

Evaluarea stării emoționale pe baza poziției corpului uman, prin metode ale viziunii computerizate

Delia Mitrea

Universitatea Tehnică din Cluj-Napoca, Facultatea de Automatică și Calculatoare
Str. Barițiu Nr.26-28
Delia.Mitrea@cs.utcluj.ro

Sergiu Nedevschi

Universitatea Tehnică din Cluj-Napoca, Facultatea de Automatică și Calculatoare
Str. Barițiu Nr.26-28
Sergiu.Nedevschi@cs.utcluj.ro

Dorian Gorgan

Universitatea Tehnică din Cluj-Napoca, Facultatea de Automatică și Calculatoare
Str. Barițiu Nr.26-28
Dorian.Gorgan@cs.utcluj.ro

REZUMAT

Recunoașterea stărilor emoționale ale utilizatorilor umani este un subiect de interes în ceea ce privește interacțiunea om-calculator, întrucât se conformează conceptului de interacțiune naturală. Astfel, înlăturarea oricăror constrângeri în ceea ce privește comunicarea om-mașină este una dintre țintele cercetării actuale în domeniu: ideea este aceea de a face mașina invizibilă din punctul de vedere al utilizatorului uman. Limbajul corporal, reprezentat prin poziția corpului și gesturi, este unul dintre mijloacele non-verbale, indirecte de exprimare a emoțiilor, alături de expresia feței și tonalitatea vocii. Metodele specifice viziunii computerizate reprezintă o modalitate eficientă, adecvată, puțin costisitoare, de recunoaștere a stărilor emoționale pe baza poziției corpului, respectiv a gesturilor umane, în contextul interacțiunii om-calculator. Ne propunem să studiem problema recunoașterii stărilor emoționale pe baza limbajului corporal, prin intermediul viziunii computerizate, să trecem în revistă metodele existente în domeniu, să elaborăm o soluție proprie eficientă din punctul de vedere al acuratetei și timpului de execuție.

Cuvinte cheie

Interacțiune naturală, recunoașterea stărilor emoționale, limbaj corporal, poziția corpului uman, viziune computerizată.

Clasificare ACM

H.5.1. Information Interfaces and Presentation; H.5.2. User interfaces, Interaction styles; H.1.2. User/Machine Systems, Human factors; Human information processing; I.2.10. Vision and scene understanding, Shape

INTRODUCERE

Analiza și recunoașterea stărilor emoționale este un subiect de interes în cercetarea științifică actuală. Se pune atât problema asocierii unor stări emoționale personajelor virtuale, de tipul avatariilor sau agenților [1], cât și aceea de a recunoaște starea emoțională a utilizatorului uman [2]. Scopul este acela de a înlătura barierele dintre om și mașină, de a crea o realitate virtuală cât mai apropiată de viața de zi cu zi, precum și acela de a înlesni o comunicare naturală om-calculator, în contextul căreia nu numai utilizatorul se adaptează mașinii, dar și mașina se adaptează necesităților utilizatorului. Așa cum la nivelul relațiilor interumane, percepția emoțiilor interlocutorului joacă un rol important în comunicarea dintre indivizi, sesizarea automată a stării emoționale a interlocutorului este importantă și în contextul interacțiunii om-calculator,

contribuind la înlăturarea treptată a unor convenții nenaturale de comunicare.

Pe lângă mijloacele directe, verbale, de exprimare conștientă a emoțiilor, există și limbajul corporal, manifestat prin gesturi voluntare sau involuntare, prin poziția și mișcările corpului, capului, ale membrelor inferioare și superioare, care reprezintă o modalitate alternativă, mai mult sau mai puțin conștientă de exprimare, completând uneori limbajul verbal. Există însă și situații în care limbajul corporal poate contrazice limbajul verbal, exprimând adevăratele sentimente ale interlocutorului [7].

