

**ACADEMIA DE ȘTIINȚE A MOLDOVEI
INSTITUTUL DE MATEMATICĂ ȘI INFORMATICĂ**

Cu titlu de manuscris
C.Z.U: 519. 95

ALBU VEACESLAV

**RECUNOAȘTEREA ACȚIUNILOR UMANE ÎN BAZA
REȚELELOR NEURALE MODULARE**

**SPECIALITATEA: 122.03
MODELARE, METODE MATEMATICE, PRODUSE PROGRAM**

Autoreferatul tezei de doctor în informatică

Chișinău, 2016

Teza a fost elaborată în laboratorul "Sisteme informatice" al Institutului de Matematică și Informatică, Academia de Științe a Moldovei.

Conducător științific: COJOCARU Svetlana, doctor habilitat în informatică,

Referenți oficiali:

VAGHIN Vadim, doctor în științe tehnice, prof., Institutul de Energetică, Moscova.

CĂPĂȚÂNĂ Gheorghe, doctor în informatică, prof., Universitatea de Stat din Moldova.

Membrii Consiliului științific specializat:

GAINDRIC Constantin, *Președinte*, dr. hab. în informatică, prof., m.c. AȘM, IMI,

CIUBOTARU Constantin, *Secretar științific*, dr. în informatică, conf.univ., IMI,

COSTAȘ Ilie, dr. hab. în informatică, prof., ASEM

GUȚULEAC Emilian, dr. hab. în informatică, prof., UTM,

AVERKIN Alexei, candidat în șt. tehnice, conf. univ., Centrul de Calcul al Academiei de Științe a Rusiei,

BURȚEVA Liudmila, dr. în informatică, conf. cerc, IMI,

ȚIȚHIEV Inga, dr. în informatică, conf.univ., IMI.

Susținerea va avea loc la 9 noiembrie 2016, la orele 15.00 în ședința Consiliului științific specializat D 01.122.03 - 03 pe lângă Institutul de Matematică și Informatică, Academia de Științe a Moldovei, str. Academiei 5, Chișinău, MD-2028, Republica Moldova.

Teza de doctor și autoreferatul pot fi consultate la Biblioteca Centrală "Andrei Lupan" a Academiei de Științe a Moldovei și pe pagina web a C.N.A.A. (www.cnaa.md).

Autoreferatul a fost expediat la ____ octombrie 2016

Secretar științific al Consiliului științific Specializat: _____ Ciubotaru C.

Conducător științific:

Dr. hab. în informatică, prof. cerc. _____ Cojocaru S.

Autor: _____ Albu V.

1. REPERELE CONCEPTUALE ALE CERCETĂRII

Actualitatea și importanța problemei de recunoaștere a emoțiilor și gesturilor

Oamenii posedă abilitatea remarcabilă de a recunoaște obiectele foarte exact printr-o simplă privire. Cu toate acestea, atunci când studiem procesele neurale, care stau la baza acestor acțiuni, ele par a fi extrem de complicate: recunoașterea în cortexul vizual primar implică mai multe domenii și o complexitate relativ mare de procesare. Un sistem artificial, care va încerca să imite toate funcțiile sistemului de recunoaștere naturală, va fi fie prea complicat pentru elaborare sau va necesita o complexitate de calcul, care este greu de atins.

Este cunoscut faptul, că sistemele artificiale, în particular cele de recunoaștere, operează, de regulă, cu probleme simplificate și în cercetarea noastră vom modela doar principiile funcționale generale ale organizării neurale a zonelor vizuale. Astfel, vom încerca să realizăm plauzibilitatea neurofiziologică și să menținem un nivel pe cât e posibil de înalt al detalierei. Mai mult ca atât, cerința de recunoaștere a obiectelor în mișcare în timp real, adică recunoaștere nu doar a imaginilor statice, ci și a unui flux video în timp real, adaugă o complexitate suplimentară prin includerea componentei temporale în procesul de recunoaștere.

Abordarea noastră pentru soluționarea problemei complexe de recunoaștere a emoțiilor și gesturilor constă în elaborarea unei arhitecturi noi de rețea neurală artificială (RNA) pentru clasificarea gesturilor și emoțiilor umane, obținute de la camere video cu infraroșu. Semnalul de ieșire al acestor camere servește în calitate de intrare pentru rețeaua propusă, obținându-se clasificarea reacțiilor persoanelor în tipice vs. atipice în raport cu un anumit mediu. RNA propusă în lucrare poate servi drept un instrument robust de clasificare a emoțiilor și gesturilor unui subiect uman în tipice vs. atipice pentru un anumit tip de interacțiune în timp real prin utilizarea metodelor existente cu algoritmi de învățare, ce provin din arhitecturi de rețele neurale (RN) biologic plauzibile.

Stadiul actual și problemele din domeniul recunoașterii emoțiilor și gesturilor

Clasificarea emoțiilor. În domeniul recunoașterii emoțiilor, există trei probleme principale, care necesită o clarificare. Prima problemă conceptuală dificilă, menționată de mulți cercetători, este însuși conceptul de emoție. Printre întrebările care apar aici una importantă este modul în care se face distincția dintre emoție și alte fațete ale experienței umane. Lipsa unei definiții clare a emoției a cauzat multe dificultăți pentru cei care încearcă să studieze fața și emoția. Vom analiza definiții din cercetarea clasică în domeniul recunoașterii și clasificării emoțiilor, precum și cele propuse de unii cercetători contemporani pentru a alege cea mai

potrivită definiție care poate servi scopurilor noastre. O altă problemă conceptuală dificilă este specificarea exactă a emoțiilor. De unde știm dacă informațiile furnizate prin expresia feței sunt exacte? Există vreun criteriu pentru a determina ce emoție a fost experimentată efectiv? În secțiunea experimentală, am realizat o serie de experimente psihologice cu subiecți umani, în scopul de a defini emoția exactă a persoanei din păreri personale și din comentariile observatorilor umani.

Aceste două probleme sunt privite în mod independent cu cel de-al doilea și cel mai important aspect: cum să fie recunoscută emoția și acțiunea în timp real, dintr-un flux video? În scopul rezolvării acestei probleme, folosim cunoștințele din neuroinformatică pentru a construi modelul nostru. În această teză vom interpreta emoțiile ca ghiduri sau interferențe pentru comportamente și luarea deciziilor, care este posibil de evaluat prin măsurarea caracteristicilor faciale vizibile. Există o serie de modele de emoții dezvoltate pentru diferite scopuri, cum ar fi formalizare, calcul sau înțelegere. Toate modelele de emoții pot fi clasificate în discrete și continue. Modelele discrete lucrează cu seturi limitate de emoții. Cel mai cunoscut și cel mai larg utilizat model discret de emoții a fost dezvoltat de Paul Ekman [1]. El a elaborat modelul său pe parcursul mai multor ani și a încheiat cu șase emoții de bază: furie, dezgust, frică, fericire, tristețe și surpriză. Majoritatea lucrărilor în acest domeniu se reduc la recunoașterea a cinci emoții (precum dezgust, frică, bucurie, surpriză, tristețe, furie). Urmând cercetările efectuate de Ekman și Friesen [2]. Pe de altă parte o manifestare pură a acestor emoții apare destul de rar în viața de zi cu zi, starea emoțională a persoanei fiind caracterizată printr-un spectru destul de larg de expresii. De regulă, emoțiile sunt manifestate prin acțiuni minore care modifică trăsăturile faciale, cum ar fi ridicarea colțurilor buzelor în caz de bucurie sau de coborâre oblică a lor în caz de tristețe. Prin urmare, în lucrarea noastră folosim datele din propriile noastre experimente psihologice pentru a defini expresiile faciale [3]. Expresiile faciale sunt accesate pe două căi: prin referința personală a unui subiect uman sau prin judecata unui observator uman. Cu toate acestea, folosim etichetele propuse de Ekman în lucrarea sa, cu excepția celor pe care nu le-am observat de-a lungul experimentelor.

Clasificarea acțiunilor. Recunoașterea acțiunii umane este procesul de etichetare a secvențelor de imagini cu etichete de acțiune. Soluțiile robuste ale acestei probleme își găsesc aplicații în domenii cum ar fi supravegherea vizuală, recuperarea video și interacțiunea om-calculator. Un astfel de subiect de cercetare este o adevărată provocare din cauza variațiilor de performanță ale mișcării, setările de înregistrare și diferențele inter-personale. Un număr de încercări s-au făcut pentru a aborda clasificarea video în timp real cu rețelele neurale. Unul

dintre cele mai recente progrese în acest domeniu aparține lui Karpathy et al [4]:. ei au studiat performanța rețelelor neurale convoluționale (RNC) în clasificări video de scară largă. Autorii au demonstrat că arhitecturile (RNC) sunt capabile să învețe caracteristici din datele slab etichetate, ceea ce asigură performanțe mai înalte decât metodele bazate pe caracteristici și că aceste beneficii sunt surprinzător de robuste pentru detalierea conectivității arhitecturilor în timp. De asemenea, ei au sugerat că poate fi necesară o prelucrare mai atentă a mișcării camerei video (de exemplu, prin extragerea caracteristicilor din sistemul local de coordonate a unui punct de identificare). În sistemul nostru această problemă nu există, din moment ce camera este fixată, iar utilizatorul este de obicei localizat în aceeași poziție în fața camerei în infraroșu. În mod corespunzător sunt abordate și alte probleme cu aplicarea RNC profunde pentru efectuarea clasificării intrărilor video. De asemenea, datele de ieșire ale camerei cu infraroșu (hărțile detaliate) le vom folosi ca intrare în sistemul nostru, fapt ce ne va permite simplificarea procesului de recunoaștere și obținerea unei exactități sporite.

