Universitatea Politehnica Timişoara

Facultatea de Electronică şi Telecomunicații

Habilitation Thesis

Teza de Abilitare

Computational Intelligence Paradigms with Applications in Embedded Vision

Paradigme ale calculului inteligent cu aplicații în sisteme de vedere dedicate

Dr. eng. Cătălin-Daniel CĂLEANU,
Associate Professor, MSEE, PhD.

Nr. inreg. MEdN 71202/10.02.2014

2014

Contents

1 ABSTRACT

- 1.1 Abstract 3
- 1.2 Rezumat 5

2 TECHNICAL PRESENTATION

- 2.1 Overview of Activity and Results 7
- 2.2 Computational intelligence in autonomous mobile robotics 11
  - 2.2.1 Symbolic trajectory description in mobile robotics 11
  - 2.2.2 Emergent Behaviour Evolution in Collective Autonomous Mobile Robots 36
- 2.3 Artificial intelligence paradigms for human face identification 44
  - 2.3.1 Interest Operator vs. Gabor Filtering for Facial Imagery Classification 44
  - 2.3.2 Combined pattern search optimization of feature extraction and classification parameters in facial recognition 54
- 2.4 Soft computing based face expression recognition 63
  - 2.4.1 Face Expression Recognition: a Brief Overview of the Last Decade 63
  - 2.4.2 Layered Fuzzy Facial Expression Generation of Virtual Agent 70
- 2.5 3D biometrics 81
  - 2.5.1 Human Identification Using Kinect Technology 81
  - 2.5.2 A ToF 3D Database for Hand Gesture Recognition 86
- 2.6 Future research, professional and academic career development plans 93

3 REFERENCES

- 3.1 General References 96
- 3.2 List of Publications 2001 – 2014 115
- 3.3 List of grants – selection 121

1. ABSTRACT

1.1 Abstract

Within the framework of the current habilitation thesis, my scientific, professional and academic achievements are summarized over a period starting with May 2001 (the date when my PhD thesis was defended, then certified by Minister’s Order 4202/27.07.2001) until present (2014). Some of my national and international research grants, important papers, books, patents, teaching activities/materials are also detailed in the context of the actual stage of the scientific domain of electronics and telecommunications with emphasis on the innovative aspects and personal contributions.

The first part of the thesis is constituted by the present abstract (both English and Romanian versions).

The second part of the current thesis refers to the:

- Overview of activity, in which I presented the most prominent research, professional and academic achievements (list of publications and grants classified in four main research topics, newly introduced disciplines, taught courses, contribution to the development of the academic curricula, invited professor, students internship, conducting diploma and dissertation theses, endowed laboratories and library, international cooperation, management activities, etc.). The most important mentioned aspects are: a number of 53 research articles published in the above mentioned period, 12 research grants (7 as grant/contract director) and 6 books.
- Technical presentation in which four main research topics are identified:
  - *Computational intelligence in autonomous mobile robotics*. First work presented here is interested in environment representation which permits the robot to know if it goes in the right direction by acquiring a spatial models of the robot's physical environment using a non-metric/qualitative approach. Perceptual landmarks are used to generate maps and to localise the robot with respect to these landmarks. Second work presented within the framework of the above mentioned topic deals with genetic algorithm based methods for finding optimal structure for a neural network (weights and biases) and for a fuzzy controller (rule set) to control a group of mobile autonomous robots. The goal of the robots, namely catching the targets, could be fulfilled only trough an emergent social behaviour observed in our experimental results.
  - *Artificial intelligence paradigms for human face identification*. Previous works has shown that Gabor feature extraction is one of the most effective techniques employed for the human face recognition problem. However, the selection of a particular set of Gabor filters is often problematic and, also the computational requirements are considerable. We propose an alternative feature extraction method - the Interest Operator - to be applied for the facial recognition problem. On AT&T public facial database, the system has achieved an average recognition rate of 95.2 percent using Gabor Approach and 94.7 percent using the Interest Operator. The second contribution in this field is represented by a combination between an Interest Operator based feature extraction technique and a k-NN statistical classifier having the parameters determined using a pattern search based optimization technique. This approach enables us to achieve both higher classification accuracy and faster processing time.
  - *Soft computing based face expression recognition*. The aim of the first presented work is to identify key representative approaches for facial expression recognition research in the past ten years (2003-2012). The interest in creating such an overview is multifarious. By

selecting the most interesting approaches, we want to focus the attention to new techniques and methodologies that may be of high interest to the researchers in the field of facial imagery. Moreover, this selection can be a useful indicator of the areas that will constitute the future research trends. The second detailed work concerns a layered fuzzy facial expression generation of a virtual agent. In this model, social, emotional and physiological layers contribute to the fuzzy facial expression generation.

- *3D biometrics*. In the first work, using combined skeletal tracking and depth information, a biometric person identification is performed. All these features are provided by a low cost 3D acquisition system, the Kinect sensor [Kinect12]. This information is further processed using standard image processing (PCA feature extraction) and machine learning (distance-based classifier) techniques. The second work employs the Time-of-Flight (ToF) principle - employed in certain range imaging 3D cameras. According to it, the measurement distance is derived from the propagation time of the light pulse between the camera and the subject for each point of the image. Then we describe the development of UPT ToF 3D Hand Gesture Database (UPT-ToF3D-HGDB). It represents, according to the best of our knowledge, the single database of this type which is publicly available.

It is worth noticing that the main results achieved in *Computational intelligence in autonomous mobile robotics field* are identified in the Section 3. References, subsection 3.2 List of publications by [Rob*Year-No*] and are in number of 19 papers. Also four research grants tackle a similar topic (see §3.3, [SIARAS2005], [ROBOTS2004], [SYMBOLIC2003], [AI2003]). The second research direction, namely *human face identification* has been treated in 8 publications (referred by [FaceID*Year-No*] in § 3.3) and in the following grants: [VIDEO2011], [NEURAL2006], [AI2005], [TRACK2005]. The third problematic, *face expression analysis and recognition* has been investigated in a number of 8 papers (see references with acronym [Expr*Year-No*]) and 2 research grants ([EXPR2011], [EMO2010]). The results regarding the last research direction, 3D techniques for biometrics, were published in 6 scientific publications (see [Bio2013-1], [Bio2013-2], [Bio2013-3], [Bio2012-1], [Bio2006-1], [Bio2004-1] from §3.2) and one patent [Gui2012].

- My future research, professional and academic career development plans along with possible implementation means. Here, new computational intelligence paradigms are intended to be in depth studied (e.g. Liquid State Machines [Rob2011-1]) and new implementation solutions are to be proposed (e.g. those General Purpose – GPU based as presented in [Rob2012-1]). Also I intend to investigate a new research field in co-operation with researchers form “Victor Babes” University of Medicine and Pharmacy Timişoara concerning a computer-assisted diagnosis system for the improvement of the medical decision in contrast enhanced ultrasound imagery for focal liver lesions.

The third section is dedicated to the references.

1.2 Rezumat

În cadrul prezentei teze de abilitare sunt prezentate succint realizarile personale în plan ştiinţific, didactic şi academic din perioada mai 2001 (data susţinerii tezei de doctorat confirmată ulterior prin Ordinul 4202/27.07.2001) până în prezent (2014). Concret, sunt făcute referiri la granturile naţionale şi internaţionale la care am participat în calitate de director sau ca membru al echipei de cercetare, lucrări ştiinţifice, cărţi, brevete şi materiale didactice elaborate în această perioadă. Acestea sunt prezentate în contextul noilor abordări existente în literatura de specialitate, accentuându-se caracterul inovativ al acestora.

Prima parte a tezei este constituită dintr-un rezumat ce cuprinde sinteza tezei de abilitare redactat în lb. engleză şi română.

Secţiunea a doua se referă la următoarele aspecte:

- Prezentare sumară a realizărilor remarcabile obţinute în activităţile de cercetare şi didactică (listă de publicaţii şi granturi clasificată în patru direcţii de cercetare, discipline nou introduse în planurile de învăţământ, cursuri predate, contribuţii aduse la dezvoltarea syllabusurilor, profesor invitat, activităţi de practică cu studenţii, conducerea lucrărilor de licenţă şi disertaţie, dotare laboratoare şi bibliotecă, cooperare internaţională, activităţi de management etc.). De menţionat că în perioada sus-amintită au fost publicate un număr de 53 de articole ştiinţifice, am participat în cadrul a 12 granturi/contracte de cercetare câştigate prin competiţie (dintre care 7 ca şi director de proiect) şi am elaborat 6 cărţi în domenii conexe prezentei teze.
- Prezentare tehnică a patru direcţii de cercetare abordate în această perioadă:
    - *Calcul inteligent pentru roboţi mobili autonomi*. O primă lucrare prezentată abordează reprezentarea mediului înconjurător de o asemenea manieră încât să permită elaborarea unui model spaţial non-metric/calitativ. Localizarea robotului este făcută în raport cu nişte repere perceptuale. A doua lucrare arată modalitatea de a proiecta o reţea neuronală respectiv un controller fuzzy prin intermediul algoritmilor genetici cu aplicaţie în controlul unor roboţi mobili autonomi. Rezultatele experimentale arată că agenţii rezultaţi permit realizarea unor sarcini complexe pe baza unui comportament de grup introdus prin evoluţie.
    - *Paradigme ale inteligenţei artificiale pentru identificare facială*. Cercetările au arătat că extragerea trăsăturilor bazată pe filtre Gabor dă rezultate foarte bune în problema recunoaşterii faciale. Selecţia unui anumit set de filtre se poate dovedi însă problematică. Se propune o metodă alternativă bazată pe tehnica operatorului de interes ce furnizează rezultate comparabile (95.2% rată de recunoaştere folosind filtre Gabor respectiv 94.7% folosind metoda operatorului de interes, baza de date AT&T) cu costuri de calcul mult reduse. A doua contribuţie prezentată are la bază combinaţia dintre tehnica de extragere de trăsături bazată pe operatorul de interes şi un clasificator statistic de tip k-NN la care parametrii sunt determinaţi printr-o tehnică holistică de optimizare. Prin aceasta se obţin atât rate ridicate de recunoaştere cât şi timpi de procesare reduşi.
    - *Sisteme bazate pe cunostinte cu incertitudine în recunoaşterea expresiei faciale*. Sunt detaliate în primul rând abordările reprezentative studiate pe o perioadă de 10 ani (2003-2012). Motivele realizării unui astfel de studiu sunt multiple. Prin selectarea celor mai atractive abordări sunt subliniate noile tehnici şi metode folosite în acest domeniu precum şi ariile ce pot prezenta în viitor un interes deosebit, constituindu-se în direcţii viitoare de cercetare. A doua lucrare prezentată se referă la generarea - pe principiile logicii vagi (fuzzy) - a expresiei faciale pentru un agent virtual. Modelul propus conţine componentele social, emoţional şi psihologic definind propriile contribuţii în generarea fuzzy a expresiei faciale.

- *Sisteme 3D biometrice*. Prima lucrare prezentată combină informația de adâncime (3D) a imaginii cu cea referitoare la diversele segmente ale corpului (schelet) pentru a implementa o identificare biometrică a unei persoane. Aceste informații sunt obținute folosind un sistem de achziție 3D cu cost scăzut – un senzor de tip Kinect [Kinect12]. Informația este ulterior procesată prin intermediul unor tehnici standard aferente prelucrărilor de imagini si inteligenței artificiale (PCA, clasificatori bazați pe distanță). A două abordare prezentată foloseşte principiu ToF (Time-of-Flight) pentru obținerea, cu un înalt grad de acuratețe, a infomației 3D. În conformitate cu acest principiu, distanțele măsurate sunt derivate din timpul de propagare a unui puls de lumină. Se descrie în continuare dezvoltarea unei baze de date de gesturi statice şi dinamice UPT ToF 3D Hand Gesture Database (UPT-ToF3D-HGDB). Ea reprezintă singura bază de date publică de acest tip pusă la dispoziția comunității ştiințifice.

Este de notat că principalele rezultate obținute în direcția *Calcul inteligent pentru roboți mobili autonomi* sunt identificate in Secțiunea 3. Bibliografie, subsecțiunea 3.2 Lista publicațiilor prin [Rob*An-Nr*] şi sunt în număr de 19 articole. De asemenea, subiecte similare sunt tratate în cadrul participării la patru granturi (vezi §3.3, [SIARAS2005], [ROBOTS2004], [SYMBOLIC2003], [AI2003]). În legatură cu cea de a doua direcție de cercetare, vizând *recunoaşterea identității unui individ*, au fost publicate 8 articole (referite prin [FaceID*An-Nr*] în § 3.3) şi a constituit subiectul următoarelor granturi: [VIDEO2011], [NEURAL2006], [AI2005], [TRACK2005]. Problematica *analizei si recunoaşterii expresiei faciale* a fost investigată în cadrul a 8 articole (vezi referințe bibliografice cu acronim [Expr*An-Nr*]) şi a două granturi ([EXPR2011], [EMO2010]). Rezultatele ultimei direcții de cercetare prezentate, *Sisteme 3D biometrice* au fost publicate în 6 lucrări ştiințifice [Bio2013-1], [Bio2013-2], [Bio2013-3], [Bio2012-1], [Bio2006-1], [Bio2004-1] şi constituie subiectul unui brevet [Gui2012].

- Prezentarea planurilor de evoluție şi dezvoltare cu privire la cariera profesională, ştiințifică şi academică precum şi modalități concrete de punere a acestora în practică. Îmi propun studiul unor noi paradigme ale calculului inteligent (de exemplu Liquid State Machines [Rob2011-1]) precum şi a unor noi soluții de implementare pentru acestea (de exemplu cele bazate pe GPU de uz general [Rob2012-1]). De asemenea îmi propun să investighez o nouă direcție de cercetare prin cooperarea cu cadrele didactice de la Universitatea de Medicină şi Farmacie „Victor Babeş” din Timişoara în vederea realizării unui sistem de diagnoză asistată în ecografia de contrast pentru leziuni ale ficatului.

Ultima secțiune este dedicată referințelor bibliografice.

2. TECHNICAL PRESENTATION

2.1 Overview of Activity and Results

In May 2001 my PhD thesis, *"Face recognition using parallel neural processing and interest operator method"* was presented publicly at University 'POLITEHNICA' Timisoara. Thus, the overview of activity will be made starting from the above mentioned date up to present (2014).

During this period "Computational Intelligence" (mainly Neural Networks, Fuzzy Systems and Genetic Algorithms), "Digital Image Processing", "Embedded Systems" and "Electronic Devices and Circuits" were the main research topics. They represent as well my teaching activities. For example, during 2009 I have proposed, elaborate the syllabus and taught two new courses: "Elements of Artificial Intelligence" and "Expert Systems" within the framework of the master program "Intelligent Electronic Systems". At the bachelor level I have introduced "Embedded Systems" and "Electronic Devices", both taught in English at "Politehnica International" section of the Faculty of Electronics and Telecommunication.

The main developed applications of the above mentioned topics are Autonomous Mobile Robots Navigation, Face Recognition, Face Expression Recognition, Gesture Recognition and Biometrics, thus I chose to group the achievements into four categories, as follows.

The contributions to the field of *Computational intelligence in autonomous mobile robotics* are presented in Section 2.2 of the current thesis. First work presented here is mainly based on the results published in a chapter of book printed abroad [Book2007-1], [Rob2006-1] and [Rob2003-1] and concerns environment representation which permits the robot to know if it goes in the right direction by acquiring a spatial models of the robot's physical environment using a non-metric/qualitative approach. Perceptual landmarks are used to generate maps and to localise the robot with respect to these landmarks. Second work presented (see for more details [Rob2008-3]) within the framework of the above mentioned topic deals with genetic algorithm based methods for finding optimal structure for a neural network (weights and biases) and for a fuzzy controller (rule set) to control a group of mobile autonomous robots. The goal of the robots, namely catching the targets, could be fulfilled only through an emergent social behaviour observed in our experimental results. Other topic tackle within this framework was related to the development of an integrated environment for assisted movement of visually impaired. The results were published in several articles e.g. [Rob2008-2] or [Rob2007-3], the former achieving 5 citations in Thomson Web of Knowledge database. It is worth noticing that the main results achieved in *Computational intelligence in autonomous mobile robotics field* are identified in the Section 3. References, subsection 3.2 List of publications by [Rob*Year-No*] and are in number of 19 papers. Also four research grants tackle a similar topic (see §3.3, [SIARAS2005], [ROBOTS2004], [SYMBOLIC2003], [AI2003]). The international cooperation regarding this research direction was done with prestigious universities and research institutes. Among them:

- Complex Systems Laboratories from University of Evry, France;
- Logique des Usages, Sciences Sociales et sciences de l'Information, ENST Bretagne, France
- Fraunhofer Institute for Manufacturing Engineering and Automation IPA in Stuttgart, Germany

The contributions to the field of *Artificial intelligence paradigms for human face identification* are presented in Section 2.3 of the present thesis. The article [FaceID2007-3] published in a prestigious ISI journal was chosen as the first illustrative paper for this topic. It achieved a number of 14 citations in Thomson Web of Knowledge database. Previous works has shown that Gabor feature extraction is one of the most effective techniques employed for the human face recognition problem. However, the selection of a particular set of Gabor filters is often problematic and, also the computational requirements are considerable. We propose an alternative feature extraction method - the Interest Operator - to be applied for the facial recognition problem. On AT&T public facial database, the system has achieved an average recognition rate of 95.2 percent using Gabor Approach and 94.7 percent using the Interest Operator. The second contribution in this field is represented by a combination between an Interest Operator based feature extraction technique and a k-NN statistical classifier having the parameters determined using a pattern search based optimization technique and it was presented in details in a prestigious ISI journal [FaceID2011-1]. This approach enables us to achieve both higher classification accuracy and faster processing time. The implementation of such system is treated in [FaceID2007-4] where a detection and recognition (FDR) system is presented. Mainly the following aspects are detailed: how to acquire an image, broadcast a video stream, manipulate a database, and finally, the detection/recognition phase, all in relation with theirs possible C#/.NET solutions. Emphasis was placed on artificial neural network (ANN) methods for face detection/recognition along with C# object oriented implementation. The second research direction, namely *human face identification* has been treated in 8 publications (referred by [FaceID*Year-No*] in § 3.3) and in the following grants: [VIDEO2011], [NEURAL2006], [AI2005], [TRACK2005]. The international cooperation regarding this research direction was done with Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, China.

The contributions to the field of *Soft computing based face expression recognition* are shown in Section 2.4 of the current thesis. The aim of the first presented work, presented in details in [Expr2013-1], is to identify key representative approaches for facial expression recognition research in the past ten years (2003-2012). The interest in creating such an overview is multifarious. By selecting the most interesting approaches, we want to focus the attention to new techniques and methodologies that may be of high interest to the researchers in the field of facial imagery. Moreover, this selection can be a useful indicator of the areas that will constitute the future research trends. The second detailed work, based on a paper published in an ISI journal [Expr2010-1], concerns a layered fuzzy facial expression generation of a virtual agent. In this model, social, emotional and physiological layers contribute to the fuzzy facial expression generation. The third problematic, *face expression analysis and recognition* has been investigated in a number of 8 papers (see references with acronym [Expr*Year-No*]) and 2 research grants ([EXPR2011], [EMO2010]). The former grant „Research on facial expression recognition in complicated environments”, contract no. 222/15.04.2009 - Bilateral Inter-Governmental S&T Cooperation grant between China and Romania No.39-5 aims the investigation and development of facial expression recognition methods and principles. The research was focused on finding robust solutions for the following subsystems: face representation, face detection, face synthesis, feature selection and extraction, classification.

Several major face imagery processing topics have been addressed, e.g.:

- an overview of the current state of the art systems/algorithms/methodology was performed;

- data acquisition: some of the experiments were performed using Beihang University facial expression database and with images acquired from real environments with complex backgrounds, large variety of emotional states, occlusions;

- preprocessing: we have employed robust techniques for mean shift segmentation, background estimation, tracking;

- feature extraction: a novel method for facial expression recognition which is robust to facial occlusion has been proposed. The face to be recognized is reconstructed using robust principal component analysis (RPCA), and saliency detection is used on the difference image of reconstructed face and the face to be recognized to obtain the facial occlusion region. For improving the nonlinear alignment performance of Active Appearance Models (AAM), we apply a nonlinear manifold learning algorithm, Local Linear Embedded, to model shape-texture manifold. Experiments show that our method maintains a lower alignment residual to some small scale movements compared with traditional AAM based on Principal Component Analysis (PCA) and makes a success alignment to large scale motions;

- classification: some possibilities regarding the use of novel neural architectures (e.g. Liquid State Machine) for processing the facial expression have been analysed. Also a reweighted AdaBoost classifier has shown good results with respect some public databases of faces, e.g. JAFFE;

- optimization: we have been proposed the application of the Pattern Search Optimization for feature extraction and classification parameters and show that when the process simulation is very complex and it is not designed in a vectorised manner, Pattern Search represents an attractive alternative to the other optimization methods, e.g. genetic algorithm, as it is often computationally less expensive and can minimize the same types of functions and yields better results in terms of classification accuracy and processing speed;

- face expression synthesis: novel model of layered fuzzy facial expression generation has been proposed. A novel layered fuzzy facial expression generation language is also developed for conveniently controlling facial expression generation of virtual agent.

The international cooperation regarding this research direction was done with aeronautics and astronautics "Beihang" University from Beijing, China were I have presented some lectures on face expression recognition.

3D biometrics. In the first work, based mainly on the results published in [Bio2013-3], a combined skeletal tracking and depth information is used, in order to implement a biometric person identification. All these features are provided by a low cost 3D acquisition system, the Kinect sensor [Kinect12]. This information is further processed using standard image processing (PCA feature extraction) and machine learning (distance-based classifier) techniques. The second work, (see details in ISI Proceedings paper [Bio2012-1]), employs the Time-of-Flight (ToF) principle - used in certain range imaging 3D cameras. According to it, the measurement distance is derived from the propagation time of the light pulse between the camera and the subject for each point of the image. Then we describe the development of UPT ToF 3D Hand Gesture Database (UPT-ToF3D-HGDB). It represents, according to the best of our knowledge, the single database of this type which is publicly available. Other approaches were dedicated to the study of gesture recognition as in the paper "Fingertip-based Real Time Tracking and Gesture Recognition for Natural User Interfaces", currently proposed to be published in an ISI journal. The widespread deployment of Natural User Interface (NUI) systems in smart phones, tablets or intelligent TV sets has heightened the need for robust multi-touch, speech or facial recognition solutions. In the air gestures recognition represent one of the most appealing technology in the field. This work proposes a fingertip-based approach for hand gesture recognition. The novelty of the proposed system is related to the tracking principle, where an improved version of the multi-scale mode filtering (MSMF) algorithm has been used, and to the classification stage, where the proposed set of geometric features provides high discriminative capabilities. Empirically, we conduct an experimental study involving different hand gesture recognition performed by multiple persons against various backgrounds in which our approach achieves a global recognition rate of 95.66%. The results regarding the last research direction, 3D techniques for biometrics, were published in 6 scientific publications (see [Bio2013-1], [Bio2013-2], [Bio2013-3], [Bio2012-1], [Bio2006-1], [Bio2004-1] from §3.2) and one patent [Gui2012].

Some pf my published papers were awarded by UEFISCSU prizes: „Research Awards for Articles", PN-II-RU-PREC-ISI, 2007, PN-II-RU-PRECISI-2011-5, 2011.

The awarded research grants and contracts enable us to endow the faculty laboratories with modern equipment (more then 500 000 RON), to donate books and various publications (IEEE Embedded Systems Letters, Transactions on Evolutionary Computation, Transactions on Fuzzy Systems, Transactions on Neural Networks, etc.) to the university library.

I am an active reviewer for the following ISI journals:
- Pattern Recognition Letters, Elsevier, The Netherlands
- ETRI Journal, Electronics and Telecommunications Research Institute, South Korea
- Digital Signal Processing, USA.
- The Institution of Electronics and Telecommunication Engineers Technical Review, India.
and many ISI conferences.

From 1994 to present I am a member of IEEE - The Institute of Electrical and Electronics Engineers, Inc., New York, USA, Computational Intelligence Society.

2.2 Computational intelligence in autonomous mobile robotics

2.2.1 Symbolic trajectory description in mobile robotics

One main issue for mobile robots is their capacity to go from one point to another autonomously, without getting lost or crashing into another object [Arkin98].

It is based on three concepts:

1. Planning which computes a trajectory between the two points,
2. Navigation which gives motion orders to the robot to follow the computed trajectory,
3. Environment representation which permits the robot to know if it goes in the right direction.

Works presented here are interested in point 3 that is, in acquiring spatial models of the robot's physical environment.

Two different approaches to this problem have emerged. The first one, the metric/quantitative representation of the environment, has some disadvantages. For example, due to incorrigible wheel slippage, dead-reckoning could be unreliable. The non-metric/qualitative approach use perceptual landmarks to generate maps and to localise the robot with respect to these landmarks. Works presented here are interested in the non-metric approach, trying to perform a qualitative description of a structured indoor environment.

These problems are tackled by the Simultaneous Localisation and Mapping (SLAM) introduced by Leonard and Durrant-Whyte [Leonard91], [Smith97] in robotics. SLAM is still today a very active field of research [Meyer02], [Filliat02a], [Filliat02b]. This problem is regarded as one of most significant for a true autonomy of the robots. Crucial questions still remain satisfactorily unanswered in spite of great progress in this field and the existence of robust methods to map static, very structured and limited sized environments.

[Kulic03] use a robot motion planning based on behavioural cloning. In a first phase, the robot is trained under operator's control to locate unmoving obstacles avoidance through a simulator. In that phase, the evaluated variables are stored in a log file. The second phase, called learning phase, machine learning program generates the differential equations defining the operator's trajectory, i.e. the clone. Finally, the verifying phase, the robot is controlled by the clone. These developmental phases are repeated changing both problem domain representation and learning system according to the cloning system criterion.

The problem of mapping can be generally regarded as the fact of giving to an autonomous robot the capacity to move in an environment. Thus, the problem of mapping goes further than simple construction of a plan gathering the obstacles in a given zone. [Thrun02] gives a general survey of the mapping problem. He points out the six key aspects of the mapping problem:

- The effects of the noise in the measurements (Wheel slippage, localisation error introduced by integration of data from wheel encoders, drift of inertial systems are three examples among many others.),
- The high dimensionality of the entities that are being mapped (How many parameters describe the environment, its major topological elements like corridors, crossings, doors, rooms, etc.?),
- The correspondence problem, also known as the data association problem (Do the measurements made by the sensors at different points in time in the environment correspond to the same object?),
- The perceptual aliasing (Two different places from the environment can be perceived in an identical way by the sensors.),
- The environment changes over time,
- The robotic exploration that is the task of generating robot motion in the pursuit of building a map.

Related works

Related works can be found in the fields of Image Based Navigation systems, shape understanding using sensor data, vision based homing. Vision for mobile robot navigation did have specific development during the last twenty years. [DeSouza02] give a complete survey of the different approaches. For indoor navigation, systems are classified in three groups: map-based navigation using predefined geometric and/or topological models, map-building-based navigation constructing by themselves geometric and/or topological models, and mapless navigation using only object recognition and actions associated to these objects [Gaussier97].

Kuipers' works [Kuipers91] defined symbols as distinct places situated at equal distances from the nearby obstacles. Connections between these places link symbols and represent free path [Choset01]. Fig. 2.2.1 shows an example of the Voronoii graph of an environment. The labelled vertices represent the symbols while edges connecting the symbols are the path the robot can use.

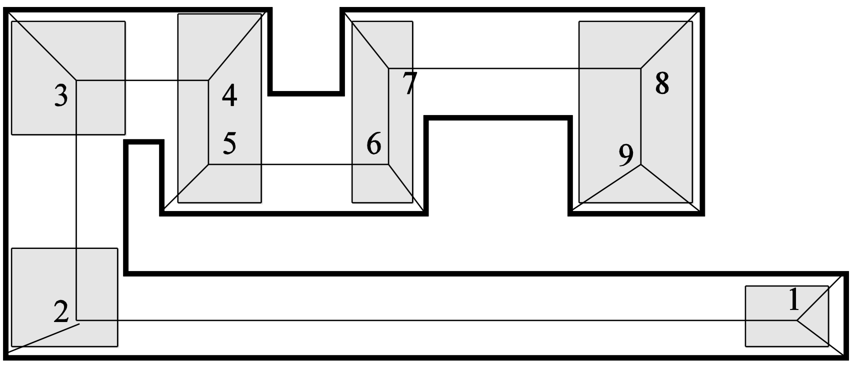


Fig. 2.2.1. Voronoii diagram with labelled vertices.

Indeed, assume that the robot has to move in this simple environment (Fig. 2.2.1) according to a mission given by the user, if the robot goes from label 1 to label 9, the most important areas are those filled, where the robot changes its direction. Between them, when there are no changes in the environment, it is useless to preserve the whole set of information from this part of the way. On the contrary, it is necessary to find a method of mapping of the filled zones which can describe them unambiguously.

In Image Based Navigation systems, several great classes of systems can be identified from the literature. The first one uses conventional telemeters and vision to find and identify objects in the environment (Wichert, 1996). The second one is the class of the systems coupling more or less directly sensor data to motor control thanks to a supervised learning process. Among them neural networks systems used as classifiers are noticeable. These systems begin to classify the environment into global classes such as "corridor, corner, room, crossing ..." [Al Allan95], [Pomerleau93] are often followed by a second processing unit that outputs a navigation command. In addition to restrictions related to the supervised learning, these classes give only a global description and are of least interest in cluttered and complex environments. The third class includes the systems which compare current sensor data and predefined models both at a low level (edges, planes ...) – see [Kim94] - and at a high level (door, room, object ...). These systems use mainly vision sensors (cameras) that provide a huge amount of data that must be reduced to be processed in real time. The elements extracted from the data are compared to reference models known a priori. The fourth class evoked here includes the systems trying to geometrically build environment models before deciding an optimised path plan [Crosnier99].

In the field of shape understanding using sensor data, environment interpretation stresses the use of natural landmarks to ease the navigation and the pose estimation of a mobile robot. Among other works, one can pinpoint [Simhon98a] which is interested in defining islands of reliability for exploration. He proposes strategies to couple navigation and sensing algorithms through hybrid topological metric maps. [Oore97] consider the problem of locating a robot in an initially unfamiliar environment from visual input. In the same way, [MacKenzie94] involve a methodology to bind raw noisy sensor data to a map of object models and an abstract map made of discrete places of interest.

Several implementations of vision based homing systems are presented in [Franz97]. A method aiming at highlighting salient features as, for example, landmarks between these two views and deriving a decision is used in [Hong91]. In these works, a homing system extracts landmarks from the view and allows a robot to move to home location using sequence target locations situated en route between its current location and home. Other works are biologically inspired. [Judd98] showed that ants store series of snapshots at different distances from their goal to use them for navigating during subsequent journeys. Judd and Collett experimented their theory with a mobile robot navigating through a corridor, homing successive target locations. Weber (Weber & al., 1999) proposes an approach using the bearings of the features extracted of the panoramic view leading to a robust homing algorithm. This algorithm pairs two landmarks situated into two snapshots to derive the homing direction. The bearings-pairing process uses a list of preferences similar to neighbourhood rules.

Symbolic processing methods are described in Tedder's works [Tedder01]. This formal approach is often called structural or syntactic description and recognition. The general method for perception and interpretation proposes to symbolically represent and manipulate data in a mapping process. [Tedder01] solve the problem in modelling the 3D environment as symbolic data and in processing all data input on this symbolic level. The results of obstacle detection and avoidance experiments demonstrate that the robot can successfully navigate the obstacle course using symbolic processing control. These works use a laser range finder. A way for defining suitable landmarks from an environment as the robot travels is a research problem pointed out by Fleisher and al. in [Fleisher03]. An automatic landmark selection algorithm chooses as landmarks any places where a trained sensory anticipation model makes poor predictions. The landmark detection system consists of a sensory anticipation network and a method of detecting when the difference between the prediction of the next sensor values and the current measured values can reveal the presence of a landmark. This model has been applied to the navigation of a mobile robot. An evaluation has been made according to how well landmarks align between different runs on the same route. These works show that the robot is able to navigate reliably using only odometry and landmark category information.

In [Lamon01], a method is proposed for creating unique identifiers called fingerprint sequences for visually distinct significant features in panoramic images. This localisation system proves that the actual position of a robot in an environment can be recovered by constructing a fingerprint sequence and comparing it with a database of known fingerprints.

The proposed work goes on the way proposed by [Tedder01] and [Lamon01]. According to these works, our contribution applies mainly on a method to extract clues of interest among raw distance data delivered by a 2D panoramic laser range finder installed on the robot. These clues of interest, i.e. the landmarks, are gathered in a sequence that we call a fresco. We consider that the trajectory of the robot can be described by the set of the frescoes. To do that, we have to select the frescoes that bring new information. The originality of this work stays in the simple but efficient criteria used for the construction and the validation of the fresco but mainly to select the most pertinent frescoes along the route of the robot. In addition to this qualitative approach, one must consider that the system will have to be embarked on a vehicle, which vibrates, runs at variable speeds on a non-uniform ground. This leads to constraints of speed, size, robustness, compactness and cost, implying various choices both at the design and at the development levels of the system. The methods used have been chosen as simple as possible to reduce the cost and the complexity of the processing. Nevertheless the method must be robust compared with the robot movements, the sensor accuracy and the variations of the complexity of the environment.

Test-bed perception system

The application field of this work is a middle-cost mobile robot sent in an apartment to do service for a user. Hence, the environment is of a structured and not engineered indoor type environment. At this point, the problem is two-fold. Firstly, through the Human-Machine Interface (HMI), a mission must be entered and its development must be explained to the user. Secondly, the robot has to be programmed to execute the mission. Building a description of the route as close as a human could do has at least two advantages. This description, on one hand, is requested by the HMI and, on the other hand, at the execution level, it can be a way to take into account the stumbling blocks highlighted by the conventional navigation systems.

The size of the non holonomous robot is (width x length) 0.50m x 0.75m. Its linear and angular speeds are up to 1 ms^{-1} and 2.45 $rads^{-1}$. Placed at the geometrical centre of the robot with practical/maximum ranges equal to 3m/10m, a panoramic 2D telemeter captures a circular environment. It has been decided to consider a 36m² squared environment to ease the reconstruction process (measurements at the corners are valid according to the maximum range of the telemeter). Only 256 measurements over the 1024 the telemeter is able to deliver are used by the fresco construction process. At a $1ms^{-1}$ speed, the translation displacement error remains lower than 10cm for one complete rotation of the telemeter. In 100 ms, the rotation of the robot remains lower than 23°. Experiments in the following have been made with measurements coming from both a simulated laser range finder and the real telemeter.

We will then consider that:

- there is a lack of accuracy of the telemetry measurements due to the vibrations caused by the jolts,
- most part of the environment is composed of co-operative targets (low material absorption coefficient, acceptable level of the reflected signal up to a 80° incident angle),
- reference position of the laser coincides with the main axis of the robot,
- data sequencing compensates the effects of the clockwise (CW) or counter clockwise (CCW) rotations of the robot so that the 256 horizontal distance information are regularly arranged on 360°,
- precision is greater than 20 cm for every measurements.

According to these considerations, we chose to digitise the environment on a 32 x 32 cells grid which covers the area seen by the telemeter, each cell representing a 0.1875m x 0.1875m square. The terms "grid" or "cellular space" will be considered as equivalent in the following.