Metodele aferente viziunii computerizate reprezintă o modalitate eficientă de a prelucra și interpreta informația vizuală, precum și de a selecta în mod automat aspectele relevante în ceea ce privește caracterizarea stării emoționale pe baza limbajului corporal. Există o varietate de tehnici care pot deservi acest scop: Transformata Wavelet [10], metodele de scheletonizare a imaginii corpului uman și evaluarea orientării segmentelor componente [4], analiza formei 2D a conturului corpului uman [6], sunt doar câteva exemple ce vor fi detaliate în cele ce urmează. Ne propunem, astfel, să analizăm problema caracterizării stării emoționale pe baza poziției corpului uman și să adoptăm o soluție de timp real și de mare acuratețe, bazată pe metode specifice viziunii computerizate. Metoda dezvoltată va putea fi integrată ulterior în sisteme de tipul mediilor inteligente, ce se adaptează în mod automat necesităților utilizatorului, exemple elocvente fiind casele inteligente, sălile de clasă inteligente.

LIMBAJUL CORPORAL. PROBLEMA RECUNOAȘTERII STĂRILOR EMOȚIONALE PE BAZA GESTURILOR CORPORALE

Limbajul corporal reprezentat prin poziția și mișcările specifice ale corpului, capului, membrelor, direcția privirii, este purtător de informație în ceea ce privește starea emoțională a indivizilor umani. O poziție înclinată, aplecată a corpului poate releva o stare de oboseală, supărare, îndoială; o poziție dreaptă a corpului poate releva o stare de normalitate; dacă aceasta este însoțită de mișcările bruște ale membrelor superioare sau de o expresie specifică a feței, poate exprima o stare de nervozitate. Astfel, limbajul corporal se compune din gesturi corporale, voluntare sau involuntare, la realizarea cărora participă întregul corp. În Tabelul 1 sunt ilustrate următoarele posibilități de interpretare a limbajului corporal (limbaj comportamental non-verbal) [7]. Astfel, comportamentul non-verbal este unul imediat, instinctiv și mai puțin controlat, spre deosebire de cel verbal, punând

mai bine în evidență adevăratele atitudini și sentimente ale interlocutorilor [7].

Tabelul 1. Gesturi corporale și semnificațiile acestora

Gest corporal	Semnificație
Mers hotărât, poziția corpului rămânând dreaptă	Încredere
În picioare, poziție dreaptă a corpului, mâinile pe șolduri	Pregătit de acțiune; posibil agresivitate
Așezat, cu picioarele încrucișate, atingându-se ușor	Plictiseală
Brațele încrucișate pe piept	Apărare
Plimbare cu mâinile în buzunar	Nepăsare
Mâna la bărbie	Evaluare, gândire
Atingere ușoară a nasului	Respingere, îndoială, lipsă de sinceritate
Frecarea ochiului	Îndoială, neîncredere
Ținerea capului în mână, privirea în jos	Plictiseală sau supărare
Așezat cu mâinile la ceafă sprijinind capul, picioarele încrucișate	Superioritate/relaxare
Palma deschisă	Sinceritate, deschidere, inocență
Privirea în pământ, fața întoarsă	Neîncredere

Percepția stărilor emoționale se poate face și în mod automat, prin metode computerizate. Există o ramură a științei calculatoarelor, numită calcul afectiv, care are ca obiectiv dezvoltarea unor dispozitive și metode pentru recunoașterea stărilor afective și a emoțiilor umane. Această ramură este, de fapt, una interdisciplinară, acoperind și domeniile psihologiei și științelor cognitive. Originile acesteia se regăsesc în anii 1995, odată cu elaborarea lucrării lui Rosalind Picard [2]. Detecția informației emoționale debutează prin intermediul unor senzori pasivi, care captează date despre starea fizică sau comportamentul utilizatorului. Aceste date de intrare sunt similare cu cele pe care indivizii umani le utilizează atunci când își percep reciproc emoțiile. O altă direcție a acestei științe este cea corespunzătoare proiectării unor dispozitive computaționale, capabile de simularea stărilor emoționale. Procesarea vorbirii emoționale presupune recunoașterea stării emoționale a utilizatorului prin analiza șabloanelor de vorbire. Detecția și analiza expresiei faciale sau a gesturilor corporale este realizată tot cu ajutorul senzorilor și al detectorilor [2].