Scopul principal al tezei

Scopul principal al cercetărilor prezentate este de a dezvolta un instrument de clasificare a reacțiilor umane (inclusiv emoții și acțiuni) în tipice și atipice în timp real, într-un anumit mediu. Acest instrument oferă observații statistice și măsurători ale stărilor emoționale umane în timpul unei sesiuni de interacțiune cu un produs software (implementat într-o platformă hardware puțin complementa). Folosind metode de computer vision și algoritmi de machine learning, emoțiile sunt înregistrate, recunoscute și analizate pentru a se obține un feedback statistic al emoțiilor per total ale unui număr de obiective într-un anumit interval de timp. În mod similar, vom clasifica acțiunile subiecților umani, pe care un utilizator le poate efectua în timpul interacțiunii cu o componentă al unui complex de software/hardware și vom oferi o clasificare a acțiunilor sale. Feedback-ul, produs de sistemul propus, poate oferi măsuri importante pentru utilizator ca răspuns la un sistem ales. Un exemplu de aplicare a acestei cercetări este un sistem ce constă dintr-o cameră de luat vederi încorporată într-un dispozitiv care este utilizat în mod frecvent, cum ar fi un ATM. Noi folosim înregistrările camerei de luat vederi pentru a captura starea emoțională a clienților (fericit, trist, neutru, etc) și de a construi o bază de date a utilizatorilor și a emoțiilor înregistrate pentru a fi analizate ulterior. În scopul studiului, am dezvoltat și testat un complex de software/hardware, pe care îl folosim pentru a efectua experimente psihologice.

Obiectivele lucrării

Principalele obiective de cercetare ale lucrării pot fi formulate în modul următor:

1. Dezvoltarea unui instrument de clasificare a emoțiilor și acțiunilor unui subiect uman în două grupuri (tipice vs. atipice) pentru un anumit tip de interacțiune. Noi propunem arhitectura rețelei neurale pentru clasificarea gesturilor și emoțiilor umane, obținute de la camerele cu infraroșu. Semnalul de ieșire al camerei servește ca o intrare în rețeaua propusă, care clasifică reacțiile umane în tipice vs. atipice în timpul unei interacțiuni cu un anumit tip de mediu. Aici, termenul ”reacție” se referă la combinația emoțiilor și mișcărilor corpului, efectuate de un subiect uman.

În scopuri de cercetare, am ales un număr foarte limitat de stări emoționale și modele comportamentale pentru studierea doar a tipului de astfel de comportament standard: interacțiunea unui utilizator cu un echipament ATM tipic, deoarece el ne oferă modele foarte distincte de expresii faciale ”tipice” și ”atipice”. Pe parcursul acestui studiu, am observat comportamentul subiecților umani în timpul interacțiunii standard cu ATM versus interacțiune non-standard.

Analiza automatizată a acestor comportamente cu tehnici de învățare ne-a permis de a forma un complex de rețele neurale convoluționale (RNC) pentru a face o concluzie cu privire la comportamentul unui utilizator prin clasificarea atât a mișcărilor corpului cât și a caracteristicilor faciale. Un astfel de feedback poate oferi măsuri importante pentru răspunsul utilizatorului în timpul unei interacțiuni cu orice sistem ales cu un număr limitat de gesturi implicate. Noi folosim camere cu infraroșu pentru a detecta automat caracteristicile și mișcărilor membrelor, în scopul de a clasifica comportamentul utilizatorului în tipic sau atipic pentru tipul de acțiune pe care îl efectuează.

Scopul acestei lucrări este de a analiza acțiunile persoanei în timpul interacțiunii cu o interfață a utilizatorului și punerea în aplicare a algoritmului, care va fi capabil să clasifice comportamentul uman (normal vs. anormal) procesând în timp real semnalul senzorului infraroșu.

2. Analiza abordărilor existente de ultimă generație, utilizate atât pentru clasificarea emoțiilor cât și a gesturilor, pentru a alege una, care va fi robustă, cu performanțe ridicate și va permite recunoașterea caracteristicilor selectate; dezvoltarea și testarea a două tipuri de algoritmi, care ar putea fi aplicați într-un astfel de sistem și compararea rezultatele acestor studii. Alegerea a două tipuri de rețele neurale este condiționată de faptul, că sunt analizate două tipuri de caracteristici (trăsături faciale și gesturi) în mod simultan, ceea ce presupune volume de calcul substanțiale. Noi sugerăm utilizarea

rețelelor neurale profunde în combinație cu rețeaua de bază a funcției radiale (detaliile vor fi furnizate în capitolul doi). Cu toate acestea, al doilea tip de rețele neurale ar putea fi utilizat independent pentru acest tip de probleme.

3. Efectuarea experimentelor comportamentale, în scopul de a evalua felul în care sistemul propus poate detecta în mod eficient comportamentul normal vs. anormal al unui client în timpul interacțiunii cu ATM și deducerea concluziei cu privire la aplicabilitatea sistemului propus în scopuri industriale/comerciale.

Metodologia cercetării

Pe tot parcursul studiului, vom introduce *două metode principale de cercetare*, ce vor fi utilizate pentru a construi software-ul. Ambele metode provin din teoria rețelelor neurale, de aceea în capitolul doi o vom expune-o în detalii, oferind notația matematică detaliată pentru fiecare parte a modelului, inclusiv algoritmul de învățare. Algoritmii de învățare pe care îi folosim pentru cele două părți ale sistemului sunt foarte similari, diferă însă în unele detalii. Noi folosim unele concepte din domeniul machine learning, deoarece acesta a constituit cea mai mare parte a acestui studiu.

Noutatea și originalitatea științifică a tezei constă într-o arhitectură nouă de rețele neurale modulare, constituite din două părți separate combinându-le rezultatele în scopul efectuării unei clasificări a datelor obținute de la senzori infraroșii. Acesta este un prim sistem de acest fel aplicat atât pentru recunoșterea emoțiilor faciale, cât și a acțiunilor umane.

Mai exact, noi propunem o combinație a celor mai recente tehnici biometrice cu abordarea RN pentru analiza în timp real a emoțiilor și comportamentului. Tehnici de recunoaștere a emoțiilor și acțiunilor au fost prezentate separat în mai multe studii în ultimii cinci ani. Cu toate acestea, o abordare holistică nu a fost prezentată până acum. Mai mult decât atât, noi prezentăm algoritmul nostru într-un cadru de aplicare pentru soluționarea unei problemei concrete.

Importanța teoretică

Rezultatele cercetării fundamentează soluționarea următoarelor probleme: stabilirea arhitecturii instrumentarului pentru clasificarea fiabilă a emoțiilor și gesturilor unui subiect uman în tipice vs. atipice; stabilirea posibilității și eficienței utilizării învățării profunde în cadrul unei abordări integrate pentru identificarea expresiilor întregului corp uman în timp real.

Din acest punct de vedere, contribuția noastră este de dublă: noi oferim o nouă arhitectură de rețea neurală, constituită din două părți separate, ce combină rezultatele sale

pentru a efectua clasificarea intrărilor senzorilor cu infraroșu. Din cunoștințele noastre, acesta este primul sistem de acest gen, fiind aplicat la recunoașterea acțiunii umane și emoțiilor. Unele părți ale acestui sistem (cum ar fi procesarea video, recunoașterea emoțiilor cu rețelele convoluționale, etc.) au fost puse în aplicare înainte, dar realizarea este nouă. Mai mult decât atât, algoritmi existenți au fost modificați în mare măsură (de exemplu, algoritmul convențional SOM) pentru scopurile acestui studiu.

Valoarea aplicativă a lucrării

Aplicațiile acestei abordări sunt posibile în diverse domenii, inclusiv sisteme de securitate, sisteme de supraveghere cu camere video, date biometrice etc.

Problema științifică importantă soluționată constă în elaborarea unei metode multimodale de clasificare a reacțiilor umane (unind emoțiile și acțiunile) în tipice și atipice în raport cu un mediu dat, fapt care asigură funcționarea eficientă în timp real a unor sisteme de monitorizare a acțiunilor umane.

Rezultatele științifice principale înaintate spre susținere

Performanța integrală a sistemului, bazată pe rezultate experimentale, poate fi rezumată după cum urmează:

- 1) Este propusă arhitectura modului de bază al rețelei, ce cuprinde hărți cu auto-organizare (SOM) de funcții de bază radiale (RBF), este prezentată baza ei matematică. Abordarea propusă este nouă din punct de vedere al arhitecturii sistemului și implementării algoritmului de învățare. După câte știm, această arhitectură nu a fost niciodată aplicată în soluționarea problemei de recunoaștere a emoțiilor.
- 2) A fost demonstrată posibilitatea de a adapta arhitectura rețelei neurale convoluționale la un nou tip de procesare a semnalelor de intrare (infraroșu). S-a demonstrat că un astfel de tip de arhitectură este capabil să rezolve sarcina noastră (procesarea acțiunilor) în timp real.
- 3) Modelul RN dezvoltat este capabil să recunoască și să clasifice emoțiile și mișcările corpului în două tipuri (tipice și atipice). Recunoașterea expresiei faciale are rata de eroare de 8% și, respectiv, cea a mișcărilor - de 14%. Combinate, acestea reprezintă rata de recunoaștere a rezultatelor 99% pe tipul selectat de acțiuni. Odată cu creșterea numărului de acțiuni sau în cazul schimbării tipului de acțiune precizia sistemului s-ar putea să scadă cu 1- 1,5%.
- 4) Sistemul propus este capabil:

- Să captureze, să recunoască și să clasifice emoțiile și acțiunile unui subiect uman într-o manieră robustă;
- Să integreze recunoașterea emoțiilor și acțiunilor asigurând monitorizarea comportamentului uman în timp real, oferind rezultate mai robuste decât sistemele existente.