Representation construction

Cyclic representation and cellular space

Landmarks such as "Opening, Closure, End_of_Closure, Angle of_Closures" used to build the qualitative description of the environment from the measurements. According to the sequential aspect of the data delivered by the laser range finder, the landmarks extraction order corresponds to the measurements order. The robot refers to two main axis: the "lengthwise axis" corresponds to the forward and rear directions of displacement, the "crosswise axis" is perpendicular to the lengthwise axis at the robot geometrical centre.

The fresco construction is divided into two main steps:

- The construction of the reliable digitised environment: cellular space building, signature extraction, crosswise, lengthwise and diagonal segments extraction, refining, reorientation.
- The landmarks extraction: Opening, Closure, End_of_Closure and Angle_of_Closures extraction, fresco construction, fresco validation.

Conventions used in the cellular space

The method uses evolution laws in the cellular space that act on every cells. For a cell called CELL the neighbourhood conventions use standard Von Neuman neighbourhood. For example, CELL_W, CELL_E, CELL_N, CELL_S are the names of the cells situated westbound, eastbound, northbound, and southbound. We add the word Great to name the cells in the second neighbourhood layer (Great West: CELL_GW, Great East: CELL_GE …). The quadrants are numbered counter clockwise in relation to the lengthwise axis: quadrant 0 is the front right one.

Construction of the digitised description

Fig. 2.2.2 summarises the operations leading to the construction of a reliable cellular space [Pradel94].

(a) Generation of the digitised environment: the very first operation performed consists in the lay-down of the distance measurements onto the grid to create the initial cellular spaces. They perform the same operations on the distance measurements issued from the sensor (part 1) and on the 45° shifted measurements set (part 2). On the grid, black cells represent the range finder impacts. Noise introduced in the measurements (measurements are made while the robot is moving) appears mainly on the form of cells agglomerations. Agglomerations also occur when measurements belong to the border between adjacent cells. Elimination of agglomerations is performed keeping only the cells situated the closest to the robot for obvious safety reasons. The method adopted for this elimination uses evolution laws close to those used in cellular automata.

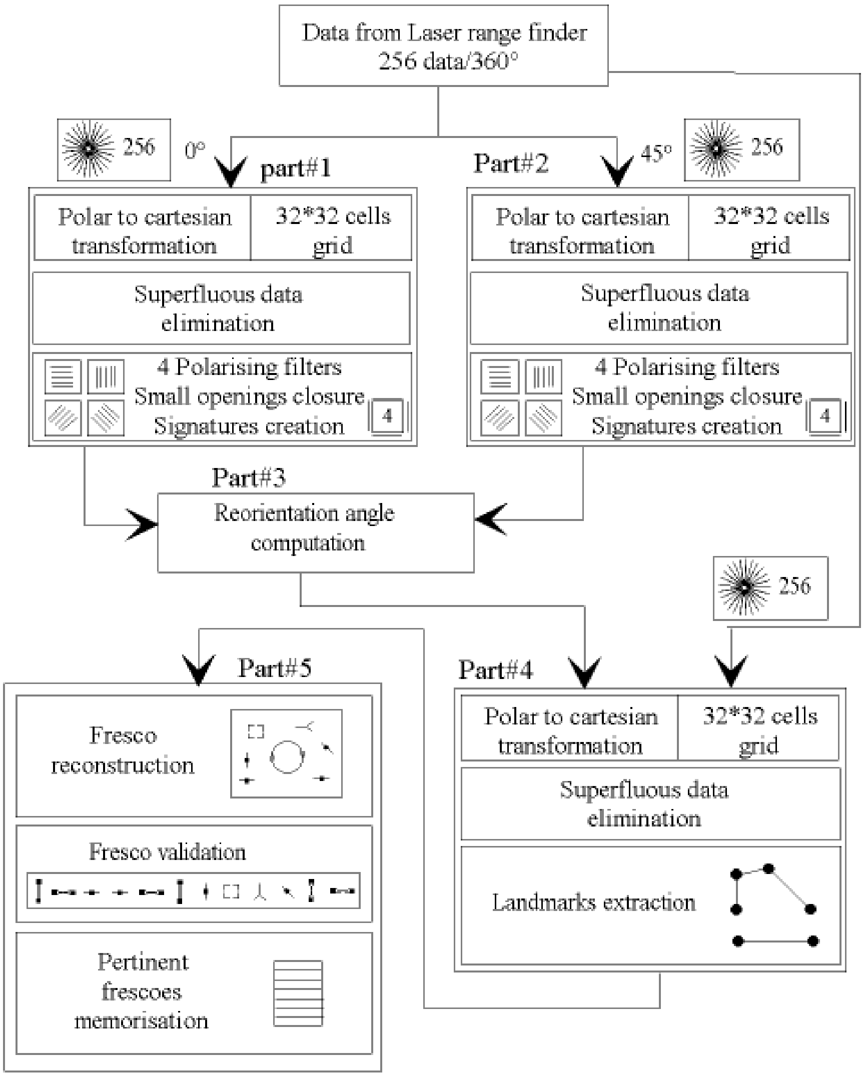


Fig. 2.2.2. Block diagram showing the operations performed in the construction of the digitised environment.

(b) Segmentation of the cellular space: the next operation is the extraction of the segments corresponding to the obstacles from the cellular space. Four directions are considered. In

addition with the lengthwise (fig. 2.2.3a) and crosswise axis (fig. 2.2.3c), a search for the segments is made onto the two diagonals (fig. 2.2.3d, f). The extraction laws leave alive a cell owning a neighbour alive in the considered direction.
(c) Reorientation of the cellular space: another origin of noise is bound to the oblique walls. These digitised oblique walls take the form of small adjacent segments with junctions without real significance. To eliminate these oblique walls and the noise they introduce we decided to use a second grid on which the measurements are laid with a 45° angular shift (Part 3). Superfluous data elimination and segmentation are also applied on this second grid.

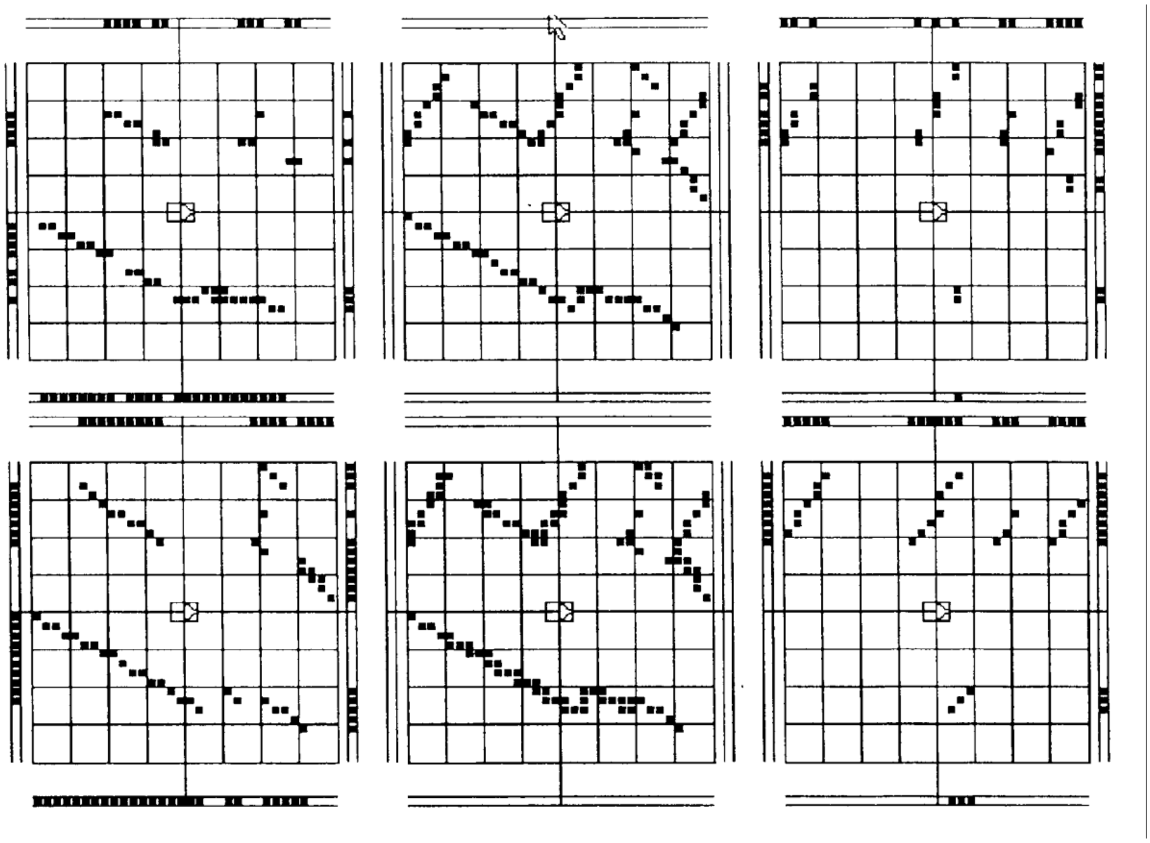


Fig. 2.2.3. Extraction of segments in the 4 filtering directions:
a (upper left): Lengthwise segmentation,
b (upper centre): Refined environment,
c (upper right): Crosswise segmentation,
d (lower left): First diagonal segmentation,
e (lower centre): Initial measurements,
f (lower right): Second diagonal segmentation

A search for the longest and the shortest continuous segments is performed [Bras95] among the projections of the environment on the crosswise and lengthwise axis in each quadrant of the cellular space according to a filtering direction (lengthwise, crosswise, diagonal 1 and 2). A reorientation angle is then computed according to:

$$\Theta = \arctan \frac{\text{ShortestCrosswiseSegment}}{\text{LongestLengthwisesegment}} \pm n * \frac{\pi}{4} \tag{2.2.1}$$

According to the direction in which the longest segment is found (i.e. the most plausible reference in the environment), adequate choices for the sign and the value of n (n in {0, 1}) lead the robot to be reoriented parallel to the longest segment ($0 \leq \Theta \leq \pi/3$) or perpendicularly to it ($\pi/3 < \Theta \leq \pi/2$). The reoriented cellular space is re-built from the initial measurements according to the reorientation angle. Fig. 4c shows the benefits of the reorientation. The reoriented cellular space is then considered as reliable and will allow the landmarks to be extracted.

(d) Landmarks extraction: as told in the introduction, the environments are described using a fresco made of ordered series of landmarks: "Opening", "Wall" also called "Closure" and "Corner" also called "Angle_of_Closures". Let us note that an "Angle_of_Closures" must be neighboured by two "End_of_Closure" landmarks. The landmarks extraction first considers the "Opening" elements that are directly extracted from the reoriented signatures.

The "Angle_of_Closures" and "End_of_Closure" landmarks are extracted from the reoriented cellular space by the following laws. The first operation consists in the "Angle_of_Closures" extraction by the following equation that is applied to every cell in the grid:

Angle_of_Closures = ((CELL & CELL_W) | (CELL & CELL_E)) & ((CELL | CELL_N) | (CELL & CELL_S)) & neg CELLdiag

with: CELLdiag meaning that the logical state of the cell is true if it belongs to a diagonal. Operators & (logical AND) and neg (logical NOT) are applied on the states of the cells.

The first ligne of this equation checks if the cell has east or west neighbours while the second line checks north and south neighbours. Therefore a cell is considered as an Angle_of_Closures if it has at least a crosswise and a lengthwise neighbour.

The second operation aims at extracting the "Lenghtwise End_of_Closure" and "Crosswise End_of_Closure" landmarks. These operations are allowed if and only if the cell does not belong to the two diagonals and is not an "Angle_of_Closures".

Fig. 2.2.4d and 2.2.4e show the "Angle_of_Closures" and "End_of_Closure" landmarks positioned on the grids. To each landmark are associated three qualitative attributes representing three properties of landmarks. The off-sight attribute is set when the landmark stands on the cellular space border. The position attribute can take the following values: crosswise, diagonal, lengthwise according its position. The certainty attribute is introduced to take into account landmarks that could come from a possible noise introduced in the digitisation process not detected by the previous laws or a still possible bad reorientation. It is false for every landmark (for instance, diagonal "End_of_Closure", "45°_angles") whose evolution cannot be known.

Fresco construction

The first step of the fresco construction gathers the landmarks space into ordered series of semantic clues and describes the environment by positioning landmarks in respect to each others. Each landmark has exactly two neighbours (the last landmark in the list has the first one as second neighbour). Building the fresco is made using the symbols presented in Table 1 which gathers the landmarks identity and attributes. The landmarks identity and attributes have been chosen according to the indoor environment in which the robot moves. This operation mainly aims at eliminating the notion of distance to the profit of a spatial series and highlights the qualitative representation of the environment. An example of fresco is given in fig. 2.2.4f. The robot is situated in the middle of the environment. To each landmark are associated three qualitative attributes representing three properties of landmarks. The off-sight attribute is set when the landmark stands close to or beyond the end of the sensor range. The position attribute can take the following values: crosswise, diagonal or lengthwise according its position related to the lenghtwise and crosswise robot axis. The certainty attribute is introduced to take into account landmarks whose evolution can be forecast. It is false for every landmark (for instance, diagonal "End_of_Closure", "45°_angles") that could come from a possible noise introduced in the digitisation process and whose evolution cannot be known [Pradel00], [Pradel01].

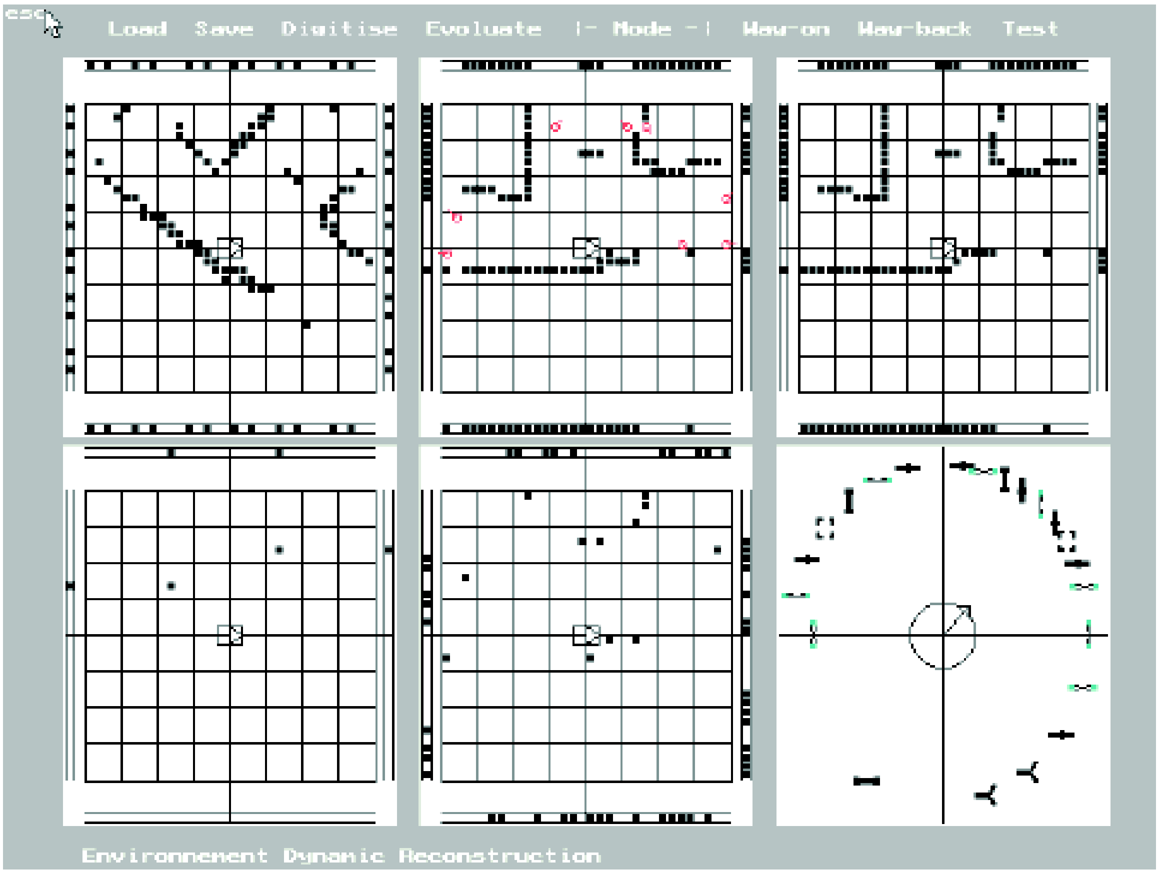


Fig. 2.2.4. Example of the digitised constructions:
a (upper left): real world from raw measurements; b (upper centre): reoriented cellular space; c (upper right): refined space after superfluous data elimination;
d (lower left): Angles_of_Closure extraction,
e (lower centre): End_of_Closure extraction;
f (lower right): fresco construction

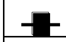
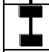
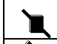
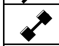
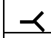
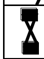
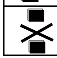
Symbol	Landmark	Position	Off-sight	Certainty
	Angle_of_Closure			True
	End_of_Closure	lengthwise		True
	End_of_Closure	lengthwise	off_sight	False
	End_of_Closure	crosswise		True
	End_of_Closure	crosswise	off_sight	False
	End_of_Closure	diagonal1		False
	End_of_Closure	diagonal1	off_sight	False
	End_of_Closure	diagonal2		False
	End_of_Closure	diagonal2	off_sight	False
	45°Angle	lengthwise		False
	45°Angle	crosswise		False
	Opening	lengthwise		True
	Breakthrough	lengthwise		True
	Opening	crosswise		True
	Breakthrough	crosswise		True

Table 2.2.1. Landmarks used in the fresco construction.

The second step focuses on the fresco validation. Assuming that there is only one description for one environment, strict laws of neighbourhood are defined. Fig. 2.2.5 shows these neighbourhood laws that can be interpreted as a set of logical assertions. An Angle_of_Closure can only have as neighbours Angle_of_Closures or End_of_Closures. For each landmark, the neighbourhood is checked. Every time a fresco is built, the whole set of these rules is applied in order to validate the fresco. If one rule failed, the fresco is not valid.

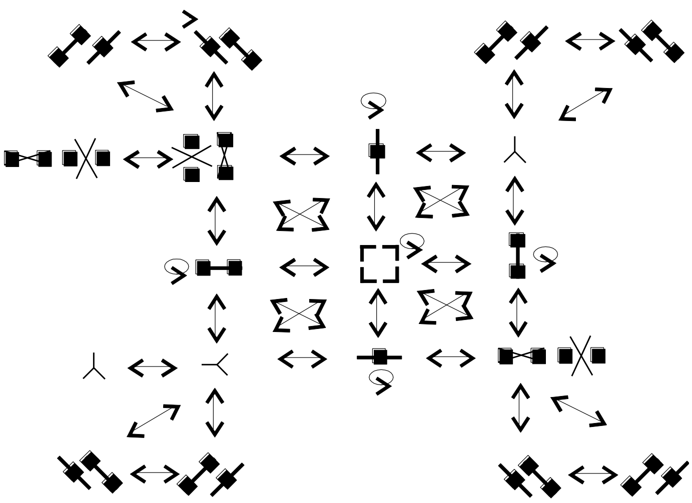


Fig. 2.2.5. Landmarks neighbourhood rules.

The validation fails mainly due to a bad landmark extraction process in a very noisy cellular space or a bad reorientation. Making the necessary corrections in the extraction laws to solve these seldom failing cases leads to an increasing of the complexity of the evolution laws, increasing not really justified by the low frequency of the failures. We consider that the loss of a fresco is not an important drawback: a failure in the validation of the fresco will be corrected by the next valid one with only slight effects on the mission of the robot and the effect of this loss is very attenuated because the process of transitions detection and environment memorisation eliminates a greater part of the frescoes. When it is validated, the fresco appears as shown in fig. 2.2.4f. A fresco will contain at most 64 landmarks symbols organised into 4 sectors of 16 symbols at most.

Symbolic trajectory description using frescoes

Building the symbolic description of the route followed by the robot is three-fold:

- how to build the qualitative descriptions (frescoes) in accordance with the robot's sensors ?
- how to describe the route by a sequence of the most pertinent frescoes ?
- how to use these frescoes with the control-command level of the robot ?

This section deals with the second point. The choice of the most salient frescoes is made using different criteria described in the following sections. Every time the laser range finder scans the environment, a fresco is built. In our case, the fresco built-in period is 300ms. Hence, if all frescoes are stored their number grows quickly and some of them are not useful. Storing all the frescoes when the robot runs in a corridor is a trivial example. All frescoes are very similar excepted at both ends. If only few frescoes are useful, how then is it possible to select them? Is a specific sequence of frescoes able to describe a part of the environment? Answering, at least partially, to these questions is the aim of this section.

Following a specific path, the total number of stacked frescoes could be large enough. Moreover, successive frescoes could be identical or slightly different. Therefore, a selection of meaningful frescoes, which offers a thoroughly environment description, is absolutely necessary. Based on these salient selected frescoes, the robot also should be able to find a return path. In local homing for example, an agent returns to a previously visited location by moving to maximize the correspondence between what it sees currently and a remembered view from the target.

In dealing with frescoes, which are basically a collection of symbolic strings, we were inspired by different methods, such as those used in spell checking, optical character recognition (OCR), molecular biology for DNA or amino-acid sequences study [Altschul91], [Karlin90] or computational intelligence.

The first two criteria proposed to evaluate a kind of distance between frescoes are called resemblance and barycentre. A new fresco is considered as bringing new information if its distance to the previous stored one regarding one of the criteria is greater than a threshold. The two next sections describe these criteria. A systematic study gives an evaluation of the thresholds to use to make the criteria effective.

Resemblance method

This criterion uses a nearby principle of that presented in [Hong91]. A correlation function allows calculating the resemblance between two frescoes. This criterion has been tested in the same environment as that used for the construction and the validation of the frescoes. The use of this criterion shows that the landmarks that are not certain make very difficult the evaluation of the resemblance so only the certain elements were kept. The resemblance between two consecutive frescoes is calculated by taking into account the difference between the number of certain landmarks in the corresponding quadrants. The resemblance between two frescoes is calculated from the difference between the number of landmarks in respective quadrants of two consecutive frescoes. The comparison of this difference with a reference threshold indicates if the current fresco should be kept or rejected because not bringing enough information.

The resemblance between two consecutive frescoes i and j is calculated as:

$$r_{ij} = |N0_i-N0_j| + |N1_i-N1_j| + |N2_i-N2_j| + |N3_i-N3_j| \quad (2.2.2)$$

where
Nk_i, k = 1 ... 4 represents the number of landmarks in quadrant k of the i-th fresco Nk_j, k = 1 ... 4 represents the number of landmarks in quadrant k of the j-th fresco.

If the resemblance r_{ij} is greater then an a priori specified threshold then the j-th fresco will be selected and memorized as sufficiently different from the rest.

Barycentre method

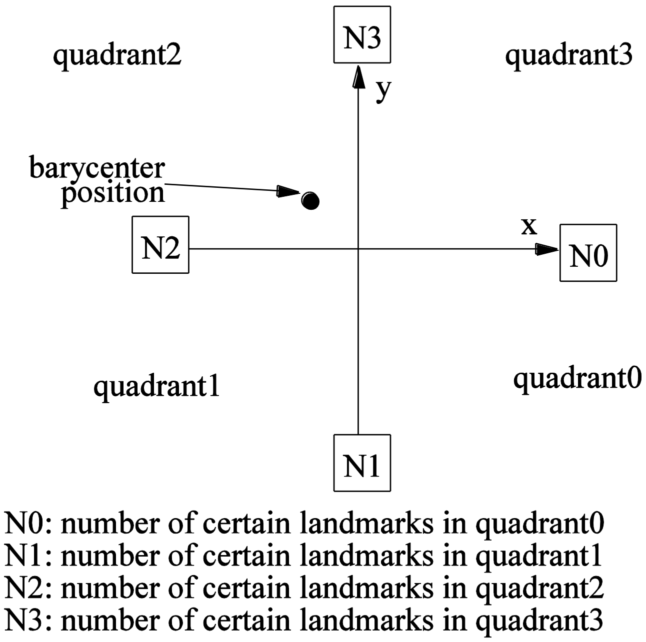


Fig. 2.2.6. Barycentre computation between certain landmarks.

This criterion is inspired by the distance of Hausdorff which measures the distance between two sets [Ahuactzin95], [Huttenlocher93]. In our case, this notion was very simplified to respect real-time constraints. It takes into account only the number of certain landmarks in every quadrant. The landmarks are positioned as indicated on the fig. 2.2.6 and the barycentre is positioned at the following coordinates:

$$xref = \frac{N0_i - N2_i}{Ntot_i}; x = \frac{N0_j - N2_j}{Ntot_j} \quad (2.2.3)$$

$$yref = \frac{N1_i - N3_i}{Ntot_i}; y = \frac{N1_j - N3_j}{Ntot_j} \quad (2.2.4)$$

$$bary_{ij} = \sqrt{(x_{ref} - x)^2 + (y_{ref} - y)^2} \quad (2.2.5)$$

where
Nk_i, k = 1 ... 4 is the number of landmarks in quadrant k of the i-th fresco,
Nk_j, k = 1 ... 4 is the number of landmarks in quadrant k of the fresco,
$Ntot_i$ and $Ntot_j$ are the total numbers of certain landmarks in the i-th/j-th frescoes respectively.

Any variation of the number of elements in a quadrant implies a movement of the barycentre. If this displacement is greater then an a priori specified threshold then the j-th fresco will be selected and memorized.

Distances based methods

Distance is usually, but not necessarily, defined on a vector space. For strings, there are also some ways for quantifying how much two strings differs, as we will see in the next sections. These metric functions attempt to ascribe a numeric value to the degree of dissimilarity between two strings.

(a) Hamming distance method: the Hamming distance (HD) could be defined only for strings of the same length [Gusfield97]. For two strings, S_1 and S_2, the Hamming distance HD(S_1, S_2) represents the number of places in which the two strings differ, (Lamon, 2001) have different characters as shown in the following example: HD ('ABCD', 'ACDB') = 3

(b) Levenshtein distance method: the Levenshtein distance (LD) realizes a more complex evaluation of two strings than the Hamming distance. It could operate with strings not necessary of the same length and represents the minimum number of elementary transformations (insertion, deletion and substitution of a symbol) needed to transform one string into another [Levenshtein66]:

$$LD(S_1, S_2) = \min(N_{ins} + N_{del} + N_{subs}t) \quad (2.2.6)$$

Closely related to it is the weighted Levensthein distance (WLD) also known as edit distance, where different costs are assigned to each edit operation [Kohonen88], [Wagner74]:

$$WLD(S_1, S_2) = \min(w_{ins}N_{ins} + w_{del}N_{del} + w_{subst}N_{subst}) \quad (2.2.7)$$

(c) N-Gram method: an N-gram is a substring of N consecutive symbols. Let N_1 and N_2 be the number of N-grams in strings S_1 and S_2, respectively let *m* be the number of matching N-grams. If one string is longer than the other, the unmatched N-grams are also counted as differences. The feature distance (FD) is defined then as [Kohonen87]:

$FD(S_1, S_2) = \max(N_1,N_2) - m(S_1, S_2)$ (2.2.8)

Similarity based methods

Finding similarities in character strings is an important problem in text processing and data mining. It has applications in genetics research as well, since strands of DNA can be expressed as very long strings of the characters.
A similarity measure is simpler than a distance. For strings S_1, S_2, finding similarities in character strings is an important problem in text processing and data mining. It has applications in genetics research as well, since strands of DNA can be expressed as very long strings of characters.
A similarity measure is simpler than a distance. For strings S_1, S_2 $\in S$, any function $s: S^2 \to \Re$ can be declared similarity. For strings, similarity is closely related to alignment.

(a)Cross correlation matching method

This function is commonly used in signal processing. For symbols, the function compares string S_1 (of length m) with S_2 (of length $l = n \geq m$) and produces a cross correlation similarity vector, CCS, of length ($l = m + n-1$) with elements CCS_i (with i = 0, 1 … l-1) given by [Gusfield97] [Haykin99]:

$$CCS_i(s_1, s_2) = \begin{cases} \sum_{j=0}^{i} S(s_{1j}, s_{2(n-1)-i-j}), if \;\; i = 0...m-1 \\ \sum_{j=0}^{m-1} S(s_{1j}, s_{2(n-1)-i-j}), if \;\; i = m...n-1, m \neq n \\ \sum_{j=0}^{(l-1)-i} S(s_{1(n-1)-(l-1-i)+j}, s_{2j}), if \;\; i = n...l-1 \end{cases} \quad (2.2.9)$$

where:

$$S(x,y) = \begin{cases} 1 \; if \;\; x = y \\ 0 \; if \;\; x \neq y \end{cases} \quad (2.2.10)$$

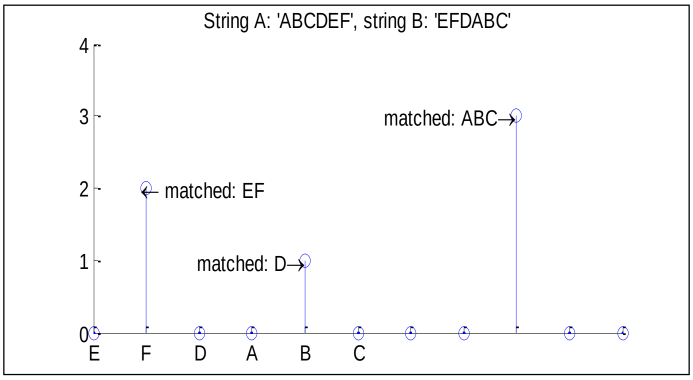


Fig. 2.2.7. Results of the cross correlation function. Peak value is obtained for the alignment of "ABC" tri-gram.

Fig. 2.2.7 gives an example of the results given by the cross correlation matching method for two strings of length equal to 6.

Neural network based method

Speaking in a neural network terminology, finding the salient frescoes is equivalent with finding prototype vectors. Self Organizing Feature Map-Neural Networks, SOFM-NN, tries to place or adapt neurons in such a way that they serve as good prototypes of input data for which they are sensitive.

(a) Classic SOFM-NN: these networks are based on unsupervised competitive learning and winner-takes-all neurons [Haykin99]. During the training phase a SOFM-NN creates a topologic correspondence between the spatial location of the neurons in the output layer and the intrinsic features of the input patterns. If there is some similarity between input vectors then neighbours neurons will fire. If two input patterns are different than output neurons situated at considerable distance or spatial location will respond. For prototype vectors calculus usually Euclidian distance is used, as elements having the smallest sum of squared distance over the data set.
The principal problem is that classic SOFM-NN training algorithm is defined for numbers and not for strings. There are numerous ways for string to numbers conversion and vice versa [Aha91], [Blanzieri99]. For our particular case, the maximum number of symbols within a fresco is 16, hex coded. So the NN input vector could be constructed by means of:

- Direct coding: each symbol had its own binary equivalent (0 = 0000, 1 = 00001 … F = 1111),
- Exclusive coding that is, the symbol is coded with an unary vector with all the components but the i-th set to zero (0 = 000 … 0001, 1 = 00 … 010, … , F = 10 … 000).

Finally, a fresco will be represented as a binary vector composed by concatenation of each binary coded constituent string.

(b) Symbolic SOFM-NN: based on distance measure for strings and calculating the prototype as a mean or median value, SOFM-NN for strings have been defined [Kohonen98]. These SOFM-NN are organized as a symbol strings array, whereby the relative locations of the strings on the SOFM ought to reflect some distance measure (e.g. LD, FD) between the strings. The idea behind the pure symbolic SOFM-NN is to define similarities or distances between data objects. In our application data objects are represented by symbolic strings. Based on these similarities/distances, finding representative prototypes (for our application, meaningful frescoes) will be the next step.
In training of pure symbolic SOFM-NN, two steps are repeated:

- Find best-matching unit (BMU) for each data item, and add the data item to the list of its best-matching unit; BMU is found using the defined similarity/distance measure,
- Update the models in SOFM nodes: find the median data item belonging to the union of the list (data list union contains all data items in the neighbourhood of the current node being update).

For computing the median data item, assume there are 3 data items (e.g. symbol strings S_1, S_2, S_3) and the following pair wise distances (Table 2.2.2):

	S_1	S_2	S_3
S_1	0	4	2
S_2	4	0	2
S_3	1	2	0

Table 2.2.2. Pair wise distances.

then compute the sum of the distances from each data item to others:

$$\begin{aligned} S_1 &: 0 + 4 + 1 = 5 \\ S_2 &: 4 + 0 + 2 = 6 \\ S_3 &: 1 + 2 + 0 = 3 \end{aligned} \quad (2.2.11)$$

The smallest sum of distances is for data item S_3, so that is the median of this data set. In the case of SOFM-NN, the distances could be weighted by the value of neighbourhood function (e.g. a Gaussian-shaped neighbourhood function).

Experimental results

Application of the resemblance and barycentre criteria in simple environment

The two criteria apply only on the certain landmarks and have been tested in two types of environments. In a first step, experiments in simple environments led us to point out the thresholds relevant ranges. In a second step, a complex environment has been used to validate these thresholds.
The problem is to find the right threshold for each criterion. A representative panel of situations is first established and systematic tests are made on each situation in which the frescoes are listed for different thresholds of the two criteria. Then a reference threshold for each criterion is fixed taking into account firstly the ratio of kept frescoes and secondly the position of these frescoes with respect to their situation along the robot's route in the considered environment. Finally, thresholds that have been defined are tested in a complex environment.

(a) Choice of different types of environment: indoor environments can be described using a limited number of situations [Al Alan95]: openings, walls, angles, room, corridor, dead-end and crossings. So far, tested situations are listed in Table 2.2.3. Fig. 2.2.8 shows the example of the "opening on the left situation". Numbers on the left of the figure show the different positions where frescoes have been constructed. In this example, frescoes are built from position 1 to position 31 (only one of five is drawn to make the figure readable).

Fig. 2.2.8. Example of situation: Opening on the left.

In the different situations, the initial numbers of frescoes are different (Table 2.2.3).

	Situation	Number of frescoes
Angle to the left	AL	31
Angle to the right	AR	31
Opening on the left	OL	31
Opening on the right	OR	31
X-crossing	CX	42

Table 2.2.3. Initial number of built frescoes.

(b) Number of pertinent frescoes vs. criterion: it is firstly interesting to observe the number of frescoes kept for different values of thresholds. For barycentre criterion, values between 0 and 2 with a step of 0.05 are tested. For resemblance criterion, values between 0 and 12 with a step of 0.5 are tested. Beyond these limits, only fresco number one is kept. As the initial number of frescoes is different in all situations, the ratio between the number of frescoes kept and the initial number of frescoes is analysed. Fig. 2.2.9 shows the results for resemblance criterion. Fig. 2.2.10 shows the results for barycentre criterion. It can be seen that curves in each figure are similar, meaning that criteria have the same response in all the environment situations. It seems then possible to find a common threshold.

