

Navigation System in Virtual Environments Using Human Gestures

Oana Mihaela Vultur, Ștefan Gheorghe Pentiu

Abstract—Human gestures express feelings, intentions, replacing or accompanying speech. Also we can use human gestures in communication with our computer or navigation in a virtual environment, like a building. In this paper I will try to present a solution for real-time gesture recognition in order to navigate and interact with a virtual world.

Index Terms—Computer Vision, Gesture recognition, Human Computer Interaction (HCI), human gestures, kinect sensor

I. INTRODUCERE

GESTURILE reprezintă mișcări ale mâinii, capului, brațelor care exprimă o idee, un sentiment, o intenție, înlocuind uneori vorbele sau dând mai multă expresivitate vorbirii [1].

Gesturile transmit informație și sunt însoțite de semnificații și conținut. Diferite studii de natură psiholingvistică au fost realizate în scopul înțelegerii comunicării de natură gestuală, studii care asigură un excelent material pentru domeniul interacțiunii om – sistem informatic [2,3,4]. Astfel, Cadoz [2] folosește termenul de canal de comunicare gestual, identificând trei tipuri de gesturi (trei roluri funcționale diferite asociate gesturilor) deci o clasificare după funcția îndeplinită:

- **gestul ergonomic** acționează asupra mediului, derivând din noțiunea de a modela lumea fizică. Este gestul ce poate fi utilizat pentru interacțiunea cu obiectele unui sistem de realitate virtuală.
- **gestul epistemic** oferă informații cu privire la temperatură, presiune, calitatea suprafeței unui obiect, formă, orientare, greutate (simțul tactil). Descoperirea mediului este realizată prin experiență tactilă.
- **gestul semiotic** produce un mesaj informațional pentru mediu, cu rolul de a comunica informație. Este tipul de gest pentru operații de tip da/nu, afirmare/negare, etc. într-un dialog cu un sistem informatic.

De asemeni, gesturile pot fi folosite și pentru a comunica cu un sistem informatic, sau pentru a naviga într-un mediu

virtual (cum ar fi o clădire, un campus universitar, un labirint, un joc). În această lucrare am încercat să descriu principiile de funcționare ale unui sistem de navigare în mediul virtual prin intermediul gesturilor, sistem virtual reprezentat de corpul E al Campusului Universitar.

II. ARHITECTURA HARDWARE ȘI SOFTWARE

A. Arhitectura hardware

Aplicația rulează pe un computer ce are un procesor Intel (R) Core (TM) 2 Quad CPU Q9650 la o frecvență de 3GHz, având 4 GB memorie RAM. S-au folosit două monitoare LCD și un senzor Kinect.

Senzorul Kinect este un dispozitiv orizontal atașat pe un mic suport prevăzut cu un pivot motorizat și este proiectat pentru a fi așezat în lungime deasupra sau dedesubtul ecranului video. Este prevăzut cu o cameră RGB, senzor de adâncime și cu o matrice de microfoane care asigură captarea integrală a mișcării corpului în spațiu, recunoașterea facială și vocală.



Fig. 1. Senzorul Kinect – părți componente.

B. Arhitectura software

Arhitectura software este compusă din:

- Sistem de operare Windows 7 Professional (32 biți)
- Microsoft Visual Studio 2010 Professional
- Microsoft Kinect SDK
- Aplicația care realizează achiziția imaginilor, scheletizarea, recunoașterea gesturilor efectuate de utilizator, transmiterea simulării apăsării de taste
- Aplicația de vizualizare 3D a corpului E din cadrul Campusului Universitar.



Fig. 2. Aplicația de vizualizare 3D a corpului E din cadrul Campusului Universitar.

III. MODELE ȘI ALGORITMI DE RECUNOAȘTERE A GESTURILOR

Recunoașterea gesturilor se face în două etape: etapa de învățare și etapa de recunoaștere.

Datele sunt furnizate de către codul aplicației ca un set de puncte, numite poziții schelet („skeleton positions”) ce compun un schelet așa cum arată în figura 3:

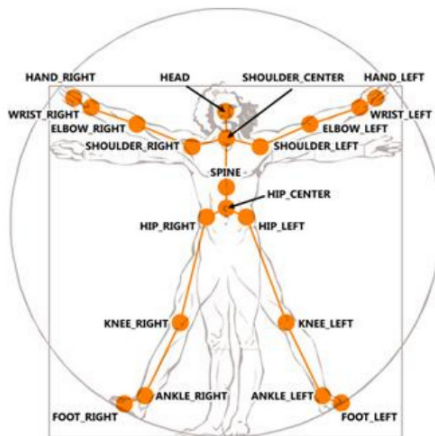


Fig. 3. Pozițiile încheieturilor scheletului relative la corpul uman[5].