5) Rezultatele experimentale demonstrează că sistemul este potrivit pentru punerea în aplicare la dispozitivele ATM. Sistemul este gata pentru teste pe teren și ar putea fi pus în aplicare pentru testare într-un terminal ATM tipic.

Aprobarea rezultatelor științifice

Rezultatele au fost aprobate și publicate în lucrările următoarelor **conferințe internaționale**:

1. Conferința a 3-a a societății matematice a Republicii Moldova. Chișinău: Institutul de Matematică și Informatică, Academia de Științe a Moldovei, 2014;
2. Tendințe de dezvoltare ale științei contemporane: viziuni ale tinerilor cercetători. Chișinău, Republica Moldova, 2015;
3. Workshop Bazele Informaticii - FOI-2015, 24-29 august 2015, Chișinău, Republica Moldova;
4. The 7th International Multi-Conference on Complexity, Informatics and Cybernetics: IMCIC 2016, 8 -11 Martie 2016, Orlando, Florida, Statele Unite ale Americii.

Publicații la tema tezei

În baza rezultatelor cercetării au fost publicate **8 lucrări științifice** (4 articole în reviste științifice recenzate și 4 în cadrul materialelor conferințelor).

Conținutul tezei și structura. Teza este scrisă în limba engleză și culeasă la calculator ca un manuscris. Teza are următoarea structură: introducere, trei capitole, concluzii generale și recomandări, bibliografie (109 surse). Teza este expusă în 121 pagini de text de bază, 5 anexe, ilustrată cu 37 de figuri și 2 tabele.

Cuvinte cheie: computer vision; rețele neurale artificiale; rețele neurale convoluționale; recunoașterea emoțiilor; recunoașterea gesturilor; rețelele cu funcții de bază radiale; hărți cu auto-organizate; machine learning; procesarea datelor camerei cu infraroșu.

2. CONȚINUTUL TEZEI

În lucrare a fost dezvoltat un model de rețele neurale pentru recunoașterea mișcărilor corpului și a expresiei faciale și pentru clasificarea lor în două tipuri (tipice și atipice). O astfel de sarcină complexă a solicitat analiza stărilor emoționale ale subiectului uman, a întregului spectru de acțiuni pe care el le efectuează în anumite circumstanțe, precum și construirea și

implementarea modelelor matematice, potrivite pentru soluționarea acestor probleme. Problema generală a fost divizată în două sub-probleme.

În prima parte sunt descrise rețele neurale modulare, care se aplică la problema recunoașterii emoțiilor. Arhitectura modulului de bază al rețelei este harta auto-organizată (SOM) [5-7] de funcționale ale modulelor cu funcții de bază radiale (RBF). În acest scop este oferită o prezentare matematică succintă cu privire la acest subiect. În prima jumătate a capitolului doi se propune o descriere matematică detaliată a abordării aplicate. Este formalizată matematica modelului și propusă o argumentare cu privire la alegerea și implementarea modelului algoritmului de învățare. Implementarea algoritmului ca model bazat pe rețea neurală este prezentată în capitolul trei.

Cea de a doua parte prezintă descrierea rețelei neurale convoluționale pe care am folosit-o pentru clasificarea acțiunilor. Noi am folosit rețele neurale convoluționale profunde pentru clasificarea în timp real a mișcărilor corpului uman. Notăția matematică detaliată a arhitecturii rețelei, algoritmului de învățare și procesul de implementare și validare a modelului au fost restrânse.

În Introducerea tezei sunt descrise obiectivele cercetării și evidențiate rezultatele principalele în domeniul recunoașterii emoțiilor și acțiunilor, din punct de vedere al în ceea ce privește aspectelor psihologice și modelelor teoretice. Aici este oferită o scurtă trecere în revistă a sistemului și a componentelor sale majore. De asemenea, au fost evidențiate astfel de probleme importante, cum ar fi relevanța subiectului, scopul și obiectivele cercetării și descrierea metodologiei pe care o folosim în teză. La fel a fost efectuată descrierea stării curente a tehnologiei în domeniul recunoașterii emoțiilor și acțiunilor și identificarea problemelor de cercetare, care există în domeniu. Problema științifică importantă, care este rezolvată în cadrul tezei, este descrisă succint, din punct de vedere al semnificației sale teoretice și valorii aplicative a lucrării. O astfel de introducere generală are ca scop să facă lucrarea accesibilă pentru un spectru larg de cititori, cu experiență și cunoștințe diferite, din moment ce această lucrare combină rezultatele obținute atât din psihologie cât și din și matematica aplicată.

Capitolul 1 (Teoria și modelele de recunoaștere a emoțiilor și acțiunilor) este o parte introductivă. Acesta conține descrierea fundalului și prezentarea generală a studiilor importante cu privire la subiectele aferente. În prima jumătate a capitolului 1 este descrisă baza psihologică pentru modele de recunoaștere a emoțiilor și a acțiunilor. În a doua jumătate a capitolului este examinată literatura de specialitate cu privire la recunoașterea obiectelor. Pe

parcursul ultimelor decenii a fost propus un număr mare de modele de recunoaștere a obiectelor. Acestea diferă din mai multe puncte de vedere, de exemplu, prin numărul sau tipul de emoții pe care le recunosc sau în tehnicile de învățare automată. Într-un singur manuscris este dificil de a descrie toate modelele existente, astfel, este oferită doar o scurtă prezentare a principalelor tipuri de modele de recunoaștere ale obiectelor în conformitate cu abordarea pe care o folosesc. Un accent deosebit este pus pe abordările matematice existente în construcția de rețele neurale, aferente obiectului de studiu.

Capitolul 2 (arhitectura rețelei neurale și algoritmi de învățare)

Capitolul al doilea prezintă arhitectura modelelor de rețele neurale propuse pentru recunoașterea emoțiilor și acțiunilor. Acest capitol ar putea fi împărțit în două părți logice. Prima parte descrie rețele neurale modulare, pe care le aplicăm la recunoașterea emoțiilor. Cea de a doua parte prezintă descrierea rețelei neurale convoluționale pe care o folosim pentru clasificarea gesturilor. De asemenea, în prima parte se descrie arhitectura modulului de bază al rețelei: harta cu auto-organizare (SOM) de funcționale ale modulelor cu funcții de bază radiale (RBF). Este efectuată o introducere matematică cu privire la acest subiect. Prima jumătate a capitolului doi include descrierea matematică detaliată a abordării aplicate: formalizarea matematică a modelului și argumentarea alegerii și implementării modelului algoritmului de învățare. În continuare în același capitol este descris al doilea tip de arhitectură de rețele neurale (RN) pe care îl folosim în experimentele noastre. A doua parte descrie un algoritm, care este capabil, fără pre-procesarea avansată a caracteristicilor și nici o modelare specială sau învățare a lor, să extragă în mod automat și să învețe caracteristici importante din datele, pe care le clasifică eficient în una din cele două clase de gesturi. În această parte, noi descriem arhitectura pe rețele neurale convoluționale profunde, pe care le folosim pentru clasificare: notațiile matematice, algoritmul de învățare și procesul de implementare și validare a modelului.

Arhitectura modelului nostru se bazează pe noțiunea de hartă auto-organizată (SOM), propusă de Kohonen. Acest tip de rețele neurale este instruit folosind învățarea nesupravegheată pentru a produce o hartă bidimensională a spațiului de intrare a probelor de antrenare. Calitatea SOM de utilizare a unei funcții de proximitate pentru păstrarea proprietăților topologice ale spațiului de intrare este aplicată în simulările noastre pentru a crea harta similară cortexului IT. Algoritmul convențional SOM are o serie de restricții, principala din care este capacitatea sa de a se ocupa numai cu datele vectorizate. Pentru a rezolva această problemă, au fost propuse o serie de modificări ale SOM convenționale. Noi

am folosit una dintre aceste modificări ca bază pentru construirea modelului nostru. Această arhitectură are un număr de avantaje. În primul rând, fiecare modul în SOM modificat are capacitatea de procesare a informației și poate forma o hartă dinamică, care constă dintr-un ansamblu de module funcționale. În al doilea rând, RBF-SOM combină algoritmi de învățare supravegheați și nesupravegheați: la nivel RBF, rețeaua este antrenată de un algoritm de învățare supravegheat, adică cu propagarea înapoi la nivelul modulului RBF, în timp ce nivelul SOM superior este descris într-o manieră nesupravegheată. Pentru scopurile acestui studiu am folosit module de rețea RBF. Utilizarea RBF în locul MLP adaugă următoarele proprietăți la o astfel de rețea păstrând în același timp capacitatea de a forma o hartă dinamică: 1) nu este nevoie de un algoritm pentru a evita minimele locale; 2) rețeaua poate recunoaște obiectul și poate păstra reprezentarea lui în centrul său interior. Algoritmul generalizat pentru procesarea SOM a modelelor funcționale poate fi aplicat și în acest caz. Arhitectura SOM a modulului RBF are o structură ierarhică: este alcătuită din două nivele, pe care le vom numi nivelul RBF și nivelul SOM al rețelei. La primul nivel, arhitectura rețelei noastre reprezintă rețelele RBF k , care sunt modificările rețelei Poggio și Edelman [8]. Din moment ce fiecare modul reprezintă o anumită "caracteristică funcțională" determinată de arhitectura model, nivelul SOM în SOM de RBF reprezintă o hartă a acestor caracteristici. Rețeaua propusă rezolvă o problemă de aproximare într-un spațiu de dimensiuni mari. Recunoașterea unui obiect este echivalentă cu găsirea unui hiperplan în acest spațiu, care oferă cea mai bună ajustare la un set de date de antrenare. Datele de antrenare reprezintă un vector cu coordonatele de proiecții 2D ale obiectelor 3D, luate la fiecare grad de rotație. Pentru a investiga capacitatea de a clasifica obiectele 3D complexe, cum ar fi fețe, vom extinde modelul nostru de SOM cu RBF prin adăugarea unui modul de pre-procesare ierarhică, prezentată în ierarhii de filtre cu diferite grade de rezoluție și straturi de pooling (Fig.1).