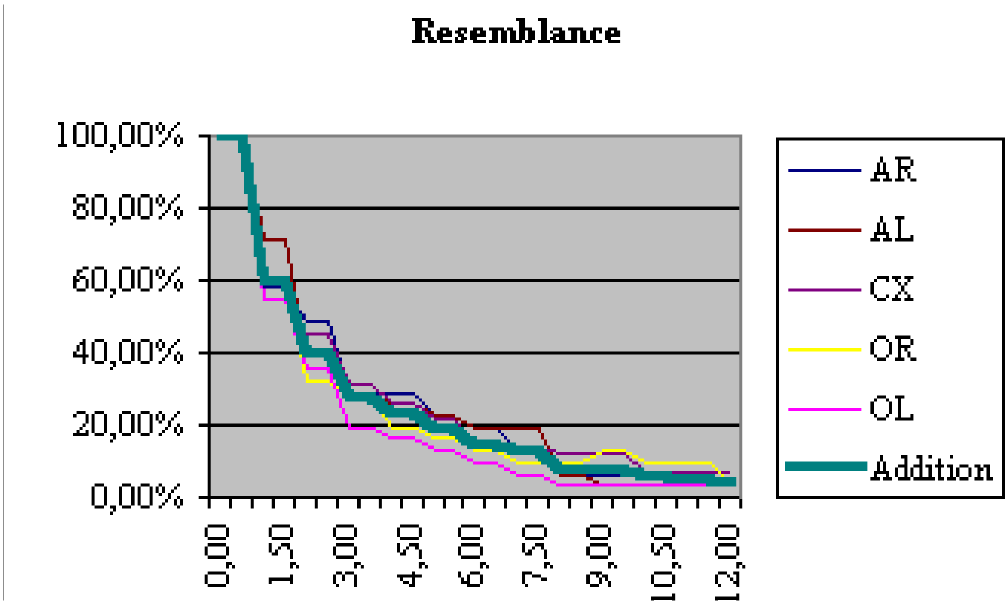


Fig. 9. Percentage of frescoes selected by resemblance criterion vs. threshold value (AR/AL: angle on the right/left, CX: X-crossing, LA: lab, OR/OL opening on the right/left, Sum: add up).

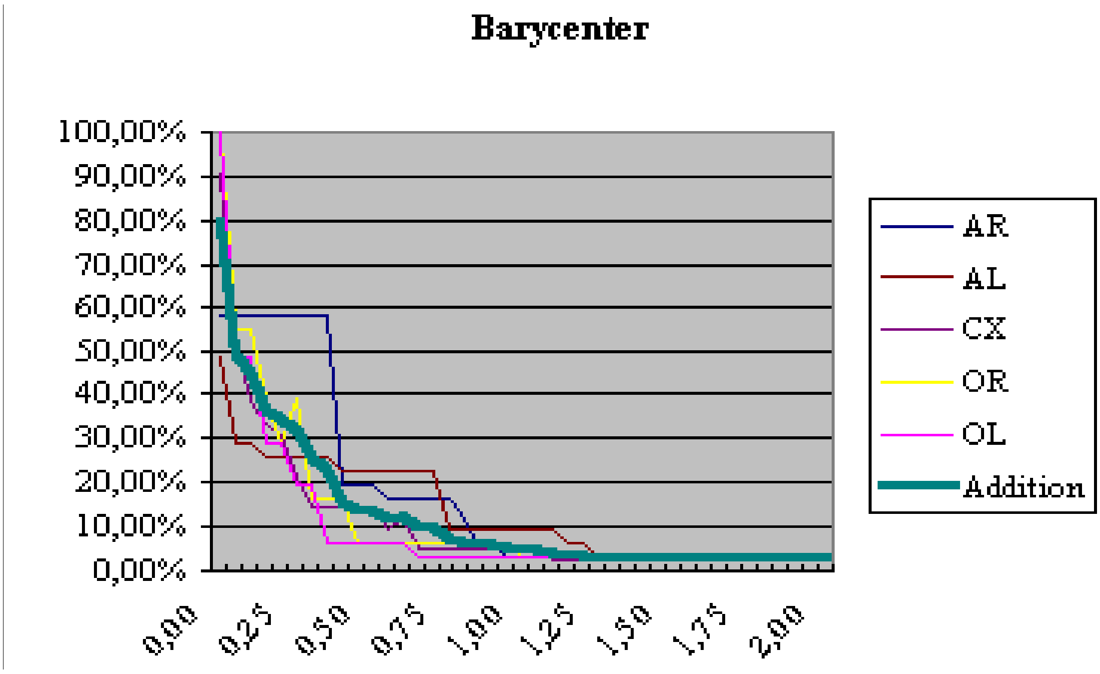
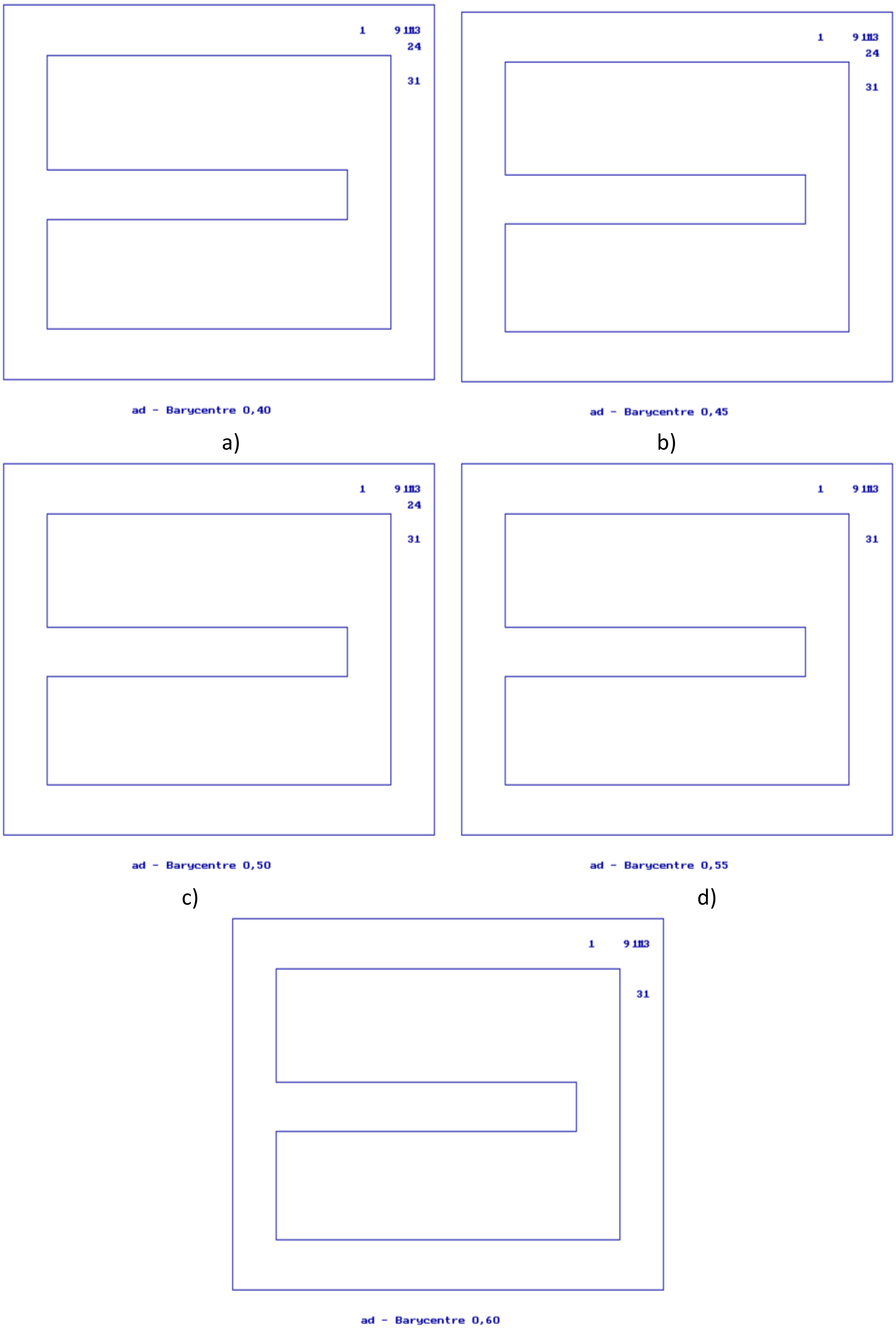


Fig. 10. Percentage of frescoes selected by barycentre criterion vs. threshold value (AR/AL: angle on the right/left,CX: X-crossing, LA: lab, OR/OL opening on the right/left, Sum: add up).



e)

Fig. 2.2.11. Pertinent frescoes vs. barycentre criterion (AL situation); a) threshold=0.40; b) threshold=0.45; c) threshold=0.50; d) threshold=0.55; e) threshold=0.60.

It also can be noted that curves decrease quickly for low thresholds values. In fig. 2.2.8, frescoes between 1 and 10 represent the same part of the environment with very slight differences. The objective is to keep a reasonable part of frescoes between 10% and 20% in the first approximation. That means thresholds values comprise between 5 and 7 for resemblance criterion and between 0.4 to 0.6 for barycentre criterion.

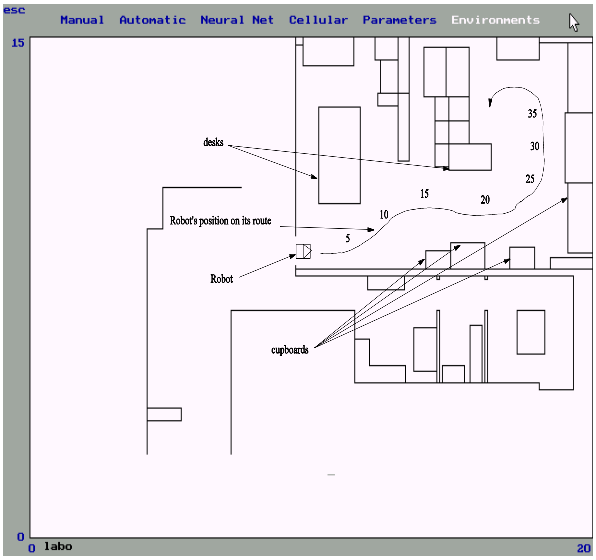
(c) Positions of pertinent frescoes: for both criteria, it is interesting to visualise which frescoes are considered as pertinent (fig. 2.2.11). Frescos number 1 and 31 represent the beginning and the end of the trajectory: they appear for all thresholds. Frescoes 9, 11, 13 and 24 represent the heart of the turning. They are very close considering Euclidean distance but they differ in term of orientation. Fresco number 24 disappears for thresholds equal to 0.55 or 0.60. The value 0.50 is the central threshold value for barycentre criterion. A similar analysis has been conducted for all other situations. In the same way, the resemblance criterion leads to the same conclusion with 6.0 as central threshold.

Application of the resemblance and barycentre criteria in complex environments

A complete trajectory has been studied in a complex environment (fig. 2.2.12 a)). The two criteria have been applied. The variations of the thresholds have been limited to the range determined by the tests in simple environments: 5 to 7 for resemblance and 0.4 to 0.6 for barycentre. Fig. 2.2.13 shows the percentage of selected frescoes for both criteria. For barycentre criterion, there is no significant difference between the complex and the simple environments. For resemblance criterion, the ratio is greater in the complex environment than in the simple ones. Nevertheless, for a threshold equal to 7.0, the ratio becomes close to the ratio obtained in simple environments.

Application of the other criteria

Against the frescoes acquired (fig. 2.2.12 b)) from the lab environment [Hoppenot03], the above mentioned possibilities of salient frescoes selection were implemented and compared.



E00C004000000F6EC700006AC4500000E000020000900000000000000000000040
E000C005F740009002A6000000600000E002090000000000000000000000000040
E000CC004F74000000004C0B00060000E029000000000000000000000000000040
6E00CC0674F704000000004CA6006000E2900000000000000000000000000000F0
0CC0004F74000000004CA6006000EC40000000000000000000000000009B6E00
C05F704F700CE6F0000064CA060E0029000000000000000000000000000F006A6E
0C500F700492C0007000F0004C00EC04000000000000000000000009000BF040E00
05C000F0704004C00700000F04C0E00050000000000000000000F00F000040E00
040000000F6EC76A6F4C7000F4CE000500000000484000009000900 4CE66E0C
E0C00400000000F6EC0B0F4C70000F4CE00005000000990000000000000004CE6
E00C00040000000000004C70F04C760000E000050004C40000000000000004CE6F
0E000C0004000000000004C7F4C000000E000504C4000000000000000064CEF40
E0000C00000400000000000004CE6F4C0E000054C44C40000000004C4F0F004000
0A000060000C04000000004CE6F500EC0044C40000F0F000000FF0004C4F0
0000500000EC4F000000600E0CE00066E60F0F00000FF00006E84FF0000000
0000400000E0C00400F00006E0C000076E6F0F000BB000006E29FF000000000
000000E67600C0000000B0F04C00EC00040F0B094C4000484090 0B000000000
00600E0C0000076E6F04C0EC000400F0F04840000484FFF000000000000000
004C4400E000C0504400F6E0C000070F0F6E606A60FFF0000000000000000000
400E0000C05044000F6EC00002BFF06E6000BBFFF000000000000000000000000
E000C0504400F6E0C000070F0F6E606A60FFF000000000000000000000000500
E0000C500044000004C00070F91E6006E06F000000000000000000000000000500
6000E650080004400004C0070F5E6006E005000000000000000000000000005E0
BB000E6F600000000FF004C000000000E0050000000000000000000000000F00
B00000E6F6000000000FF004C0000000E005000000000000000000000000090B0B

a) b)

Fig. 2.2.12. a) Test environment: the lab; b) Frescoes acquired by the robot from the environment shown in the left.

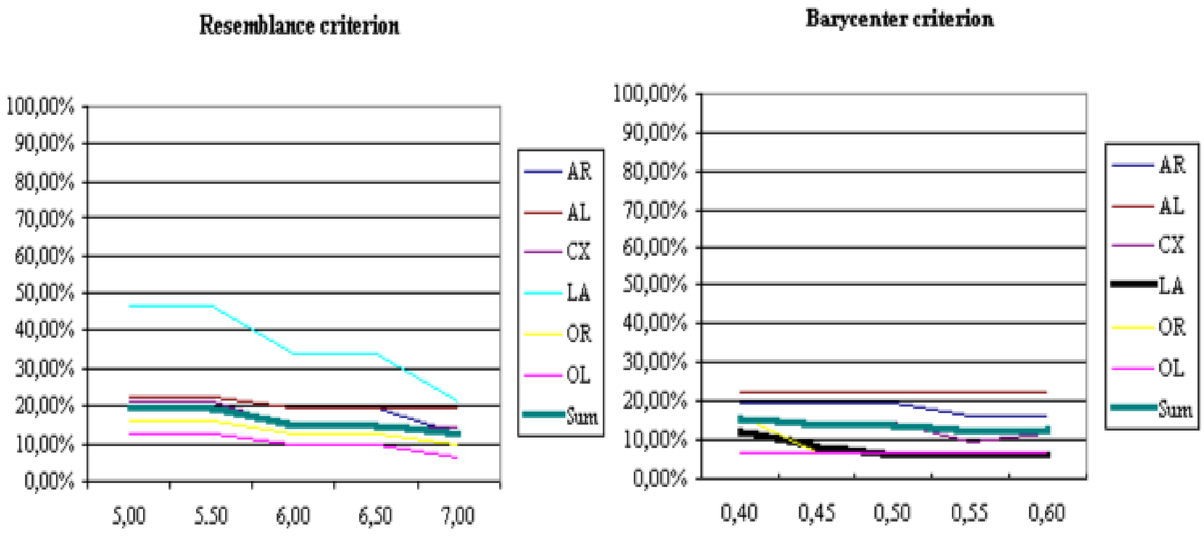
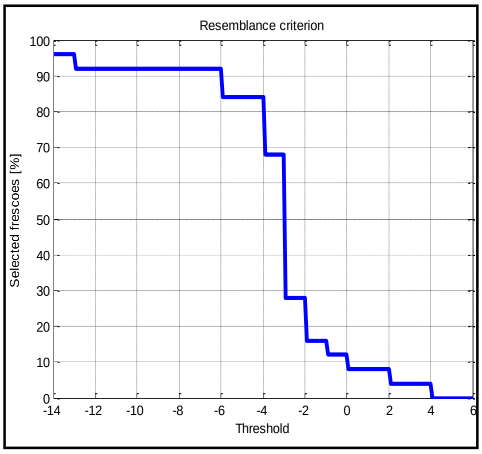


Fig. 2.2.13. Comparison of percentage of frescoes selected by resemblance/barycentre criterion in complex (LA) and simple environments vs. threshold.

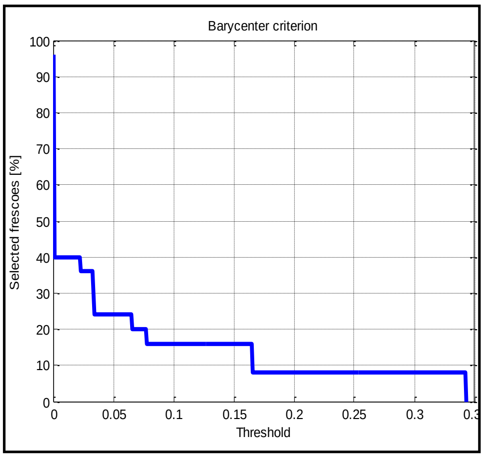
The current numbers of the selected frescoes are synthetically presented in Table 2.2.4.

Method	Index
R	2 3 8 22 23 24 25
B	4 9 11 13 21 23
H	9 10 13 15 17 18 20
L	9 11 15 17 19 22
C	9 11 15 17 18 19 25
N	1 3 7 8 13 15 17

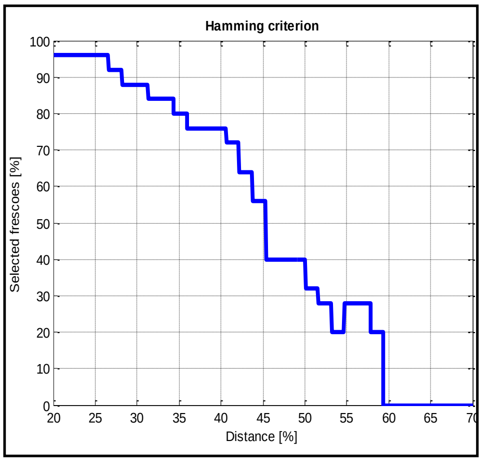
Table 2.2.4. Indexes of selected frescoes with R-Resemblance, B-Barycentre, H-Hamming, L-Levenshtein, C-Cross-correlation, N-Neural Network.



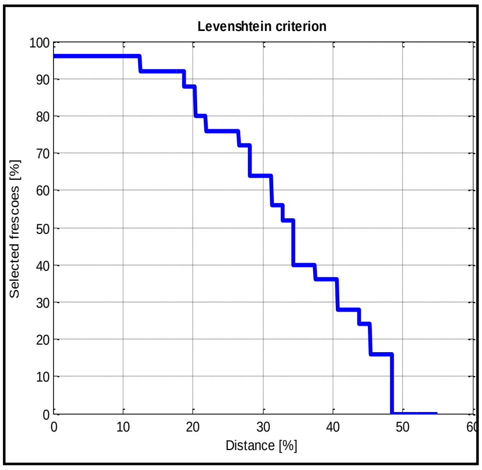
(a)



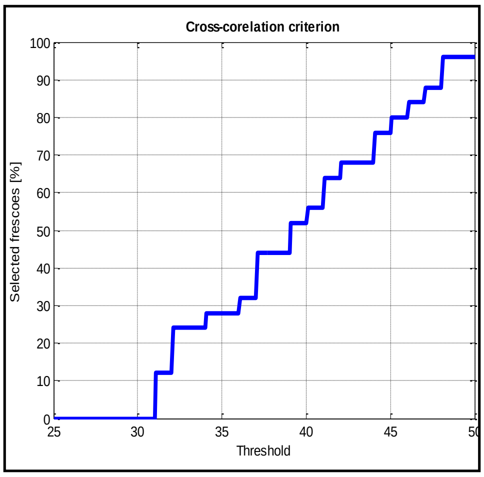
(b)



(c)



(d)



(e)

Fig. 2.2.14. The dependence percent of selected frescoes – threshold. (a) Resemblance criterion; (b) Barycentre criterion; (c) Hamming criterion; (d) Levenshtein criterion (e) Cross-correlation criterion;

In fig. 2.2.14 the dependence percentage of selected frescoes vs. threshold is depicted. Fig. 2.2.15 show the salient frescoes selected by each method. An acceptable percent of the selected meaningful frescoes should be around 30% or less from the total amount of frescoes. In absolute values this mean around 7 frescoes selected.

a) b) d) e) e) f)

Fig. 2.2.15. Selected salient frescoes using the criteria of: a) resemblance; b) barycentre; c) Hamming; d) Levenshtein; e) cross-correlation; f) SOFM-NN.

Future works and conclusion

Human beings, as well as insects [Collet92], use resemblance (or dissimilarity) to compare views of the environment rejecting those that do not bring up new elements without using metrics, only using the occurence of landmarks. In this work, we propose a qualitative method inspired of homing methods [Weber99] to construct the environment surrounding an indoor mobile robot equipped with a 2D telemetry sensor. Every times distance measurements are made, landmarks are extracted and organised into series called frescoes. From this point, distance information are no longer used. In order to derive the pertinent frescoes that can describe the trajectory of the robot, we plan to use a pairing-like method. The first criterion that is primarily being investigated uses a resemblance between two frescoes. The landmarks are bounded and a correlation function measures the difference between consecutives frescoes. The second criterion is based on the difference between the barycentre positions of consecutive frescoes [Huttenlocher93]. Those frescoes separated by a difference higher than a threshold are considered as pertinent to describe the robot's route. In both cases the differences are compared with thresholds that are experimentally set up. Despite the criteria simplicity, the results in the very changing test environment (Fig. 2.2.12a) show that the thresholds experimentally trimmed in simple environments are well fitted to a complex environment. But the resemblance and barycentre methods have the disadvantage of not taking into account the qualitative aspect of landmarks but only the quantitative one. Lets consider an hypothetically example, in which two consecutive frescoes are completely different but has the same number of landmarks/quadrant. Both methods will give an inappropriate answer, resulting in meaningful frescoes losses, because both operate with number, not type, of landmarks. The value of the selection threshold for the resemblance method is also difficult to be anticipated because of rapid variation of the number of selected frescoes in the region of the optimal threshold. It could be easily observed the poor performance of this criterion: the marked frescoes are somehow similar and not representative. The barycentre method is similar with the previous one: in selecting the salient frescoes only number of landmarks from quadrants are counted. It differs in respect of computing the difference between strings and it seems to give slightly better results.

The Hamming distance compares two strings/frescoes character by character, the distance representing the number of different characters found. Here the selection threshold has been expressed in percentage form. The principle underlying Hamming distance is totally different from the previous two methods: it takes into account the qualitative aspect of strings and, as a consequence, is a better solution. In spite of this fact one might consider it giving unsatisfactory results. Let's take a fragment from two successive frescoes, for example: … 0004F74000 … and … 004F740000 … It is clear that these two consecutive frescoes contain basically the same landmark. The 1-character left shift is an environment perspective changing due robot movement along the trajectory. Although, HD score is very high, as the two consecutive frescoes were completely different, resulting in a possible selection of both strings.

This kind of problem is not present in the case of Levenshtein distance. LD computes the distance, in this case, as a simple insertion operation, the distance being kept at minimum. It appears that this method is the best solution for the problem of salient frescoes selection. The computationally cost represents the main disadvantage of LD. One might observe that frescoes are padded with lots of zeros representing empty cells. In order to reduce the computation time, these empty spaces might be ignored. We called the result fast Levenshtein distance, fLD, which produce, in terms of selected frescoes, the same results as LD, but in a significantly shorter time. Almost the same results as LD are given by the cross-correlation principle. Due to alignment underling principle of these methods, the perspective modification of landmarks in a fresco is not seen as a fundamental change.

The SOFM-NN implemented has an input layer of 256 (64 symbols/fresco x 4 bits) neurons and 7 output neurons. Thus, the training is constituted of 25 binary vector having 256 elements. The network has been trained for 5000 epochs. After the learning phase, the seven weight vector corresponding to the output neurons should represent the essential frescoes selected from the input set.

Using a SOFM neural network for salient frescoes selection turns out to be improper. Among possible explanations are:

- The reduced size of training elements; The 25 considered set of frescoes are not enough to form appropriate prototype vectors. Thus, prototype vectors are not entire identically with some of the 25 training frescoes.
- There is no sufficiently redundancy in the 25 frescoes selected.
- The conversion process frescoes -> binary vectors -> real numbers and vice-versa generates errors.

Within the framework of mobile robots navigation, six methods for salient frescoes selection were described and tested. Of the six, the Levenshtein distance and cross-correlation defined for strings approaches produced the most accurate results and had some benefits in interpreting the score in meaningful ways (see Table 2.2.5). The good results given by these approaches could be explained based on theirs ability in dealing with frescoes perspective modification.

	R	**B**	**H**	**L**	**C**	**N**	**Score**
R	-	1(23)	0	1(22)	1(25)	2(3,8)	5
B	1(23)	-	2(9,13)	2(9,11)	2(9,11)	1(13)	8
H	0	2(9,13)	-	3(9,15,17)	4(9,15,17,18)	3(13,15,17)	12
L	1(22)	2(9,11)	3(9,15,17)	-	5(9,11,15,17,19)	3(7,15,17)	14
C	1(25)	2(9,11)	4(9,15,17,18)	5(9,11,15,17,19)	-	2(15,17)	14
N	2(3,8)	1(13)	3(13,15,17)	3(7,15,17)	2(15,17)	-	11

Table 2.2.5. Common selected frescoes. Based on these common frescoes a score for each method is computed.
Legend: R-Resemblance, B-Barycentre, H-Hamming, L-Levenshtein, C-Cross-correlation, N-Neural Network.

One application field is service robotics (e.g., supplying help to old or handicapped persons.). It can be easily foreseen that the robot will have to make return journeys in the user's flat. The problem is then fourfold: i) the journey must be described using a human-like language, ii) series of frescoes are inferred from the route description, iii) navigation uses these series to lead the robot to the target point, iv) the robot has to return to its starting point and must retrieve its route using only the pertinent frescoes recorded when on the way on?

Point 1 was studied in [Saïdi06]. From a high-level route description, the robot's journey is built. The problem was extended to a group of robots. To solve point 4, selected frescoes describing the way on are stored in robot's memory (LIFO). After having been processed its task, the robot has to return on a route that is not exactly the same than the way on. Therefore, the current fresco does not correspond exactly to the stored frescoes (the 180° rotation is, obviously, taken into account): the fresco and one situated on the top of the LIFO do not correspond. A first method consists in shifting left or right the current fresco to better fit to one of the stored frescoes [Pradel00]. Another method consisting in gathering landmarks into representative sets (alcove, cupboard ...) and using all possible transformations of the current fresco is too time consuming. On the contrary, a method grounded on the study of the evolution of very small groups of landmarks is more promising, simple and low resource consuming. On the other hand, with this method, the robot must anticipate the future environments. This anticipation, even if it needs a complete description of all transforms of a fresco, is simpler when the fresco is split into small groups of landmarks. Anticipating frescoes from the current one and comparing them with the stored frescoes seems to be a promising method that will allow the robot to choose the right return way. First results show that the robot is able to return to its starting point in various environments. Nevertheless, the method must be validated in complex and changing environments.

Present and future works focus on points 2, 3. Another perspective is to use a single vision sensor (CCD camera) instead of the laser range finder, extracting distances from images to build a structure similar to frescoes.

2.2.2 Emergent Behaviour Evolution in Collective Autonomous Mobile Robots

This work deals with genetic algorithm based methods for finding optimal structure for a neural network (weights and biases) and for a fuzzy controller (rule set) to control a group of mobile autonomous robots. We have implemented a predator and prey pursuing environment as a test bed for our evolving agents. Using theirs sensorial information and an evolutionary based behaviour decision controller the robots are acting in order to minimize the distance between them and the targets locations. The proposed approach is capable of dealing with changing environments and its effectiveness and efficiency is demonstrated by simulation studies. The goal of the robots, namely catching the targets, could be fulfilled only trough an emergent social behaviour observed in our experimental results.

Introduction

Collective autonomous mobile robots systems represent nowadays the subject of much research [Hawkes03], [Kurabayashi99], [Seo07]. Systems of distributed robots or software agents have obvious advantages such as faster operation, higher efficiency, and better reliability than a single robot system. Novel complex tasks could be solved in parallel either through explicit cooperation, competition or some combination thereof [Stone00].

Lots of algorithms were proposed for efficient control of a group of robots/agents accomplishing a wide variety of tasks. Among them distributed auction algorithm [Wein90], the graph matching algorithm [Kwok02], network simplex algorithm [Orlin96]. More recent, the computational intelligence paradigms (artificial neural networks (ANN) [Zhu06], fuzzy systems (FS) [Zhang03], genetic algorithms (GA) [Leigh05], etc.) try to incarnate the unique behaviours of living creatures in nature onto artefacts like robots.

Our work exploit the combination of several soft computing techniques, ANNs, fuzzy systems and evolutionary algorithms, in the development of a pursuing system in order to demonstrate emergent characteristic of the artificial life into machine learning. The resulting evolutionary artificial neural network (EANN) and evolutionary fuzzy logic controller (EFLC) represent important tools in the evolutionary robotics domain [Wang06]. Numerous researches about the collective autonomous mobile robot evolutionary control in the predator-prey pursuit system have been carried out. Parker and Parashkevov [Parker05] employed a cyclic genetic algorithm (CGA) for evolving single loop control programs for legged robots. The design proved successful for the evolution of a controller that allowed a robot to efficiently search for a static target in a square area. Also they demonstrate the capability of CGAs with conditional branching to generate a controller the predator in a predator-prey scenario. Nolfi and Foreano [98] investigated the role of co-evolution in the context of evolutionary robotics simulating a pursuit system with two robots (a predator and a prey) in real environments. McKennoch *et. al.* [McKennoch04] studies software agents in a predator-prey environment when the movements of prey agents are evolved upon a Mamdani type fuzzy inference system. It has been shown that probabilistic predation and starvation forces, along with simulated communication activity act upon agents, causing them to cluster.

Analysing the literature, some of the weak points of the control paradigms could be pointed out, e.g. when the gene structure was represented by fuzzy membership functions a long time is required in order to resolve the pursuing problem, when reinforcement learning is employed there are difficulties of learning when the rewards of taken actions are not instantly computed [Jeong99]. Also an environment which is dynamically changing represents in most cases an important issue. Therefore, in this paper, to resolve those problems, we apply the EANN and EFLC paradigms for modelling the robots evolving in a dynamic virtual environment.

The Architecture

In recent years, there has been an increasing interest in the utilization of unconventional control strategies such as Computational Intelligence techniques. ANN, FS and GAs are among the most important of them. These control methods derive their advantages from the fact that they do not use any mathematical model of the system and they can deal with complex engineering problems which are difficult to solve by classical methods [Konar05]. By augmenting an ANN or a FS with GAs-based schemes [Yao93], [Seiffert01], [Kumar04], [Herrera05], hybrid evolutionary systems are created with added reasoning, adaptation, and learning ability [Akbarzadeh00], [Tiponuț02].

A brief description of these paradigms along with their implementation for the predator-prey problem are the presented below.

The Paradigms

a) Genetic Algorithms. GAs are stochastic optimization search algorithms based on mechanics of natural selection and evolutionary genetics. If a binary representation for the members of a population is adopted then the basic element processed by a GA is a string formed by concatenating sub-strings, each of which is a binary coding of a parameter. The main operations involved in the generation of new populations mainly consist of: coding the parameters, initialization of the population, evaluate the fitness of each member, selection, reproduction/crossover and mutation.

b) Fuzzy Systems. Fuzzy logic aims to model the imprecise modes of human reasoning and decision making, which are essential to make rational decisions in situations of uncertainty and imprecision. Functioning on the principles of fuzzy logic, the fuzzy systems deal with uncertainty and noise, are inherently robust. These properties led to the use of the fuzzy systems in numerous control problems. The resulting system, called Fuzzy Logic Controller (FLC) capture the knowledge/experience of a human expert linguistically in the form of fuzzy sets, fuzzy logic operations, and the most important aspect, fuzzy rules. They are usually of the form: IF X_1 is A_{i1} AND X_2 is A_{i2} AND… X_n is A_{in} THEN Y_1 is B_{i1} where X_1… X_2 are inputs, Y_1 is the output and A_{in} is the input membership function and B_{i1} is the output membership function.

c) Artificial Neural Network. An ANN is a mathematical model or computational model based on biological neural networks methods that have become very popular recently, involving mapping of input-output vectors for cases where no theoretical model works satisfactorily. ANNs acquire the knowledge trough learning which imply the adjustment of the free network parameters, typically weights and biases, as a response to the environment stimuli.

The Virtual Environment

The virtual environment of the pursuit system is a rectangular (size adjustable at the beginning of the each experiment) lattice where up-and down and right-and-left are connected each other so if a robot or a prey move continuously toward one direction, it comes back to its origin (fig. 2.2.16).

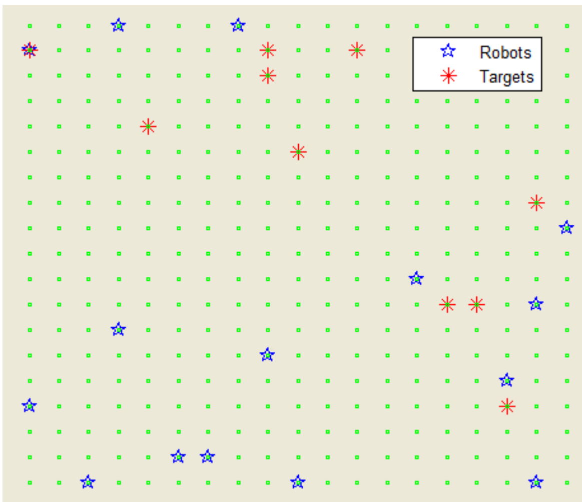


Fig. 2.2.16. A lattice environment.

The robots and the preys contain sensory organs that allows them to perceive limited space of the environment (fig. 2.2.17).

0	0	0	0	0	-	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	-	0	-
0	0	0	0	0	0	0	0	0
0	0	0	0	R	0	1	0	0
0	0	1	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0

Fig. 2.2.16. Both robots and preys have a limited knowledge of the environment. Here, the seeing distance is 2 for the robots (*r_radius*) and 1 for the preys (*t_radius*). The lattice points are coded as follows: "1" for robots, "-1" for preys, "0" empty space.

The robots and the preys could move only in north, south, west or east directions according to their own speed (the same for all robot-type, respectively prey-type, units).

The Controllers

a) Preys. The algorithm used for preys control is fairly simple but very effective: choose the next location as the lattice point situated in the movement range which maximize the total distance (city-block type) from the robots. If there are more points with the same distance, choose one of between them randomly.
b) Robots. First, a rule base algorithm was considered: choose the next location as the lattice point situated in the movement range which minimize the distance (city-block type) from the preys.

Also two hybrid approaches have been implemented: hybridization between Fuzzy Logic and GAs leading to genetic fuzzy systems (GFSs) or evolutionary fuzzy controllers and the hybridization between ANN and GAs leading to evolutionary ANN as depicted in fig. 2.2.17. Details of theirs implementations are given below.

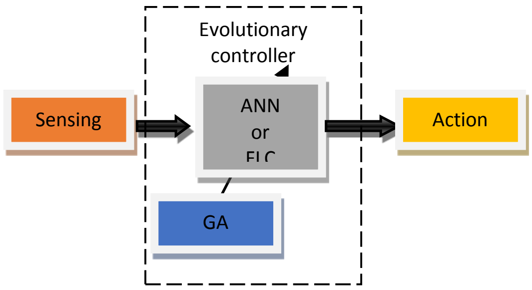
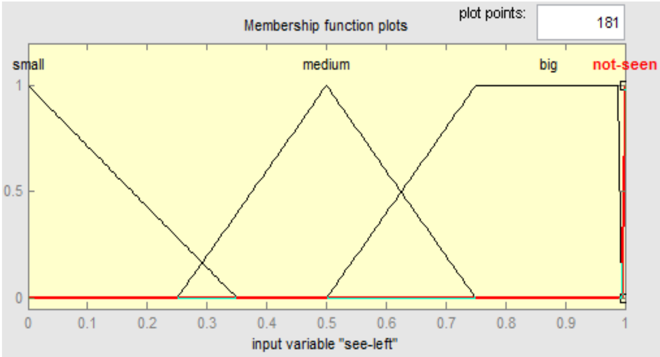


Fig. 2.2.17. Structure of the pursuit system.

