

R E Z U M A T

al tezei de doctorat

DYNAMIC HAND GESTURES RECOGNITION FOR

HUMAN COMPUTER INTERFACES

elaborată de ing. Gheorghe Daniel POPA

Conducător științific Prof.univ.dr.ing. Marius OTEȘTEANU

Teza este redactată în limba engleză și abordează un subiect de interes major – interacțiunea om-calculator bazată pe recunoașterea gesturilor dinamice.

1. Motivația și obiectivele tezei

Dezvoltarea calculatoarelor și a sistemelor ce utilizează procesoare pe parcursul ultimelor decenii a făcut posibilă pătrunderea acestora în mai toate aspectele vieții moderne. A apărut astfel necesitatea dezvoltării unor interfețe om-computer cât mai naturale. Gesturile, fiind un important mijloc de comunicare între oameni, pot sta la baza dezvoltării unor interfețe naturale om-computer. Din acest motiv, recunoașterea gesturilor a captat – și continuă să capteze – atenția multor cercetători în domeniul interacțiunii om-computer (*Human Computer Interaction* – HCI). Interfețele om-computer bazate pe dispozitive care urmăresc vizual mișcările utilizatorului se bucură de o largă popularitate, iar teza „*Dynamic Hand Gestures Recognition for Human Computer Interfaces*” propune noi soluții pentru implementarea unor astfel de interfețe.

Teza are ca obiectiv global dezvoltarea unui sistem de recunoaștere a gesturilor dinamice efectuate cu mâna pentru interfețe om calculator. Pentru a permite utilizarea metodei pe orice calculator cu dotări medii fără a necesita echipamente suplimentare costisitoare, recunoașterea gesturilor trebuie să se realizeze exclusiv pe baza informației vizuale, achiziționate cu ajutorul unei camere web.

În ultimii ani camerele web au devenit din ce în ce mai răspândite, pe fondul îmbunătățirii calității imaginilor furnizate de acestea, concomitent cu scăderea prețurilor.

În prezent astfel de camere sunt integrate la marea majoritate a dispozitivelor mobile comercializate, respectiv în cazul sistemelor *desktop*, *webcam*-urile pot fi achiziționate sub formă de echipamente periferice la un preț scăzut, fiind larg utilizate pentru aplicații precum comunicații video prin internet.

Deși subiectul recunoașterii gesturilor pe baza imaginilor achiziționate cu o singură cameră a preocupat comunitatea științifică încă din anii '90, marea majoritate a soluțiilor disponibile fie nu sunt suficient de robuste, fie au cerințe foarte mari de putere de calcul, fie necesită utilizarea unor camere cu performanțe mult superioare celor atinse de *webcam*-urile uzuale din prezent, având astfel o aplicabilitate limitată. Calitatea secvențelor video achiziționate cu ajutorul *webcam*-urilor s-a îmbunătățit semnificativ în ultimii ani, dar imaginile furnizate sunt în continuare relativ zgomotoase, iar numărul de cadre pe secundă este redus, adesea apărând fenomenul de *motion blur* în special la mișcări rapide și/sau în condiții de iluminare redusă. Acești factori sporesc gradul de dificultate al recunoașterii și urmăririi unui obiect de interes, iar în cazul unui sistem de recunoaștere a gesturilor dinamice efectuate cu mâna, urmărirea mâinii este o etapă esențială pentru recunoașterea cu succes a gesturilor.

Obiectivele asumate în teză vizează fiecare din cele 3 componente principale ale unui sistem de recunoaștere a gesturilor dinamice: preprocesarea imaginilor, urmărirea mâinii și recunoașterea gesturilor pe baza traiectoriei.

Preprocesarea imaginilor presupune prelucrarea imaginilor provenite de la camera web și extragerea de caracteristici pe baza cărora să se poată realiza urmărirea mâinii. La nivelul acestei etape un prim obiectiv este identificarea caracteristicilor relevante care să permită urmărirea mâinii. Caracteristicile avute în vedere sunt culoarea, caracterul de obiect de prim plan al mâinii în cazul în care se efectuează gesturi dinamice, precum și caracteristici geometrice și de proporționalitate. Pentru extragerea caracteristicilor sunt necesare operații de filtrare, conversii în diferite spații de culoare, segmentare în funcție de culoare, segmentare prim-plan/fundal.