METODE EXISTENTE PENTRU RECUNOAȘTEREA STĂRILOR EMOȚIONALE PRIN INTERMEDIUL VIZIUNII COMPUTERIZATE

Viziunea computerizată oferă metode performante, atât sub aspectul timpului de procesare, cât și sub cel al acurateței prelucrării adecvate recunoașterii limbajului corporal caracteristic stărilor emoționale. Avantajul este acela că utilizatorul nu este afectat în nici o privință, senzorul fiind camera video, imaginile preluate de către aceasta fiind analizate prin intermediul sistemului computerizat. Metodele de analiză și recunoaștere specifice acestei ramuri a viziunii computerizate se bazează pe următoarele etape: 1. Detecția corpului uman în scena încărcată de obiecte;

2. Descompunerea corpului uman în părți componente și analiza poziției, orientării, respectiv formei acestor componente. În [2] autorii propun segmentarea corpului uman, urmată de calculul unor metrici capabile de a ne informa atât cu privire la volumul ocupat de corpul respectiv, cât și cu privire la cantitatea mișcării corporale, aceste aspecte fiind în relație cu starea emoțională. Cantitatea mișcării, reprezentată prin produsul (masă corporală) * (viteză de mișcare), respectiv indexul de contracție, acesta fiind determinat prin dimensiunea dreptunghiului sau a cercului care circumscrie corpul uman, reprezintă exemple elocvente ale acestor măsuri. În [6] se realizează detecția conturului corpului uman, urmată de analiza formei acestui contur și descompunerea sa în segmente corespunzătoare părților componente ale corpului uman. Detecția unui punct de separație, corespunzător unei încheieturi, se face pe baza determinării unui punct de minim local al unei curbe Spline. În [8] detecția conturului corpului uman implică următoarele etape: (1) extragerea muchiilor; (2) conectarea muchiilor și conversia acestora la linii drepte; (3) formarea segmentelor; (4) recunoașterea părților componente ale corpului uman. Astfel, se extrage așa-numitul schelet, compus din trunchiul individului, membrele superioare și respectiv cele inferioare ale acestuia, apoi se estimează orientarea fiecăreia dintre aceste segmente corporale. O altă metodă, ce determină scheletul de tip stea al corpului uman („star skeleton”) este propusă în [4]. După o etapă de preprocesare, bazată pe binarizarea imaginii, se realizează apoi eliminarea golurilor interioare prin intermediul operației de dilatare, urmată de o operație de eroziune, apoi având loc extragerea conturului. Scheletul de tip stea, ce unește centrul individului cu extremitățile acestuia, format din trunchi, mâini, picioare și cap, se obține prin determinarea acelor puncte de pe contur ce corespund maximelor locale ale distanței conturului față de centru. În scopul unei caracterizări multiscalare a variației posturii corpului uman, în [10] autorii utilizează transformata Wavelet. Aceasta este urmată de determinarea unor armonici sferice, corespunzând unor momente ce caracterizează distribuția 3D a voxelilor. Variațiile acestor momente corespund variației poziției părților componente ale corpului uman. Metodele de recunoaștere propriu-zisă, ce utilizează drept intrări fie trăsături referitoare la părțile componente ale corpului uman, fie trăsături multiscalare, sunt reprezentate atât prin clasificatori de tipul celor Bayesieni [10], cât și prin modele din teoria încrederii [9].

METODA PROPUȘĂ PENTRU CARACTERIZAREA POZIȚIEI CORPULUI UMAN ȘI RECUNOAȘTEREA STĂRILOR EMOȚIONALE

Caracterizarea și recunoașterea stărilor emoționale pe baza poziției corpului uman, respectiv a posturii părților componente ale acestuia, o realizăm efectiv prin aplicarea unei metode care furnizează atât acuratețe, cât și rapiditate a răspunsului. Metoda se bazează pe aplicarea unui algoritm de skeletonizare, urmat de estimarea orientării părților componente ale scheletului de tip stea, constituit prin unirea centrului siluetei umane cu extremitățile capului, membrilor superioare și respectiv ale celor inferioare. Această estimare a orientării se poate face prin intermediul histogramei orientării muchiilor, sau

în mod direct, obținându-se un vector de trăsături format din valorile orientărilor segmentelor individuale. Acest vector de trăsături este apoi furnizat la intrările unui clasificator de tip Rețea Bayesiană sau Perceptron Multinivel.