EFLC: we have employed a Mamdani type FLC with four input linguistic variables: *see_left, see_right, see_up, see_down* each of them having four linguistic values: *small, medium, big* and *not_seen* (fig. 2.2.18 a), b), c)). The output of the controller, called *action,* has five values: *moveleft, moveright, moveup, movedown* and *stop*. As a result, a robot could move N, S, E or W with the specified speed or it could not change its initial position.



a) Membership functions for the linguistic input variables.

-	0	0	0	1	0	0	0	0
0	0	-	0	0	0	0	0	0
0	0	0	0	0	0	-	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	R	0	1	0	0
0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	0

b) *see_up* and *see_down* regions for a specific robot R.

-	0	0	0	1	0	0	0	0
0	0	-	0	0	0	0	0	0
0	0	0	0	0	0	-	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	R	0	1	0	0
0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	0

c) *see_left* and *see_right* regions for a specific robot R.

Fig. 2.2.18. If more than a target is seen in a region, the city-block normalized distance to the closest target is used for the fuzzyfication process. If there are no targets inside the visual area of the robot, then the *see-left/right/up/down* variable is *not_seen* with the membership value 1.

Although GAs were not specifically designed for learning, but rather as global search algorithms, they could be efficiently employed in the search of the rule sets space. In order to encode rules within a population of individuals the so called “Pittsburgh approach” [De Jong93] – The Chromosome = Set of rules – have been choose for finding optimal rules.

EANN: the second controller has its parameters obtained with the help of combination between an ANN and a GA. Learning in ANNs is achieved by varying the connection weights iteratively so that the network is trained to perform certain tasks. Usually this is done using supervised learning, but in the present case a training set is not available. A prominent feature of EANNs is that they can evolve towards an optimal one in a virtual environment without outside interference.

In our experiments each robot is driven by a Multilayer Perceptron ANN (MLP) with the number of input neurons as a function of *r_radius* parameter (see fig. 2.2.16): *input_neurons* = [(2**r_radius* + 1)^2 − 1]. The number of hidden neurons was chosen at ¼ from the number of input neurons. Five output neurons were considered as ANN outputs. They generate one of the following actions: moving left, right, up, down or stop. We encode, by a genotype representation, just ANN's weights&biases (W&B) not the entire network topology. In the binary form, each connection of W&B is represented by a bit string of certain length. These strings are concatenated to form what is called a chromosome (fig. 2.2.19).

Experimental Results

Simulation

The following steps were performed during the simulation procedure:

1. Initialization. Define the lattice dimension, the properties for robots and preys (number, visual range, speed, controller type and its parameters).
2. Construct the robots brain and code the phenotype.
3. Move the *N* robots and the *T* preys during fixed time. Movements performed in this interval, *s = l*speed/2*, where *l* is the side-length of the square lattice, are enough to explore half of the environment. Robots can obtain fitness values according to proper actions performed. The fitness is defined as the sum of the minimum distance between a robot and a prey calculated at each step:

$$f_i = \sum_{j=1}^{S} \min(d_j(i,t)) \qquad (2.2.12)$$

for all $i \in \{1,...,N\}, t \in \{1,...,T\}$.

4. When run time is completed, calculate the objective function of the system based on the cumulated fitness of each robot. The objective function to be *minimized* represent in our case the total distance between all the robots and the targets:

$$Objective = \sum_{j=1}^{N} f_i \qquad (2.2.13)$$

5. With the rate proportional to the fitness values, next generation of robots are re-produced by crossover and mutation. The GA parameters used are:

- Generation gap, GGAP = 0.2; Fitness-based reinsertion used to implement an elitist strategy whereby the four most fit individuals always propagate through to successive generations. (*N* × GGAP) new individuals are produced at each generation.
- Crossover rate, XOVR = 1; Because we are using discrete recombination, for the breeding of offspring, the crossover rate is not used and, hence XOVR = 1 above.
- Mutation rate, MUTR = 0.03; Represents the rate that the least unit of genetic information can be mutated.

6. Verify the stop criteria (minimum value for the objective function, maximum number of generations). If it is not met go to Step 2.

Evaluation

The first experiment involved *static targets* (fig 6). After 123, respectively 87, generations the evolved ANN, respectively FLC, type controllers fulfilled the goal of minimizing the distance between the robots and the targets (catching all preys) as is depicted by fig. 6 b). The evolution of the objective function trough generations in presented in fig. 7.

The second part of the experiment implies *moving preys* (algorithm described in §2.3 a)). For a 20x20 lattice we set for all *N* = 30 robots *r_radius* = 3 and for *T* = 10 preys *t_radius* = 1 as presented in fig. 2.2.16. The number of rules for FLC was set to 10. Both controllers (EANN and EFLC) have been evolved during the same number of generation, 2000. Then, the evolved structures were saved and the simulation was restarted, this time without creating new generations. The performance of the group was evaluated (Table 2.2.6).

	SIMULATION STEPS			
CONTROLLER	**200**	**400**	**600**	**790**
Random	3	3	4	4
Rules	3	4	5	5
EFLC	4	5	6	6
EANN	3	7	8	10

Tab. 2.2.6. The number of caught preys (from a total of 10 preys).

The results presented in tab. 2.2.6 show that, due to smart evasion target algorithm, an efficient pursuing strategy requires cooperation between agents. The EANN method showed better evolution and excellent emergent behaviour results, outperforming the EFLC or rule base controller. The strategy learned by the ANN based robots was to wait outside de visual range of the prey and pursuit only when there are at least three robots in its vicinity. The evolved robots never attack the targets in groups smaller than three. Despite of the fact that EFLC developed a meaningful rule base, e.g. "IF *see_up* IS *medium* AND *see_right* IS *small* AND *see_down* IS *big* AND *see_left* IS *not_seen* THEN *action* IS *moveright*" its chasing efficency is considerable lower. Possible reasons are: the number of the rules and the fact that the rules do not include information about the units of the same type (robots). They refer exclusively to the position of targets.

Conclusion

Our results demonstrate that the evolutionary design of an ANN controller is viable for relatively complex tasks.

Contribution of this research is applying an evolutionary approach for tuning neural and fuzzy controllers in a predator-prey scenario. The robots have learned to chase based on a collective behaviour which improved dramatically the ability to capture the preys, despite theirs evading strategy. We find that the GA-tuned neural controller outperforms the hand-tuned or GA-tuned fuzzy controller by about 170 percent. In the future, we will adapt this simulation from a software environment to a hardware environment.

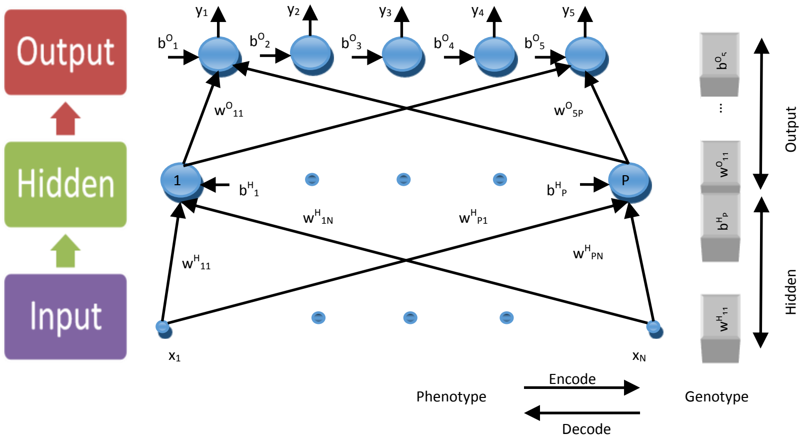
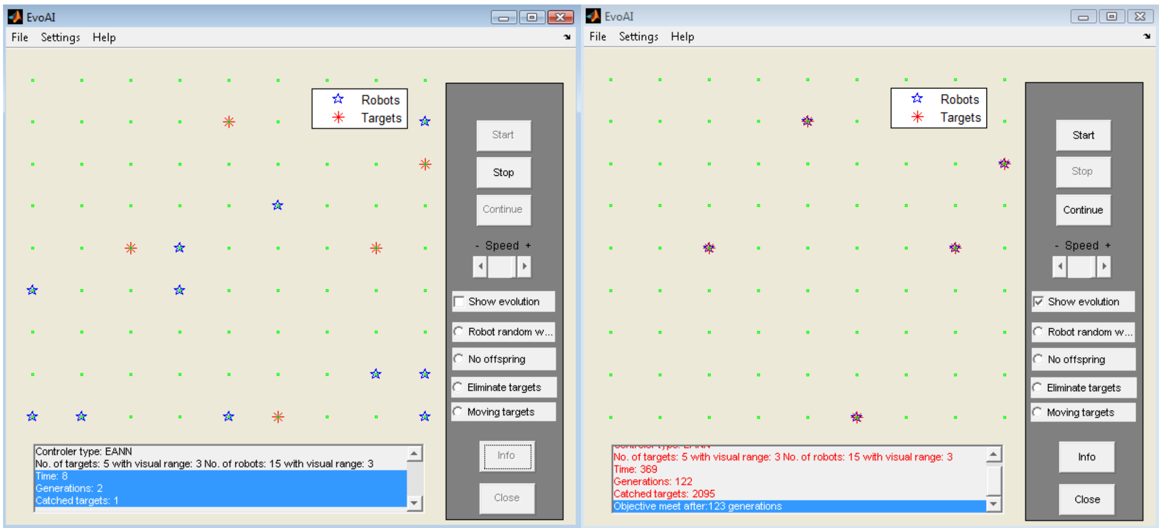


Fig. 2.2.19. Mapping between the weights&biases of the ANN (problem space) and a chromosome (representation space).



a) b)

Fig. 2.2.20. Fixed targets case, EANN controller for the robots. a) At the very beginning of the experiment the robots are spread and move randomly across the map b) After 123 generations all robots have learn how to fulfil the goal of minimizing the distance between them and the targets: complete overlap with the targets.

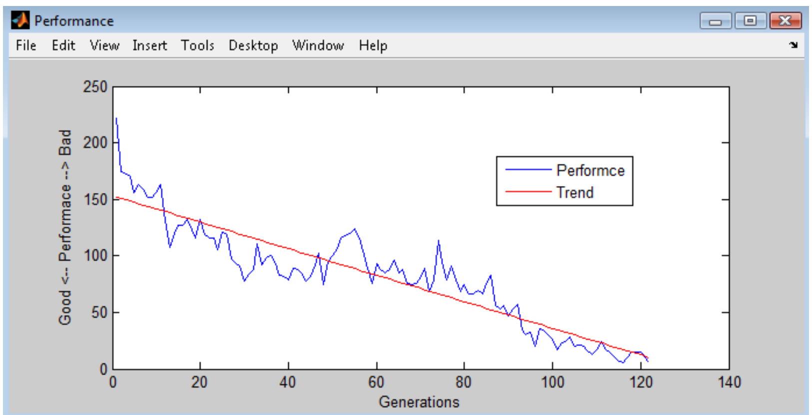


Fig. 2.2.21. The evolution of the performance over the generations

2.3 Artificial intelligence paradigms for human face identification

2.3.1 Interest Operator vs. Gabor Filtering for Facial Imagery Classification

Previous work has shown that Gabor feature extraction is one of the most effective techniques employed for the human face recognition problem. However, the selection of a particular set of Gabor filters is often problematic and, also the computational requirements are considerable. We propose an alternative feature extraction method - the Interest Operator - to be applied for the facial recognition problem. This method has already been successfully used in the mobile robots navigation, stereoscopic vision and automatic target recognition. Experimental results presented in this section indicate that classifiers, both neural (Multilayer Perceptron) and statistical (k Nearest Neighbour), using the Interest Operator–based feature extraction, are capable to achieve almost the same classification rate as the Gabor-wavelet-based methods but in one order of magnitude lower processing time. A special care has been put on the selection of the feature extraction filters and classifiers parameters. Then, on AT&T public facial database, the system has achieved an average recognition rate of 95.2 percent using Gabor Approach and 94.7 percent using the Interest Operator.

Introduction

Face recognition is nowdays an important issue due to its potential applications in different areas: biometrics, information security, law enforcement and surveillance, smart cards, access control and so forth. The details of these applications are referred to in the surveys [Barrett98], [Pentland00], [Zhao03]. Despite the huge interest and the corresponding research effort, human face imagery classification is still an open topic. That is because a computer based face recognition system must take into account the pose variation of faces, presence or absence of structural components, facial expression, orientation, lightning differences, occlusions and self shadowing of facial features.

Many paradigms are trying to deal with the above mentioned difficulties and implement a reliable feature extraction and classification scheme. Among them *Geometric Features based Matching* and *Eigenfaces* represents some of the oldest approaches. The basic idea behind *Geometric Features* principle is to describe the overall configuration of the face by a vector of numerical data representing the relative position and size of the main facial features: eyes and eyebrows, nose and mouth [Brunelli92]. *Eigenfaces* are a set of orthonormal basis vectors computed from a collection of training face images. They provide a basis of low dimensional representation of the facial images and are optimal in the minimum least square error sense [Turk91].

Some of the most promising methods incorporate *Support Vector Machines* with binary tree recognition for multi-class recognition. More on this topic in [Guo01]. [Cesar02] approached facial feature recognition as a problem of *matching inexact graphs* where the graphs were built from regions and relationships between regions in an image.

More recent, other approaches have been employed. For example, in [Wei04] new face recognition algorithm based on the matching of relative image *gradient magnitudes* between images has been proposed. First, a face localization step finds some candidate poses of the face in the image through a fast k-NN search. Then, a face similarity measure is computed from the normalized correlation between the relative gradients. [Hu04], based on a single frontal face image, use a *2D-to-3D integrated face reconstruction* approach Then, realistic virtual faces with different pose, illumination and expression are synthesized based on the personalized 3D face to characterize the face subspace. Finally, face recognition is conducted based on these representative virtual faces. Ekenel and Sankur [Ekenel05] proposed multiresolution analysis for facial recognition. They employ multiresolution analysis to decompose the image into its subbands prior to the subspace operations such as principal or independent component analysis. Based on Gabor kernels placed in fiducial points on the face, excellent results were reported using *elastic bunch graph matching* method [Wiskott99]. In this paper we propose an alternative feature extraction method - the Interest Operator - to be applied for the facial recognition problem and show comparative results with a reference method based on features extracted using Gabor kernels.

Feature extraction

In the feature extraction problem, the task is to find an efficient way to represent the pixel data, usually trough a mapping function, from the original image space to a lower dimensional feature vector space. In the following, two methods for feature extraction are considered: the Gabor approach - as a reference in obtaining best result for the problem of facial recognition and the Interest Operator technique – as a suitable alternative to Gabor filters.

The Gabor approach

Gabor filters, first introduced in [Gabor46], are joint entropy minimizing frequency sensitive filters. Later extended for two dimensions [Daugman85], their use in vision systems is also biologically motivated, as the kernels of Gabor wavelets are similar to the 2D receptive field profiles of the mammalian cortical simple cells and exhibit desirable characteristics of spatial locality and orientation selectivity [Liu01].

In this paper, the following form of a normalized 2-D Gabor filter function, in the continuous spatial domain, has been employed:

$$\psi(x,y;f,\theta)=\frac{f^2}{\pi\gamma\eta}e^{-\left(\frac{f^2}{\gamma^2}x'^2+\frac{f^2}{\eta^2}y'^2\right)}e^{j2\pi f x'} \qquad (2.3.1)$$

$$x'=x\cos(\theta)+y\sin(\theta) \qquad (2.3.2)$$

$$y'=-x\sin(\theta)+y\cos(\theta) \qquad (2.3.3)$$

where f is the frequency of a sinusoidal plane wave, ϑ is the anti-clockwise rotation of the Gaussian envelope and the sinusoid, γ is the spatial width of the filter along the plane wave, and η the spatial width perpendicular to the wave, as it was presented in [Kyrki02].

In order to represent face images using Gabor filters, we have placed a square grid over the face region in the image. At each grid point on the image we have convolved the image with Gabor kernels (fig. 2.3.1).

Fig.2.3.1. A rectangular placement for the fiducial points.

The set of convolution coefficients for kernels of different orientations and frequencies at one image pixel is called a *jet*. The responses of convolutions in an image ξ(*x*, *y*), around a given pixel situated at location (*x*, *y*), are given by:

$$r_{\xi}(x, y; f, \theta) = \psi(x, y; f, \theta) * \xi(x, y) = \int\int_{-\infty}^{\infty} \psi(x - x_{\tau}, y - y_{\tau}; f, \theta)\xi(x_{\tau}, y_{\tau})\mathrm{d}x_{\tau}\mathrm{d}y_{\tau}$$

(2.3.4)

and it represents feature vectors to be further classified.

The Interest Operator

The *Interest Operator* feature extraction technique was first introduced by Moravec [Moravec81] for the purpose of mobile robot navigation. Later, it has been used by Nasrabadi and Choo for a Hopfield neural network based stereoscopic vision system [Nasrabadi94] and by Wang et. al in automatic target recognition [Wang98]. In [Căleanu00] and [Căleanu01], it has been employed in the facial recognition problem. More recent, Zhong-Qiu Zhao et al. reports good results using the interest operator for the facial imagery classification [Zhao04].

The equations used in this paper for the directional variance features, computed for a given image block are:

$$\sigma^2 = \frac{1}{P\,x\,Q}\sum_{y=0}^{P-1}\sum_{x=0}^{Q-1}[p(x, y) - \mu]^2 \quad (2.3.5)$$

$$\sigma^2{}_H = \frac{1}{P\,x\,Q}\sum_{y=0}^{P-1}\sum_{x=0}^{Q-1}[p(y, x+1) - p(y, x)]^2 \quad (2.3.6)$$

$$\sigma^2{}_V = \frac{1}{P\,x\,Q}\sum_{y=0}^{P-1}\sum_{x=0}^{Q-1}[p(y+1, x) - p(y, x)]^2 \quad (2.3.7)$$

$$\sigma^2{}_{135} = \frac{1}{P\,x\,Q}\sum_{y=0}^{P-1}\sum_{x=0}^{Q-1}[p(y+1, x+1) - p(y, x)]^2 \quad (2.3.8)$$

$$\sigma^2{}_{45} = \frac{1}{P\,x\,Q}\sum_{y=0}^{P-1}\sum_{x=0}^{Q-1}[p(y+1, x) - p(y, x+1)]^2 \quad (2.3.9)$$

where μ is the mean of the pixels values within a block:

$$\mu = \frac{1}{P\,x\,Q}\sum_{y=0}^{P-1}\sum_{x=0}^{Q-1}p(y, x) \quad (2.3.10)$$

and $\{p(y,x), 0 \le y \le P\text{-}1, 0 \le x \le Q\text{-}1\}$ represents the pixels in a $P \times Q$ block. Thus, for each image block, five directional variances will be extracted, resulting the features vectors set: $\Theta_i = \{\sigma, \sigma_H, \sigma_V, \sigma_{45}, \sigma_{135}\}_i$, $i = 1 \ldots L$, where L is the total number of distinct blocks in which an image is divided.

Classification

In the automated face recognition problem, we are given a database of image samples of known individuals. The task is to design a system that for any input image, the system identifies the input with one of the known individuals. Thus, the classification problem involves designing a function to map feature vectors to the appropriate class label. To make the comparison more relevant, we used two classifiers. The first classifier is the k-Nearest Neighbour classifier, founded on statistical nonparametric probability density estimation. The second one is the Multilayer Perceptron belonging the category of neural network classifiers. Both are briefly described below.

3.1 Statistical classifier

K-Nearest Neighbour (k-NN) classification is a very simple, yet powerful statistical classification method. Using a suitable distance, e.g. Euclidean (p = 1) or city-block (p = 2) distance:

$$d(\mathbf{x},\mathbf{y}) = \left(\sum_{i=1}^{N} \left| x_i - y_i \right|^p \right)^{\frac{1}{p}} \qquad (2.3.11)$$

one has to find k closest training points to the vector that should be labelled. Then, by specifying a rule, e.g. majority rule or consensus rule, it will be decided how to classify the sample. It can be shown that the performance of a k-NN classifier is always at least half of the best possible classifier for a given problem. More on this topic could be found in [Mitchell97].

Neural Network classifier

Various Artificial Neural Networks architectures were employed for the human face classification problem. Among them, the Multilayer Perceptron [Căleanu00], Convolutional Neural-Network [Lawrence97], Probabilistic Decision-Based Neural Networks [Lin97], Radial Basis Function Neural Networks [Sato98] or Fuzzy ART Neural Networks [Pessoa99].
In our work, we choose the Multilayer Perceptron (MLP) as it is currently the most general-purpose, and (not coincidentally) commonly used neural-network paradigm. The MLP (fig. 2.3.2) learns to generate a mapping from the input pattern space to the output pattern space by minimizing the error between the actual output produced by the network and the desired output across a set of pattern vector pairs.

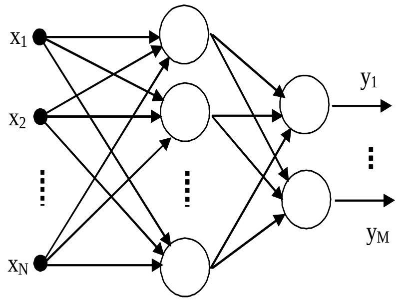


Fig. 2.3.2. The architecture of a single hidden layer Perceptron neural network.

As the training algorithm, a modified conjugate gradient (CG) type was selected. In the classical conjugate gradient algorithms a search is performed along conjugate directions, not only in the steepest descent direction (negative of the gradient) as in gradient descent method [Haykin99]:

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \eta_k \mathbf{p}_k \quad (2.3.12)$$

where $\mathbf{p}_k$ represents the search directions computed in the following manner:

$$\mathbf{p}_0 = -\mathbf{g}_0 \quad (2.3.13)$$

$$\mathbf{p}_{k+1} = -\mathbf{g}_k + \beta_k \mathbf{p}_k \quad (2.3.14)$$

where $\mathbf{w}_k$ is a vector of current weights and biases, $\mathbf{g}_k$ is the current gradient, and η_k is the learning rate.

In the experimental part we employed the scaled conjugate gradient (SCG). It is based on conjugate directions, but this does not perform a line search at each iteration. SCG needs to calculate Hessian matrix which is approximated by:

$$E''(\mathbf{w}_k)\mathbf{p}_k = \frac{E'(\mathbf{w}_k + \sigma_k \mathbf{p}_k) - E'(\mathbf{w}_k)}{\sigma_k} + \lambda_k \mathbf{p}_k$$

(2.3.15)

where E' and E'' are the first and second derivative information of global error function $E(\mathbf{w}_k)$. The terms σ_k and λ_k represent a parameter controlling the change in weight for second derivative approximation and, respectively, a parameter for regulating the indefiniteness of the Hessian. A detailed description of the SCG algorithm can be found in [Moller93].

Methodology

All experiments have been carrying out using AT&T Laboratories Database of Faces (formerly 'The ORL database'). It contains ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The size of each image is 92x112 pixels, with 256 grey levels per pixel. A preview image of a part of the AT&T Database of Faces is presented in fig. 2.3.3.

Fig. 2.3.3. A preview image of a part of the AT&T Database of Faces.

In all experiments the available data were divided in half: 40 persons x 5 images for training purpose and the rest of 200 images for the testing purpose. The training and test images have been randomly selected. However, when the effect of a certain filter/classifier parameter was analyzed, in order to evaluate its usefulness in capturing relevant information, the data sets have been freezed. Otherwise, the results would have been also influenced by the data selection process.

The system design. Experimental results

In order to achieve better recognition rates with low computational requirements, the following sections are dedicated to the study of filters and classifiers parameters influence. These useful values will be used later in creating the final configuration for the recognition system.

Selection of the features extraction filter parameters

A) *Gabor filter*

In the case of Gabor filters, first we analyzed the influence of Gabor filter *dimension* and the *number of fiducial points* over the classification accuracy. With the best values for these parameters, we want to learn further which *frequencies* and which *orientations* are useful for the face recognition problem. We carried out experiments to show independent contributions of each frequency and orientation on the recognition performance. Thus, by using the most important subset of frequency and orientation parameters, we can speed up the feature representation phase and have a more compact form of feature vectors.

A1) *The influence of Gabor filter dimension and the number of fiducial point over the classification accuracy*.

At each fiducial point, we have convolved a facial subimage with Gabor kernels having seven different frequencies and eight different orientations, given by the following two relations. It was shown [Kyrki01] how orientations must be spaced uniformly, that is:

$$\theta_k = k\frac{2\pi}{n}, k = \{0,...,n-1\} \qquad (2.3.16)$$

where θ_k is the kth orientation and *n* is the number of orientations to be used. In our case *n* = 8. In the selection of discrete frequencies f_k , exponential sampling must be used [Kamarainen02], that is:

$$f_k = a^{-k} f_{\max}, k = \{0,...,m-1\} \qquad (2.3.17)$$

where f_k is the kth frequency, f_{max} is the highest frequency desired, and *a* is the frequency scaling factor (*a* > 1). In our case *a* = 2 for octave spacing and *m* = 7.
The magnitude of complex outputs of Gabor convolutions are used as feature descriptors, giving a feature vector size of 56 (8 orientations x 7 frequencies) at each image fiducial point.
The experimental results varying both filter dimension (*gwx*, *gwy*) and the number of fiducial points *(px*, *py*) are presented in Table 2.3.1 and Table. 2.3.2. In order to minimize the effect of the randomly weights initialization, five experiments, using the same data set, has been carried out for the case of the MLP classifier.

		px py				
		6 | 7	7 | 9	8 | 10	9 | 11	10 | 12
gwx	*gwy*					
7	7	26.5	20.0	17.5	16.5	15.0
9	9	19.0	14.5	15.5	15.5	12.5
11	11	17.5	19.0	18.5	16.0	-

Tab. 2.3.1. Test error [%] for the k-NN classifier.

		px py				
		6 | 7	7 | 9	8 | 10	9 | 11	10 | 12
gwx	*gwy*					
7	7	32.5	25.7	27.1	22.5	18.3
9	9	21.1	16.5	15.0	17.0	11.6
11	11	22.5	18.9	17.9	18.7	-

Tab. 2.3.2. Mean value over five experiments for the test error using the MLP classifier.

The experiments show that the lowest test error rate is obtained for *gwx* = *gwy* = 9 and *px* = 10, *py* =12, independently of the chosen classifier. A possible explanation is that, at these values we have the best image coverage. These values will be kept for the following experiments.

A2) *The influence of Gabor filter frequency and orientation over the classification accuracy*.

Table 2.3.3 and Table 2.3.4 illustrate the independent contributions of each frequency and orientation on the recognition performance. Also different combinations of features were tested.

x (frs = 1/x)	KNN [%]	MLP [%]
2	58.0	39.9
4	23.5	15.3
8	19.0	22.4
16	13.0	16.7
32	13.0	16.8
64	13.0	16.3
128	13.0	17.0
4 8 16	10.0	13.8
8 16 32	11.5	15.9
16 32 64	13.0	16.4
8 16 32 64	11.5	15.6
4 8 16 32 64	10.5	11.9
4 8 16 32 64 128	10.5	13.8
2 4 8 16 32 64 128	12.5	11.6

Tab. 2.3.3. Frequency importance.

It could be observed that the best results for the k-NN classifier are obtain for f^{-1} = 4, 8, 16. For the MLP case, the best results are obtained when all frequency features are employed. A similar effect on both classifiers could be remarked when changing frequency. For the following experiments, the f^{-1} = 4, 8, 16, 32, 64 set will be used, as a compromise between classification accuracy and the computational expense. By comparison with the previous case A1, the processing time is significantly reduced while the test error is 2.5% decreased for the k-NN case.
For filters orientation case, by selecting only the four most representative orientations for the total amount of eight orientation previously used, the processing time has been cut in half. The performance of k-NN was also improved slightly, by 0.5%. See Table 2.3.4 for details.

orientations	KNN [%]	MLP [%]
0	17.5	21.3
pi/8	11.5	16.6
2*pi/8	13.5	15.3
3*pi/8	13.5	13.2
4*pi/8	18.5	20.0
5*pi/8	10.5	14.3
6*pi/8	14.0	17.6
7*pi/8	11.5	15.6
pi/8 3*pi/8 5*pi/8 7*pi/8	10.0	11.6

Tab. 2.3.4. Orientation importance.

B) *Interest Operator filter*

In the case of Interest Operator, we examine the influence of the image block dimension and the resolution effect over the recognition rate.

B1) *The influence of the image blocks dimension over the classification accuracy*.

Unlike the standard approach, we take into account a non rectangular window for features extraction. As one can see from Table 2.3.5, the image blocks dimension affect somehow different the k-NN and the MLP classifiers.

Image block dimension [pixels]	7 x 6	10 x 8	12 x 10	14 x 12
k-NN – test error [%]	12.0	13.5	7.5	8.5
MLP – test error [%]	14.3	12.2	11.0	9.8

Tab. 2.3.5. The influence of the image blocks dimension over the classification accuracy.

B2) *The influence of the image resolution over the classification accuracy*.

The influence of the image resample factor was further studied. We examine the situation of a 75%, 50% and 25% resample factor and found the best results by reducing by half the original image size (Table 2.3.6). In comparison with the B1) case, where full resolution images were used (112x92 pixels), it could be observed here an 1.5% test error improvement for the k-NN classifier. The processing time has been also cut in half, so resampling shown to be useful for the case of Interest Operator filter.

Image block dimension [pixels]	7 x 6	10 x 8	12 x 10	14 x 12
k-NN – test error [%]	10.5	6.0	7.0	12.5
MLP – test error [%]	12.2	10.7	11.4	13.1

Tab. 2.3.6. The influence of the image blocks dimension over the classification accuracy for the facial images size reduce by half (56x46 pixels).

5.2 Selection of the classifiers parameters

A) *k-Nearest Neighbour*

The choice of the metric to measure the similarity between input vectors depends mainly on the specific features of the data. Table 2.3.7 presents the influence of the distance over the classification accuracy. Also choosing an appropriate value for k is a problem: each domain suiting different values of k. Most often this is achieved using a trial and error process. In our case k = 1 and the 'cityblock' distance provides best results.

Distance	Euclidean	Cityblock	Cosine	Correlation
Gabor	10.0	6.5	12.0	12.0
Interest Operator	6.0	6.0	8.0	8.0

Tab. 2.3.7. The influence of the k-NN similarity measure over the test error value.

B) *Multilayer Perceptron*

Neural Networks are known to be optimal discriminators, in that they are able to reach the Bayes limit in any classification problem. This statement is only true, unfortunately, under the condition that we have wisely chosen the NN's architecture, and managed to make it learn correctly.
The most influential MLP parameter in achieving a good classification performance is the number of hidden neurons. Unfortunately, there is no evidence that there is a certain procedure allowing us to estimate apriori the optimum number of the hidden neuron or hidden layers. Instead of using sophisticated and not always reliable methods, we determined appropriate numbers of hidden neurons by making several trials (Table 2.3.8).

Hidden neurons	40	60	80	120
Gabor	12.7	11.7	12.6	11.9
Interest Operator	12.5	11.5	10.6	9.8

Tab. 2.3.8. The influence of the hidden neuron number over the classification accuracy.

Also a two hidden layer MLP network was tested but it shows no significant classification accuracy improvement.

5.3 Interest Operator vs. Gabor kernels - comparative results

The final experiment using the Gabor kernels, respectively, the Interest Operator employ the useful found values for the number of lattice points, resolution, frequency, orientation - see section A1 and A2 for the Gabor case - and, image blocks dimension and resolution - see section B1 and B2 for the Interest Operator case.
Unlike previous cases, the training and testing data are generated *randomly*, for each of the five experiments. The results are presented in Table. 2.3.9.

Features extraction technique	Test error [%]	min. max. std.	Test error mean value [%]	Total processing time [min.]
k-NN				
Gabor	6.0 4.0 5.5 2.5 6.0	2.5 6.0 1.5	4.8	2.4
Interest.op.	6.0 3.0 7.0 5.0 5.5	3.0 7.0 1.5	5.3	0.2
MLP				
Gabor	4.5 5.5 5.5 9.5 4.5	4.5 9.5 2.1	5.9	2.3
Interest.op.	7.5 4.5 6.5 10 5.0	4.5 10.0 2.2	6.7	0.4

Tab. 2.3.9. The influence of the random training and testing data selection over the classification accuracy.

From the classification perspective, it could be noted the higher recognition rate for the case of k-NN classifier, whereas MLP came the second (4.8% compared to 5.9%). From the feature extraction technique point of view, the Interest Operator offers almost the same results (4.8% vs 5.3%) as the Gabor approach, while the magnitude of the processing time is reduced by an order.

Conclusions and future work

Previous work in the facial imagery classification suggests that Gabor-kernel-based methods can achieve among the highest recognition rate. We proposed a much faster alternative based on the Interest Operator. First, we focus on the individual, independent features extraction filters parameters contribution to the recognition performance. Then we extract useful frequencies and orientation, dimension and resolution and combine them, in order to obtain representative and smaller features vectors. Also, an experimental study on the influence of the k-NN and MLP classifiers key parameters has been carried out.
Although simple, the proposed approach had proven to be efficient: compared with Gabor approach, the Interest Operator offers almost the same results in about one order of magnitude lower processing time. In addition, it could be remarked that the k-NN classifier performs slightly better than MLP neural-network-based solution.
For future work, we plan a more sophisticated filter and classifier parameter selection methodology using fast sub-optimal algorithms or evolutionary optimization for a holistic approach.

2.3.2 Combined pattern search optimization of feature extraction and classification parameters in facial recognition

Constantly, the assumption is made that there is an independent contribution of the individual feature extraction and classifier parameters to the recognition performance. In our approach, the problems of fea- ture extraction and classifier design are viewed together as a single matter of estimating the optimal parameters from limited data. We propose, for the problem of facial recognition, a combination between an Interest Operator based feature extraction technique and a k-NN statistical classifier having the parameters determined using a pattern search based optimization technique. This approach enables us to achieve both higher classification accuracy and faster processing time.

Introduction

Facial recognition systems are computer-based security sys- tems that are able to automatically identify human faces. The prob- lem of recognizing faces has been largely investigated, due to its wide application. Current research directions include recognition from outdoor, non-frontal facial images, greater understanding of the effects of demographic factors on performance, improved mod- els for predicting identification performance on very large galleries, integration of morphable models into face recognition perfor- mance, etc. (see Phillips et al., 2006, for more details). The follow- ing initiatives are recent applications of the above mentioned research directions. The US-VISIT program requires visitors of the United States to provide fingerprints and a digital photograph at their port of entry. Then, interfaces with the Automated Biometric Identification System database check and see if the visitor is a ‘‘per- son of interest’’. Similarly, the Real ID Act of 2005 includes an inte- grated computer chip in every driving license issued after May 2008. The chip contains a digital photograph, which could be used for facial recognition purposes.

The wide variation in appearance, e.g. the pose, the illumina- tion, the person’s facial expression, the image size and quality, whether the person is using cosmetics, is wearing glasses, etc., makes the recognizing of people from images of their faces very challenging and the problem of facial recognition, an open topic. In dealing with the above mentioned difficulties, several face rec- ognition techniques have been employed. In the following, some of the most recent approaches will be briefly pointed out. For a comprehensive survey of the facial recognition techniques see Zhao et al. [Zhao03], Zhao and Chellappa [Zhao02], Gross et al. [Gross01]. These techniques might be classified into two categories: holistic and feature-based matching methods. In holistic approaches face recognition is obtained using a single feature vector that repre- sents the whole face image. First, some of the holistic based meth- ods will be discussed.