Ieșirea de rețea reprezintă harta de activare, activarea fiecărui modul arată apartenența expresiei detectate către una dintre cele cinci emoții de bază. În cadrul acestui studiu, am selectat cinci emoții de bază, care sunt localizate pe un plan pătrat, divizat în 25 părți. Modulul câștigător reprezintă emoția cea mai plauzibilă. Această abordare permite definirea emoției sau emoțiilor celor mai plauzibile (deoarece modulul cel mai activ poate fi definit între două emoții).

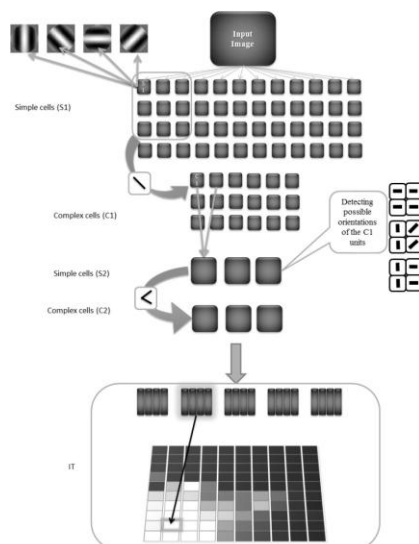


Figura 1. Arhitectura modului de pre-procesare a celulelor simple și complexe

În acest studiu, am folosit doar cinci emoții, utilizarea unui număr mai mare de etichete de emoții fiind de asemenea posibil (Fig.2).

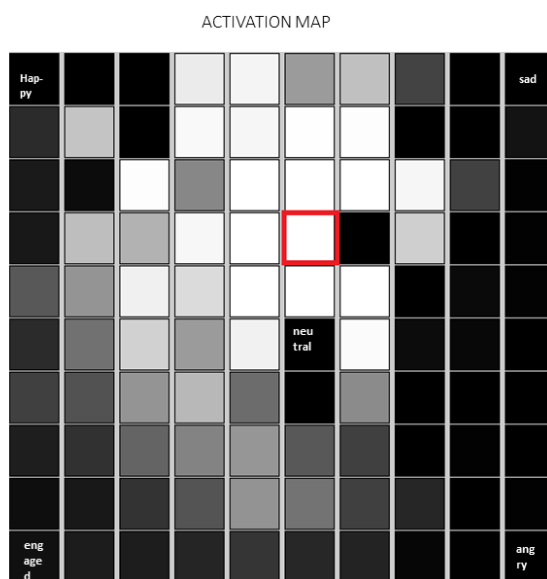


Figura 2. ieșirea rețelei RBFxSOM

Pentru a rezolva problema recunoașterii acțiunilor umane, folosim rețele neurale convoluționale [9] (RNC), a căror arhitectură este extrem de eficientă pentru clasificarea unei mari cantități de date. Prin termenul ”acțiuni” aici înțelegem mișcări ale părților corpului care au căzut în câmpul receptiv al camerei cu infraroșu, cu excepția feței.

O rețea neurală profundă [10] (RNP) este o rețea neurală artificială cu mai multe unități de straturi ascunse între straturile de intrare și de ieșire. Similar cu RN, RNP superficiale pot modela relații complexe non-lineare. Arhitecturi RNP, de exemplu, destinate

detectării și analizei obiectelor, generează modele în care obiectul este exprimat sub forma unei compoziții de straturi ale primitivelor imaginii. Straturile suplimentare permit compoziția caracteristicilor din straturilor inferioare, oferind posibilitatea de modelare a datelor complexe, cu mai puține unități decât o rețea cu performanțe similare de adâncime mică.

Arhitectura unui RNC poate fi descrisă în felul următor. O regiune de intrare mică merge spre neuronii de intrare și apoi se conectează la primul strat convoluțional ascuns (Fig.3).

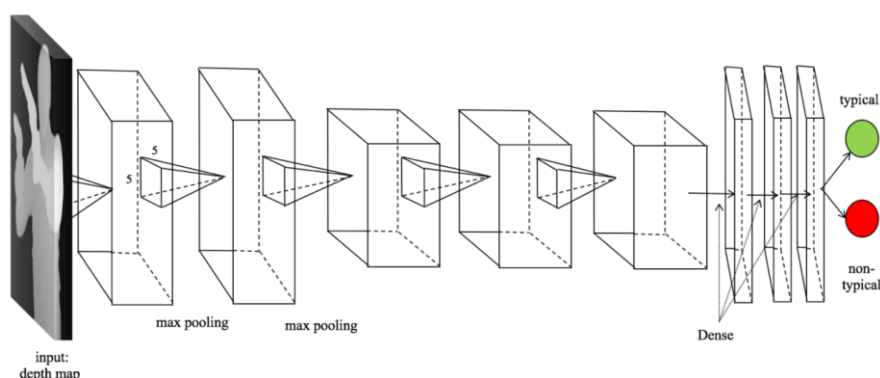


Figura 3. Arhitectura RNC

Intrarea în RNC-ul este o hartă profundă normalizată; ieșirea este o clasificare a acțiunii de intrare (tipic vs. atipic). Între stratul de intrare și cel de ieșire putem vedea un set de filtre ce pot fi învățate, ele sunt activate în timpul prezentării unui anumit tip particular de caracteristică în regiunea pixel din intrare. La aceasta fază, RNC face o deplasare de invarianță, care este efectuată de hartă caracteristicilor. Stratul de subeșantionare este următorul. Acolo avem două procese: calculul mediei locale și eșantionarea. Ca rezultat, vom obține rezoluția în scădere a hărții caracteristice. Pentru a îndeplini această sarcină RNC are nevoie de învățare supravegheată. Înainte de a începe experimentul, am selectat un set de clipuri video etichetate cu diferite experiențe emoționale. Sistemul analizează imaginile și găsește caracteristicile similare. Apoi, sistemul creează o hartă, în care aranjează videoclipurile în conformitate cu caracteristicile similare. Astfel, imaginile cu emoții similare formează o anumită clasă. Pentru a testa sistemul, vom adăuga alte înregistrări video și corecta sistemul atunci când face concluzii necorespunzătoare.

Modelul propus este format din trei straturi convoluționale, urmate de straturi de tip max-pooling, și trei straturi conectate complet cu un clasificator final prezentat cu MLP (cu

două ieșiri de bază, ce corespunde unui comportament tipic și atipic). Datele de intrare au fost prezentate ca ieșiri filtrate și normalizate a camerei cu infraroșu.

Pentru a procesa un video în timp real, noi propunem o arhitectură de flux dublu, care include rețelele spațiale și temporale (Fig.4). O astfel de RNC, antrenată pe un cadru multiplu de flux optic dens este capabilă de a obține performanțe foarte bune, în ciuda datelor limitate de antrenare, care sunt foarte oportune în cazul nostru. Pentru prima dată acest tip de procesare a fluxului dublu de imagini video a fost propus de [11].

În model, mișcarea este reprezentată în mod explicit folosind câmpul de deplasare a fluxului optic, calculat pe baza ipotezelor că intensitatea și netezimea fluxului sunt constante.

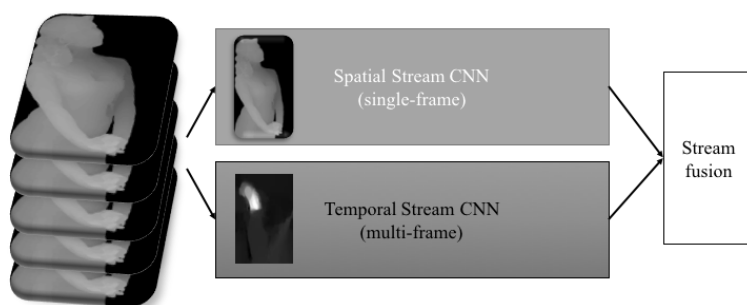


Figura 4. Arhitectura de flux dublu pentru clasificarea video

Pentru procesarea capturilor video în timp real propunem o arhitectură de flux dublu, care include rețelele spațiale și temporale. Este evident, că dintr-o secvență video este posibil să se separe componentele spațiale și temporale. Partea spațială va avea grijă de informațiile referitoare la scene și obiecte, care sunt reproduse prin cadre individuale. Partea temporală reproduce mișcarea observatorului (camerei) și a obiectelor prin intermediul mișcării prin cadre. Figura 4 ilustrează arhitectura propusă pentru clasificarea video în baza acestor principii, adică separarea în două fluxuri. Utilizarea scorurilor SoftMax combinate prin fuziune cu întârziere într-o RNC profundă face ca fiecare flux care urmează, să fie pus în aplicare. Noi luăm în considerare calculul mediei și antrenarea unei mașini suport vector (SVM) liniară multi-clasă ca metodă de fuziune.

