIBUTED SYSTEM FOR REAL TIME TRAFFIC MONITORING AND ANALYSIS, Advances in Electrical and Computer Engineering, Suceava", 2005.