The main idea behind one of the most popular face recognition algorithm, the Principal Component Analysis (PCA), also known as Karhunen–Loeve expansion, is to find the vectors which best ac- count for the distribution of face images within the entire image space. The method was first introduced by the work of [Turk91]. Later on, Moon and Phillips [Moon01] introduced a generic modular PCA-based face recognition system. Independent component analysis (ICA), a generalization of PCA, minimizes both second-order and higher-order dependencies in the input data and attempts to find the basis along which the data are statistically independent. Bartlett et al. [Bartlett02] have proved that ICA representation outperforms the PCA/eigenface representation. Other examples of the holistic category are Linear Discriminant Analysis (LDA) and the Bayesian Intrapersonal Classifier. The first one finds the vectors in the underlying space that best discriminate among classes, see [Lu03]. The Bayesian Framework pro- poses a probabilistic similarity measure based on Bayesian belief that the image intensity differences are characteristic of typical variations in the appearance of an individual (see [Moghaddam01]).

Because of the restrictions imposed by global methods (frontal views of faces, with no occlusions and images acquired at a fixed scale), there has been a shift to feature-based approaches, as seen from the literature. In [Brunelli93] face recognition was performed by independently matching templates of three facial regions (both eyes, nose and mouth). The configuration of the components during classification was unconstrained since the system did not include a geometrical model of the face. A face detection and recognition system based on Support Vector Machines (SVMs) was presented by [Heisele01]. In all experiments the component-based systems outperformed the global systems whole face pattern, since the input features significantly simplify the task of face recognition. A final example of a feature- based approach is the Elastic Bunch Graph Matching (EBGM), introduced by Wang et al. [Wang05]. Here, faces are represented as graphs, with nodes positioned at fiducial points. Each node contains a set of complex Gabor wavelet coefficients at different scales and orientations.

Recently, new trends in face recognition have become evident. The use of 3D data of the face improves the accuracy and robust- ness with regard to facial pose and lighting variations. See Xu et al. [Xu09a], Lee et al. [Lee04], Bronstein et al. [Bronstein05], for representative work in this direction. Also the face recognition via sparse representation offers a new promising mathematical framework which leads to highly robust and scalable algorithms, as shown by Destrero et al. [Destrero09] and Wright et al. [Wright10]. From this perspective of parsimony for inference, our approach is highly sparse.

Taking into account all the above mentioned approaches, it might be concluded that the assumption of the independent contri- bution to the recognition performance of the individual feature extraction and classifier parameter optimization is always present. Obviously, this hypothesis is not true. The novelty of our approach consists in the design of the feature extraction and the classifier to- gether as a single problem of estimating the best parameters for these modules. This is made possible through the Pattern Search technique. We show that minimizing an objective function using the Pattern Search algorithm is an attractive alternative because it is often computationally less expensive than other optimization methods and provides both higher accuracy and faster processing time to the facial recognition system.

Pattern search

Pattern search methods are a particular class of direct search methods firstly analysed by Torczon [Torczon97] for unconstrained opti- mization. They were also extended to bound and linear constrained optimization by Lewis and Torczon [Lewis99]. The research about pat- tern search methods is still quite flourishing: several generaliza- tions and extensions have been proposed by Audet and Dennis ([Audet03], [Audet06]). Such methods operate by searching a set of points called a pattern, which expands or shrinks depending on whether any point within the pattern has a lower objective function value than the current point. The search stops after a minimum pattern size is reached.

To put the problem in a formal manner, we consider it of the form:

$$\min_{x \in X} f(x) \tag{2.3.17}$$

where $x \in \mathbf{X}$ is the vector of design parameters, $f: X \to R$ is the cost (objective) function, and $X \subset R^n$ is the constraint set, defined as:

$$\boldsymbol{X} \triangleq \{x \in \Re^n | l^i \leqslant x^i \leqslant u^i, i \in \{1, \ldots, n\}\} \tag{2.3.18}$$

with $-\infty \leq l^i < u^i \leq \infty$ for all $i \in \{1, \ldots, n\}$ In our case, the x vector contains the design parameters form both feature extraction and classification modules. The objective function is represented by the error of the facial recognition system regarding the test data set, usually referred as test error.

Two types of pattern search algorithms have been considered: the Generalized Pattern Search (GPS) algorithm and the Mesh Adaptive Search (MADS) algorithm implemented using Global Optimization Toolbox [MathWorks10] and NOMADm Optimization Software [Abramson10]. The algorithms differ in the way of computing the set of points forming the mesh. The GPS algorithm uses fixed direction vectors, whereas the MADS algorithm uses a random selection of vectors to define the mesh.

Let k ∈ N denote the iteration number, and let $x_k \in \mathbf{X}$ denote the current iterator. The pattern search algorithms have in common the fact that after a finite number of iterations, they search for a lower cost function value than $f(x_k)$ on the points in the set:

$$L \triangleq \{x \in \boldsymbol{X} | x = x_k \pm \Delta_k s^i d_i, i \in \{1, \ldots, n\}\} \tag{2.3.19}$$

where $\Delta_k > 0$ is a scalar called the mesh size factor, and $s \in R^n$ is a fixed parameter that can be used to take the different scaling of the design parameter components into account.

Once the direction vectors have been defined, the GPS and MADS algorithms form the mesh by multiplying the pattern vectors by a scalar, the mesh size Δ_k, defined by the length and direction $s^i d_i$. The pattern search algorithms have a rule that selects a finite number of points in **X** on a mesh defined by:

$$M(x_0, \Delta_k) \triangleq \{x_0 + m\Delta_k s^i d_i | i \in \{1, \ldots, n\}, m \in Z\} \tag{2.3.20}$$

where $x_0 \in \mathbf{X}$ is the initial value of the iterator. If a mesh point $x' \in M(x_0, \Delta)$ with $f(x') < f(x_k)$ has been found, then the search continues with $x_{k+1} = x'$ and $\Delta_k = \Delta_{k+1}$ only if the complete poll option is disabled. In this case the algorithm stops polling the mesh points as soon as it finds a point whose objective function value is less than that of the current point.

If a complete poll is considered, the algorithm computes the objective function values at all mesh points. All points in L_k are tested for a decrease in f(•). If $f(x') \geq f(x_k)$ for all $x_k \in L_k$, then the search continues with $x_{k+1} = x_k$ and a reduced mesh size factor. Hence, the search continues on a finer mesh. The search stops if the mesh M (x0, •) has been refined a user-specified number of times.

Feature extraction and classification modules

In the feature extraction problem, the task is to find an efficient way to represent the pixel data, usually trough a mapping function, from the original image space to a lower dimensional feature vec- tor space.

The Interest Operator feature extraction technique was first introduced by Moravec [Moravec81] for the purpose of mobile robot navigation. Later, it was used by Nasrabadi and Choo [Nasrabadi98] for a Hopfield neural network based stereoscopic vision system and by Wang et al. [Wang98] in automatic target recognition. Căleanu ([Căleanu00], [Căleanu01]) employed it for the first time in the facial recognition problem. Zhao et al. [Zhao04] have reported good results using the Interest Operator for the facial imagery classification. Căleanu et al. [Căleanu07] have shown that the proposed feature extraction approach proved to be efficient and accurate compared with one of the best face recognition method, namely the Elastic Bunch Graph Matching/Gabor kernels approach. It offers almost the same classification error rates in about one order of magnitude lower processing time. Recently, Zhao et al. [Zhao08] present an improved Interest Operator in the form of a multi-orientation and multiscale interest filter. Xu et al. [Xu09b] propose two new versions of Interest Oper- ators for face recognition which are used to calculate the pixel intensity variation information of overlapping blocks produced from the original face image.

The Interest Operator computes, for a given image block, some directional variance features (eqs. (2.3.5) – (2.3.10)).

Fig. 2.3.4 presents the effect of the Interest Operator (IO) when applied to a human face. Table 2.3.10 provides a comparison between the performances of IO and other widely used feature extraction techniques: Fisher Linear Discriminant (FLD), Principal Component Analysis (PCA) and Gabor filter (GAB). The effect of combining the IO with 2D-FLD (two-dimensional Fisher discriminant analysis) or 2D-PCA (two two-dimensional PCA) can be examined in the column 2 of Table 2.3.10.

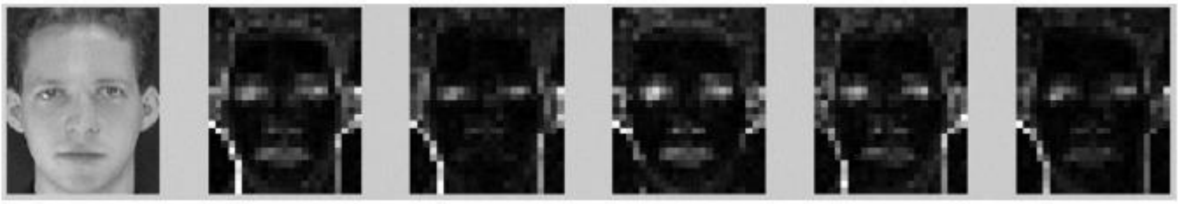


Fig. 2.3.4. Original and variance images.

Database	AT&T	AR	AT&T
Classification scheme	Committee of RBF NN	Maximum similarity	k-NN
Test error [%]	5.1 (FLD) 8.9 (PCA) – 13.0 (IO)	18.3 (2D-FLD) 21.2 (2D-PCA) 6.5 (IO+2D-FLD) 8.1 (IO+2D-PCA)	4.8 (GAB) – – 5.3 (IO)
Reference	Zhao et al. (2004)	Xu et al. (2009b)	Căleanu et al. (2007)

Tab.2.3.10. IO – alone or in combination – versus other feature extraction techniques.

In the automated face recognition problem, we are given a database of image samples of known individuals. The task is to design a system that for any input image, identifies the input with one of the known individuals. Thus, the classification problem involves designing a function to map feature vectors to the appropriate class label. In the present work a statistical classifier was employed, namely the K-Nearest Neighbour (k-NN) classifier. It is a very simple, yet powerful statistical classification method. Using a suitable distance, e.g. Euclidean (p = 1) or city-block (p = 2) distance:

$$d(\mathbf{x},\mathbf{y}) = \left(\sum_{i=1}^{N} |x_i - y_i|^p\right)^{\frac{1}{p}} \tag{2.3.21}$$

one has to find k closest training points to the vector that should be labelled. Then, by specifying a rule, e.g. majority rule or consensus rule, it will be decided how to classify the sample. It can be shown that the performance of a k-NN classifier is always at least half of the best possible Bayesian classifier for a given problem. More on this topic could be found in [Mitchell97].

Experimental results

The datasets

In our experiments we utilize the following two datasets: AT&T laboratories database of faces (see [AT&T02] for details) and the UMIST face database (see [Graham98], for details). In the AT&T database there are 10 different images per subject for 40 distinct subjects. For some of the subjects, the images were taken at different time periods, by varying lighting slightly, changing facial expressions and facial details. All the images were taken against a dark homogeneous background (Fig. 2.3.5). In all cases, five training images per person (thus 200 total training images) were randomly taken for training and the remaining images (200 total images) are taken for testing.

The second one, the UMIST database consists of 575 images of 20 people with varied poses from profile to frontal views. There are a maximum of 48 images per subject and a minimum of 19 (see Fig.2.3.6). Subjects cover a range of race/sex/appearance. Ten images per person are randomly chosen to produce a training set of 200 images. The remaining 375 images are used to form the test set.

The images from these databases are cropped to the size of 112 x 92, a standardized image size commonly used in face recognition tasks. The test and training set did not have any images in common. When the effect of a certain filter/classifier parameter was analysed, in order to evaluate its usefulness in capturing relevant information, successive evaluations used the same randomly chosen test and train images. Otherwise, the results would have been also influenced by the data selection process.

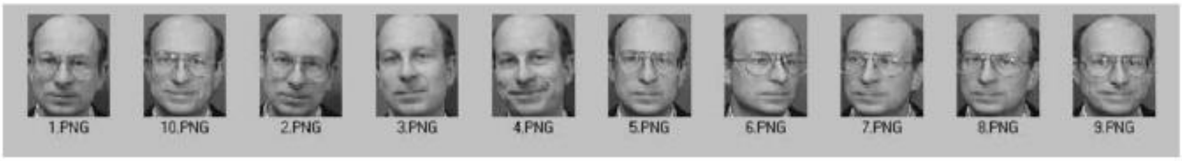


Fig. 2.3.5. Images of a subject from the AT&T database. The database has a number of 10 images for each of the 40 subjects.

Fig.2.3.6. Images of a subject from the UMIST database. The database has variable number of images for each of the 20 subjects.

The heuristic approach

Through this approach, we experimentally examine the influ- ence of the individual, independent feature extraction filters and classifier parameters to the recognition performance.

In the case of feature extraction module, we experimentally search the optimum values of the image block dimensions. As it was previously pointed out, a facial image is divided into image blocks from where the response of Interest Operator is computed. We take into account a rectangular window for feature extraction (Table 2.3.11). Also the influence of the image resampling factor is further studied. We examine the situation of a 75%, 50% and 25% resample factor and found the best results by reducing by 50% the original image size.

For the case of classification module, the choice of the metric to measure the similarity between input vectors depends mainly on the specific features of the data. Also choosing an appropriate value for k is a problem: each domain suiting different values of k. Most often the optimal values for the distance type d and the number of neighbours are achieved using a heuristic/trial and error process (Table 2.3.12, respectively Table 2.3.13). In our case k = 1 and the 'cityblock' distance provided best results.

Image block dimension, $w_y \times w_x$ pixels	7×6	10×8	12×10	14×12
Test error [%], AT&T database	10.5	6.0	7.0	12.5
Test error [%], UMIST database	4.5	3.5	2.9	5.9

Table 2.3.11. The influence of the image blocks dimension over the classification accuracy.

Type of distance, d	Euclidean	Cityblock	Cosine	Correlation
Test error [%], AT&T	6.5	6.0	8.0	8.0
Test error [%], UMIST	5.6	2.9	6.4	6.1

Table 2.3.12. The influence of the k-NN similarity measure over the test error value.

Number of neighbours, k	1	3	6	10
Test error [%], AT&T database case	6.0	10.0	11.5	15.0
Test error [%], UMIST database case	2.9	4.8	14.1	24.0

Table 2.3.13. The influence of the k-NN number of neighbours over the test error value.

The pattern search approach

The second part of the experiments aims to apply the Pattern Search technique in order to find the optimal parameter vector, x = [w_Y w_x d k] which minimizes an objective function:

$$f(x) = a_1 \cdot e_t + a_2 \cdot t_p \tag{2.3.22}$$

where e_t and t_p denote the classification error over the test data and the total processing time. Almost the same test error has been noted for different combinations of the feature extraction window size; adding t_p in small amount (a2 « 1) into the cost function favours finding larger values for w_y and w_x. As a result, the test error is kept at the same value while the processing time is decreased. The two coefficients from the objective function were chosen as the weight of e_t accounts for about 90% from the total value of f(x), respectively 10% for the weight of t_p. Practically, taking into account the average numerical value for the total processing time of the evaluation and the error magnitude, this means a_1 = 0.99 and a_2 = 0.01.

The Pattern Search algorithm was set to make a complete poll of the mesh points using the mesh adaptive search method for 30 iterations. See Table 2.3.14 for the additional setting of the optimization algorithm and Fig. 2.3.7 for the evolution of the objective function.

Starting point	7 7 5 3	InitialMeshSize	5
Lower Bounds	5 5 1 1	InitialPenalty	30
Upper Bounds	20 20 10 4	PenaltyFactor	300
TolMesh	0.001	CompletePoll	on
MaxIter	30	PollingOrder	success
MeshContraction	0.75	SearchMethod	MADSPositiveBasis2N
MeshExpansion	4	CompleteSearch	on

Table 2.3.14. Options structure for the optimization procedure.

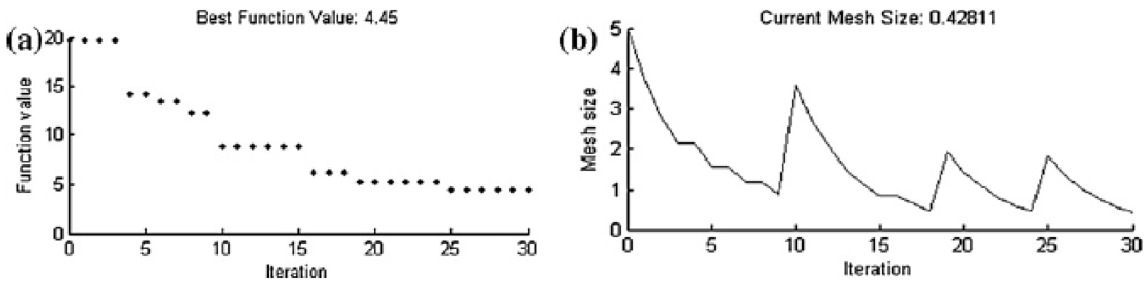


Fig. 2.3.7. (a) The objective function is constantly decreasing over iteration steps and (b) the mesh size was increased after a successful polling, otherwise it was refined.

Heuristics vs. pattern search – comparative results

The final experiment compares the system accuracy using the best parameters, found by the trial and error respectively Pattern Search algorithm, for the feature extraction and classifier modules. The training and testing data sets are distinct and randomly selected in five consecutive experiments (four runs in [Li99] and three runs in [Lawrence97]). The results are presented in Table 2.3.15.

Optimization technique	Test errors [%]	min/max/std.dev.	e_t [%]	t_p [s]
AT&T				
Heuristic	6.0/3.0/7.0/5.0/5.5	3.0/7.0/1.5	5.3	55
Pattern Search	4.0/1.5/2.5/2.0/4.5	1.5/4.5/1.3	2.9	37
UMIST				
Heuristic	2.1/2.4/1.3/4.0/4.0	1.3/4.0/1.2	2.8	66
Pattern Search	3.5/1.6/1.3/2.1/0.8	0.8/3.5/1.0	1.9	40

Table 2.3.15. The influence of the optimization technique and the random training and testing data (results for 5 tests in ''Test errors [%]'' column) selection over the classification accuracy.

Conclusion

The identification of individuals by recognition of faces has been considered. A system is described, based on a new feature extrac- tion approach – the Interest Operator, for the recognition of human faces. Unlike the independent view of feature extraction and clas- sification, this method treats the problems of feature extraction and classifier design as the single problem of estimating the best parameters from limited data. It uses an advance optimization technique, the Pattern Search, in order to design the feature extraction and classification stages.

It has been proved that using a Pattern Search optimization both the classification error and the processing time have been substantially reduced. When the process simulation is very com- plex and it is not designed in a vectorised manner, Pattern Search represents an attractive alternative to other optimization methods, e.g. genetic algorithm, as it is often computationally less expensive and can minimize the same types of functions. Experimental re- sults of facial recognition have proven that the pattern search ap- proach outperforms the heuristic approach, both in accuracy and processing time. Although the principal concern of the present pa- per is the concept proofing of Pattern Search based global optimi zation for the feature extraction – classification ensemble, the experiments showed excellent results when the proposed system, PS (IO+KNN), was applied to the particular problem of facial recognition. The test error obtained in standard experimental conditions for the AT&T and UMIST public face databases was 2.9% respec- tively 1.9%, among the lowest reported in literature (Table 2.3.16).

In order to obtain higher recognition rates we will examine, as one of the future research direction, a more complex structure, possibly parallel and hierarchic, for the classification stage.

Method	Test error [%]	Results obtained by
AT&T		
Pseudo-2D-HMM	5.00	Lawrence et al. (1997)
KLT+CN	5.30	Lawrence et al. (1997)
SOM+CN	3.80	Lawrence et al. (1997)
KPCA	2.50	Kim et al. (2002a)
SVM+NN	2.10	Kim et al. (2002b)
UMIST		
PCA+LDA+NN	2.62	Zhang et al. (2004)
MLA+NM	3.73	Zhang et al. (2004)
2DCT+AML	1.53	Satonaka and Uchimura (2006)
PCA+AML	2.27	Satonaka and Uchimura (2006)
2DCT+EM+AML	2.86	Satonaka and Uchimura (2006)
DF-LDA	2.00	Lu et al. (2003)

Table 2.3.16 Results using AT&T and UMIST face databases.

2.4 Soft computing based face expression recognition

2.4.1 Face Expression Recognition: a Brief Overview of the Last Decade

The huge research effort in the field of face expression recognition (FER) technology is justified by the potential applications in multiple domains: computer science, engineering, psychology, neuroscience, to name just a few. Obviously, this generates an impressive number of scientific publications. The aim of this work is to identify key representative approaches for facial expression recognition research in the past ten years (2003-2012).

INTRODUCTION

The scientific study of facial expression showed that it represents one of the most powerful and immediate means for emotions and intentions communication [Tian03]. Due to its potential applications e.g., natural human-machine interface, behavioral science, clinic practice, automatic facial expression recognition has attracted much attention which has been manifested through a large number of scientific publications. As a consequence, numerous methods were proposed. For comprehensive surveys of the past efforts in the field, readers are referred to [Pantic00], [Fasel03] or [Tian05]. Despite these efforts, recognizing facial expression with a high accuracy remains to be difficult and still represents an active research topic [Matuszewski11].

The present article is intended to be a high-level survey over the facial expression recognition research that has been carried out in the past ten years, between year 2003 and 2012. Due to length restrictions, only a small sample of recognition techniques is explicitly referred to. More exactly, we have selected 10 papers that we felt were important and different from each other. Also the novelty of the proposed technique, the number of the citations and the number of downloads were taken into account in the selection process.
The interest in creating such an overview is multifarious. By selecting the most interesting approaches, we want to focus the attention to new techniques and methodologies that may be of high interest to the researchers in the field of facial imagery. Moreover, this selection can be a useful indicator of the areas that will constitute the future research trends.

TAXONOMY AND DESCRIPTION OF APPROACHES

In the last decade, the field of expression analysis has grown immensely. The proposed approaches can be classified from many perspectives, which will be analyzed below.

Taxonomy

From the feature extraction point of view, the techniques aiming the recognition of face expression can be categorized into methods that use appearance features and methods that use geometric features, although some hybrid-based approaches are also possible. The use of a face model/template constitutes a distinct approach, and it is typically referred as a template-based method.

In the first category, the face images are processed by an image filter or filter banks on the whole face or some specific regions of the face to extract changes in facial appearance.

Typically, the entire facial images or some specific facial regions are convolved with some filters, e.g. Gabor wavelets, and the extracted filter responses at the fiducial points (manually selected, most of the times) form vectors that are further used for facial expression classification. Also, principal component analysis (PCA), variants of independent component analysis (ICA) or local binary pattern (LBP) are most often used in a holistic manner.

In the geometric feature extraction system, the shape, distances, angles or the coordinates of the fiducial points form a feature vector that represents facial geometry.
It seems that the highest recognition rates have been obtained when both the responses methods were combined [Tian02].

From the temporal perspective, facial expression recognition techniques are represented by static (typically using still images) and dynamic (image sequences) approaches.

Other categories are the global techniques - which analyze the texture of the whole face without having explicit knowledge about the location of single facial features - versus local approaches - which try to extract local features of the face or to fit any holistic face model containing a set of feature points to the face.

Finally, we could classify the methods according to the image data type in 2D and 3D approaches.

Description of Approaches

The facial action coding system developed by Ekman [Ekman78] is a source of inspiration for many research papers. In the first paper that has been analyzed, Cohen et al. [Cohonen03] present a method for recognizing the emotions through facial expressions presented in a video sequence is discussed. The novelty of this work consists in introducing the Tree-Augmented-Naive Bayes (TAN) classifier which incorporates the dependencies between features. This is in opposition to the Naïve Bayes (NB) assumption in which features are considered to be independent as in work of Sebe et al. [Sebe02].

The method presented by Ma and Khorasani [Ma04] is based on a combination of a two-dimensional discrete cosine transform (2D-DCT) and constructive one-hidden-layer feedforward neural network. The novelty of the approach comes from the proposed pruning technique which substantially reduces the size of the neural network while improving the generalization capability and the recognition rate.

A novel low-computation discriminative feature space is introduced in [Shan05] by Shan et al. It is based on Local Binary Patterns (LBP) features which could be extracted faster. Moreover, it is shown that LBP features are robust to low resolution. As possible solutions for the classification stage, template matching and Support Vector Machine (SVM) were considered and compared. Also, a comparison with the previous mentioned work of Cohen et al. [Cohen03] (geometric features +TAN) proved the superiority of the proposed technique.

Since the year 2006, little work has been done in 3D based face expression recognition. One of the first attempts is represented by the work of Zeng et al. [Zeng06] who employed a 3D face tracker for feature extraction. In the same year, in the approach proposed by Wang et al. [Wang06], the classification of the prototypic facial expressions is performed by extracting primitive 3D facial expression features, and by calculating the feature distribution. Other innovative 3D solutions are the two approaches of Dornaika et al. [Dornaika07] and Kotsia & Pitas [Kotsia07] which use AUs together with a 3D face model named Candide, initially proposed by Ahlberg [Ahlberg01]. Some of the most important advantages of these techniques are the texture independence, the view independent (since the used tracker simultaneously provides the 3D head pose and the facial actions) and a simple learning phase which only need to fit second-order auto-regressive models to sequences of facial actions. For an excellent 3D facial expression survey please consult the work of Fang et al. [Fang11].

One of the most interesting approaches is presented by Panning in [Panning08]. The authors propose a novel approach from multiple perspectives e.g., it use facial feature detection in color image sequences and the feature extraction is initialized without manual input. For expression classification a three layer feed forward artificial neural network is employed. In the next work, Buciu et al. [Buciu09] present a comparison of ICA approaches whereas the classification stage is implemented using either a Cosine Similarity Measure (CSM) or a SVM classifier.

In the paper proposed by Jabid and Chae [Jabid10], a new appearance based feature descriptor, called the local directional pattern (LDP), has been proposed to represent facial geometry. Template matching and support vector machine are used for classification. A novel facial expression recognition technique based on a sparse representation is proposed by Jia in [Jia11]. It is also shown that a multi-layer sparse representation provides better experimental results than the conventional sparse representation.

The last selected paper, proposed by Valstar and Pantic [Valstar12], enables the detection of much larger range of facial expressions.

IMAGE ACQUISITION AND PRE-PROCESSING

Facial data can be acquired from a database, a live video stream or other sources, in 2D or 3D, both in a static or dynamic mode. The most popular type of pictures is the 2D grey scale facial images. Typically, this step is followed by some pre-processing (noise removal, light compensation, detection, normalization, tracking, etc.) operations. In the context of the selected papers, these two steps are detailed in the following.

In the work of [Cohen03] the images are acquired from a video stream using a face tracking algorithm and a 3D wireframe model (16 surface patches embedded in Bézier volumes) proposed by Tao and Huang [Tao98].

A normalization process was employed by Ma et al. [Ma04] in which the centers of the eyes and mouth are taken as the reference points. A fixed distance d between the centers of the eyes represents a first normalization criterion in [Shan05]. An interesting observation is further formulated and applied as a second pre-processing step. It refers to the face dimensions: the width of selected face is roughly 2d and the height is roughly 3d. No illumination compensation is required, due to the LBP's gray scale invariance.

In [Wang06], the surface feature analysis is based on the triangle meshes of faces, which are created by a 3dMD static digitizer [3dMD], which uses the principle of light pattern projection. In another 3D approach [Dornaika07], the local facial actions and deformations are calculated using an appearance-based tracker which simultaneously computes the 3D head pose and the facial actions.

The work of [Pannings08] is innovative from the perspective of pre-processing step due to the particular modality in implementing the face detection stage. More exactly, in order to minimize the false positives detections in complex backgrounds, a combination of Haar-like-Feature detection and skin color detection has been proposed.

The following pre-processing steps were proposed in [Buciu09]: a registration operation based on the manually identified position of eyes followed by a rotation of the image to horizontally align the face according to eyes; the final 60 x 45 pixels facial image was obtained by cropping and downsampling operations.

The images used in [Jabid10] were cropped from the original C-K database one using the positions of two eyes and resized into 150×110 pixels. The height of the image is 2.7d with level of eye located 2d apart from bottom boundary, where d represents the distance between the eyes. No attempt was made for illumination compensation, since LDP is robust in illumination change.

The fixed distance of 60 pixels between the eyes represents a normalization criterion for the work presented in [Jia11]. The final cropped face had the width of two times this distance and the height roughly three times.

The first step of the system proposed in [Valstar12] is represented by the well-known Viola-Jones face detector [Viola04]. For the next step, characteristic facial point detection, the solution proposed by Vukadinovic and Pantic [Vukadinovici05] was chosen. Registration and smoothing operations represents other steps in the processing scheme proposed by Valstar in [Valstar12].

FEATURE EXTRACTION

One of the most critical aspects for any successful facial expression recognition system is extracting the best features to describe the physical phenomena. The efficiency (minimizing within-class variations of expressions while maximizing between-class variations, low-dimensional feature space, etc.) and effectiveness (can be easily extracted from the raw face image) representation of the facial images would provide robustness during recognition process.

From this perspective, the work of Cohen et al. [Cohen03] uses the features proposed by Sebe et al. [Sebe02] which consist of 12 facial motion measurements relative to a 3D wireframe model (16 surface patches embedded in Bézier volumes) proposed by Tao and Huang [Tao98].

The 2-D discrete cosine transform (DCT) compression technique was employed in [Ma04] to the difference image obtained by subtracting a neutral image from a given expression image. In the frequency domain, only the coefficients of the lower frequencies (having large amplitudes) were considered as input vectors for the following classifier.

An extended LBP (neighborhood of different sizes and uniform patterns) is proposed in [Shan05] as feature extraction method. In this way a face is divided into small regions from which LBP histograms are extracted and concatenated.

Research on face expression recognition has mainly relied on 2D images and, obviously, has certain limitations. See Liu and Ward [Liu05] for arguments in the favor of 3D facial imagery processing. The next two approaches employed 3D facial data. Wang et al. [Wang06] use a geometric feature based facial expression descriptor in the 3D Euclidian space. Dornaika et al. [Dornaika07] inferred the facial expression from the temporal representation of a 3D model which includes head pose (three rotations and three translations) and a facial action vector. Obviously, the system performance will be dictated by the ability to accurately track the local facial actions/deformations.

Panning [Panning08] consider, in an innovative manner, both static (10 distances) and transient (3 regions for texture analysis) feature types. The set of distances are measured between static feature points and the transient features are detected by the appearance of edges in predefined areas (forehead, the nose bridge, and the nasolabial fold).

The work of Buciu [Buciu09] take ICA approach as baseline for feature extraction and compare it with five additional ICA flavors.

Jabid et al. [Jabid10] proposes the LBP operator which encodes the micro-level information of edges, spots, and other local features in an image. It. computes the edge response values in different directions and uses these to encode the image texture. After computing all the LDP code for each pixel, the input image is represented by an LDP histogram which represents a descriptor of that image.

Jia et al. [Jia11] divide the image into several 20x20 pixel patches, then an enhanced LBP operator is used as feature descriptor. Because some face regions provide more important information than others, different face regions should be given different weights.
The features computed in [Valstar12] from the 20 tracked fiducial points are: the positions of these points, the distances between pairs of points and the angle that the lines make with the horizontal axis. Finally, some temporal information is added.

CLASSIFICATION

Expression categorization is a process of assigning observed data to one of predefined facial expression categories by a classifier. A wide range of classifiers, covering parametric as well as non-parametric techniques, has been applied to the automatic expression recognition problem: nearest neighbor classifiers, Fisher's linear discriminant, ANNs, HMMs, SVMs or random forests are typical examples .

The classification scheme of [Cohen03], called Tree-Augmented-Naive Bayes (TAN) classifier, is based on an acyclic graphical model with the class and features as the nodes, and the dependencies represented by the directed edges in the graph between the nodes, forming in this way a Bayesian network. The simple template matching with weighted Chi square statistic and SVM are adopted as classification solutions in [Shan05] and further compared. The generalization performance for SVM using RBF kernel provided a substantial increase in the recognition rate when compared with template matching or the previous presented classification scheme, TAN [Cohen03].

In [Ma04], for implementing the classification stage, the application of an adaptive constructive one-hidden layer feedforward neural network to face expression recognition was considered. In this approach new hidden units/layers are added incrementally. Another ANN based approach is presents in [Panning08] where, for classification purpose, a full connected feed-forward artificial neural network is trained and used. It is implemented using the open source FANN library [Nissen].

The work presented in [Wang06] used in the experiments four popular classifiers: Quadratic Discriminant Classifier (QDC), Linear Discriminant Analysis (LDA), Naive Bayesian Classifier (NBC), and Support Vector Classifier (SVC) with RBF kernel.

In order to evaluate the affect state, the authors of [Dornaika07] calculate the cosine of the angle between two vectors which represents the actual expression and the

synthesized universal expression trajectories. Based on this parameter, a normalized similarity measure is determined and used in the recognition process. In [Buciu09], two different classifiers are employed. The first is represented by the CSM classifier which is based on the nearest neighbor rule; the second classifier is the Support Vector Machine. SVM makes binary decisions, thus, in the work of Jabid [Jabid10], the one-against rest technique is adopted.
The key concept of the sparse representation expression recognition presented in [Jia11] is the evaluation of the classes from the training samples for the minimum reconstruction error based on the sparse coding coefficients. Furthermore, in order to deal with subtle facial expressions, a multi-layer sparse representation method is proposed to improve recognition performance of multi- intensity expressions.

The approach presented in [Valstar12] perform a SVM based AU recognition. The choice is motivated by the high dimensionality of the feature space which, in this case, will not affect the training time. For the problem of AU temporal model detection, two approaches were compared: multiclass SVM (mc-SVM) and Hybrid SVM-HMM.

EXPERIMENTAL RESULTS

The development of robust facial expression recognition algorithms requires well labeled databases of sufficient size that include carefully controlled variations of pose, illumination and resolution. These databases can be categorized into 2-D image, e.g., JAFFE database [Lyons99] or MMI database [Pantic05], 2-D video (arguable the most used database of this type is Cohn-Kanade AU-Coded database [Kanade00]), 3-D static (BU-3DFE [Yin06]) and 3-D dynamic (BU-4DFE [Yin08]).

Some researchers have employed custom/in-house database. For example, the experiments described in [Cohen03] are based on data collected from video sources in which a set of five people are displaying six emotions. Two types of evaluations were performed: person dependent and independent experiments. The average expression recognition accuracies reported were 83.31% and 65.11% respectively. Another custom database was created in [Ma04]. It contains images of size 128 x 128 with 256 gray levels expressing just four affects: smile, anger, sadness, and surprise are the four specific facial expressions of interest. The reported testing recognition rate was 93.75%.

In [Panning08] for training purpose the authors interactively placed landmarks on images of 15 different people of the “Feedtum Facial Expression Database” from the Technical University of Munich [Wallhoff06]. The reliability of the feature detection and tracking system was tested on image sequences of the Smart-Kom database [Smartkom03]. On recall of the training data as well as other additional 30 samples (which not had been included for learning) it achieved good generalization with almost 100% correct classification.

The next paragraph refers mainly to the result reported against Cohn-Kanade (C-K) Facial Expression Database. The comparisons provided in the work of [Shan05] showed that the LBP-based SVM (87.6%) outperform both Gabor-based SVM (86.9%) and Geometric Feature + TAN (73.2%) on C-K database. The principal advantage of this approach lies in the simplicity of LBP histogram which requires much less computational resource. The recognition rates reported by [Dornaika07] were obtained using two different training sets, the C-K database and a custom data set consists of five 30 second videos. From 101 played expressions there were 14 misclassified expressions leading to a recognition rate of 86.14%.