Calculul s-a efectuat în sistemul Python. Modelul a fost instruit cu date de antrenare, iar evaluarea modelului s-a realizat pe datele de test cu validarea încrucișată k-fold (detalii sunt prezentate în compartimentul următor). Calculul s-a efectuat pe serverul Amazon EC2 (<https://portal.aws.amazon.com>).

Unitățile de procesare graphică (GPU) externe sunt folosite din următorul motiv: a fost stabilit, că o singură unitate GPU GTX 580 care conține numai 3 GB memorie, limitează mărimea maximă a rețelei ce poate fi antrenată pe ea. Pentru rețele cu dimensiuni mari,

precum este a noastră, este necesar de implicat mai multe unități GPU. GPU contemporane posedă un avantaj important exprimat prin capacitatea lor pentru paralelizare încrucișată, întrucât acestea sunt capabile să citească și să înscrie direct pe memoria alteia, fără a se adresa către memoria mașinii gazdă. Urmând experiențele cunoscute, noi aplicăm o schemă de paralelizare, care plasează o jumătate din neuroni pe fiecare GPU, respectând condiția că GPU comunică numai pe anumite nivele. Aceasta înseamnă, spre exemplu că neuronii de pe nivelul 3 primesc semnalul de intrare de pe toate hărțile nucleului de nivelul 2. Totuși, neuronii de pe nivelul 4 primesc semnalul de intrare numai de la hărțile de nivelul 3, care se află pe același GPU. Modelul de comunicare poate fi ajustat astfel, ca consumul de resurse pentru interconexiune să devină o fracțiune acceptabilă din volumul total de calcul.

Validarea modelului rețelei neurale a fost realizat prin tehnica validării încrucișate leave-one-out (LOOCV – leave one out cross validation). Folosirea validării încrucișate leave-one-out a fost necesară pentru estimarea corectă a nivelului de regularizare optimal și a parametrilor (ponderi de conexiune) a rețelei neurale obținute. Validarea încrucișată este un model de tehnică de validare pentru evaluarea modului în care rezultatele analizei statistice se vor generaliza într-un set independent de date. Validarea încrucișată leave-one-out este un caz particular de validare încrucișată leave-p-aut. Validarea încrucișată leave-p-aut (LpOCV) implică folosirea a p observații în calitate de set de validare și a celorlalte observații în calitate de set de antrenare. Acest fapt este repetat pentru toate căile de reducere a eșantionului original într-un set de validare a observațiilor p și un set de antrenare. Validarea încrucișată LpO necesită învățare și validare de Cnp ori (unde n constituie numărul observațiilor în eșantionul original). În validarea încrucișată leave-one-out presupunem $p = 1$. Totuși, pentru scopul nostru, validarea încrucișată leave-one-out s-a dovedit a fi foarte lentă. Prin urmare, validarea rezultatelor rețelei RNC a fost realizată prin tehnica validării încrucișate de k ori [12]. În validarea încrucișată de k ori, eșantionul original este divizat aleatoriu în k subeșantioane de dimensiuni egale. Un singur subeșantion din numărul k de eșantioane este reținut în calitate de informație de validare pentru testarea modelului, iar subeșantioanele rămase $k - 1$ sunt folosite în calitate de informație de antrenament. Procesul validării încrucișate este ulterior repetat de k ori, cu fiecare dintre cele k eșantioane folosite doar o singură dată în calitate de informație de validare. Ulterior se poate face media rezultatelor de k ori (sau pot fi combinate în alt mod) pentru a produce o singură estimare. Avantajul acestei metode față de sub-eșantionarea aleatorie repetată (vezi mai jos) constă în faptul că toate observațiile sunt folosite atât pentru antrenare, cât și pentru validare, iar fiecare observație

este folosită pentru validare numai o singură dată. De regulă este folosită validarea încrucișată de 10 ori, dar în general k rămâne un parametru nefixat. Când $k=n$ (numărul de observații), validarea încrucișată de k ori constituie exact ceea ce se numește validarea încrucișată leave-one-out.

Rezultatele celor două ieșiri a RN au fost unite într-un mod bazat pe reguli, permițându-ne să combinăm ieșirile în două subsisteme separate.

Capitolul 3 (Aplicații de cercetare și experimente psihologice) conține două secțiuni de bază: descrierea etapei experimentale pentru recunoașterea emoțiilor și, respectiv, recunoașterea acțiunilor.

Majoritatea cercetărilor se axează pe detectarea expresiilor faciale într-un cadru izolat, unde fiecare țintă este analizată separat. În lucrare este prezentăm un cadru colectiv pentru a analiza emoțiile de grup și comportamentul uman general.

Scopul cercetării noastre constituie folosirea camerelor cu infraroșu pentru capturarea imaginii utilizatorului. În studiul dat noi utilizăm Kinect API pentru a înregistra o bază de date cu mai multe obiective. Totodată și alte modele de camere pot fi folosite în acest scop. De asemenea, va fi creată o bază de date ce va conține emoțiile etichetate manual în scopul analizei și evaluării.

Noi propunem o arhitectură hibrid pentru o analiză complexă a evenimentelor. Analiza în timp real a reacțiilor umane (expresiile faciale și gesturile) este realizată cu ajutorul celor mai performante tehnici de învățare automată, descrise în capitolul doi. Măsurările rezultante sunt comparate cu datele statistice, înregistrate anterior și cu datele oferite de observatorul uman.

În scopul studiului noi am montat un suport hardware, care constă dintr-un terminalul ATM pe partea superioară a căruia a fost instalată o cameră cu infraroșu. A fost elaborat un simulator de procesare a cardurilor, similar cu cel folosit în terminale ATM. Softul i-a permis utilizatorului să realizeze una dintre cele patru operațiuni standard: depunerea banilor, extragerea banilor, transferul banilor pe alt cont și transferarea banilor pe contul de telefonie mobilă. Modelele pentru crearea soft-ului au fost luate de la câteva bănci populare și au fost combinate pentru a reprezenta un „terminal ATM standard”. Softul a fost scris în Java.

Utilizatorii au fost divizați în două grupuri: exemple de test pozitive și exemple de test negative. Subiecții din grupul „pozitiv” urmau să efectueze operațiuni standard prin intermediul terminalului ATM, în modul în care procedează de obicei, fără alte recomandări adiționale. Folosind acest complex de hardware și soft, subiectul uman, care a luat parte în

experimente, a fost rugat să întreprindă un număr de operațiuni (depunerea banilor, extragerea banilor, transferarea banilor pe alt cont și transferarea banilor pe un cont de telefonie mobilă). Utilizatorii grupului „negativ” au fost rugați să întreprindă operațiuni non-standard (să încerce să spargă terminalul, să imite starea de ebrietate, să însceneze un jaf asupra unui alt utilizator etc.)

Toate acțiunile au fost înregistrate video și au fost prezentate unui grup de observatori, care au etichetat acțiunile ca „tipice” sau „atipice”. Datele obținute au fost folosite în continuare pentru antrenarea rețelei neurale.

Grupul de utilizatori participanți la test au fost rugați să întreprindă operațiuni standard prin intermediul terminalului ATM pentru a verifica performanța sistemului în timp real.

Informația vizuală în sistemul propus este prezentată în câteva etape (Fig.5):

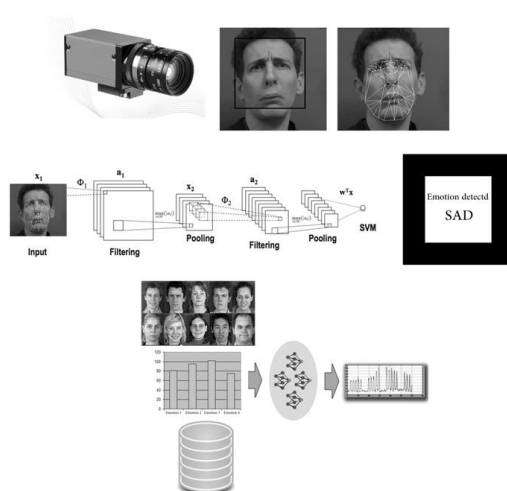


Figura 5. Prezentare generală a procesului de recunoaștere facială

1. În primul rând, folosim camere video cu senzori 3D de tip Microsoft Kinect pentru a detecta trăsăturile feței cu scopul de recunoaștere și clasificare a emoțiilor și a gesturilor.
2. În al doilea rând, aplicăm tehnici de viziune computerizată pentru extragerea caracteristicilor și recunoașterea modelelor.
3. Aplicăm tehnicile de învățare automată (rețelele neurale) pentru detectarea emoțiilor și clasificarea acestora.
4. Utilizăm datele statistice înregistrate din tranzacțiile mașinii și log-uri pentru antrenarea sistemului nostru. Antrenăm o rețea neurală modulară împreună cu înregistrările emoțiilor pentru a efectua analiza evenimentelor. Putem utiliza rețelele antrenate pentru a analiza acțiunile utilizatorului în timp real.

5. În timpul interacțiunii utilizatorului cu sistemul, putem urmări acțiunile frauduloase în timp real și întreprinde măsuri de securitate pentru a preveni infracțiunea sau fraudă.

Au fost realizate două serii de experimente. Primul grup de experimente a fost realizat pentru a evalua cât de efectiv sistemul dat poate detecta comportamentul normal și anormal al utilizatorului în timpul interacțiunii cu terminalul ATM. În scopul experimentului a fost elaborat un soft de simulare a terminalului ATM, care a fost folosit în terminalul autonom. În timpul etapei de interacționare, reacțiile utilizatorilor au fost înregistrate de o cameră video, montată pe partea superioară a terminalului (Fig.6).