Algoritmii de skeletonizare implică următoarele etape:

1. Binarizarea imaginii – transformarea acesteia într-o imagine alb-negru. Pentru generalitate, utilizează algoritmul de binarizare cu prag adaptiv pe baza distanței Manhattan [5].
2. Aplicarea operatorului morfologic de dilatare [5] a obiectului de interes din imagine (corpul uman), în scopul eliminării golurilor din interiorul obiectelor.
3. Aplicarea operatorului morfologic de eroziune [5], pentru aducerea obiectului de interes la dimensiunile inițiale.
4. Determinarea centrului de masă al obiectului de interes [5].
5. Detecția conturului obiectului de interes [5], prin scăderea imaginii erodate din imaginea obținută prin binarizare.
6. Determinarea distanțelor punctelor de pe contur la centrul de masă și găsirea punctelor de pe contur corespunzând maximelor locale ale distanței față de centru [4].
7. Unirea centrului de masă cu punctele de maxim local și determinarea scheletonului de tip stea [4].

Există următoarele posibilități de caracterizare efectivă a posturii individului:

a. Binarizarea imaginii, apoi detecția de muchii și construirea histogramei orientării muchiilor, urmată de compararea histogramei corespunzătoare diferitelor stări emoționale [5].

b. Detecția părților componente ale corpului uman – trunchiul, membrele superioare și cele inferioare, în felul următor: trunchiul unește punctul superior de maxim local al conturului cu centrul acestuia, picioarele unesc punctele inferioare de maxim local cu centrul, iar celelalte două segmente reprezintă membrele superioare [4]. Poziția normală a scheletonului uman fiind cea verticală, se calculează orientările (unghiurile) segmentelor față de orizontală, valorile acestora formând un vector de trăsături presupus a conține informație relevantă în ceea ce privește caracterizarea stărilor emoționale. Acest vector de trăsături este descris în cele ce urmează:

$$V_t = \{\text{orientare_trunchi}, \text{orientare_ms}_1, \text{orientare_ms}_2, \text{orientare_mi}_1, \text{orientare_mi}_2\} \quad (1)$$

Orientare_{ms₁} și orientare_{ms₂} corespund orientărilor membrelor superioare, iar orientare_{mi₁} și orientare_{mi₂} corespund orientărilor membrelor inferioare. Ambele metode au fost experimentate și comparate. S-a experimentat și posibilitatea recunoașterii automate a stărilor emoționale cu ajutorul clasificatorilor. A existat o fază de învățare a clasificatorului, pe baza unui set de antrenament conținând trăsături ale unor indivizi cu stări emoționale cunoscute, respectiv faza de clasificare propriu-zisă, constând în recunoașterea unei stări emoționale necunoscute. Clasificatorii aleși spre utilizare au fost cei bazați pe rețele neuronale artificiale, aceștia fiind inspirați din percepția umană, capabili de a reprezenta orice tip de funcție de separație între clase. S-a utilizat metoda Perceptronului Multinivel (Multilayer

Perceptron – MLP), în conjuncție cu algoritmul de învățare bazat pe propagare inversă (“back-propagation”) [3]. Rețelele Bayesiane de Încredere (Bayesian Belief Networks) reprezintă o metodă de clasificare bazată pe construirea unui graf de dependențe între trăsăturile individuale, respectiv între trăsături și parametrul clasă [3]. Aceste dependențe se exprimă prin probabilități condiționate. S-a experimentat și acest clasificator, acuratețea acestuia comparându-se cu cea a perceptronului multinivel.