The experiments reported in [Buciu09] have been performed using two databases of facial expression images: C-K AU-coded facial expression database in which thirteen persons expressed six basic emotions and Japanese female facial expressions (JAFFE). The best results obtained with leave-one-out are 87.6% for C-K database using fastICA + SVM polynomial and 81% for JAFFE using extended infomax and SVM-RBF. Jabid et al. [Jabid10] carried out a 7-fold cross-validation scheme where each dataset is randomly partitioned into seven groups separately. Six groups were used as a training dataset to train the classifiers or model their templates, while the remaining groups were used as testing datasets. The results of [Jia11] were reported against C-K+ (Extended Cohn-Kanade Dataset) using a 10-fold cross validation testing scheme. The best reported result is the recognition rate of 87% using the expression recognition based on sparse representation. In the work of [Valstar12], four databases were used (the C-K database, the MMI facial expression database, the DS118 data set of spontaneous facial displays, and the triad data set of spontaneous human behavior) for various sets of experiments When evaluated for the C-K database (six-basic-emotion detection system on 153 videos), a 91.7% recognition rate was reported and a 95.3% accuracy for MMI database (244 videos). These results are among the best ever reported and show that appearance-based approaches do not necessarily outperform geometric-feature-based approaches.

To date, there are few publicly available 3D databases designed specifically for expression analysis. This is probably the reason for having fewer reported results. From our survey, the test of [Wang06] is on the six prototypic expressions using the data from BU-3DFE database captured from 60 subjects with two high intensity models for each expression. The highest correct recognition rate of 83.6% was obtained using LDA classifier.

CONCLUSION

This paper's objective is to survey and discuss recent advances in face expression recognition. More exactly, face expression recognition systems have improved a lot over the past decade. It is almost impossible to cover all of the published work so we have selected 10 papers, published since 2003 till date, that we think were relevant for this topic. The main emphasis has been on a review of the recent developments in: data acquisition and pre-processing, feature extraction and selection, and subsequent classification. Experimental results reported against different facial databases were also presented.

This analysis enables us to:

- Identify the weak points and define the desirable characteristics of such systems. The lack of a basis for benchmarks (comprehensive, accessible reference set of expression displays) and associated evaluation procedures for all different efforts in the research on machine analysis of human face expression recognition represents arguable the most important current issue. In our opinion, the next big problem is that many of the systems still require manual intervention. The challenge is to make the system fully automatic. Some unrealistic assumptions are still present in many papers e.g., frontal view image of faces without hair, glasses and all facial expression displayed using six basic emotion categories.
- Predict the future trends in the domain of facial imagery. In our opinion, the most important research direction concerns the development of the 3D dynamic datasets using new 3D sensors such as structure light cameras or time-of-flight cameras. This direction is

quite promising for real-time segmentation. Also the inclusion of the temporal information will improve accuracy and robustness. In the future, the focus will shift from considering posed expression recognition to the development of methods for spontaneous expression recognition along with the deployment in a system with real-time capability. More work has to be done with regard to the integration of other communication channels such as voice and gesture.

We conclude by saying that the technology of facial expression recognition has enormous market potential and, in the near future, will enhance most human-computer interfaces.

2.4.2 Layered Fuzzy Facial Expression Generation of Virtual Agent

To realize intelligent and comprehensive facial expression generation of virtual agent, a novel model of layered fuzzy facial expression generation is proposed. In this model, social, emotional and physiological layers contribute to the fuzzy facial expression generation. Then, a layered fuzzy facial expression generation system is founded, where influences of the three layers and expression personality are considered for intelligent decision of facial expression generation based on fuzzy theory. A novel layered fuzzy facial expression generation language is also developed for conveniently controlling facial expression generation of virtual agent. The system is evaluated by subjects and compared with other systems, showing that it is efficient for intelligent facial expression generation of virtual agent.

Introduction

Virtual agent has been developed as a humanoid representation of user or system-related assistant to achieve effective human computer interaction as it is capable of communicational modalities that can be easily recognized by the users. Facial expression is of major importance for virtual agent to convey emotions, cognitive states, etc. Intelligent and lifelike facial expression generation of virtual agent can improve the efficiency of human computer interaction and make the user feel at ease when using a system.
Automatic facial expression generation of virtual agent has been actively researched along with the rapid progresses in computer science, human computer interaction, affective computing, etc. In the past, most researchers have been developing methods to generate authentic facial expressions of emotions. Recently, some researchers have focused on intelligent facial expression generation of virtual agent. Ochs et al. introduced emotional intelligence into an animated character to express felt emotions and expressed emotions [Ochs05]. Khanam et al. proposed a performance-driven facial animation system [Khanam07], which intelligently blends the detected expression and the expression implied by the context to generate enhanced facial expression. Arellano et al. presented an affective model that determines the emotional state of a character according the personality traits and the experienced emotions [Arellano08]. These researches have introduced social context, display rules and personality into the intelligence in facial expression generation of virtual agent.

However there are two limits in the above works of facial expression generation of virtual agent. The first limit is that most works of facial expression generation are merely related to emotions. Generally speaking, they are in the "Emotion View". However, facial expressions may convey not only emotions, but also cognitive or physiological states. For example, one may wink just because he is too tired or to give a hint to someone. Virtual agent's facial expressions should be more comprehensive to accommodate the complex environment in human computer interaction.

The second limit is that facial expression generation is mostly monotone, or in the "Invariable View". They usually correlate one model of facial expression animation to one emotion. However, human tends to act more complicatedly to express one emotion. For example, human displays kinds of facial expressions to express happiness, such as smile with mouth open or closed, symmetrically or asymmetrically, even with head wobbled.

This work aims at intelligent and expressive facial expression generation of virtual agent to achieve harmonious and affective human computer interface. Based on the cues of sources and characteristics of facial expressions, we propose a novel framework of layered fuzzy facial expression generation, in which the social, emotional and physiological layers contribute to the facial expression generation and fuzzy theory helps to generate mutative and rich facial expressions.

Sociality Aspect for Facial Expression Generation

Facial displays in social settings have been actively studied since the late 1970s. Facial displays are sensitive to the sociality of the situation. For example, smiles may occur more frequently when individuals are in social contact with others than when they are not facing or interacting with others. Buck argued that social factors can facilitate or inhibit facial expression depending upon the nature of emotion being expressed and the expresser's personal relationship with the other [Buck91]. Fridlund contended that facial expressions are inherently social [Fridlund94]. For example, even when someone is alone he is holding an internal dialogue with another person, or imaging himself in a social situation. Furthermore, display rules may lead individual to attenuate, amplify, inhibit or cover the involuntary expression with the sign of another emotion [Ekman69].

Emotion Aspect for Facial Expression Generation

Human expresses emotions through facial expressions in communication. In our daily life, we smile when we are happy, and cry when sad, so we call the facial expression that reflecting emotion the "emotional facial expression" or "facial expression of emotion". Ekman and Friesen have studied the basic emotional facial expressions of happiness, sadness, anger, disgust, fear, and surprise in the cross culture studies [Ekman71]. There are many factors that influence mapping emotional state to facial expression, such as the type and intensity of emotion, and how emotional state elicited [Picard97].

Physiology Aspect for Facial Expression Generation

Kinds of facial expressions may occur when physiological state changes. For example, headache state may lead to facial expressions such as furrowed eyebrows, closed eyes, slow eye blinks, lip pursuing, and facial grimacing, and so on. Physiological states such as hunger or pain can also influence the activation of emotional states [Picard97]. For example, hunger can increase irritability, and pain can spur anger. Also, changes in brain blood temperature along with other physiological changes may lead to the feeling of excitation or depressed [Lisetti00].

Model of Layered Fuzzy Facial Expression Generation

Based on the fuzziness of facial expressions, sociality, emotion, and physiology aspects for facial expression generation, the model of layered fuzzy facial expression generation is proposed. As seen in fig.2.4.1, physiological layer, emotional layer and social layer are three aspects with increasing cognition, they determine the fuzzy facial expression generation.

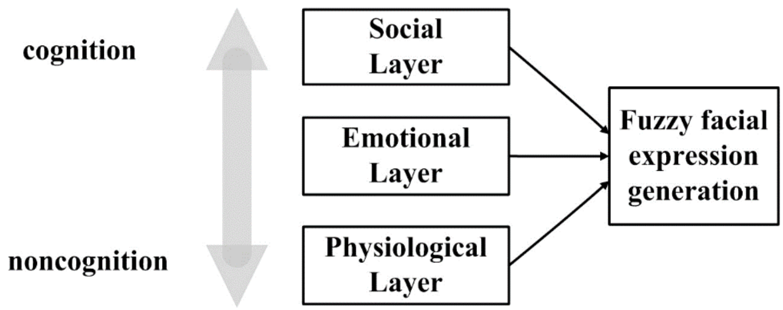


Fig. 2.4.1. Model of layered fuzzy facial expression generation

The social layer includes social factors for facial expression generation, such as social intents and display rules. When, where and how to express facial expressions is restricted by the social rules. A felt emotion may be masked by a fake emotion due to some display rules. For example, in public, whatever a waiter felt, he should show a smile of politeness to the customer.

The emotional layer includes emotional factors for facial expression generation, such as emotions and moods. Basic emotions may lead to universally recognized emotional facial expressions. The type and intensity of the emotion, the mood, and the expression personality will influence the emotional facial expression.

The physiological layer includes physiological factors for facial expression generation, such as physiological variables and states. The physiological variables influence the emotional expressions or lead to physiological expressions. Physiological states such as pain may lead to a grimace. Facial reflexes are also aroused by physiological layer.

The fuzzy facial expression generation deals with the factors in the three layers and generates facial expression fuzzily. In one hand, the mappings from the factors in the layers to the social, emotional, and physiological expressions are fuzzy. In another hand, the intensity control of facial expression is fuzzy. Thus, facial expressions generation can be smart and expressive.

In the LFFEG model, social expressions, emotional expressions and physiological expressions can be generated from the social layer, emotional layer and physiological layer respectively. Social expressions are the facial expressions such as smile of politeness, social wink and social plea regardless of the emotion behind. Emotional expressions are the facial expressions elicited by kinds of emotions such as happiness, sadness and so on. Physiological expressions are the facial expressions elicited by physiological activities, alike facial reflexes, including quick expressions such as startle, horror and other expressions such as frown, blink and gape. Social expressions, emotional expressions and physiological expressions may overlap each other. For example, a wink may be elicited by social, emotional or physiological factors.

Layered Fuzzy Facial Expression Generation System

System Overview

Based on the model of layered fuzzy facial expression generation, a layered fuzzy facial expression generation system is founded, as shown in Fig.4.2.2. The social layer, emotional layer, physiological layer and personality are the inputs of the layered fuzzy facial expression generation which is composed of 5 modules and output facial expression animation.

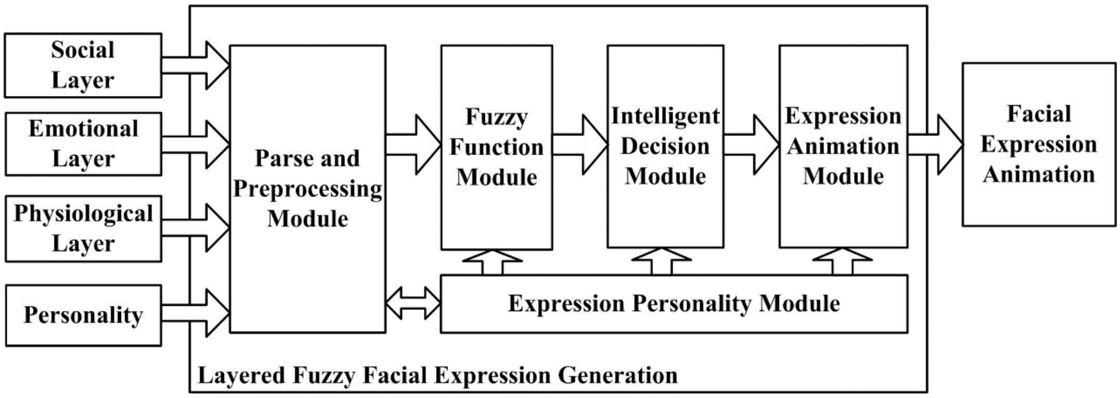


Fig. 4.2.2. Overview of the Layered Fuzzy Facial Expression Generation System

The outputs of the three layers are defined as follows:
(1) Time: t;
(2) Social layer: $S(t)=\{S_P(t), S_I(t), S_E(t), S_S(t), S_R(t)\}$, where S_P: priority, S_I: social intent, S_E: social expression, S_S: social situation, S_R: social display rules;
(3) Emotional layer: $E(t)=\{E_P(t), E_S(t), E_M(t)\}$, where E_P: priority, E_S: specific emotions, E_M: mood;
(4) Physiological layer: $P(t)=\{P_P(t), P_E(t), P_V(t), P_S(t)\}$, where P_P: priority, P_E: physiological expressions, P_V: physiological variables, P_S: physiological states.
The personality of virtual agent is defined as: $PS=(O, C, E, A, N)$, where O, C, E, A, N represent the values of openness, conscientiousness, extraversion, agreeableness, and neuroticism in the five factor model [McCrae92].

Parse and Preprocessing Module

Parse fuzzy vectors for fuzzy function module

(1) Social layer: Given the social space $S=\{s_1,s_2,\ldots,s_l\}$, where s_i is the possible social intent (plea, agreement, disagreement, bored, interest) or social expression (normal, smile, wink). Then define the fuzzy vector $F_S(t)$:

$$F_S(t)=\left(fs_1(t), fs_2(t),\cdots, fs_l(t)\right) \tag{2.4.1}$$

where $fs_i(t)$ is the degree of membership of s_i in time t.
(2) Emotional layer: Given the emotion space $E=\{e_1,e_2,\ldots,e_m\}$, where e_i is the possible emotion of 4 categories[12](①pre-emotions: comfort, distress; ②basic emotions: joy, anger, fear, sadness; ③primary cognitive emotions: liking, disliking, etc.; ④secondary cognitive emotions: hope, resentment, pity, etc.). Then define the fuzzy vector $F_E(t)$:

$$F_E(t)=\left(fe_1(t), fe_2(t),\cdots, fe_m(t)\right) \tag{2.4.2}$$

where $fe_i(t)$ is the degree of membership of e_i in time t.
(3) Physiological layer: Given the physiological space $P=\{p_1, p_2, \ldots, p_n\}$, where p_i is the possible physiological state (pain, tiredness) or physiological expression (sneeze, yawning, startle). Then define the fuzzy vector $F_P(t)$:

$$F_P(t)=\left(fp_1(t), fp_2(t),\cdots, fp_n(t)\right) \tag{2.4.3}$$

where $fp_i(t)$ is the degree of membership of p_i in time t.

Parse Influences for intelligent decision module

(1) Social layer: $I_S(t)=\{\theta_{pos}(t),\theta_{neg}(t),w_{pos}(t),w_{neg}(t)\}$
Display rules can be specified by some of the above four variables, θ_{pos} is the threshold of positive expression, θ_{neg} is the threshold of negative expression, w_{pos} is the weight of positive expression, and w_{neg} is the weight of negative expression.
Social situations (alone, in public, with close friends, with family members, with casual acquaintances, with people of higher status, with people of lower status, and with children[13]) are also translated into corresponding display rules. For example, *alone*={0.5, 1.0, 1.0, 1.3}, denoting that in alone situation, the threshold of positive expression is 0.5, and the negative expression will be weighted by 1.3.
(2) Emotional layer: $I_E(t)=\{E_M(t)\}$
If mood E_M is not provided from the emotional layer, it can be calculated depending on the intensities of positive and negative emotions over the last n time periods.

$$E_M(t)=\begin{cases} positive & \sum_{i=t-n}^{t-1} w_i^+ I_i^+ > \sum_{i=t-n}^{t-1} w_i^- I_i^- \\ negative & otherwise \end{cases} \tag{2.4.4}$$

where I_i^+ is the intensity of positive emotions at time i, and I_i^- is the intensity of negative emotions at time i, w_i^+ is the weight of positive emotions at time i, and w_i^- is the weight of negative emotion at time i.

Mood will influence the expressiveness of emotions. If mood is positive, the intensity of positive expression will increase while the intensity of negative expression will decrease. If mood is negative, the intensity of positive expression will decrease while the intensity of negative expression will increase.
(3) Physiological layer: $I_P(t)=\{P_V(t)\}$

Physiological variables are adrenaline, blood pressure, blood sugar, dopamine, endorphine, energy, heart rate, respiration rate, temperature, vascular volume. As these physiological variables' numerical values are of different ranges, both normalized values (in [0,1]) and linguistic values (such as low, normal, high) are adopted here to make it convenient.

Physiological variables will influence the emotional facial expression generation. For example, high levels of endorphine can increase the expressiveness of positive emotions or decrease the expressiveness of negative emotions, or trigger a state of happiness.

Preprocessing priorities for intelligent decision module

The priorities of the layers are most important to indicate their impact on the facial expression generation. Normally, social layer, emotional layer, and physiological layer have high, middle, and low priorities respectively. However, when the variables in different layers change, the priorities may be alterable and calculated as below when they are not specified.

(1) Social layer: $S_P(t) = 0.4e^{0.9/k(\theta_{pos}+\theta_{neg})}$

where *k*=1 normally and can be adjusted according to the expression personality.

(2) Emotional layer: $E_P(t) = 0.3e^{1.2\max\{e_i(t)\}}$

(3) Physiological layer: $P_p(t) = 0.2e^{1.6\max\{P_S(t), 2|P_V(t)-0.5|, P_E(t)\}}$

Fuzzy Function Module

The fuzzy function module is composed of 4 components, as seen in Fig.2.4.3. Firstly, input fuzzy vector is mapped to the facial expression through fuzzy expression mapping. Then, the maximum intensity of input variable is mapped to the intensity of corresponding facial expression through the components of fuzzification, fuzzy control rule and defuzzification.

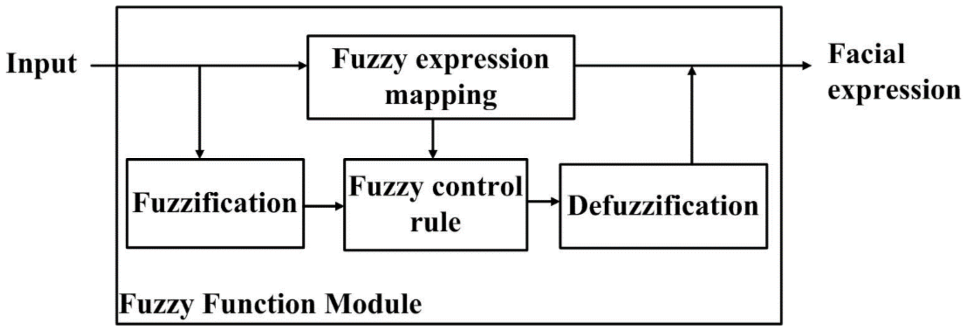


Fig. 2.4.3. Overview of the fuzzy function module.

Fuzzy expression mapping

For the emotional layer, given the emotional facial expression space *EF*={ef_1, ef_2,…, ef_p}, where ef_i indicates any mode of facial expression. The fuzzy mapping matrix R_E is defined as the fuzzy relation $\tilde{R}_E$ from the emotion space *E* to the emotional facial expression space *EF*.

$$R_E = (re_{ij})_{m \times p} \tag{2.4.5}$$

where $re_{ij} = \tilde{R}_E(e_i, ef_j) \in [0,1]$ indicates the correlation degree of (e_i, ef_i) to $\tilde{R}_E$.

Then fuzzy emotional expression vector F_{EX} can be obtained:

$$F_{EX} = F_E \circ R_E = (fef_1, fef_2, \cdots, fef_p) \tag{2.4.6}$$

where fef_i is the membership of the facial expression ef_i, $\circ$ means the compositional operation of the fuzzy relations.

Given the fuzzy emotional expression vector F_{EX}, a certain emotional expression EX_E can be selected according to the maximum membership principle. In case of multiple facial expressions having the same maximum membership, one facial expression will be chosen with certain possibilities according to the expression personality. Thus, the facial expression may be variable when the inputs are the same.

Similarly, we can obtain the social and physiological expressions according to the fuzzy expression mappings for the social and physiological layers respectively.

Fuzzification

Fuzzification is the process to transform crisp values into degrees of membership for linguistic terms of fuzzy sets. The fuzzy linguistic values of input intensities are defined as very low, low, middle, high and very high. The fuzzy linguistic values of output facial expressions are defined as very small, small, middle, large and very large.

Fuzzy control rule

The mapping from linguistic value of input intensity to linguistic value of facial expression intensity is realized through fuzzy control. For example, emotion e_4 (surprise) can be fuzzily expressed by facial expression ef_5 or ef_6. Here are two rules:
(1) If e_4 is very low, then ef_5 is small or ef_6 is very small.
(2) If e_4 is very high, then ef_6 is very large.

Defuzzification

Defuzzification is the process to produce a quantifiable result in fuzzy logic. After the fuzzy intensity control, linguistic values of facial expression intensities are defuzzified to crisp values of intensities. Here, the Center of Gravity method of defuzzification is used.

Expression Personality Module

Expression personality is the personality traits in facial expression generation. We adopt the arousal-valence-expressiveness (AVE) space [Mao07] for emotion expression mapping. Arousal denotes the intensity of the emotion. Valence denotes whether the emotion is positive or negative. Expressiveness denotes the intensity of emotional facial expression. Expression personality can be reflected from the expressiveness distribution of emotions. For example, emotion with low expressiveness is mapped to the facial expression with low intensity, such as sadness, hate and love; emotion with high expressiveness is mapped to the facial expression with high intensity, such as fury and surprise.

The expression personality can be adjusted according to the extraversion and agreeableness of personality. The values of extraversion include extravert, neutral, and introvert. The values of agreeableness include agreeable, neutral, and disagreeable. Both extraversion and agreeableness are neutral default. If extravert is specified, the expressiveness of the emotions will increase. If introvert is specified, the expressiveness of the emotions will decrease. If agreeableness is specified, the expressiveness of the negative emotions will decrease. If disagreeable is specified, the expressiveness of the negative emotions will increase.

Intelligent Decision Module

Intelligent decision module determines the final facial expression according to the priorities of different layers, as seen in Fig. 2.4.4. The flow of facial expression determination is from the layer1 to layer3 with decreasing priorities. If expression is specified in one layer, influences of the layer and higher layers will be made. Otherwise, default expression will be influenced by the three layers. Note that default expression can be variable according to the expresion personality.

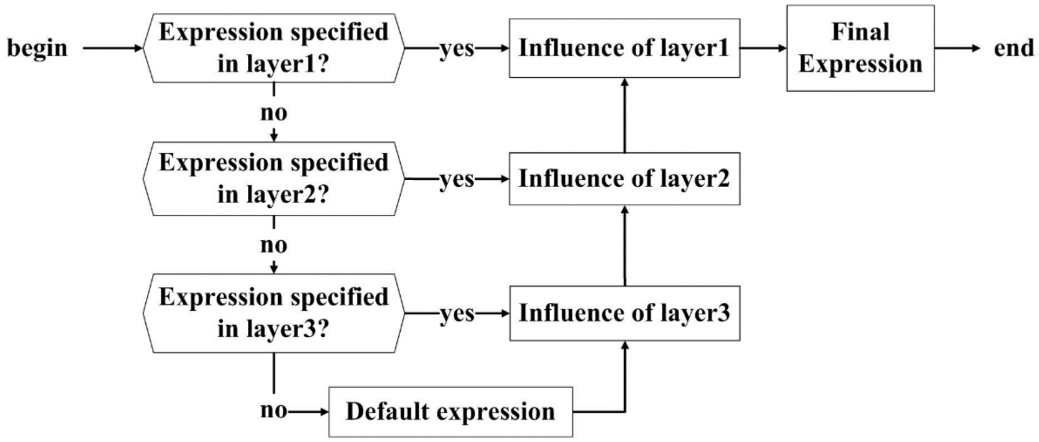


Fig. 2.4.4. Flow of facial expression determination

Expression Animation Module

Expression animation module generates facial expression animations based on Xface toolkit [15]. The facial expression animation adopts example-based expression synthesis scheme. We also developed advanced XfaceEd to generate various keyframes of facial expressions based on MPEG-4 standard. Some keyframes of facial expressions such as disgust, surprise, disliking, pain and tiredness are shown in Fig. 2.4.5.

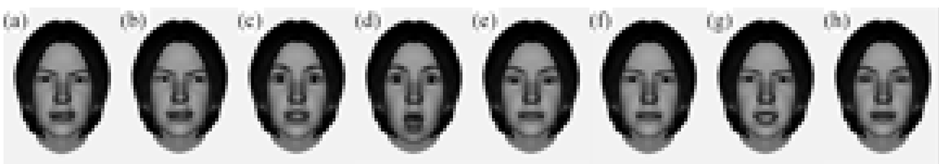


Fig. 2.4.5. Some keyframes of facial epressions.
(a)disgust1;(b)disgust2;(c)surprise1;(d)surprise2;(e)disliking1;(f)disliking2;(g)pain;(h)tiredness

Once facial expression (related to a keyframe of facial expression) is determined by the intelligent decision module, facial expression animation can be generated by the expression animation module. Note that different facial expression animations may have different durations. For example, the duration of startle is much shorter than that of surprise. For some facial expressions such as agreement, disagreement, sneeze, yawning, etc., head rotation is also added in facial expression animation. The facial animation can be controlled by the SMIL-AGENT script.

Layered Fuzzy Facial Expression Generation Language

The LFFEGS is realized based on extensible markup language (XML), which provides an easy way to control the agent's behavior. Previous efforts in the XML-based languages are Human Markup Language (HML), Multimodal Presentation Markup Language (MPML) [Prendinger04], Synchronized Multichannel Integration Language (SMIL) [Not05], etc. Here, Layered Fuzzy Facial Expression Generation Language (LFFEGL) is developed to realize the LFFEGS.

Document-type definition (DTD) is a set of rules that defines the grammer of an XML document. The DTD for the LFFEGL is shown in Fig.2.4.6. The tag of "time" specifies the time information of facial expression animation. The tags of "social", "emotional" and "physiological" relate to the social layer, emotional layer and physiological layer respectively.

```
<!ELEMENT lffegl(time)>
<!ATTLIST lffegl agent          CDATA     #REQUIRED
                 personality    CDATA     #REQUIRED>

<!ELEMENT time(social,emotional,physiological)>
<!ATTLIST time   begin          CDATA     #REQUIRED
                 end            CDATA     #IMPLIED>

<!ENTITY social(normal | smile | plea | agreement | disagreement | wink | bored |
interest)?>
<!ATTLIST social priority             CDATA     #IMPLIED
                 situation            CDATA     #IMPLIED
                 positive_threshold   CDATA     #IMPLIED
                 negative_threshold   CDATA     #IMPLIED
                 positive_weight      CDATA     #IMPLIED
                 negative_weight      CDATA     #IMPLIED>
……
<!ENTITY smile EMPTY>
<!ATTLIST smile  intensity            CDATA     #REQUIRED>
……

<!ENTITY emotional(happyfor | pity | resentment | gloating | joy |distress | hope |
fear | satisfaction | fearsconfirmed | relief | disappointed | pride | selfreproach |
appreciation | reproach | gratitude | anger | gratification | remorse | liking | disliking |
surprise | disgust | sadness | sorryfor)?>
<!ATTLIST emotional priority          CDATA     #IMPLIED
                    mood              CDATA     #IMPLIED
                    positive_weight   CDATA     #IMPLIED
                    negative_weight   CDATA     #IMPLIED>
……
<!ENTITY pity EMPTY>
<!ATTLIST pity  intensity             CDATA     #REQUIRED>
……

<!ENTITY physiological(adrenaline | bloodpressure | bloodsugar | dopamine |
endorphine | energy | heartrate | respirationrate | temperature | vascularvolume |
pain | tiredness | sneeze | yawning | startle)?>
<!ATTLIST physiological priority      CDATA     #IMPLIED>
……
<!ENTITY bloodpressure EMPTY>
<!ATTLIST bloodpressure  intensity    CDATA     #REQUIRED>
……
```

Fig. 2.4.6. DTD for LFFEGL

An example of LFFEGL Script is shown in Fig. 2.4.7. In the first time step, physiological layer has the main impact, and painful expression is displayed. In the second time step, the social layer has higher priority than emotional layer, and joy expression influenced by public situation is displayed.

```
<lffegl agent="alice" personality="extravert">
  <time begin="0s">
    <physiological priority="0.9">
        <pain intensity="1"/>
    </physiological>
  </time>
  <time begin="3s">
    <social situation="public">
    </social>
    <emotional mood="positive">
        <joy intensity="0.5"/>
    </emotional>
  </time>
</lffegl>
```

Fig. 2.4.7. An example of LFFEGL Script

Evaluation and Results

Evaluation of the LFFEGS

The interface of the LFFEGS is shown in Fig. 2.4.8. The list of key frames of facial expressions can be seen in the top left of the interface, and the work space of the LFFEGL is positioned in the bottom left. Fuzzy facial expression animation can be generated through LFFEGL script, as seen in the right region of the interface.

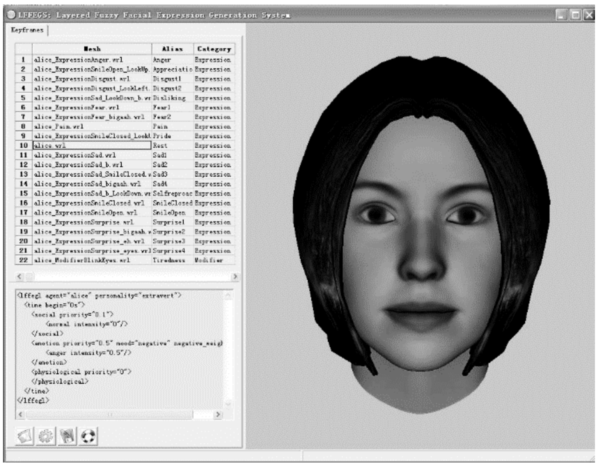


Fig. 2.4.8. Interface of the LFFEGS

To evaluate the LFFEGS, 35 scripts were written in LFFEGL to generate fuzzy facial expression animations with middle intensity. 20 subjects were asked to evaluate the fuzzily generated facial expression animations through the scripts and give the score of satisfaction from 1 (low) to 5 (high). The average scores of social expressions, emotional expressions and physiological expressions reached more than 3, denoting that most facial expressions are expressive.

Then, the subjects were allowed to use the system freely and asked to fill a questionnaire with 5 statements. For each statement, subjects were asked to evaluate how much they agreed with it, on a scale from 1 (low) to 5 (high). The results are shown in Fig.2.4.9, denoting that social, emotional and physiological expressions are acknowledged as necessary by most people. The LFFEGS and LFFEGL are also well evaluated, showing that the proposed framework for layered fuzzy facial expression generation of virtual agent is successful and efficient.

Questionnaire	
It is necessary to display social expressions, such as smile of politeness.	4.4
It is necessary to display emotional expressions, such as happiness, sadness.	4.5
It is necessary to display physiological expressions, such as tiredness, painful.	4.3
The LFFEGL is easy to use.	3.9
Overall I was pleased with the LFFEGS.	3.9

Fig. 2.4.9. Questionnaire results

Comparison of facial expression generation systems

A comparison of facial expression generation systems was given in the items of Social Expression, Emotional Expression, Physiological Expression, Fuzzy Expression, and Expression Personality, as seen in Table 2.4.1. In the LFFEGS, the inclusion of social, emotional and physiological expressions is more comprehensive than other systems. The fuzzy expression and expression personality are also the highlights of the LFFEGS.

Systems	Social Ex.	Emotional Ex.	Physio. Ex.	Fuzzy Ex.	Ex. Personality
Ochs's[3]	masked Ex.	OCC's	—	Yes	—
Khanam's[4]	—	Basic	—	Yes	—
Arellano's[5]	—	OCC's + Basic	—	—	FFM+PAD
LFFEGS	wink, etc.	OCC's + Basic	tiredness, pain, etc.	Yes	FFM+AVE

Table 2.4.1. Comparison of facial expression generation systems. Note: Ex.=Expression, Physio.=Physiological

Conclusion

In this paper, we proposed a novel model of layered fuzzy facial expression generation and developed the corresponding facial expression generation system and language. In the LFFEG model, the influences of the social, emotional and physiological factors are considered in different layers, and facial expressions are fuzzily generated. In the LFFEG system, fuzzy theory, expression personality and intelligent decision are realized to generate personalized fuzzy facial expression. The LFFEG language provides an efficient way for facial expression generation of virtual agent.

2.5 3D biometrics

2.5.1 Human Identification Using Kinect Technology

Face recognition is one of the most widely studied problems in computer science due to various advantages, such as universality, robustness, permanence and accessibility [Yang11]. It becomes increasingly important in many applications, including urban surveillance, home security, and healthcare [Lin11]. Our approach aims human identification using soft biometrics (face, skeleton) extracted from 2D and 3D video sources. The proposed solution is based on low cost hardware [Kinect12].

INTRODUCTION

For many human identity recognition applications, facial imagery represents a key criterion. Face recognition is still a vividly researched area because the current state-of-the-art person identification systems have good recognition performance for structured environments, when the user presents a frontal view, neutral expression under consistent lighting conditions whereas the performances degrade sharply with variations in facial expression, position, head pose, or illumination [Zhao05].

Most current state-of-the-art facial and body recognition systems are based on 2-D images or videos, which offer good performance only for the data captured under controlled conditions [Lei11]. As a result, there is currently a shift towards the use of 3-D data to yield better recognition performance [Schwarz11]. However, it requires more expensive data acquisition systems and sophisticated processing algorithms. 3D face recognition is a promising technology because it is expected to provide greater recognition rates than the two-dimensional approach, e.g. by overcoming the limitations due to viewpoint, shape or light variations [Medioni03], [Hesher03]. Supplementary, the distance and curvature information could contain key discriminative information offering certain advantages over traditional intensity based techniques.

However, we think that there may be useful information in the 2D that is not in the 3D shape, such as skin color, freckles, and other such features. Thus, the appropriate issue may not be 3D versus 2D, but instead be the best method to combine 3D and 2D. Thus, in our paper, using combined skeletal tracking and depth information, we obtain the 2D image of the face region. All these features are provided by a low cost 3D acquisition system, the Kinect sensor [Kinect12]. This information is further processed using standard image processing (PCA feature extraction) and machine learning (distance-based classifier) techniques.

THE KINECT TECHNOLOGY

Today, compact and relatively inexpensive systems for the capturing 3D images with high accuracy are available.

Stereo vision - a method of rendering objects with the added depth information - is arguable the most used method. Despite the numerous attempts [Labayrade02], [Leyes11], [Wang11] it still has some major disadvantages, e.g. the correspondence problem.

Currently, the Time-of-Flight (ToF) methods are intensively studied and implemented in various hardware solutions [Kolloroz08], [den Bergh11], [Breuer07]. In this situation, the distance is computed from the propagation time of the light beam between the camera and the scene's objects for each point of the image. Typical examples for such 3D video cameras are Mesa Imaging's Swiss Ranger or the PMD [vision]® CamCube 3.0 [PMD2012]. Still, the hardware price is too high for consumer appliances.

In a third principle, called "structured light", a narrow band of light is projected onto a 3D shape. In the same time the scene is observed by a camera [Slavi10], [Scharstein03], [Gupta11]. Because of the distance between objects and the light source, the appearance of the light band will suffer modifications. In this way it is possible to calculate the depth information for a particular scene. This principle is used in some 3D sensors, e.g. Microsoft Kinect. This range sensor was employed in our work mainly because it is inexpensive and widely available.