Înregistrările obținute au fost evaluate mai târziu de observatori umani, iar emoțiile prezentate în aceste înregistrări au fost clasificate în „tipice” și „atipice”.

Pentru a înregistra emoțiile care nu au fost prezentate în timpul primei serii de experimente (în conformitate cu opinia observatorilor umani și a sentimentelor subiective ale participanților la experiment), noi am înregistrat emoțiile prezentate de către aceiași subiecți în timpul observărilor înregistrărilor video de scurtă durată. Pentru a păstra uniformitatea datelor, am înregistrările video au fost demonstrate prin intermediul aceluiași echipament, care a fost folosit în timpul etapei experimentale cu terminalul ATM.

Douăzeci de subiecți sănătoși, cu vârsta cuprinsă între 21-37 ani, cu vedere normală, sau corectată la normal, au participat în cadrul experimentului. În același timp, datele din cele două serii de experimente au fost procesate prin intermediul unei camere cu infraroșu și folosite ca date de intrare pentru modelul rețelei neurale. Fiecare subiect a realizat 10 sesiuni cu softul de simulare a terminalului ATM și 5 sesiuni video.

În timpul etapei de interacțiune, reacțiile utilizatorilor au fost înregistrate de o cameră video, montată pe partea de sus a terminalului. Câmpul de vedere a camerei includea corpul întreg, de la cap în partea de sus până la genunchi în partea de jos. Totuși, în scopul acestei părți a experimentului, a fost procesată și analizată doar fața (colțul din dreapta sus a figurii).

Pentru a evalua performanța modelului rețelei neurale, am realizat experimentele de simulare, introducând aceleași date preluate de la camerele cu infraroșu. De asemenea, am folosit datele furnizate de aceeași subiecți umani, înfățișând alte emoții. În total au fost demonstrate a câte 7 emoții de la fiecare subiect.

Înregistrările rezultante au fost clasificate în mod aleatoriu în subseturi de antrenare și testare. În timpul simulării, rețeaua a clasificat comportamentul „tipic” a utilizatorilor terminalelor ATM cu o precizie de 86%

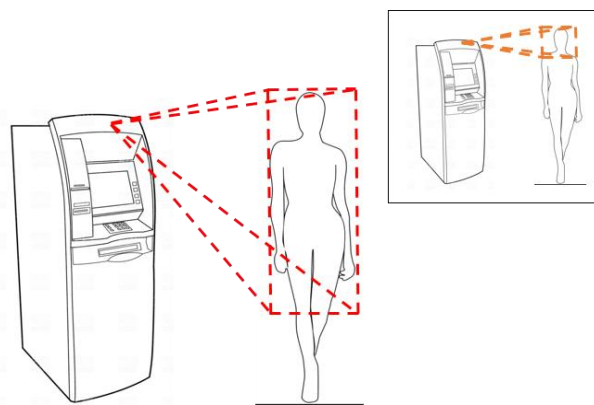


Figura 6. Configurarea experimentală a echipamentului

În timpul celei de a doua serii de experimente, am folosit una dintre abordările de recunoaștere a gesturilor și urmărire a corpului: clasificarea mișcărilor corpului. O tehnică de clasificare pentru această metodă este recunoașterea modelului: cu alte cuvinte o cameră video cu infraroșu recunoaște acțiunile umane: legănare, sărituri, gesturi cu mâna etc. Printre primii reprezentanți de succes ale acestei tehnologii se enumeră camera Kinect Microsoft. Kinect folosește lumină structurată și tehnici de învățare automată după cum urmează:

- Harta adâncimii este construită prin analiza unui punct model de lumină laser infraroșie.
- Părțile corpului sunt deduse folosind algoritmul de învățare automată *randomized decision forest*, creat în baza a peste 1 milion de exemple de antrenare.
- Începe cu 100.000 imagini de adâncime cu schelete cunoscute (dintr-un sistem de captare a mișcării).
- Transformă imaginile de adâncime într-o imagine a părții corpului.
- Transformă imaginea părții corpului într-un schelet.

În cadrul cercetării noastre nu folosim tehnica de clasificare propusă de Kinect, acesta fiind folosit doar în calitate de senzor infraroșu.

Experimentele psihologice. Noi am efectuat o serie de experimente pentru a evalua cât de eficient sistemul propus poate detecta comportamentul normal și anormal al clientului în timpul interacțiunii cu terminalul ATM. În scopul experimentului a fost elaborat și utilizat un soft de simulare a terminalului ATM instalat într-un terminal autonom. În timpul etapei de interacțiune, mișcările corpului utilizatorilor și expresiile faciale au fost înregistrate cu o cameră video instalată în partea de sus a terminalului. Aceste înregistrări au fost mai târziu evaluate de observatori umani, iar comportamentul reprezentat pe aceste înregistrări a fost clasificat în „tipic” și „atipic”.

Pentru a păstra uniformitatea datelor, înregistrările video au fost afișate prin intermediul aceluiași echipament, care a fost folosit în timpul etapei experimentale cu terminalul ATM. În cadrul experimentului au participat treizeci subiecți sănătoși, cu vîrsta cuprinsă între 21-37 ani, cu vedere normală, sau corectată la normal. În același timp, datele din cele două serii de experimente au fost procesate prin intermediul unei camere cu infraroșu și folosite ca date de intrare pentru algoritmul RNC. Fiecare subiect a realizat 10 sesiuni cu softul de simulare a terminalului ATM și 5 sesiuni video.

În timpul fiecărei sesiuni, recunoașterea mișcărilor părții superioare a corpului (în câmpul de vedere a camerei montate pe partea de sus a unui terminal ATM tipic) a fost realizată împreună cu clasificarea și recunoașterea trăsăturilor feței. Dintre treizeci de subiecți, am folosit 22 în calitate de exemple de comportament „normal” și 8 în calitate de comportament „anormal”. Figura 7 prezintă eșantioane de date obținute în timpul experimentelor noastre.

Imaginile de ecran sunt preluate de la camera cu infraroșu, montată în partea de sus a terminalului ATM, care arată subiectul uman din aceeași poziție, ce ar putea fi prezentată de o cameră de supraveghere obișnuită. Imaginile de ecran sunt preluate de la camera cu infraroșu, montată în partea de sus a terminalului ATM, prezentând două acțiuni consecutive: mâinile în jos (lipsa acțiunii) și o mână în sus (introducerea codului PIN)

Performanța generală a sistemului este descrisă în termenii de ieșire: dacă acesta clasifică comportamentul utilizatorului (emoții + gesturi) ca fiind „tipice” (Fig.8) sau „atipice” (Fig.9). Este necesar să menționăm faptul că sunt posibile mai multe combinații de algoritmi. RBFxSOM constituie, vorbind la general, un algoritm mai simplu decât RNC. Îl folosim în paralel cu RNC doar pentru două scopuri majore:

- Prin utilizarea acestui algoritm, obținem o hartă continuă de trăsături caracteristice, care este mai ușor de interpretat în comparație cu clasificarea în doar două grupuri de emoții (tipice și atipice).
- Nu trebuie să antrenăm RNC de două ori, fapt ce este mult mai ieftin.
- Totuși, această etapă poate fi omisă dacă avem suficient timp și capacitate de a antrena RNC de două ori: pentru ambele clasificări ale emoțiilor și mișcărilor corpului.

În acest capitol este prezentată descrierea aplicației pentru rețelele propuse și comparată performanța acestor două abordări. Am folosit în acest studiu un echipament special

pentru obținerea datelor infraroșii (Kinect API), cu toate că oricare altă cameră cu infraroșu ar putea fi folosită pentru implementarea sistemului.

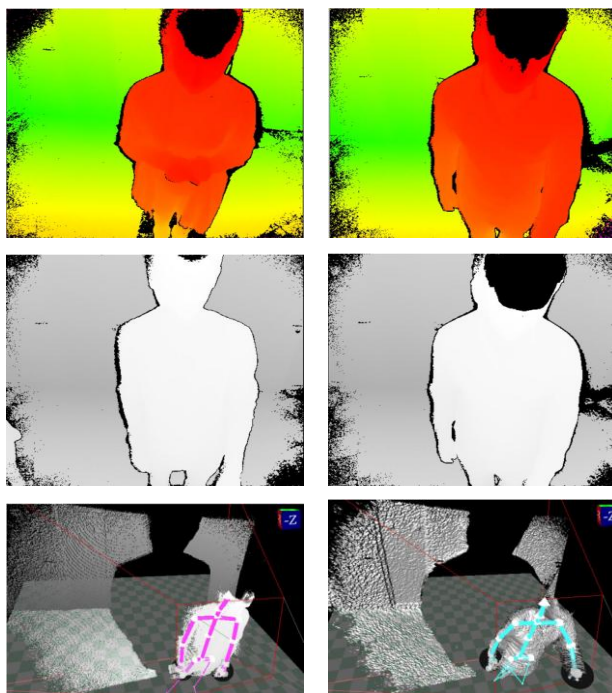


Figura 7. Eșantioane de date obținute în timpul experimentelor.

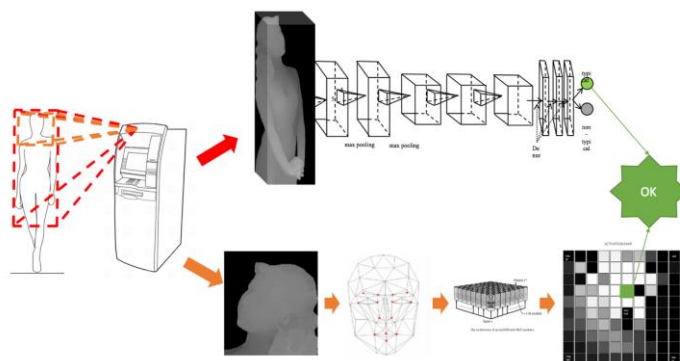


Figura 8. Răspunsul sistemului la comportamentul utilizatorului în timp real. Cazul A: comportamentul utilizatorului este calificat ca fiind „tipic”.