Etapă de învățare constă în construirea setului de învățare, urmată de scrierea gesturilor în fișier. În etapa de învățare sunt urmărite șase puncte de interes din cadrul scheletului: încheietura mâinii drepte, cotul drept, umărul drept, încheietura mâinii stângi, cotul stâng, umărul stâng. Sunt memorate în fișier coordonatele x,y corespunzătoare celor șase puncte de interes urmărite.

În etapa de recunoaștere sunt comparate gesturile noi cu secvența de gesturi din setul de învățare, aplicând algoritmul DTW (Dynamic Time Warping). Acest algoritm calculează distanța minimă DTW dintre două secvențe de coordonate – secvența model și secvența candidat (a fi un gest). DTW găsește cea mai bună potrivire (*best match*) dintre cele două secvențe.

Învățare/Recunoaștere:

- procesare imagine
- segmentare

- scheletizare
- citire coordonate puncte de interes
- transformarea coordonatelor spațiului scheletului în coordonate relative la schelet.

Transformarea în coordonate relative la schelet este:

$$centru.x = (um\bar{a}rStg.x + um\bar{a}rDr.x) / 2$$

$$centru.y = (um\bar{a}rStg.y + um\bar{a}rDr.y) / 2$$

$$d = \sqrt{(um\bar{a}rStg.x - um\bar{a}rDr.x)^2 + (um\bar{a}rStg.y - um\bar{a}rDr.y)^2}$$

$$pct[i].x = (pct[i].x - centru.x) / d$$

$$pct[i].y = (pct[i].y - centru.y) / d$$

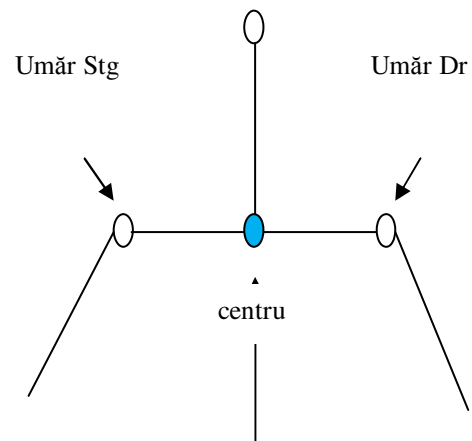


Fig. 4. Partea superioară a scheletului uman – umărul stâng, umărul drept, precum și centrul (funcție de care se face normalizarea coordonatelor x, y ale umărului stâng, respectiv ale umărului drept)

IV. DICȚIONARUL DE GESTURI UTILIZAT

Vom folosi șapte gesturi pentru a naviga în mediul virtual. Acestea sunt:

- 1) Gestul efectuat cu mâna dreaptă (intitulat VINO) pentru a ne deplasa înaintea
- 2) Gestul mâinii drepte (intitulat DU-TE) pentru a ne deplasa în spate
- 3) Gestul efectuat cu mâna dreaptă (balansarea mâinii drepte) pentru a ne deplasa la dreapta
- 4) Gestul efectuat cu mâna stângă (balansarea mâinii stângi) pentru a ne deplasa la stânga
- 5) Ridicarea mâinii stângi pentru a urca la un nivel superior din clădire
- 6) Coborârea mâinii stângi pentru a merge la un nivel inferior din clădire
- 7) Gestul efectuat cu ambele mâini pentru a înainta cu o viteză mai mare.

Gesturile sunt ilustrate în figurile 5 - 11:

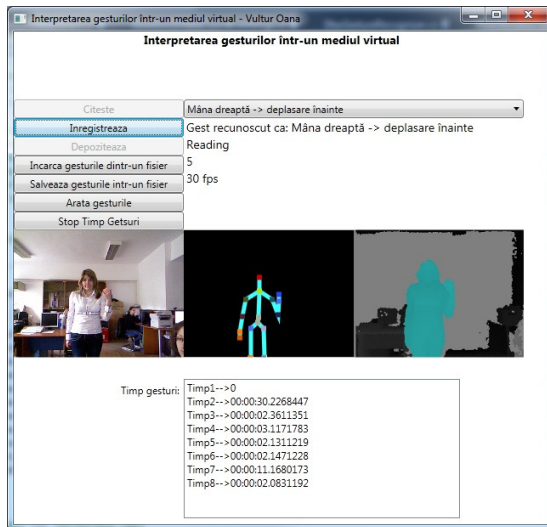


Fig. 5. Gestul efectuat cu mâna dreaptă pentru a ne deplasa înainte.