The Hardware Platform

Fig. 2.5.1 depicts the main external components of the Kinect sensor. It has as outputs an infrared structured-light laser projector, a LED indicator, and a motor to control tilt-in base; as inputs, four microphones, two cameras (RGB and IR), and one accelerometer [19], [20].

B. The Software Platform

The Kinect Software Development Kit (SDK) represents the software part of the Kinect for Window package. Microsoft Visual Studio 2010 and its various supported programming languages, e.g. C#, Visual Basic or C++, might be used in developing SDK based applications. The newly release Kinect for Windows SDK version 1.6 (October 2012) offers improved skeletal information, high quality speech recognition, and the ability to support up to four Kinect devices connected to a single computer [Webb12], [Catuhe12].

The Kinect for Windows SDK provides the tools and APIs, both native and managed, that one need to develop Kinect-enabled applications for Microsoft Windows [Kinect SDK12].

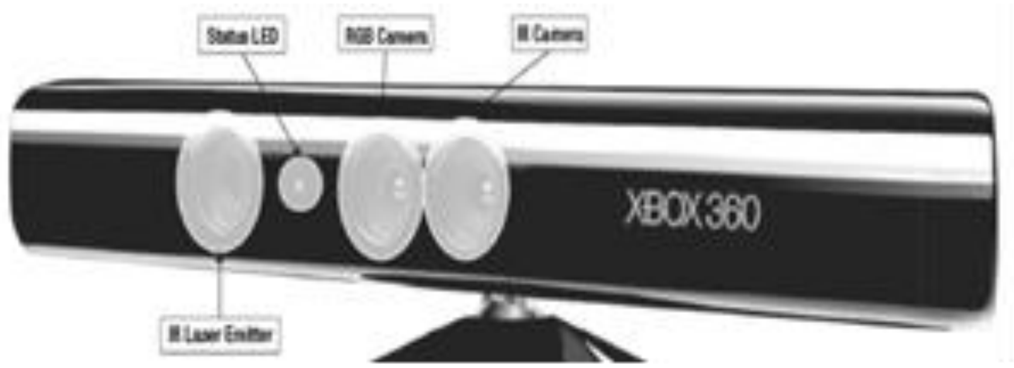


Figure 2.5.1. The Kinect sensor architecture.

HUMAN IDENTIFICATION PROCEDURE

In order to perform person identification using soft biometrics we developed an application which is able to:

- Capture, record and store color/depth data
- Play the recorded data
- Perform 3D skeletal tracking
- Generate 2D facial images from the scene
- Perform human identification
- Implement speech recognition

Persons standing in front of the sensor are detected using skeletal function. The skeletal function is able to detect up to six persons and track maximum two persons. The data from the skeletal tracking are provided to the application as a set of points such as head, elbow, spine, hand, foot, knee etc.

The head point of the skeleton is used to detect person's face. Once the algorithm detected the head, the depth data is used to map the skeleton depth. This is necessary because from the depth data of the skeleton point we can retrieve the X and Y coordinates of the head point.

The color data is copied according to the X and Y coordinates of the head, coordinates determined earlier using the skeleton and depth data from the Kinect sensor. Only the face of the person is copied into binary array files.

Then, as a preprocessing steps, we first normalize the current face image taken in real time by the Kinect sensor to 100x100 pixels and converted it to 8 bit gray scale. Also we apply histogram equalization on the gray scale images. It will increase the global contrast of images, especially when the usable data of the image is represented by close contrast values. This allows for areas of lower local contrast to gain a higher contrast. Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values [acharya05].

The recognition algorithm uses a function implemented in the Emgu CV [EmguCV], called EigenObjectRecognizer. With the help of this function, the faces existing in the database are compared with the ones received in real time from the Kinect sensor.

OpenCV (Open Source Computer Vision Library) contains a set of useful functions for the fields of computer vision and machine learning, developed by Intel and now supported by Willow Garage. The library is cross-platform, it aims real-time digital image processing, and it is free for use under certain license conditions [Bradski08], [Bradski12].

The Emgu CV allows calling OpenCV routines from .NET compatible languages, e.g. C# [Solis12], so it is a cross-platform .NET wrapper to the OpenCV set of functions.

The EigenObjectRecognizer class from Emgu CV used for the recognition algorithm creates an object recognizer using the specific training data and parameters, using a procedure similar with [Turk91a] and [Turk91b]. A new face image is transformed into its eigenface components by a simple operation:

$$\omega_n = \mu_n(\Gamma - \Psi) \tag{2.5.1}$$

for n = 1,..., M' where M' represents the most significant eigenvectors, Ψ - the average face of the set - defined by $\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n$ and μ_n are the eigenfaces:

$$\mu_n = \sum_{k=1}^{M} v_{nk}\Phi_k = Av_n, n = 1,...,M \tag{2.5.2}$$

The weights form a vector $\Omega^T = [\omega_1, \omega_2, ...\omega_{M'}]$ that describes the contribution of each eigenface in representing the input face image. Further, the method determines which face class provides the best description of an input face image by calculating the face class k that minimizes the Euclidian distance:

$$\varepsilon_k^{\ 2} = \left\| (\Omega - \Omega_k)^2 \right\| \tag{2.5.3}$$

where Ω_k is a vector describing the kth face class.

It will always return the most similar object. It has as parameters the images used for training, each of them having the same size, the labels corresponding to the images and the criteria for recognizer training. The algorithm returns the name of the person form the database having the most similarities with the one in front of the sensor. If the person is not recognized (not enough similarities are found) the algorithm returns an empty string, which means the person is unknown. In order to increase the recognition rate, multiple faces (from different angles) of the same person could be added to the database.

EXPERIMENTAL EVALUATION

The human identity recognition application was implemented using Microsoft Visual Studio 2010 and Microsoft Kinect SDK v1.4. In the following we will refer only to the face based identification part of our system.

The first component of the application displays Kinect data as following: color data (the top-left corner), depth data (the top-right corner), skeletal tracking data (the bottom-left corner), sensor's elevation angle (the bottom-right corner), as is depicted in fig .2.5.2.

When the person is unknown, a label named "Unknown person" appear above the persons head. The person remains unknown until it is added into the data based and the recognition algorithm is started (fig. 2.5.3).

In order to add an unknown person to the database, its name must be filled in the corresponding text box and the "Add face" button should be pressed. Now the person exists in the database and can be rec0ognized by the application (fig. 2.5.4). The name of the recognized person appears above its head but also in the right side of the application.

It is very important that the person to be recognized to stay at the optimum distance from the sensor, that is between 1m and 3m. The optimum distance is about 1.5 m.

The application shows real time capabilities when running on a mid-level laptop, HP Compaq 8510w, configured as Intel Core 2 Duo T7500 processor, 2GB DDR2 RAM and 256MB NVIDIA Quadro FX 570M GPU with Microsoft Windows 7 32 bit.

Video sequence and the final application could be downloaded from http://www.ea.etc.upt.ro/Kinect.html

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This paper has presented a possibility to implement a real time system performing face based human identification using the Kinect technology. A solution for obtaining the face images using the combination of the 3D information provided by the skeletal tracker and 2D information provided by the RGB camera was proposed. Then, an appearance-based method (eigenface - a PCA approach) in conjunction with a distance based classifier solution was chosen for implementing the feature extraction and classification stages. In order to perform these operations, a .NET wrapper was employed to enable calling OpenCV image processing library functions.

Future work should address the possibility to use the newly introduced Microsoft Face Tracking SDK. Using it is possible to calculate the head pose and the face expression in real time (fig. 2.5.5).

We would also like to improve the classification stage by following the recent trends in computational intelligence: the use of biologically inspired architectures, e.g. reservoir computing [Vandoorne11], liquid state machines (LMS) [Jaeger07] and echo state networks (ESN). For example, [Woodward11] and [Grzyb09] present a face recognition/facial expression application that uses ESN and LSM architectures respectively, achieving high recognition rates and robustness to noise. These approaches use either 2D or 3D information in recognition.

Lastly, we would like to include other biometric traits (skeleton, voice, hair color, eye color and skin color, as well as the existence of beard, moustache and glasses) for further increasing the human identification accuracy [Dantcheva10].

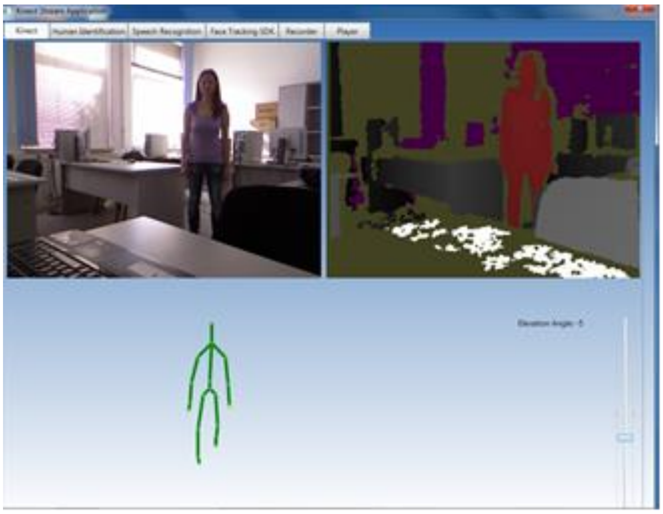


Figure 2.5.2. Kinect data acquisition.

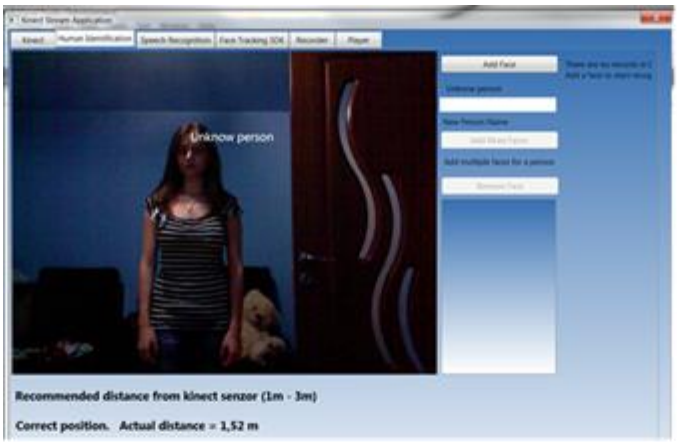


Figure 2.5.3. Detected person not present in the facial database.

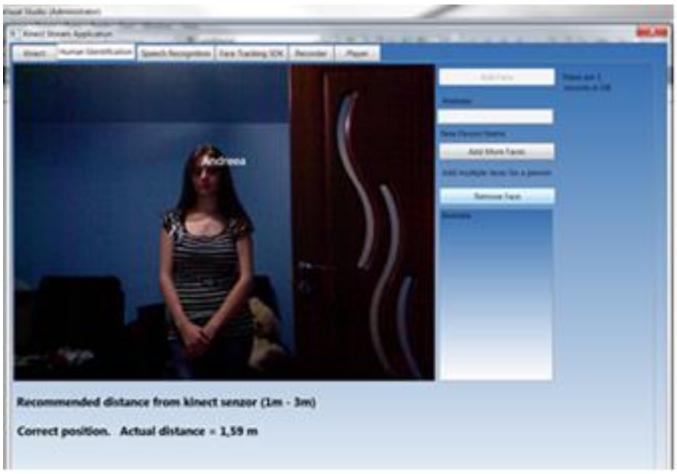


Figure 2.5.4. Successfully recognized person. In order to increase the recognition rate, multiple faces of the same person could be added to the database.

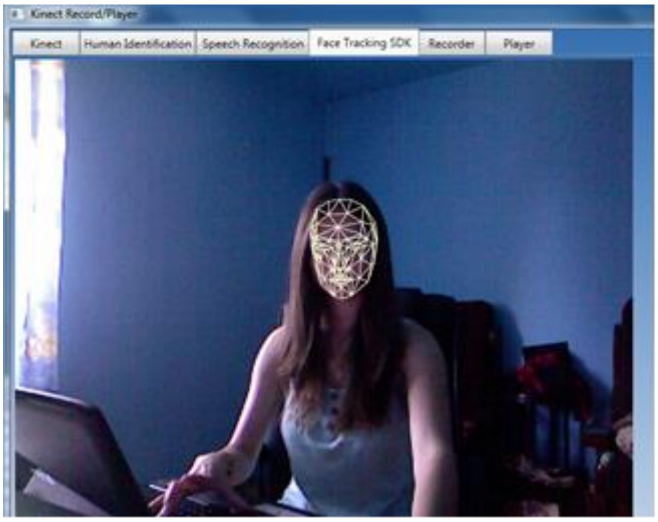


Figure 2.5.5. SDK Face Tracking offering a Candide-3 3D face model [Ahlberg12].

2.5.2 A ToF 3D Database for Hand Gesture Recognition

Although 3D information presents, in the context of hand gesture recognition (and not only), multiple advantages in comparison with the 2D counterpart, up-to-date there are very few 3D hand gesture databases. Recently, the Time-of-Flight (ToF) principle - employed in certain range imaging 3D cameras – became more and more attractive. According to it, the measurement distance is derived from the propagation time of the light pulse between the camera and the subject for each point of the image. In this paper, we describe the development of UPT ToF 3D Hand Gesture Database (UPT-ToF3D-HGDB). It represents, according to the best of our knowledge, the single database of this type which is publicly available.

Introduction

The use of hand as an interfacing method has become more and more popular in the last years. Nowadays we use hand as a direct input device, we have touch screens in our pockets embedded in smart phones with no keys, touristic information points based on touch screens, but the ultimate frontier is to use hand gestures without direct contact to interact with the surrounding devices. The Computer Vision based techniques can offer such non-contact solutions. These solutions have applications in diverse fields from virtual reality to healthcare, from vehicles interfaces to desktop and tablet PC applications.

According to the investigated references, and best of our knowledge, there are a few publicly available gesture image databases. Even fewer represent 3D datasets. These facts justify our endeavor in developing a publicly available 3D dataset for hand gesture recognition acquired using the novel ToF principle implemented in PMD[vision]® CamCube 3.0 video camera [PMD13].

Related Work

This Vision based approaches are divided into two categories: 3 D hand model based approaches and appearance based approaches.

Some appearance based approaches use low-level features like the centroid of the hand region, principle axes defining an elliptical bounding region of the hand, and the optical flow/affine flow of the hand region in a scene. In the last years it is noticeable a new trend: more and more approaches use invariant local features like SIFT [Wang08], SURF [Bao11], ARPD [Chuang11], Haar like [Chen07] features. Invariant features allow modeling the hand as collection of characteristic parts. Key points or characteristic regions are extracted. Using such features the hand gesture is decomposed in simpler parts which are easier to recognize. This approach has major advantages: even if some parts are missing a gestures still can be recognized, so there is robustness to partial occlusions, changes in view point and considerable deformations. A different popular approach, but with major drawbacks, is to look for skin colored regions in an image. The skin color detection is very sensitive to lighting conditions. Efficient and practicable solutions are obtained under controlled and known illumination; the challenge is to learn flexible skin models and to adapt them over time [Bretzner02].

Another approach is to use the eigenspace. In [Black96], this approach with three major improvements (a large invariance to occlusions, some invariance to differences in background from the input images and the training images and the ability to handle both small and large affine transformations of the input image with respect to the training images) is used for hand tracking.

The 3 D hand model based approaches use quadrics as shape primitives [Stenger06], [Kerdvibulvech09] truncated quadrics, surface mash constructed via PCA from training examples [Heap96], deformed hand triangulated surface [de La Gorce11]. Recent works use the depth information provided by the camera. One of the first who worked with range data to recognize hand gesture is Malassiotis [Malassiotos01], the 3D information was acquired following a structured light approach. [Bay04] applied a similar method by using also structured light to generate depth maps in combination with a skin color model to eliminate the background. Their model consists of a polygonal skin, driven by an underlying skeleton. In [Oikonomidis11] a Kinect sensor is use to initialize and track a full DOF hand model. [Perrin04] worked with a target-oriented laser beam and two mirrors to perform 3D tracking. The authors determine distance measurements by using the absolute value of the reflected light, but the system cannot be used in real time. A recent work is presented by Breuer [Breuer07]. They describe a system based on Swissranger SR2 TOF camera which provides depth information. To reconstruct the hand, a principle component analysis and a special hand model are used. In [Kollorz7] Kollorz presents a system which uses a PMD-Sensor with a resolution of 160 × 120 pixels and a viewing angle of 40 degrees. The 12 static hand gestures recognized by the system show good separation potential along the two image axes, the classification is based on the projections of the hand onto the x and y axis and depth feature is included to distinguish between gestures which have same projections, but different alignments. Bergh introduced [denBergh11] a hand gesture interaction system based on an RGB camera and a ToF camera. The hand is detected based on adaptive skin color detection and depth, while classification is based on 2D Haarlets.

Publicly available datasets:

2D datasets

- Cambrige [Cambrige] - Gesture data base

This database has 9 different gesture classes generated by 3 different shapes with motion added. In this database there are 900 images sequences, 100 images for each class. These images were taken under 5 different illumination conditions and the background is uniform. Image format – jpg.

- Jochen Triesch [Triesch] Static Hand Posture Database I

This database consists of 10 hand postures, more precisely letters a, b, c, d, g, h, i, l, v, y from ASL, available for 3 backgrounds (light, dark, complex). The hand postures are performed by 24 persons. Image format – pgm.

- Jochen Triesch [Triesch] Static Hand Posture Database II

In this database there are color images for 12 hand postures, performed by 19 persons against simple backgrounds (dark and light) and complex ones. The total number of images is about 1000. Image format – tiff.

- Sebastien Marcel Static Hand Posture Database [Triesch]

Consist of six hand postures perform by 10 persons, the color images have a complex background. Image format – ppm

- Sebastien Marcel Dynamic Hand Posture & Hand Gesture Database [Triesch]

This dynamic hand posture database has 4 hand gestures color image sequences: Click, Rotate, Stop-Grasp-Ok and No, the image format is pnm and the resolution is 58x63 pixels, while the Hand Gesture Database consist of 2D hand trajectories in a normalized body-face space for 4 hand gestures performed by 10 persons several times.

- The National Center for Sign Language and Gesture Resources - Video Sequences of American Sign Language (ASL) [ASL]

This contains annotated QuickTime movies of American Sign Language (ASL) sentences. The signing was captured simultaneously from four different cameras, the frame rate is 60 frames per second, and image resolution is 648x484 pixels. Therefore, the database contains samples of upper-body image sequences. However, the focus is entirely on sign language communication and not affective body expression.

3D datasets

- Bosphorus Hand Database [Bosphorus]

The Bosphorus Hand Database was build to be used for research on hand biometry. The hand geometry and the hand texture is acquired with a commercial scanner (hands are placed flat on the glass platen) when fingers are apart from each other. The total number of images is 4846, from 918 subjects; for 642 subjects 6 images/person, three right-hand images and three left-hand images were acquired, for 276 subjects only three left-hand images were acquired. 160 among 918 subjects have hand images with time lapses of several months.

- Sébastien Marcel - InteractPlay Database [InteractPlay]

The database consists of 16 gestures performed by 20 different persons, actually it is a hand gesture database made of 3D hand trajectories. Most of gestures involve one hand and some implies both hands (fly, swim and clap). The use of gloves with distinct colors permits to avoid occlusion problems that occur when both hand are used. The person performing the gesture wears gloves of different colors and a sweat-shirt of a specific color.

A new database on hand gestures

The recent technological achievements enable the capture of 3D images with high accuracy using very compact and relatively inexpensive systems. On the other hand, the increase of research activities in the domain of hand posture and gesture recognition generates new methods and algorithms, some of them being able to process 3D data type. Unfortunately, up to date there are very few platforms for evaluating the algorithms' performances. For this purpose, we have developed a ToF 3D gesture database. It represents, according to the best of our knowledge, the single publicly available database of this type.

ToF Principle

There are several principles for 3D image acquisition. Among them, stereo vision is arguable the most used method. Despite the numerous attempts, it still has some major disadvantages, e.g. the correspondence problem. Another 3D acquisition possibility is represented by the scanning systems. They imply mechanical components, are cost-intensive and have low real-time capabilities.

In a ToF 3D system, a modulated optical signal sent out by a transmitter illuminates the scene to be measured. In our situation, the employed modulation technique is called continuous wave [Beheim86] and typically fmod is 20 KHz. The reflected light is detected, in our situation, by a Photonic Mixer Devices (PMD) sensor, which is able to determine the phase difference between the sent and received optical signal per every single pixel. More exactly, the distance to the target is given by the following relation:

$$d = \frac{c \cdot \varphi}{4 \cdot \pi \cdot f_{mod}} \qquad (2.5.4)$$

where c is the speed of light and ϕ the phase shift [Lange00].
In this situation, the complete 3D information is captured in parallel, without needing excessive calculation power.

The Hardware

For image aqusition we used a high resolution 3D video camera produced by PMD Technologies GmbH Company, PMD[vision]® CamCube 3.0 [PMD13]. With the optical sensor having a resolution of 200x200 pixel, it is currently the highest resolution ToF 3D camera worldwide (fig. 2.5.6).

Figure 2.5.6. The ToF 3D acqusition system: PMD[vision]® CamCube 3.0 [1].

The Software

Although Linux/Windows application programing interface (PMDSDK2) and MATLAB interface (PMDMDK) are available from the camera provider, we have used, for the purpose of data acquisition, a more elaborated visualization tool called CamVis3 (fig. 2.5.7).

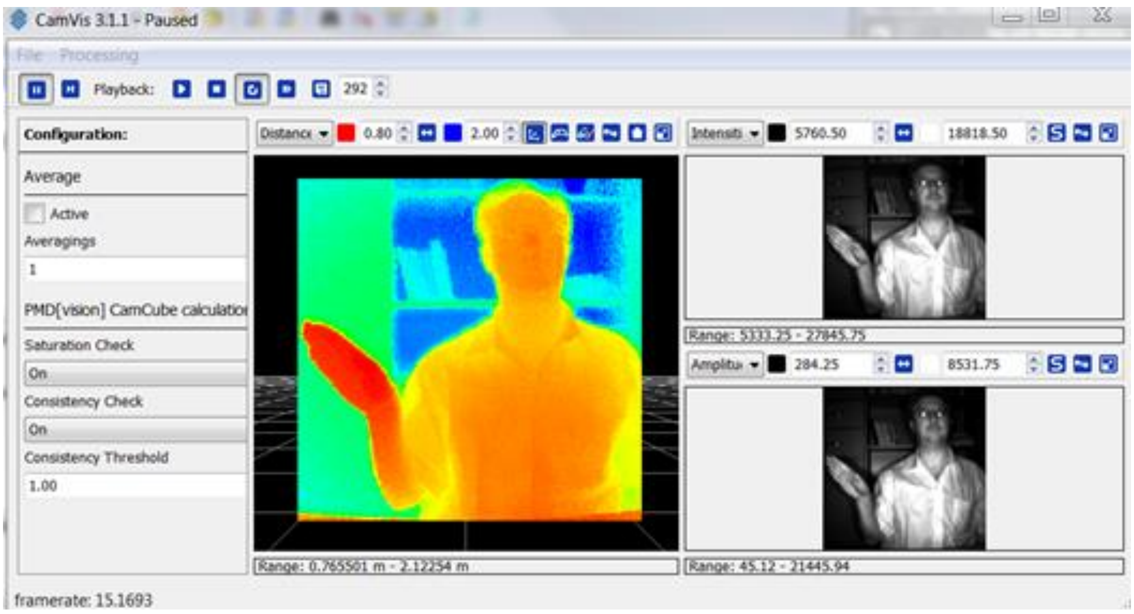


Figure 2.5.7. The ToF 3D data acquisition tool, CamVis3.

The captured data stream contain, beside the distance values, the grayscale information and the amplitudes of the signal.

UPT ToF 3D Hand Gesture Database

A new dataset, UPT-ToF3D-HGDB, is under construction at the Faculty of Electronics and Telecommunications from "POLITEHNICA" University from Timişoara. Having in mind that a high quality ToF video camera is still an expensive equipment (more than 5K €), the dataset is being constructed specifically to support researchers whom are willing to experiment 3D techniques. Current release, available at the following URL:

http://www.ea.etc.upt.ro/UPT-ToF3D-HGDB.html

contain 10+ subjects expressing six static hand poses (fig. 2.5.8) and four dynamic hand gestures.

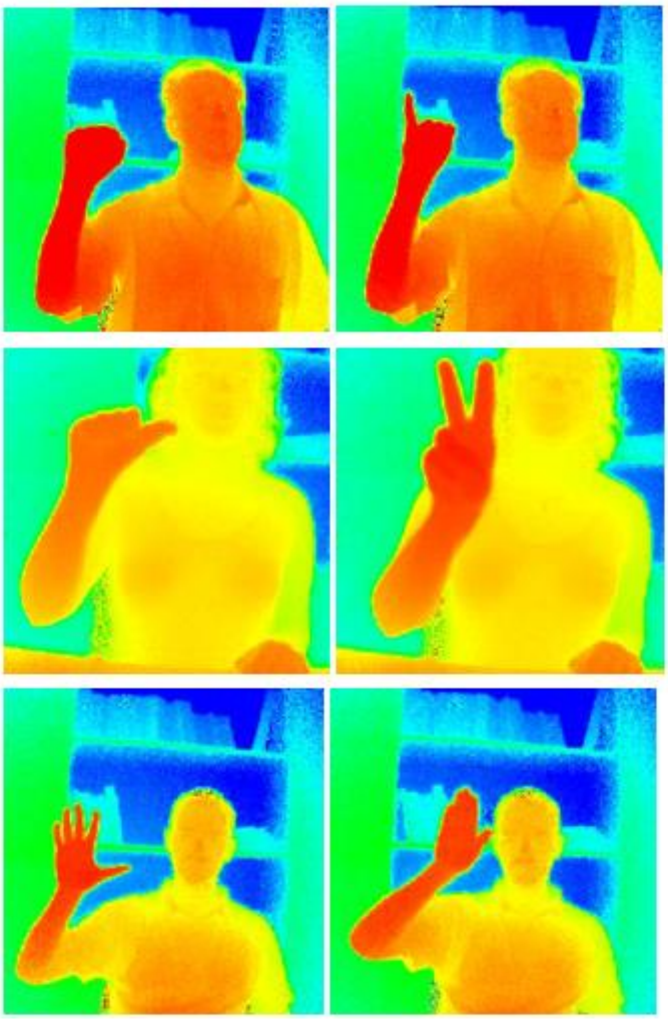


Figure 2.5.8. Samples form UPT-ToF3D-HGDB. Here are presented six hand postures performed by three distinct persons. The depth information is color coded: very close objects are displayed in red, objects in the middle in green and far away objects in blue.

The very simple segmentation operation represents a big advantage of this kind of data.

One of the typical applications for this image database is to use it as a training set for learning algorithms. The same database could also be used for the testing phase.

CONCLUSION

Recently, a number of ToF 3D approaches have been reported in the literature [Kollorz07], [Schwarz10], [Bigdelou12]. Unfortunately, their performances could not be easily compared due to the lack of a publicly available database.

The UPT ToF 3D Hand Gesture Database has been developed with the intention of providing a common benchmark for the 3D algorithms.

Besides increasing the number of database individuals, we will concentrate in the future, on the development of specific ToF 3D algorithms which have to exploit the advantage of the additional depth values while cope with the intrinsic low resolution.

2.6 Future research, professional and academic career development plans

I will refer in the following to three future research topics I would also like to propose to the potential Ph.D. students. They also constitute draft applications for national/European funding grants.

- *Innovative Methods for Biometric Person Identification*

The main purpose of this research is to propose and investigate new methods aiming biometric person identification solely carried out by facial and body biometric traits. In this way, unlike other biometrics, our system doesn't require direct physical contact or exact positioning in front of the identification device. The prospective research proposes contributions in the following fields: computational intelligence, image processing (segmentation, shape estimation), parallel computing (GPU highly optimized architectures), new generation of sensing technology (employment of time-of-flight camera, structured light and infrared sensors) and real time implementation (multicore SoC, FPGA). Although some of the above mentioned principles typically involve very expensive commercial systems we would like to explore – as one of the major challenges of the project - real-time 3D imaging solutions at a low cost (consumer hardware). This will increase the system availability and will wider the range of possible applications (surveillance & security systems, human computer interfaces) in various domains (domotics, robotics, banking sphere, law enforcement, game industry).

The current state-of-the-art person identification systems have good recognition performance for structured environments, when the user presents a frontal view, neutral expression under consistent lighting conditions. One element of difficulty is that the performance falls off drastically with variations in pose, illumination, background, aging, and expression. Another problem is represented by the availability of the data sets containing both 2D images or videos and 3D models of face and body. For example, there are some public 2D face databases and 3D face databases but as far as we are aware, there is only one database who provide 3D data in addition to 2D appearance information and it provide only facial data.

Our exploratory research aims video based human identification solely carried by a bag of facial and body traits. Thus, we propose the followings:
- A novel biometric (face&body) database will be created using a combination of 2D colour high resolution video and 3D low resolution image data captured using low cost/consumer hardware
- The development of a robust, real-time, video segmentation framework along with an adequate robust shape estimator
- Embed contextual information extracted from the image, within all processing steps. This will imply the development of a dynamic, context driven variable importance selection. At classification level, this goal will be approached in the framework of recursive partitioning and random forests, a recently dominant trend in pattern classification and computer vision
- Enhancements (new implementation and training algorithm for the read-out layer) of the existing 3rd generation neural network architecture in order to make them more suitable for sensorial information fusion (face, body shape and gait).

- The investigation of several embedded platforms suited for such biometric system, including GPU, Multi-Core Processor combining ARM and DSP cores and FPGAs which could also provide real-time capabilities.

- *Bio-Inspired Affective Computing with Applications in Automotive Engineering*

Affective Computing is a field of Artificial Intelligence concerned with understanding, recognizing, and utilizing human emotions in the design of technological systems. It has become a very hot research topic in human–computer interaction because it helps improve the quality and effectiveness of human to computer communications. Despite the advances in computing and related technologies, there is no artificial system which can match the human capabilities in dealing with emotions. The aim of the project is to investigate and promote new bio-inspired solutions, i.e. neural networks, fuzzy systems or evolutionary computing in the implementation of artificial systems that can process emotions. Various domains can benefit from the project research outcomes, e.g. robotic systems, games, learning technologies, medicine and psychology. Within the framework of the current project we are particularly interested in applying the above mentioned principles in natural driver-car communication. In the near future, the main research domains in the automobile development will definitely include emotional factors and affective states as crucial elements for enhanced safety and comfort. Our research will contribute in the development of a unified bio-inspired technology which, based mainly on the visual information, will address the problem of affect (emotions, fatigue, stress, nervousness, etc.) assessment.

The importance of the specific problem studied and the potential impact of the proposed objectives for science, society or technology are given by the possible applications of our research:

- *Automotive industry*. Emotional and affective states factors (anger, sadness, happiness, disgust, fear, irritation, surprise, interest/gaze, fatigue, eye/head movement) are crucial elements for enhanced safety and comfort.
- *Robotic systems*. Robots capable of processing affective information exhibit higher flexibility while one works in uncertain or complex environments. Also the evaluation of the emotional state is important in delivering the appropriate service. Companion devices, such as digital pets, use affective computing abilities to enhance realism and provide a higher degree of autonomy.
- *Learning technologies/e-learning*. In this context we could mention the efforts to create natural communication interfaces of perceptual or conversational type.
- *Medicine and Psychological* health services. Counselling can benefit from affective computing applications when determining a client's emotional state. Such systems could also provide help in the diagnosis and treatment of autism, infant aplasia.

- *Computer-assisted diagnosis system for the improvement of the medical decision in contrast enhanced ultrasound imagery for focal liver lesions*

Contrast enhanced ultrasound (CEUS) is a relatively new ultrasound method that brings the advantage of contrast agent administration also to ultrasound (not only to CT and MRI, as in the past). The great advantage of contrast enhanced ultrasound is that that is a real time method that permits the detailed visualization of the vascularity of a lesion. The technique was first developed for the characterization of focal liver lesions, although today it has several applications.

The aim of the project is to create an intelligent CAD system for assisting the diagnosis of FLL from CEUS images. The development of the system will be based on the specific contrast enhanced ultrasound patterns of the different lesions (hemangioma, focal nodular hyperplasia, liver metastases, hepatocellular carcinoma, etc.). To date, there is no such a CAD system commercially available for CEUS studies. Several versions of a prototype system are described in the literature and constitute a good starting base for our project. It is also a good proof of the interest for the proposed work and its feasibility. A more detailed definition of their work in given in the next section, concerned with state of the art description. To our best knowledge, there is no other CAD system dedicated to focal liver lesion diagnosis in CEUS imagery. To accomplish this objective, we will experiment new solutions on all processing levels. We will capitalize on our previous work in image processing and computer vision and bring solutions from this filed to the narrower domain of biomedical image processing. A block diagram of the proposed system is given in figure 2.6.1.

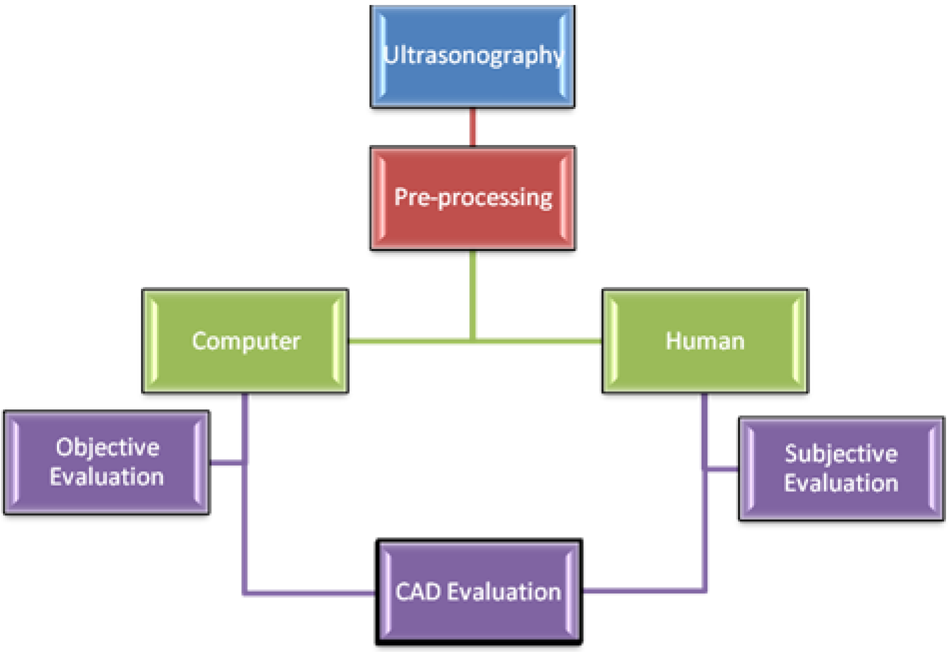


Fig. 2.6.1. Block diagram of the proposed system

The scientific results achieved during the research will be submitted for publication in prestigious ISI journals and conferences. Parts of the research could be also disseminated in my taught courses of “Elements of Artificial Intelligence”, “Expert Systems” or “Embedded Systems”. Further, additional topics focused on modern AI paradigms (e.g. Liquid State Machine (LSM)/Echo-State Network (ESN)), novel digital image techniques (3D images, recursive partitioning and random forests) or advanced embedded architectures (GPU, Multi-Core Processor combining ARM and DSP cores and FPGAs) will proposed to be studied at the doctoral school level.

3. REFERENCES

3.1 General References

[3dMD] 3dMD Ltd., http://¬www.3dmd.com.