Urmărirea mâinii presupune identificarea poziției acesteia în fiecare cadru utilizând atât datele provenite din etapa de extragere a caracteristicilor cât și pozițiile anterioare ale mâinii și informații derivate din acestea (viteza și direcția de deplasare). La nivelul acestei etape obiectivul principal este dezvoltarea unui algoritm robust de urmărire, care să ruleze în timp real. Un obiectiv secundar este dezvoltarea unei metode de

inițializare a tracker-ului care să minimizeze riscul unor inițializări accidentale, și totodată să permită o modelare inițială mai precisă a țintei.

Ultima etapă este cea în care traiectoria provenită de la tracker este prelucrată și se decide dacă utilizatorul a făcut un gest semnificativ sau nu, respectiv, atunci când este cazul se identifică care gest a fost efectuat. La nivelul acestei etape obiectivele sunt de definire a unui set de gesturi, care să faciliteze recunoașterea, dar totodată să poată fi executate natural, dezvoltarea unui algoritm de filtrare a traiectoriilor provenite de la tracker și dezvoltarea unei metode de recunoaștere a gesturilor pe baza traiectoriei filtrate. Filtrarea traiectoriilor este necesară pentru netezirea acestora, deoarece traiectoriile furnizate de tracker sunt afectate de zgomot (au un aspect rugos) datorită unei multitudini de factori cum ar fi imperfecțiunile camerei, variația condițiilor de iluminare sau modificarea părții mâinii expusă spre cameră datorită unei rotații.

2. CONTINUTUL TEZEI

Teza de doctorat este structurată în 6 capitole și Bibliografie (117 titluri).

În primul capitol, *Introduction*, este definită problema abordată, după care sunt prezentate motivația, obiectivele și structura tezei. Capitolul este încheiat de lista lucrărilor științifice publicate de autor în domeniul de cercetare abordat în teză. Lista cuprinde 7 lucrări științifice: 3 lucrări publicate în reviste de specialitate și 4 lucrări comunicate la conferințe internaționale.

Capitolul 2, *Non Parametric Density Estimation*, prezintă o sinteză a tehnicilor non-parametrice de estimare a densității de probabilitate. În prima parte a capitolului sunt prezentate noțiuni fundamentale de analiză statistică pentru seturi de date uni-variaabile și multi-variaabile și estimatorii non-parametrici de bază: histograme, poligoane de frecvență și histograme deplasate mediate.

O secțiune importantă a acestui capitol este dedicată estimatorilor pe bază de nucleu. Sunt prezentate cele mai larg utilizate nuclee: uniform, triunghiular, Epanechnikov, normal, bi-ponderat și tri-ponderat. De asemenea este evidențiată problema alegerii lățimii de bandă a nucleului cu soluțiile cunoscute din literatura de specialitate.

În încheierea capitolului 2 sunt prezentate câteva concluzii referitoare la avantajele metodelor non-parametrice și problemele ridicate de implementarea practică a acestora.

Capitolul 3, Video Tracking, abordează problematica urmăririi obiectelor în secvențe video. În prima parte a capitolului sunt prezentate principalele categorii de soluții cunoscute din literatură pentru urmărirea video în general: soluții bazate în principal pe reprezentarea și localizarea țintei (algoritmul mean-shift) și soluții bazate în principal pe filtrare și asocierea datelor (filtre cu particule). În continuare sunt prezentate soluții particularizate pentru urmărirea mâinii cunoscute din literatura de specialitate, cu evidențierea avantajelor și dezavantajelor fiecărei categorii de soluții.

Noțiunile teoretice și soluțiile cunoscute din literatura de specialitate prezentate în capitolele 2 și 3 constituie o bază pentru dezvoltarea soluțiilor noi propuse în capitolele următoare.

Capitolul 4 Robust Integration of Multiple Features for Hand/Finger Tracking prezintă soluțiile propuse în cadrul tezei pentru problemele legate de urmărirea degetului/mâinii și de inițializarea urmăririi. Sunt identificate caracteristicile alese (culoare, prim-plan) pentru a sta la baza procesului de detecție și de urmărire a degetului și soluțiile alese pentru extragerea acestora din secvențele de imagini.

În sub-capitolul 4.2 este prezentată soluția pe care am propus-o pentru inițializarea semi-automată a urmăririi mâinii folosind un automat finit. Pentru detecția mâinii sunt utilizate multiple caracteristici: culoare, prim-plan, forma și proporționalitate, iar pentru a evita inițializările nedorite se impun câteva constrângeri spațiale și temporale.