REZULTATE EXPERIMENTALE

În ceea ce privește experimentele realizate, am considerat imagini de tip bitmap, în reprezentarea 24 biți/pixel ale unor indivizi ce exprimă următoarele stări emoționale: supărare, îngândurare, furie, stare normală. Fiecare clasă conține 10 instanțe (imagini ale unor indivizi distincți). Câteva rezultate ale aplicării algoritmului de skeletonizare sunt ilustrate în Figura 1.









Imagine originală	Skeleton	Stare emoțională
		normal
		supărat
		furios
		gânditor

Figura 1. Reprezentarea stărilor emoționale prin intermediul scheletonului corpului uman

S-a luat în considerare doar bustul indivizilor umani, întrucât s-a estimat că poziția acestuia este relevantă în mare măsură în ceea ce privește caracterizarea stării emoționale. Pozițiile membrelor inferioare, care sunt specifice diferitelor stări emoționale, vor fi analizate într-o abordare ulterioară. În figura următoare sunt ilustrate histogramele orientărilor muchiilor construite pe imaginile binarizate ale indivizilor caracterizați printr-o anumită stare emoțională. Se observă, atât din forma scheletonului, cât și din cea a histogramei orientării muchiilor, că în poziția normală orientarea trunchiului individului este apropiată de direcția verticală (90°), în timp ce în celelalte stări emoționale poziția trunchiului este, de obicei, înclinată. Unghiul format cu axa orizontală are valoare maximă în cazul supărării (în jur de 30°), poziția corpului fiind mai puțin înclinată în cazul individului gânditor, și apropiată de poziția verticală în cazul stării de furie.

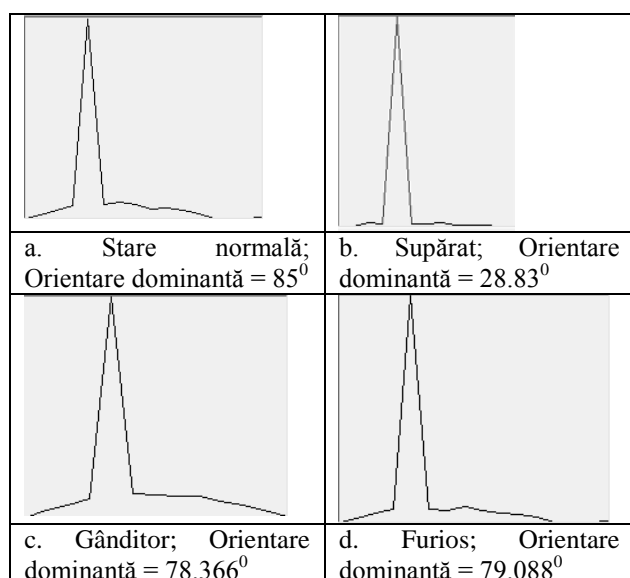


Figura 2. Reprezentarea stărilor emoționale prin intermediul histogramei orientării muchiilor

În ultimul caz observăm, în schimb, o variație semnificativă a pozițiilor membrilor superioare față de starea normală, starea de furie fiind caracterizată, astfel, prin încordarea trunchiului și printr-o mobilitate crescută a membrilor. Clasificatorii au fost experimentați în mediul Weka 3.5 [11]. Vectorul de trăsături a fost constituit din valorile a trei unghiuri: unghiul format de trunchiul individului cu axa Ox, precum și unghiurile formate de membrele superioare cu această axă. S-a utilizat metoda Perceptronului Multinivel, cu rata de învățare 0.2 și momentul $\alpha = 0.9$, urmărindu-se atât realizarea unui proces de învățare care să sesizeze toate variațiile importante, dar care să evite supra-antrenarea (luarea în considerare a zgomotelor), cât și trecerea rapidă peste platourile suprafeței de învățare. Metoda Rețelelor Bayesiene de Încredere a fost utilizată în conjuncție cu metoda de căutare genetică și cu estimatorul BMA (Bayesian Mean Average), care determină valorile probabilităților condiționate din rețea pe baza teoriei bayesiene [11]. Rezultatele aplicării acestor clasificatori sunt ilustrate în tabelul de mai jos:

Tabelul 2. Parametrii de acuratețe privind clasificarea stărilor emoționale

Metoda	Rata de clasificare corectă	Senzitivitate	Specificitate	Aria sub curba ROC
MLP	95.9%	90.3%	100%	97.8%
Rețele Bayesiene	83.33%	66.7%	100%	88.9%

Se observă că acuratețea clasificării este mai bună în cazul metodei Perceptronului Multinivel (MLP), rezultând faptul că acesta este mai adecvat setului de date caracteristice stărilor emoționale.