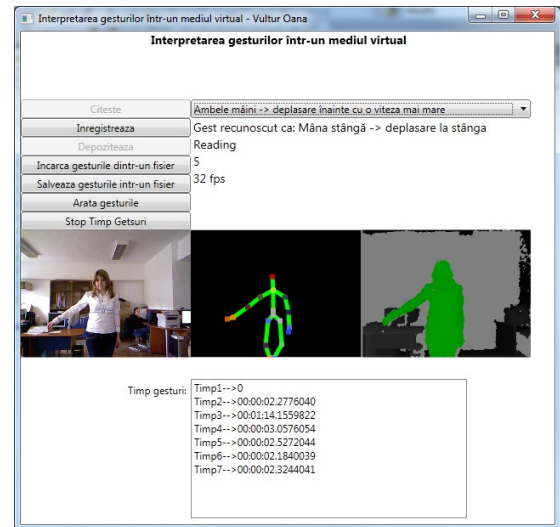


Fig. 8. Gestul efectuat cu mâna stângă (balansarea mâinii stângi) pentru a ne deplasa la stânga.

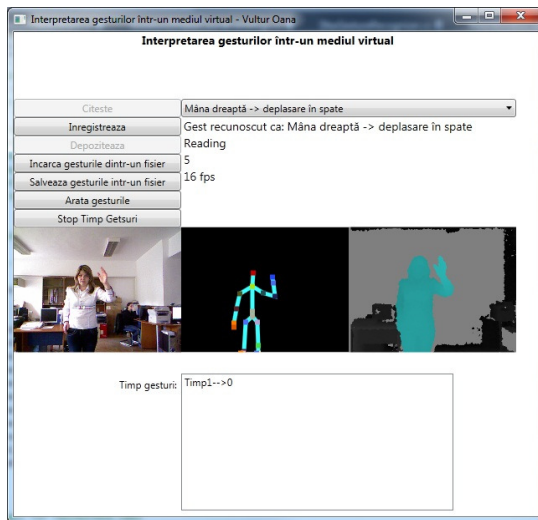


Fig. 6. Gestul efectuat cu mâna dreaptă pentru a ne deplasa în spate.

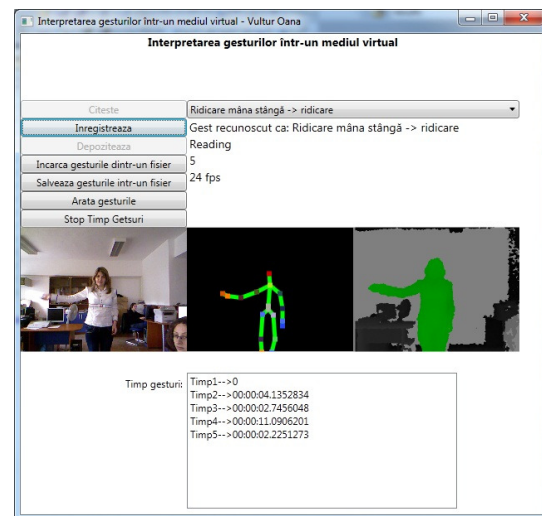


Fig. 9. Ridicarea mâinii stângi pentru a urca la un nivel superior din clădire.

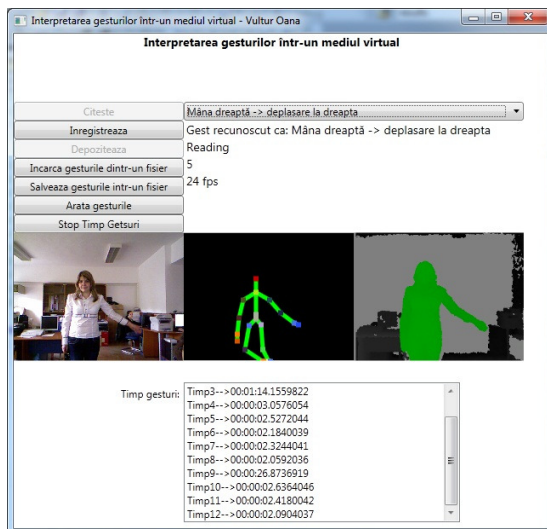


Fig. 7. Gestul efectuat cu mâna dreaptă (balansarea mâinii drepte) pentru a ne deplasa la dreapta.

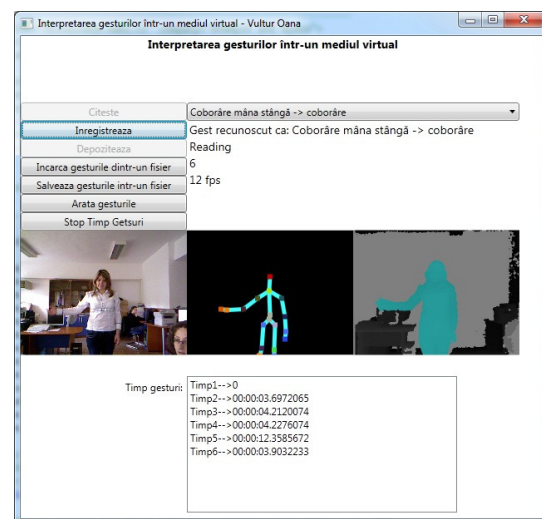


Fig. 10. Coborârea mâinii stângi pentru a merge la un nivel inferior din clădire.