Pentru a verifica performanța sistemului, am petrecut două serii de experimente psihologice. Fiecare serie de experimente a avut ca scop testarea unei părți ale sistemului: fie modulul de recunoaștere a emoțiilor, fie modul de recunoaștere a acțiunilor. Fiecare serie de experimente a constat din două părți: prima parte a avut ca scop testarea performanței arhitecturii propuse și colectarea datelor pentru antrenarea sistemului, în timp ce a doua parte a avut ca scop obținerea răspunsului de la observatorul uman.

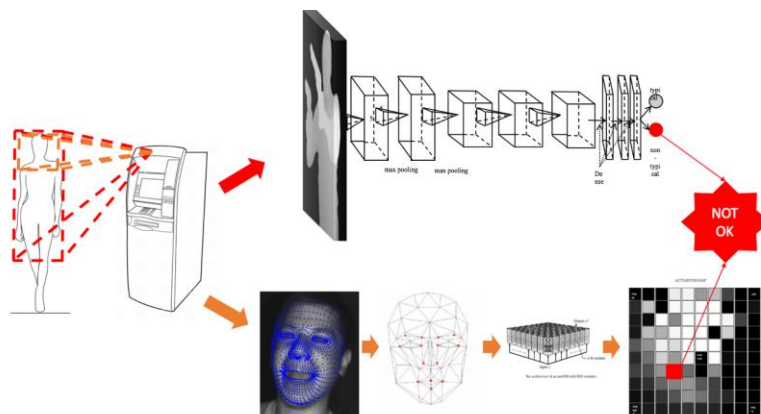


Figura 9. Răspunsul sistemului la comportamentul utilizatorului în timp real. Cazul A: comportamentul utilizatorului este calificat ca fiind „atipic”.

Prin urmare, a fost construit și testat un sistem video de securitate complex, potrivit pentru terminalele ATM. Sistemul este gata pentru testele în teren și ar putea fi implementat cu scop de testare într-un terminal ATM standard.

3. CONCLUZII ȘI RECOMANDĂRI

1) A fost propus un instrument pentru clasificarea eficientă a emoțiilor și gesturilor unui subiect uman în „tipice” și „atipice” în timpul unui anumit fel de interacțiune. Am descris cele două tipuri de arhitecturi de rețele neurale pentru clasificarea gesturilor și emoțiilor umane, obținute prin intermediul camerelor cu infraroșu. Aceste arhitecturi ar putea fi folosite în paralel (pentru o procesare mai rapidă și mai robustă), sau poate fi folosită numai RNC pentru procesarea trăsăturilor caracteristice. Alegerea arhitecturii depinde de circumstanțele în care sistemul poate fi folosit. Acest studiu poate fi considerat ca o încercare de a face încă un pas spre implementarea acestui fel de arhitecturi: noi am aplicat două tipuri de arhitecturi ierarhice modulare pentru sarcina de recunoaștere a emoțiilor și acțiunilor umane și le-am folosit pentru a soluționa problema reală de clasificare a comportamentului uman în „tipic” și „atipic” în raport cu o anumită sarcină.

2. Abordarea propusă poate fi folosită într-o varietate de aplicații. Pentru simplitate noi ne limităm doar la un fel de interacțiune (utilizator de terminal ATM). Totuși, acest fel de sarcină de clasificare este foarte utilă pentru un număr variat de aplicații, unde numărul de gesturi umane este limitat, așa precum clienți ai diferitelor tipuri de mașini automatizate. Pentru această categorie de utilizatori, algoritmul poate fi folosit pentru detectarea comportamentului obișnuit/fraudulos pentru a reduce volumul de muncă a operatorilor de televiziune cu circuit închis (CCTV), sau supraveghere video, care monitorizează aceste mașini. Spre exemplu, acest sistem poate fi utilizat pentru monitorizarea lucrătorilor din jur,

în cazul în care acțiunile acestora sunt restricționate semnificativ: linie de asamblare, lucrări de construcție pe clădiri înalte, sub sol, în mine. Un alt exemplu poate fi monitorizarea vigilenței, atenției șoferului/pilotului. Cu ajutorul acestui sistem am putea clasifica acțiunile corecte și incorecte, identifica așa stări nedorite precum pierderea atenției, boală, oboseală etc.

3. Rezultatele prezentate în această lucrare demonstrează că sistemul propus menține rata de recunoaștere similară cu cele oferite de cele mai noi tehnologii utilizate în domeniile viziunii computerizate și a recunoașterii de emoții. Mai mult ca atât, un sistem de o asemenea complexitate, care îmbină atât recunoașterea emoțiilor, cât și recunoașterea acțiunilor, nu a fost prezentat până acum. Arhitectura efectuează recunoașterea mișcărilor corpului și expresiile faciale și le clasifică în două categorii (tipice și atipice) cu o precizie generală (marja de eroare de 8% și 14% respectiv). Aceste rezultate au fost obținute independent și combinate ulterior prin intermediul unui algoritm de clasificare simplu bazat pe reguli. Combinarea rezultatelor a contribuit la îmbunătățirea performanței subsistemelor în cazul în care acestea funcționează în mod independent.

4. Pentru a îmbunătăți performanța sistemului, modelul propus solicită o cantitate enormă de date de antrenare, care nu poate fi obținută cu ușurință. Prin urmare, continuarea firească a studiului dat ar putea consta în efectuarea unor teste de teren suplimentare pentru a obține mai multe date de antrenare și a îmbunătăți performanța.

5. Rezultatele experimentelor demonstrează că sistemul este potrivit pentru implementarea în terminalele ATM. Sistemul este gata pentru efectuarea testelor de teren și ar putea fi implementat în scop de testare într-un terminal ATM tipic.

Prin urmare, putem concluziona că toate obiectivele studiului dat au fost realizate, iar sarcinile tehnice au fost implementate cu succes. Putem concluziona că sistemul propus este capabil pentru:

- captarea, recunoașterea și clasificarea emoțiilor și acțiunilor unui subiect uman într-un mod eficient;
- integrarea recunoașterii și emoțiilor permite monitorizarea comportamentului uman în timp real, oferind rezultate mult mai eficiente decât sistemele existente.

Cercetarea descrisă în studiul dat, constituie doar o mică parte din domeniul imens al recunoașterii umane. Noi ne-am referit doar la posibilitățile care se deschid la moment prin intermediul implementării studierii aprofundate a sistemelor rețelelor neurale în aplicațiile industriale. Continuarea firească a acestei cercetări va consta în elaborarea unei rețele neurale care va fi capabilă să recunoască un spectru mai larg de acțiuni. O astfel de rețea ar fi potrivită

pentru aplicații în care numărul de gesturi umane sunt limitate, așa precum clienți ai diferitor tipuri de mașini automatizate, sisteme CCTV, sau de supraveghere video, operatorii care monitorizează utilizatorii de mașini de bilete în stațiile de metrou și persoanele care se autoservesc, șoferii pe distanțe lungi, lucrătorii liniilor de asamblare și mulți alții.

BIBLIOGRAFIE

1. Ekman P. Basic Emotions. In: Handbook of Cognition and Emotion. New York, NY: John Wiley and Sons Ltd., 1999, ch. 3, p. 45–60.
2. Ekman P., Friesen W. Facial action coding system: A technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto, 1978.
3. Albu V., Cojocaru S. Measuring human emotions with modular neural networks and computer vision based applications. In: Computer Science Journal of Moldova, 2015, vol.23, vol. 1, no. 67, p. 40-61.
4. Karpathy A., Toderici G., Shetty S., Leung T., Sukthankar R., Fei-Fei L. Large-scale Video Classification with Convolutional Neural Networks. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, p. 1725-1732.
5. Kohonen T. Self-organizing maps. Series: Springer Series in Information Sciences, vol. 30, Berlin: Springer-Verlag, 2001. XX+502 p.
6. Furukawa T. SOM of SOMs. In: Neural Networks, May 2009, vol. 22, no. 4, p. 463-478.
7. Tokunaga K., Furukawa T. Modular network SOM, In: Neural Networks, January 2009, vol. 22, no. 1, p. 82-90.
8. Poggio T., Edelman S. A network that learns to recognize three-dimensional objects, In: Nature, 1990, vol. 343, p. 263 - 266.
9. LeCun Y., Boser B., Denker J.S., Henderson D., Howard R.E., Hubbard W., Jackel L.D. Backpropagation Applied to Handwritten Zip Code Recognition, In: Neural Computation, 1989, vol. 1, p. 541-551.
10. Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, In: Biological Cybernetics, 1980, vol. 36, no. 4, p. 193-202.
11. Albu V. Measuring human emotions with modular neural networks. In: The proceedings of the 7th International Multi-Conference on Complexity, Informatics and Cybernetics: IMCIC 2016, March 8 - 11, 2016, Orlando, Florida, USA, p. 26-27.