CONCLUZII ȘI POSIBILITĂȚI DE DEZVOLTARE

S-a elaborat o metodă bazată pe viziune artificială pentru recunoașterea stărilor emoționale ale indivizilor umani, acuratețea și timpul de răspuns fiind satisfăcătoare. Metoda bazată pe estimarea orientării segmentelor

individuale ale corpului uman s-a dovedit a fi mai performantă în ceea ce privește acuratețea decât cea bazată pe histograme ale orientării muchiilor. Metoda a fost experimentată pe imagini statice, dar poate fi ușor extinsă la cazul imaginilor dinamice, reducându-se la analiza imaginilor statice individuale ce fac parte dintr-o secvență de imagini. Studiul valorilor specifice ale parametrilor și a dinamicii acestora în cazul individului aflat în mișcare, caracterizat printr-o anumită stare emoțională, este un subiect de interes pentru dezvoltările viitoare. În abordările ulterioare, se vor lua în considerare și orientările membrilor inferioare, analizându-se relevanța acestor trăsături în ceea ce privește caracterizarea stărilor emoționale. Independența metodei față de orientarea individului este un alt obiectiv important: aceasta se poate obține considerând orientările relative ale segmentelor corporale, și nu orientările absolute ale acestora. De asemenea, se va trata problema detecției individului uman în scene complexe din lumea reală, elaborându-se metode adecvate pentru recunoașterea obiectelor, respectiv extragerea fundalului.

REFERINȚE

1. Camurri A., A. Coglio, An architecture for emotional agents, IEEE Multimedia, oct 1998, pp. 24-33.
2. Clay A., N. Couture, L. Nigay, Emotion capture based on body postures and movements, Proceedings of the International Conference on Computing and e-systems, Hammamet, Tunisia 2007, pp. 1-20.
3. Duda R., Pattern Classification (2nd ed), Wiley Interscience, 2003.
4. Fujiyoshi H., Real-time human motion analysis by image skeletonization, IEEE Trans. Inf.&Syst., Vol. E78-D, Nr. 1, ian. 2004, pp. 113-120.
5. Jain A.K., Fundamentals of Digital Image Processing, Englewood Cliffs, NJ: Prentice Hall, 1989.
6. Jaume A., J. Varona, M. Gonzalez-Hidalgo, R. Mas, F. Perales, Automatic human body modeling for vision based motion capture, Short Communication Proceeding, WSCG'2006, Czech Republic, pp. 12-17
7. Martinovich-Lardner C., How body language can be used to understand others, Michigan Bar Journal, oct. 2002, pp. 36-39.
8. Mojarrad M., Abassi Dezfouli M., Rahmani A., Detection and recognition of the human body composition in the images, Proceedings of the world academy of science, engineering and technology, Vol. 4, oct. 2008, pp. 872-874.
9. Girondel V., L. Bonnaud, A. Caplier, M. Rombaut, Static human body posture recognition using the belief theory, Proceedings of IEEE International Conference on Image Processing –ICIP, Genoa, Italia, 2005, Vol. 2, pp. 45-48.
10. Werghi, N., Xiao, Y., Recognition of human body posture from a cloud of 3D data points using Wavelet Transform coefficients, Procs of 5-th IEEE International Conference on Automatic Face and Gesture Recognition, 2002, Washington, pp. 70-75.
11. Weka 3, Data Mining Software in Java, <http://www.cs.waikato.ac.nz/ml/weka/>, 2004