În sub-capitolele 4.3 – 4.5 este prezentată metoda pe care am propus-o pentru urmărirea degetului. Robustețea metodei de urmărire propuse este asigurată prin combinarea mai multor caracteristici de bază (culoare, prim-plan și caracteristici geometrice) pentru a genera o reprezentare sparse, urmată de mai multe etape succesive de filtrare și *clustering*. Nivelurile superioare de procesare ale algoritmului includ o metodă de predicție a poziției și dimensiunilor degetului – pe baza evoluției acestuia în cadrele anterioare – și identificarea degetului urmărit dintre mai mulți

candidați (dacă este cazul) pe baza distanței între poziția și dimensiunile candidaților față de cele obținute prin predicție.

În finalul capitolului, sunt prezentate scenariile de test folosite pentru evaluarea performanțelor algoritmilor propuși și rezultatele obținute, fiind evidențiate avantajele acestora, inclusiv prin comparații cu alte soluții cunoscute.

Capitolul 5, *Real-Time Dynamic Hand Gesture recognition*, prezintă soluțiile noi propuse în cadrul tezei pentru recunoașterea gesturilor pe baza traiectoriei mâinii extrasă de un algoritm de urmărire. Este subliniată importanța modului de definire a gesturilor pentru simplificarea procesului de recunoaștere. Gesturile recunoscute sunt definite ca succesiuni de 2 – 4 segmente cvasiliniare. Prelucrarea traiectoriei include operații de netezire, eliminare a valorilor aberante (filtru median) și segmentare folosind algoritmul *mean-shift* cu nucleu 2D, anizotrop. Recunoașterea gesturilor se realizează foarte eficient pe baza reprezentărilor sintactice ale traiectoriei și prototipurilor gesturilor. Rezultatele experimentale scot în evidență eficiența algoritmului de recunoaștere și a metodei de filtrare și segmentare a traiectoriei.

Ultimul capitolul, *Conclusions and Future Developments*, prezintă mai întâi câteva concluzii referitoare la noile soluții de urmărire a degetului/mâinii și de recunoaștere a gesturilor bazate pe traiectorie prezentate în teză. În continuare, este prezentă o sinteză a contribuțiilor care stau la baza tezei. În final, sunt enumerate câteva sugestii pentru continuarea cercetării.

3. CONTRIBUȚII

Teza de doctorat propune soluții noi și eficiente pentru rezolvarea problemelor legate de principalele componente ale unui sistem de recunoaștere a gesturilor. Contribuțiile originale pot fi grupate în 2 categorii: contribuții referitoare la procesul de urmărire a degetului/mâinii și contribuții referitoare la recunoașterea gesturilor.

Principalele contribuții referitoare la procesul de urmărire a degetului/mâinii sunt:

- o metodă de inițializare a urmăririi degetului/mâinii care oferă rezultate repetabile și evită inițializările accidentale, nedorite; mai multe caracteristici sunt utilizate într-un automat finit pentru inițializarea urmăririi;

- o metodă de urmărire a degetului robustă care funcționează în timp real; Soluția propusă utilizează multiple caracteristici, care sunt supuse unei succesiuni de etape de prelucrare pentru a detecta degete ale mâinii și pentru a identifica degetul urmărit.

Principalele contribuții referitoare la procesul de recunoaștere a gesturilor pe baza traiectoriei mâinii sunt:

- o metodă de definire a setului de gesturi bazată pe construirea gesturilor din succesiuni de segmente de traiectorie cvasiliniare cu reprezentări simbolice simplificate;
- o soluție de prelucrare a traiectoriilor utilizând mai multe etape de netezire a traiectoriei, urmate de segmentarea acesteia pe baza algoritmului mean-shift; algoritmul de prelucrare a traiectoriilor propus, împreună cu metoda de definire și reprezentare simbolică a gesturilor conduc la o recunoaștere facilă a gesturilor bazată pe comparație sintactică.

Noile soluții propuse pentru procesul de urmărire au fost testate, în condiții variate de fundal și iluminare și cu expunere la multipli factori perturbatori. De asemenea noile soluții propuse pentru recunoașterea gesturilor au fost testate pe un număr mare de traiectorii obținute cu ajutorul unui algoritm de urmărire, conducând la recunoașterea cu succes a gesturilor chiar și în condițiile unor traiectorii cu rugozitate accentuată.