12. Kamnitsas K., Chen L., Ledig C., Rueckert D., Glocker B. Multi-Scale 3D Convolutional Neural Networks for Lesion Segmentation in Brain MRI. In: Proceedings of MICCAI Brain Lesion Workshop 2015, Munich, Germany, 2015. <http://hdl.handle.net/10044/1/27804> (visited on March, 29, 2016)

LUCRĂRILE AUTORULUI LA TEMA TEZEI

1. Albu, V. Measuring customer behavior with deep convolutional neural networks. BRAIN. Broad Research in Artificial Intelligence and Neuroscience, Volume 1, Issue 2, April 2016, pp.74-79. E-ISSN 2067 – 3957 **(ISI)**.
2. Албу, В.А., Хорошевский, В.Ф. КОГР система когнитивной графики. Разработка, реализация и применение. В: Известия Академии Наук СССР. Техническая кибернетика. 1990, nr. 5, pp. 105-118.
3. Averkin, A., Albu, V., Ulyanov, S. and others. Dynamic object identification with SOM-based neural networks. In: Computer Science Journal of Moldova, 2014, nr. 22 1/64, pp. 110-126. ISSN 1561-4042 **(B+)**
4. Albu, V., Cojocaru, S. Measuring human emotions with modular neural networks and computer vision based applications, *Computer Science Journal of Moldova*, v.23, n.1 (67), 2015, pp.40-61. ISSN 1561-4042 **(B+)**
5. Albu, V. Neural network based model for emotion Recognition. In: *Proceedings of the Workshop on Foundations of Informatics*. FOI-2015, August 24-29, 2015, Chisinau, Republic of Moldova, pp.423-434
6. Ulyanov, S., Albu, V., Barchatova, I. Intelligent robust control system based on quantum KB-self-organization: quantum soft computing and Kansei / affective engineering technologies. The third conference of mathematical society of the Republic of Moldova. Chisinau: Institute of mathematics and Computer Science, Academy of sciences of Moldova, 2014, pp. 571-582. ISBN: 978-9975-68-244-2
7. Albu V. Measuring human emotions with modular neural networks. In: Proceedings of the 7th International Multi-Conference on Complexity, Informatics and Cybernetics: IMCIC 2016, March 8 - 11, 2016, Orlando, Florida, USA, pp.26-27.
8. Albu, V. Measuring human emotions with modular NNS and computer vision applications. În: *Tendințe contemporane ale dezvoltării științei: viziuni ale tinerilor cercetători*. Teze ale Conferinței Științifice Internaționale a Doctoranzilor. Martie, 2015, AȘM, Chișinău, p.14.

ABSTRACT

of the thesis “Human actions recognition with modular neural networks” submitted by Veaceslav Albu for fulfillment of the requirements for the Ph.D. in Computer Science, specialty 122.03 – Modeling, mathematical methods, software. The thesis was elaborated at the Institute of Mathematics and Computer Science of the Academy of Sciences of Moldova, Chisinau, in 2016. The thesis is written in English and contains Introduction, 3 chapters, general conclusions and recommendations, bibliography of 109 titles. The main text amounts to 121 pages. This work includes: 37 figures, 2 tables, and 5 annexes. The results are published in 8 scientific papers.

Keywords: Deep Neural Networks, Computer Vision, Emotion Classification, Gesture Classification.

The area of the present studies is the field of emotion and action recognition using modular neural networks.

The aim and objectives of this research is to develop a tool for classification of human reactions (including facial features and body movements) into typical and non-typical in a certain environment. This tool provides statistical observations and measurements of human emotional states during an interaction session with a software product (or, optionally, with a hardware plus software complex).

Scientific novelty is a novel modular neural network architecture, constituted from two separate parts and combine the results to introduce the classification of the infrared sensor inputs, which is the first system of this kind, being applied both to emotion and human action recognition.

The important solved scientific problem is elaboration of a multimodal method for classification of human reactions (joining emotions and actions) into typical and non-typical in a certain environment, that ensures an effective functioning of systems destined to human actions monitoring in real time.

Theoretical significance. Our research solutions provide ground for solving of following problems: formulation of the tool’s architecture for robust classification of emotions and gestures of a human subject into typical vs. non-typical; the substantiation of the possibility and efficiency of using deep learning in an integrated approach for the detection of expression of the whole body in real time.

Practical value: this kind of classification task is very useful in different applications, where the number of gestures of the human is limited, such as: customers at the various types of automated machines, drivers, assembly line workers, hospital patients etc.

ADNOTAREA

tezei “Recunoașterea acțiunilor umane în baza rețelelor neurale modulare” înaintate de către Albu Veaceslav pentru obținerea titlului de doctor în informatică la specialitatea 122.03 – Modelare, metode matematice, produse program. Teza a fost elaborată în Institutul de Matematică și Informatică al AȘM, Chișinău, anul 2016. Teza este scrisă în limba engleză și constă din introducere, trei capitole, concluzii generale și recomandări, bibliografie ce cuprinde 109 titluri. Lucrarea conține 121 pagini text de bază, 37 figuri, 2 tabele, 5 anexe. Rezultatele principale sunt publicate în 8 lucrări științifice.

Cuvinte cheie: Rețele neurale adânci, computer vision, clasificarea emoțiilor, clasificarea gesturilor.

Domeniul de studiu al tezei îl constituie rețelele neurale modulare.

Scopul și obiectivele cercetării ține de elaborarea unui instrumentar pentru clasificarea reacțiilor umane (care includ aspecte faciale și mișcări ale corpului) în două clase: tipice și atipice pentru anumit mediu. Acest instrument oferă observații și măsurări statistice ale stărilor emoționale umane în timpul unei sesiuni de interacțiune cu un produs software (sau, opțional, a interacțiunii cu un complex hardware și software).

Noutatea și originalitatea cercetării o constituie arhitectura nouă a rețelei modulare neurale, care constă din două părți separate, combinându-le rezultatele în scopul efectuării unei clasificări a datelor obținute de la senzori infraroșii. Acesta este un prim sistem de acest fel aplicat atât pentru recunoașterea emoțiilor faciale, cât și a acțiunilor umane.

Problema științifică importantă soluționată constă în elaborarea unei metode multimodale de clasificare a reacțiilor umane (unind emoțiile și acțiunile) în tipice și atipice în raport cu un mediu dat, fapt care asigură funcționarea eficientă în timp real a unor sisteme de monitorizare a acțiunilor umane.

Semnificația teoretică. Rezultatele cercetării fundamentează soluționarea următoarelor probleme: stabilirea arhitecturii instrumentarului pentru clasificarea fiabilă a emoțiilor și gesturilor a unui subiect uman în tipice vs. atipice; stabilirea posibilității și eficienței utilizării învățării profunde în cadrul unei abordări integrate pentru identificarea expresiilor întregului corp uman în timp real.

Valoarea practică: soluționarea acestei probleme de clasificare este extrem de utilă pentru diverse aplicații, în care numărul de gesturi umane este limitat, precum cel al utilizatorilor mașinilor automate de cel mai variat tip, conducători auto sau cei de trenuri, muncitori la linii de asamblare, pacienți în spitale, aflați în stare de imobilitate etc.

АННОТАЦИЯ

диссертации “Распознавание действий человека на основе модулярных нейронных сетей” представленной Вячеславом Албу на соискание ученой степени доктора наук в области информатики по специальности 122.03 – Математическое моделирование, методы, программное обеспечение. Диссертация была написана в Институте математики и информатики при Академии наук Молдовы (Кишинёв), в 2016 году, на английском языке и содержит: введение, три главы, общие заключения и рекомендации, библиографию из 109 названий, 121 страницу основного текста, 37 рисунков, 2 таблицы, 5 приложений. Полученные результаты опубликованы в 8 научных статьях.

Ключевые слова: глубинные нейронные сети, компьютерное зрение, классификация эмоций, классификация жестов.

Областью исследований диссертации являются модулярные нейронные сети.

Целью диссертации является разработка инструментария для классификации реакций человека (включающих в себя выражение лица и движения тела) на два вида: типичные и нетипичные для определенной среды. Этот инструментарий предоставляет возможность проведения статистических наблюдений и измерений эмоционального состояния человека при его взаимодействии с некоторым программным комплексом (или, как вариант, с аппаратно-программным комплексом).

Научная новизна и оригинальность диссертации выражены в новой архитектуре модулярной нейронной сети, которая состоит из двух отдельных частей, результаты которых объединяются для осуществления классификации данных, полученных от инфракрасных датчиков. Это первая система такого рода применяемая как для распознавания лицевых эмоций, так и человеческих действий.

Решена важная научная проблема, которая заключается в создании мультимодального метода классификации человеческих реакций (объединяющих эмоции и действия) на типичные и нетипичные по отношению к данной среде, что обеспечивает эффективное функционирование в режиме реального времени систем мониторинга человеческих действий.

Теоретическая значимость полученных результатов состоит в обосновании решения следующих задач: создание архитектуры комплекса для надежной классификации действий на типичные и нетипичные, доказательство возможности использования глубинного обучения в рамках интегрированного подхода для распознавания выражений человеческого тела в целом в режиме реального времени.

Прикладная ценность: решение задачи классификации находит применение в ряде приложений, в которых количество жестов ограничено, например различного типа автоматы, водители автомобилей или поездов, работники сборочных линий, пациенты, находящиеся в неподвижном состоянии.

ALBU VEACESLAV

**RECUNOAȘTEREA ACȚIUNILOR UMANE ÎN BAZA
REȚELELOR NEURALE MODULARE**

SPECIALITATEA: 122.03

MODELARE, METODE MATEMATICE, PRODUSE PROGRAM

Autoreferatul tezei de doctor în informatică

Aprobat spre tipar: 30.05.2016

Formatul hârtiei 60x84 1/16

Hârtie ofset. Tipar ofset.

Tiraj 50 ex.

Coli de tipar 1.7

Comanda nr.

Valinex S.R.L.

Str.Florilor, 30/A, of.26b, mun. Chișinău, MD 2068