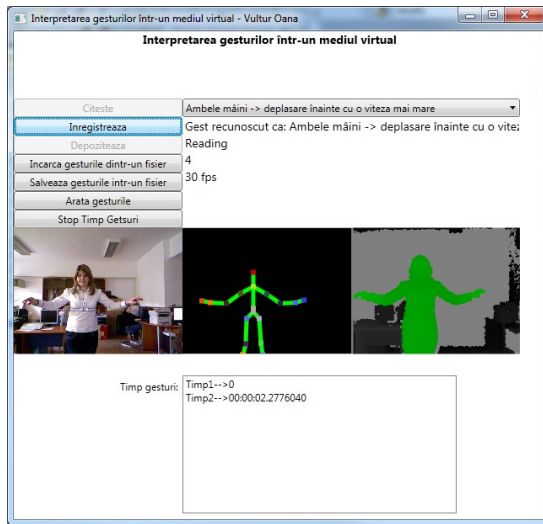


Fig. 11. Gestul efectuat cu ambele mâini pentru a înainta cu o viteză mai mare.

V. FUNCȚIONAREA APLICAȚIEI

Funcționarea aplicației este ilustrată în figura 12:

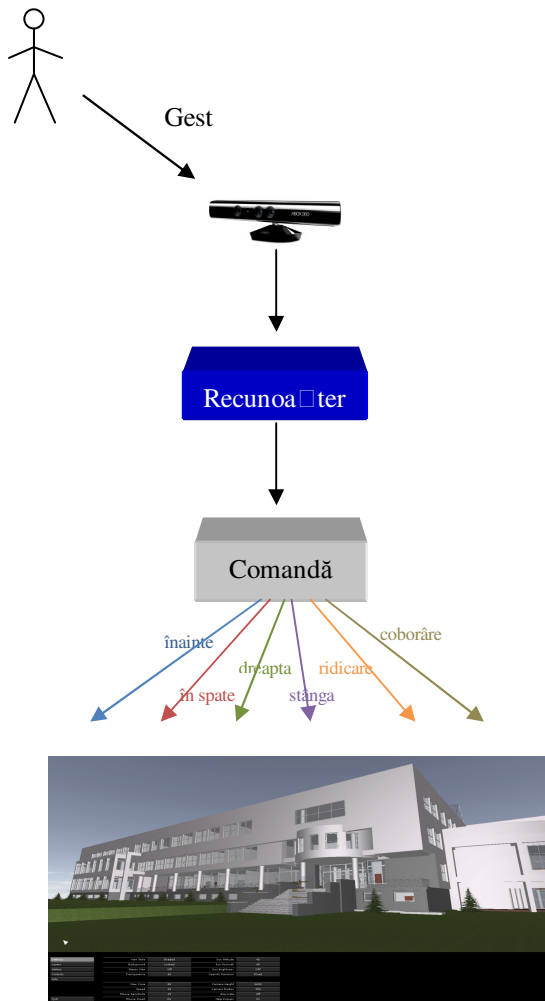


Fig. 12. Funcționarea aplicației (lanțul gest-recunoaștere-comandă).

Gesturile efectuate de utilizator sunt recunoscute cu ajutorul algoritmului amintit în secțiunea III, după care este simulată apăsarea tastelor (W, A, S, D, PageUp, PageDown, SHIFT+W), această simulare fiind transmisă ferestrei care deține focusul (în cazul nostru aplicația de vizualizare 3D a corpului E al Campusului Universitar).

REFERINȚE

- [1] Dicționarul explicativ al limbii române, Academia Română, Institutul de Lingvistică "Iorgu Iordan", Editura Univers Enciclopedic, 1998.
- [2] C. Cadoz, „Les réalités virtuelles”, Dominos, Flammarion, 1994.
- [3] A. Kendon, „How gestures can become like words”. In Potyatos, F. (ed), Crosscultural perspectives in nonverbal communication, p 131-141. Toronto, Canada: Hogrefe, 1988.
- [4] David McNeill, „Hand and Mind. What Gestures Reveal about Thought”, 1992.
- [5] Microsoft Kinect for Windows SDK beta – Programming Guide
- [6] P. Senin, „Dynamic Time Warping Algorithm Review”, 2008 pp.3-9