[Abramson10] Abramson, M.A., 2010. Nomadm optimization software. URL <www.gerad.ca/ NOMAD/Abramson/nomadm.html>.

[Acharya05] T. Acharya and A. Ray, "Image Processing: Principles and Applications", New York: Wiley, 2005.

[Aha91] Aha, D.W.; Kibler, D. & Albert, M.K. (1991). Instance-based Learning Algorithms, Machine Learning, vol. 6(1), pp. 37-66, Springer Netherlands, ISSN: 0885-6125, 1991.

[Ahlberg01] J. Ahlberg, "CANDIDE-3 -- an updated parameterized face," Report No. LiTH-ISY-R-2326, Dept. of Electrical Engineering, Linköping University, Sweden, 2001.

[Ahlberg12] J. Ahlberg, CANDIDE - a parameterized face, http://www.icg.isy.liu.se/candide/, 2012.

[Ahuactzin95] Ahuactzin, J; E. Mazer, E. & Bessière, P, (1995). L'algorithme fil d'Ariane, Revue d'intelligence artificielle, vol. 9(1), pp. 7-34, 1995.

[Akbarzadeh00] M. -R. Akbarzadeh, K. Kumbla, E. Tunstel, M. Jamshidi, Soft computing for autonomous robotic systems, *Computers and Electrical Engineering*, Vol. 26, No. 1, 2000, pp. 5-32.

[Al Allan95] Al Allan, S.; Pradel, G.; Barret, C. & Abou Kandil, H. (1995). Neural Architectures for Mobile Robot Navigation, International Conference on Advanced Robotics and Intelligent Automation, pp. 440-445, Athens. Greece, 1995.

[Altschul91] Altschul, S., (1991). Amino acid substitution matricesfrom an information theoretic perspective, Journal of Molecular Biology, vol. 219, pp. 555-565, 1991.

[Arellano08] D. Arellano, J. Varona, F.J. Perales, "Generation and visualization of emotional states in virtual characters", Computer Animation and Virtual Worlds, Vol.19, No.3-4, pp.259-270, 2008.

[Arkin98] Arkin, G., (1998). Nomad 200 – simulator MissionLab, Behaviour-Based Robotics, MIT Press, 1998.

[ASL] http://csr.bu.edu/asl/html/sequences.html

[AT&T02] AT&T Laboratories Cambridge, 2002. Database of faces. Online. URL <http:// www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.

[Audet03] Audet, C., Dennis, J., 2003. Analysis of generalized pattern searches. SIAM J. Optim. 13 (3), 889–903.

[Audet06] Audet, C., Dennis, J.E., 2006. Mesh adaptive direct search algorithms for constrained optimization. SIAM J. Optim. 17 (1), 188–217.

[Bao11] J. Bao, A. Song, Y. Guo, H. Tang, "Dynamic hand gesture recognition based on SURF tracking", IEEE International Conference on Electric Information and Control Engineering (ICEICE), 15-17 April 2011 , pp 338 – 341.

[Barrett98] Barrett, W.A., 1998. A survey of face recognition algorithms and testing results. Systems and Computers 1, 301–305.

[Bartlett02]Bartlett, M., Movellan, J., Sejnowski, T., 2002. Face recognition by independent component analysis. IEEE Trans. Neural Networks 13 (6), 1450–1464.

[Bay04] M. Bay, Koller-Meier, L.V Gool, " Smart particle filtering for 3D hand tracking", in Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Los Alamitos, CA, USA, 2004, pp 675

[Beheim86] G. Beheim, K. Fritsch, "Range finding using frequency-modulated laser diode," Applied Optics, vol. 25, no. 9, pp. 1439-1442, 1986

[Bigdelou12] A. Bigdelou, L. Schwarz, T. Benz, and N. Navab, "A flexible platform for developing context-aware 3D gesture-based interfaces," in Proceedings of the 2012 ACM international conference on Intelligent User Interfaces (IUI '12), ACM, New York, USA, 2012, pp. 335-336.

[Black96] M. Black, D. Jepson, "Eigen tracking: Robust matching and tracking of articulated objects using a view-based representation", in European Conference on Computer Vision, 1996.

[Blanzieri99] Blanzieri, E. & Rizzi,F. (1999). Advanced metrics for class driven similarity search, Proceedings of 10th International Workshop on Database and System Application, pp. 223-227, Firenze, Italy, 1999.

[Bosphorus] http://bosphorus.ee.boun.edu.tr/hand/Home.aspx

[Bradski08]G. Bradski, A. Kaebler, "Learning OpenCV", O'Reilly, 2008.

[Bradski12] G. Bradski, "Using the Kinect Depth Camera with OpenCV", http://opencv.willowgarage.com/wiki/Kinect

[Bras95] Bras, F.; Pradel, G. & Jin, Z., (1995). Cellular Automata Applied to the Path Generation and Environment Representation for a Mobile Robot, IFAC Motion Control Conference, pp. 395-402, Munich, Germany, 1995.

[Bretzner02] L. Bretzner, I. Laptev, T. Lindeberg, “Hand gesture recognition using multiscale color features, hieracrchichal models and particle filtering”, in Proceedings of Int. Conf. on Automatic face and Gesture recognition, Washington D.C., USA, May 2002

[Breuer07] P. Breuer, C. Eckes, S Muller, “Hand Gesture Recognition with a novel IR Time-of-Flight Range Camera - A pilot study”, Proceedings of Mi-rage 2007, Computer Vision / Computer Graphics Collaboration Techniques and Applications, pp.247–260

[Bronstein05] Bronstein, A.M., Bronstein, M.M., Kimmel, R., 2005. Three-dimensional face recognition. Int. J. Comput. Vis. (IJCV) 64 (1), 5–30.

[Brunelli92] Brunelli, R., Poggio, T., 1992. Face recognition through geometrical features. In: European Conference on Computer Vision (ECCV), pp. 792-800.

[Brunelli93] Brunelli, R., Poggio, T., 1993. Face recognition: Features versus templates. IEEE Trans. Pattern Anal. Machine Intell. 15 (10), 1042–1052.

[Buciu09] I. Buciu, C. Kotropoulos, and I. Pitas, “Comparison of ICA approaches for facial expression recognition,” Signal, Image and Video Processing, vol. 3, pp. 345–361, 2009.

[Buck91] R. Buck, "Social factors in facial display and communication: a reply to Chovil and others", J Nonverb Behav, Vol.15, No.3, pp.155-161, 1991.

[Căleanu00] Căleanu, C.D., 2000. Facial Recognition using Committee of Neural Networks, In: Proc. 5th Seminar on Neural Network Applications in electrical engineering, NEUREL 2000, Belgrade, Yugoslavia., pp. 97-100.

[Căleanu01] Căleanu, C., 2001. Face recognition using parallel neural processing and interest operator method. Ph.D. thesis, University POLITEHNICA Timisoara.

[Căleanu01] Căleanu, C.D., 2001. Face recognition using parallel neural processing and interest operator method. Ph.D. thesis. University POLITEHNICA Timisoara.

[Căleanu07] Căleanu, C., Huang, D., Gui, V., Tiponuţ, V., Maranescu, V., 2007. Interest operator vs. Gabor filtering for facial imagery classification. Pattern Recognition Lett. 28 (8), 950–956.

[Cambrige] ftp://mi.eng.cam.ac.uk/pub/CamGesData

[Catuhe12] D. Catuhe, Programming with the Kinect for Windows Software Development Kit, Microsoft Press, 2012.

[Cesar02] Cesar, R., Bengoetxea, E., Bloch, I., 2002. Inexact graph matching using stochastic optimization techniques for facial feature recognition. In: International Conference on Pattern Recognition (ICPR), Quebec, Canada.

[Chatzis11] S.P. Chatzis, Y. Demiris, "Echo State Gaussian Process", Neural Networks, IEEE Transactions on, vol.22, no.9, pp.1435-1445, Sept. 2011.

[Chen07] Qing Chen , N.D. Georganas, E.M Petriu, " Real-time Vision based Hand Gesture Recognition Using Haar-like features", IEEE Transactions on Instrumentation and Measurement, 2007.

[Choset01] Choset, H. & Nagatani, K. (2001). Topological Simultaneous localization and Mapping (SLAM): Toward Exact Localization without Explicit Localization, IEEE Transaction on Robotics and Automation, vol. 2(17), pp. 125-137, 2001.

[Chuang11] Y. Chuang, L. Chen, G. Zhao, G. Chen, "Hand Posture Recognition and Tracking Based on Bag-of-Words for Human Robot Interaction", IEEE International Conference on Robotics and Automation Shanghai International Conference Center, Shanghai, China May 9-13, 2011.

[Cohen03] I. Cohen, N. Sebe, A. Garg, S. C. Lawrence, S. H. Thomas, "Facial expression recognition from video sequences: temporal and static modeling", Computer Vision and Image Understanding, Special Issue on Face Recognition, vol. 91, pp. 160-187, Issues 1-2, July-August 2003.

[Collet92] Collet, T.; Dillmann, E.; Giger, A.; Wehner, R. (1992). Visual landmarks and route following in desert ants, Journal of Comparative Physiology, Springler Verlag, vol. 170, pp. 435-442, 1992.

[Crosnier99] Crosnier, A. (1999). Modélisation géométrique des environnements en robotique mobile, French Workshop on Robotic Resarch (Journées Nationales de la Recherche en Robotique), Montpellier, France, pp. 83-91, 1999.

[Dantcheva10]A. Dantcheva, J-L Dugelay, P. Elia, "Person Recognition using a bag of facial soft biometrics (BOFSB)", MMSP 2010, IEEE International Workshop on Multimedia Signal Processing, October 4-6, 2010.

[Daugman85] Daugman, J.G.,. 1985. Uncertainty relation for resolution in space, spatial frequency and orientation optimized by 2D visual cortical filters. Journal of the Optical Society ofAmerica, 2(7), 1160–1169.

[De Jong93] K.A. De Jong, W.M. Spears, D.F. Gordon, Using genetic algorithms for concept learning, *Machine Learning*, vol. 13, 1993, pp. 161-188.

[de La Gorce11] M. de La Gorce, D. J. Fleet, N. Paragios, "Model-Based 3D Hand Pose Estimation from Monocular Video", IEEE Transactions On Pattern Analysis And Machine Intelligence, 2011

[den Bergh11] V. den Bergh, L.Van Gool, "Combining RGB and ToF cameras for real-time 3D hand gesture interaction ", Applications of Computer Vision (WACV), 2011 IEEE Workshop.

[DeSouza02] DeSouza, G.N. & Kak, A.C. (2002). Vision for mobile robot navigation: a survey, IEEE Transaction on pattern analysis and machine intelligence, Vol. 24(2), pp. 237-267, 2002.

[Destrero09] Destrero, A., Mol, C.D., Odone, F., Verri, A., 2009. A sparsity-enforcing method for learning face features. IEEE Trans. Image Process. 18, 188–201.

[Dornaika07] F. Dornaika and B. Raducanu, “Inferring facial expressions from videos: Tool and application,” Image Commun., vol. 22, pp. 769–784, October 2007.

[Ekenel05] Ekenel, H. K., Sankur, B., 2005. Multiresolution Face Recognition, Image and Vision Computing 23, 1-9.

[Ekman69] P. Ekman, W.V. Friesen, "The repertoire of nonverbal behaviour: categories, origins, usage, and coding", Semitotica, Vol.1, pp.49-98, 1969.

[Ekman71] P. Ekman, W. Friesen, "Constants across cultures in the face and emotion", Journal of Personality and Social Psychology, Vol.17, No.2, pp.124-129, 1971.

[Ekman78] P. Ekman and W. Friesen, Facial Action Coding System, Palo Alto, CA: Consulting Psychologists Press, 1978.

[Fang11] T. Fang, X. Zhao, O. Ocegueda, S. K.. Shah and I. A. Kakadiaris, “3D Facial Expression Recognition: A Perspective on Promises and Challenges", in Proc. 9th IEEE International Conference on Automatic Face and Gesture Recognition (FG), Santa Barbara, CA, Mar. 21-25, 2011.

[Fasel03] B. Fasel and J. Luettin, “Automatic facial expression analysis: a survey,” Pattern Recognition, vol. 36, no. 1, pp. 259–275, 2003.

[Filliat02a] Filliat, D. and Meyer, J.A. (2002a). Map-Based Navigation in Mobile Robots. I. A review of Localization strategies, http://citeseer.ist.psu.edu/filliat03mapbased.html.

[Filliat02b] Filliat, D. & Meyer, J.A. (2002b). Global localization and topological map-learning for robot navigation, available at http://citeseer.ist.psu.edu/filliat02global.html

[Fleisher03] Fleisher, J.; Marshland S. & Shapiro, J. (2003). Sensory anticipation for autonomous selection of robot landmarks, http://www.cs.man.ac.uk\~fleischj/research.html

[Franz97] Franz, M.; Schölkopf, B. & Bülthoff, H. (1997). Image-based Homing, Proceedings of the European Conference on Artificial Life, pp. 236-245.

[Fridlund94] A.J. Fridlund, Human Facial Expression: An Evolutionary View, Academic Press, San Diego, CA, USA, 1994.

[Gabor46] Gabor, D., 1946. Theory of communications. Journal of International Electrical Engineers, 93, 427–457.

[Gaussier97] Gaussier, P. ; Joulain, C.; S.~Zrehen, S. & Revel, A. (1997). Image-based Homing, Proceedings of the IEEE International Conference on Intelligent Robots and Systems, pp. 545-550, 1997.

[Graham98] Graham, D.B., Allinson, N.M., 1998. Characterizing virtual eigensignatures for general purpose face recognition. In: Wechsler, H., Phillips, P.J., Bruce, V., Fogelman-Soulie, F., Huang, T.S. (Eds.), Face Recognition: From Theory to Applications. NATO ASI Series F, Computer and Systems Sciences, vol. 163, pp. 446–456.

[Gross01] Gross, R., Shi, J., Cohn, J., 2001. Quo vadis face recognition. In: Third Workshop on Empirical Evaluation Methods in Computer Vision. pp. 119–132.

[Grzyb09] B. J. Grzyb, E. Chinellato, G. M. Wojcik, A. Kaminski, Facial Expression Recognition based on Liquid State Machines built of Alternative Neuron Models, Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009.

[Gui2012] GUI Vasile, Alexa F., Caleanu Catalin, David C., Popa G., Simion G., "Method for tracking hand fingers and human computer communication means", patent RO127779-A0, 2012.

[Guo01] Guo, G., Li, S.Z., Chan, K.L., 2001. Support Vector Machines for Face Recognition. Image and Vision Computing 19, 631-638.

[Gupta11] M. Gupta, A. Agrawal, A. Veeraraghavan, S.G. Narasimhan, "Structured light 3D scanning in the presence of global illumination," Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 20-25 June 2011 pp.713-720.

[Gusfield97] Gusfield, D. (1997). Algorithms on strings, Trees and Sequences, Computer Science and computational Biology, Cambridge University Press, NY, USA, 1997.

[Hawkes03] J. Hawkes, K. Kiriakidis, Community robotics, *Proceedings of the 35th Southeastern Symposium on System Theory*, 16-18 March 2003, pp. 345-346.

[Haykin99] Haykin, S. (1999). Neural neworks, A Comprehensive Foundation, second Edition, Prentice Hall, ISBN: 0 13 273350 1, 1999.

[Heap96] A.J Heap, D.C Hogg, "Towards 3-D hand tracking using a deformable model", in 2nd International Face and Gesture Recognition Conference 1996, pp 140–45

[Heisele01] Heisele, B., Ho, P., Poggio, T., 2001. Face Recognition with Support Vector Machines: Global versus Component-based Approach. Proceedings of the 8th IEEE International Conference on Computer Vision, ICCV 2001, vol. 2. Vancouver, Canada, pp. 688–694.

[Herrera05] F. Herrera, Genetic Fuzzy Systems: Status, Critical Considerations and Future Directions, *International Journal of Computational Intelligence Research*, Vol. 1, No. 1, 2005, pp. 59-67.

[Hesher03] C. Hesher, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range images", Seventh International Symposium on Signal Processing and Its Applications, 2003.

[Hong91] Hong, I. (1991). Image-based Homing, Proceedings of the IEEE International Conference on Robotics and Automation, Sacramento, USA, pp. 620-625, 1991.

[Hoppenot03] Hoppenot, P.; Pradel, G.; Căleanu, C.D.; Perrin, N. & Sommeilly, V. (2003). Towards a symbolic representation of an indoor environment, Proc. IEEE-SEE-CESA2003 - Computing Engineering in Systems Applications, CD ROM paper no. S1-R-00-0048, Lille, France, 2003.

[Hu04]Hu, Y., Jiang, D., Yan, S. Zhang, L., Zhang, H., 2004. Automatic 3D reconstruction for face recognition. In Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 843 - 848.

[Huttenlocher93] Huttenlocher, D.; Klanderman, G. & Rucklidge, W. (1993). Comparing images using Hausdorff distance, IEEE Transactions on pattern analysis and machine intelligence, vol. 15(9), pp. 850-863, 1993.

[InteractPlay] http://www.idiap.ch/resource/interactplay/

[Jabid10] M. H. K. T. Jabid and O. Chae, "Robust facial expression recognition based on local directional pattern," ETRI Journal, vol. 32, no. 5, pp. 784–794, Oct 2010.

[Jaeger07] H. Jaeger, W. Maass, J. Principe, "Introduction to the special issue on echo state networks and liquid state machines", Neural Networks, vol. 20, no. 3, pp. 287-289, 2007.

[Jeong99] I. Jeong, J. Lee, Evolving fuzzy logic controllers for multiple mobile robots solving a continuous pursuit problem, *IEEE International Fuzzy Systems Conference Proceedings*, 1999, pp. 685–689.

[Jia11] Q. Jia, Y. Liu, H. Guo, Z. Luo, Y. Wang, "A sparse representation approach for local feature based expression recognition," Multimedia Technology (ICMT), 2011 International Conference on, pp.4788-4792, 26-28 July 2011.

[Judd98] Judd, S. & Collett, T. (1998). A Mobile Robot That Learns Its Place, Nature, vol. 392, pp. 710-714, 1998.

[Kamarainen02] Kamarainen, J.-K., Kyrki, V., Hamouz, M., Kittler, J., Kälviäinen, H., 2002. Invariant Gabor features for face evidence extraction. In: Proceedings of the IAPR Workshop on Machine Vision Applications. Nara, Japan, pp. 228–231.

[Kanade00] T. Kanade, J. F. Cohn, Y. Tian, "Comprehensive database for facial expression analysis," IEEE International Conference on Automatic Face and Gesture Recognition, 2000.

[Karlin90] Karlin, S. & Altschul, S. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes, proceedings of the national Academy of Science, vol. 87, pp. 2264-2268, 1990.

[Kean11] S. Kean, Y. C. Hall, P. Perry, Meet the Kinect, Apress, 2011.

[Kerdvibulvech09] C. Kerdvibulvech, H. Saito, " Model-Based Hand Tracking by Chamfer Distance and Adaptive Color Learning Using Particle Filter" EURASIP Journal on Image and Video Processing, 2009.

[Khanam08] A. Khanam, M. Mufti, "Intelligent expression blending for performance driven facial animation", IEEE Transactions on Consumer Electronics, Vol.53, No.2, pp.578-584, 2007.

[Kim02a] Kim, K.I., Jung, K., Kim, H.J., 2002a. Face recognition using kernel principle component analysis. Signal Process. Lett., IEEE 9 (2), 40–42.

[Kim02b] Kim, K.I., Kim, J.H., Jung, K., 2002b. Face recognition using support vector machines with local correlation kernels. In: Proceedings of International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), vol. 16. pp. 97–111.

[Kim94] Kim, D. & Neviata, R. (1994). A method for recognition and localization of generic objects for indoor navigation, Proceedings of ARPA Image Understanding Workshop, Monterey, USA, 1994.

[Kinect SDK12]Kinect SDK, http://www.microsoft.com/en-us/kinectforwindows/develop/developer-downloads.aspx, 2012.

[Kinect12] KinectTM for Windows, http://www.microsoft.com/en-us/kinectforwindows, 2012.

[Kohonen87] Kohonen, T. (1987). Content-Addressable Memories, Springer Series in Information Sciences, vol. 1, Springer Berlin Heidelberg, 1987.

[Kohonen88] Kohonen, T. (1988). Self-Organization and Associative Memory, Springer Series in Information Sciences, vol. 8, Springer Berlin Heidelberg, 1988.

[Kohonen98] Kohonen, T. & Somuervo, P. (1998). Self-organizing maps of symbols strings, Neurocomputing, vol. 21, pp. 19-30, Elsevier, ISSN: 0925-2312, 1998.

[Kollorz07] E. Kollorz and J. Hornegger , "Gesture recognition with a time-of-flight camera", Dynamic 3D Imaging (Workshop in Conjunction with DAGM'07), 2007.

[Kollorz08] E. N. K. Kollorz, J, Penne, J, Hornegger, A, Barke, "Gesture recognition with a Time-Of-Flight camera", International Journal of Intelligent Systems Technologies and Applications, vol. 5, pp. 334-343, no 3/4, doi 10.1504/IJISTA.2008.021296, 2008.

[Konar05] A. Konar, *Computational Intelligence: Principles, Techniques and Applications.* Springer-Verlag, 2005.

[Kotsia07] I. Kotsia, I. Pitas, "Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines," Image Processing, IEEE Transactions on, vol. 16, Issue 1, pp. 172 – 187, Jan. 2007.

[Kramer12] J. Kramer, N. Burrus, F. Echtler, D. Herrera, M. Parker, Hacking the Kinect, Apress, 2012.

[Kuipers91] Kuipers, B. & Byan, Y.T. (1991). A robot exploration and mapping strategy based on a semantic hierarchy of spatial representation, International Journal of Autonomous Systems, vol. 8, pp. 47-63.

[Kulic03] Kulic, R. and Vukic, Z. (2003). Methodology of concept control synthesis to avoid unmoving and moving obstacles, The Knowledge Engineering Review, Vol. 18, issue 3, pp. 281-291, Cambridge University Press.

[Kumar04] M. Kumar, D. P. Garg, Intelligent learning of fuzzy logic controllers via neural network and genetic algorithm, *Proceedings of 2004 JUSFA Japan – USA Symposium on Flexible Automation*, Denver, Colorado, July 19-21, 2004.

[Kurabayashi99] D. Kurabayashi, Toward realization of collective intelligence and emergent robotics, *Conference on Systems, Man, and Cybernetics IEEE SMC '99*, Vol. 4, 12-15 Oct. 1999, pp. 748-753.

[Kwok02] K. S. Kwok, B. J. Driessen, C. A. Phillips, C. A. Tovey, Analyzing the multiple-target-multiple-agent scenario using optimal assignment algorithms, *J. Intell. Robot. Syst.*, vol. 35, 2002, pp. 111–122.

[Kyrki01] Kyrki, V., Kamarainen, J.-K., Kälviäinen, H., 2001. Content based image matching using Gabor filtering. In: Proc. of the International Conference on Advanced Concepts for Intelligent Vision Systems Theory and Applications. Baden-Baden, Germany, pp. 45–49.

[Kyrki02] Kyrki, V., 2002. Local and global feature extraction for invariant object recognition. Ph.D. thesis. Lappeenranta University of Technology.

[Labayrade02] R. Labayrade, D. Aubert, J.-P. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation", Intelligent Vehicle Symposium, vol.2, 17-21 June 2002, pp. 646- 651.

[Lamon01] Lamon, P.; Nourbakhsh, I.; Jensen, B. & Siegwart, R. (2001). Deriving and Matching Image Fingerprint Sequences for Mobile Robot Localization, http://citeseer.nj.nec.com/445679.html, Proceedings of the IEEE International Conference on Robotics and Automation, Seoul, Korea, 2001.

[Lange00] R. Lange, “3D Time-of-flight distance measurement with custom solid-state imagevsensors”, in CMOS/CCD-technology, PhD thesis, Dep. of Electrical Engineering andvComputer Science, University of Siegen, 2000.

[Lawrence97] Lawrence, I.S., Giles, C.L., Tsoi, A.C., Bock, A.D., 1997. Face recognition: A convolutional neural-network approach, IEEE Trans. Neural Networks, vol.8, 98-113.

[Lee04] Lee, J., Moghaddam, B., Pfister, H., Machiraju, R., 2004. Finding optimal views for 3d face shape modelling. In: Proceedings of the International Conference on Automatic Face and Gesture Recognition. Seoul, Korea, pp. 31–36.

[Lei11] Z. Lei, S. Liao, M. Pietikäinen, S.Z. Li, “Face Recognition by Exploring Information Jointly in Space, Scale and Orientation”, Image Processing, IEEE Transactions on, vol. 20, no. 1, pp. 247-256, Jan. 2011.

[Leigh05] R.E. Leigh, T. Morelli, S.J. Louis, M. Nicolescu, C. Miles, Finding attack strategies for predator swarms using genetic algorithms, *The 2005 IEEE Congress on Evolutionary Computation*, Vol. 3, 2-5 Sept. 2005, pp. 2422 – 2428.

[Leonard91] Leonard, J. J. & Durrant-Whyte, H.F. (1991). Simultaneous map building and localization for an autonomous mobile robot, Proceedings of the IEEE/RSJ Int. Workshop on Intelligent Robots and Systems IROS'91, pp. 1442-1447, New York, NY, USA, 1991.

[Levenshtein66] Levenshtein, L.I. (1966). Binary codes capable of correcting deletions, insertions, and reversals, Soviet Physics–Doklady, vol. 10, no.7, pp. 707–710, 1966.

[Lewis99] Lewis, R., Torczon, V., 1999. Pattern search algorithms for bound constrained minimization. SIAM J. Optim. 9 (4), 1082–1099.

[Li99] Li, S.Z., Lu, J., 1999. Face recognition using the nearest feature line method. IEEE

[Lin11] W. Lin, Video Surveillance, InTech, 2011.

[Lin97] Lin, S.H., Kung, S.Y., Lin, L.J., 1997. Face recognition/detection by probabilistic decision-based neural networks, IEEE Trans. Neural Networks, vol.8, 114-113.

[Lisetti00] C.L. Lisetti, D.J. Schiano, "Automatic facial expression interpretation: where human-computer interaction, artificial intelligence and cognitive science intersect", Pragmatics and Cognition, Vol.8, No.1, pp.185-235, 2000.

[Liu01] Liu, C., Wechsler, H., 2001. A Gabor feature classifier for face recognition. In Eighth IEEE International Conference on Computer Vision,. Vol. 2, 7-14 July, pp. 270 – 275.

[Liu05] C. Liu and J. Ward, “Advantages of 3d methods for face recognition research in humans,” in Proceedings of the Second international conference on Analysis and Modelling of Faces and Gestures (AMFG'05), W. Zhao, S. Gong, and X. Tang (Eds.), Springer-Verlag, Berlin, Heidelberg, 2005, pp. 244-254.

[Lu03] Lu, J., Plataniotis, K., Venetsanopoulos, A., 2003. Face recognition using lda-based algorithms. IEEE Trans. Neural Networks 14 (1), 195–200.

[Lyes11] T. Lyes, K. Hawick, "Implementing stereo vision of gpu-accelerated scientific simulations using commodity hardware", in Proc. International Conference on Computer Graphics and Virtual Reality (CGVR'11). Number CGV4047, Las Vegas, USA, 2011, pp76-82.

[Lyons99] M. J. Lyons, J. Budynek, S. Akamatsu, "Automatic classification of single facial images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.12, pp. 1357-1362, 1999.

[Ma04]L. Ma, K. Khorasani, "Facial expression recognition using constructive feedforward neural networks," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 34, no. 3, pp.1588-1595, June 2004.

[MacKenzie94] MacKenzie, P. & Dudek, G. (1994). Precise Positioning using Model-Based Maps, Proceedings of IEEE International Conference on Robotics and Automation, pp. 1615-1621, San Diego, USA, 1994.

[Malassiotis01] S. Malassiotis, F. Tsalakanidou, N. Mavridis, V. Giagourta, N. Grammalidis, M.G Strintzis, " A face and gesture recognition system based on an active stereo sensor" In: Proceedings 2001 ICIP, Thessaloniki, Greece, 7-10 Oct. 2001, vol.3. , pp 955–958

[Mao07] X. Mao, Y.L. Xue, Z. Li, et al., "Research on multiple facial expressions of emotions", Proc. of 5th International Conf. on Computing, Communications and Control Technologies, Orlando, FL, USA, pp.189-193, 2007.

[MathWorks10] The MathWorks Inc., 2010. Global optimization toolbox. Online. URL

[Matsumoto90] D. Matsumoto, "Cultural similarities and differences in display rules", Motivation and Emotion, Vol.14, No.3, pp.195-214, 1990.

[Matuszewski11] B. J. Matuszewski, W. Quan, L-K Shark, "Facial Expression Recognition" in "Biometrics – Unique and Diverse Applications in Nature, Science, and Technology", M. Albert (Ed.). InTech, 2011.

[McCrae92] R.R. McCrae, O.P. John, "An introduction to the five-factor model and its applications", Journal of Personality, Vol.60, No.2, pp.175-215, 1992.

[McKennoch04] S. McKennoch, S. Hoyt, L.G. Bushnell, A study of group size and communication in an evolving fuzzy-controlled population, *International Conference on Fuzzy Systems*,Vol. 2, 25-29 July 2004, pp. 895 -900.

[Medioni03] G. Medioni and R. Waupotitsch, “Face recognition and modelling in 3D”, IEEE International Workshop on Analysis and Modelling of Faces and Gestures (AMFG 2003), October 2003, pp. 232–233.

[Meyer02] Meyer, J.A. & Filliat, D. (2002). Map-based navigation in mobile robots - II. A review of map-learning and path-planning strategies, http://citeseer.ist.psu.edu/meyer03mapbased.html

[Mitchell97] Mitchell, T., 1997. Machine Learning, McGraw Hill.

[Moghaddam01] Moghaddam, B., Nastar, C., Pentland, A., 2001. A bayesian similarity measure for deformable image matching. Image Vis. Comput. 19 (5), 235–244.

[Moller93] Moller, A.F., 1993. A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Networks, 6, 525-533.

[Moon01] Moon, H., Phillips, P., 2001. Computational and performance aspects of PCA-based face recognition algorithms. Perception 30 (3), 303–321.

[Moravec81] Moravec, H., 1981. Robot Rover Visual Navigation. Ann Arbor, MI: Univ. of Michigan Research Press.

[Nasrabadi94] Nasrabadi, N.M., Choo C.Y., 1994. Hopfield Network for Stereo Vision Correspondence, In M. Gupta, G. Knopf (Eds.) Neuro-Vision Systems. Principles and Applications, IEEE Press, pp. 442-449.

[Nasrabadi98] Nasrabadi, N., Choo, C., 1998. Hopfield Network for Stereo Vision Correspondence. In: Gupta, M., Knopf, G. (Eds.), Neuro-Vision Systems. Principles and Applications. IEEE Press, pp. 442–449.

[Nissen09] S. Nissen and E. Nemerson, “Fast Artifical Neural Network,” FANN, http://leenissen.dk/fann/, 2009.

[Nolfi98] S. Nolfi, D. Floreano, Co-evolving predator and prey robots: Do ‘arm races’ arise in artificial evolution?, *Artificial Life*, Vol. 4, No. 4, 1998, pp. 311-335.

[Not05] E. Not, K. Balci, F. Pianesi, et al., "Synthetic characters as multichannel interfaces", Proc. of 7th International Conf. on Multimodal Interfaces, Torento, Italy, pp.200-207, 2005.

[Ochs05] M. Ochs, R. Niewiadomski, C. Pelachaud, et al., "Intelligent expressions of emotions", Proc. of the 1st International Conf. on Affective Computing and Intelligent Interaction, Beijing, China, pp.707-714, 2005.

[Oikonomidis11] I. Oikonomidis, N. Kyriazis, A. A. Argyros, “Efficient model-based 3D tracking of hand articulations using Kinect”, in British Machine Vision Conference (BMVC) 2011.

[Oore97] Oore, S.; Hinton, G. & Dudek, G. (1997). A Mobile Robot That Learns Its Place, Neural Computation, pp. 683-699, MIT Press, vol. 9(3).

[Orlin96] J. Orlin, A polynomial-time primal network simplex algorithm for minimum cost flows, *Proc. 7th Annu. ACM/SIAM Symp. Discrete Algorithms*, 1996, pp. 474–481.

[Panning08] A. Panning, A. Al-Hamadi, R. Niese, and B. Michaelis, "Facial expression recognition based on haar-like feature detection," Pattern Recognition and Image Analysis, vol. 18, pp. 447–452, 2008.

[Pantic00] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, pp. 1424–1445, December 2000.

[Pantic05] M. Pantic, M. F. Valstar, R. Rademaker, L. Maat, "Web-based database for facial expression analysis," in Proc. IEEE International Conference on Multimedia and Expo (ICME'05), Amsterdam, Netherlands, 2005.

[Parker05] G. Parker, I. Parashkevov, Cyclic genetic algorithm with conditional branching in a predator-prey scenario, *International Conference on Systems, Man and Cybernetics*, Vol. 3, 10-12 Oct. 2005, pp. 2923 – 2928.

[Perrin04] S. Perrin, A. Cassinelli, M. Ishikawa, "Gesture recognition using laser-based tracking system", Proceedings. FGR2004, Seoul, South Korea, 17-19 May 2004, pp 541-546

[Pessoa99] Pessoa L., Leitao, A.P. 1999. Complex Cell Prototype Representation for Face Recognition, IEEE Trans. Neural Networks, vol.10, no.6, 1528-1531.

[Petland00] Pentland, A., 2000. Looking at people: sensing for ubiquitous and wearable computing. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, 107–119.

[Philips06] Phillips, P.J., Scruggs, W.T., OToole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M., 2006. Face recognition vendor test 2006. Tech. rep., National Institute of Standards and Technology, Gaithersburg, MD 20899, USA. URL <http:// www.frvt.org/>.

[Picard97] R.W. Picard, Affective Computing, MIT Press, Cambridge, MA, USA, 1997.

[PMD12] PMD[vision]® CamCube 3.0, http://www.pmdtec.com/products-services/pmdvisionr-cameras/pmdvisionr-camcube-30

[Pomerleau93] Pomerleau, D.A. (1993). Neural Network Perception for Mobile Robot Guidance}, Kluwer Academic Publishers.

[Pradel00] Pradel, G.; Avrillon, S. & Garbuio, L. (2000). Landmark interpretation by means of frescoes in mobile robotics, Proceedings of the 6th Int. Conf. On Methods and Models in Automation and Robotics, pp. 585-592, Miedzyzdroye, Poland, 2000.

[Pradel01] Pradel, G. & Bras, F. (2001). Qualitative environment description by means of frescoes in mobile robotics, Journal européen des systèmes automatisés, vol. 9(35), pp. 1105-1128, ISSN 1269-6935, 2001.

[Pradel94] Pradel, G.; Bras, F. & Jin, Z. (1994). 2D laser telemetry-based path trend generation and real time environment symbolic representation for an autonomous mobile robot, Proceedings of the IFAC-IEEE International Conference on Machine Automation, pp. 122-134, Tampere, Finland, 1994.

[Prendinger04] H. Prendinger, S. Descamps, M. Ishizuka, "MPML: a markup language for controlling the behavior of life-like characters", Journal of Visual Languages & Computing, Vol.15, No.2, pp.183-203, 2004.

[Russell97] J.A. Russell, "Reading emotions from and into faces: Resurrecting a dimensional-contextual perspective", J.A. Russell, J. Fernandez-Dols. The Psychology of Facial Expression. Cambridge University Press, New York, NY, USA, pp.295-320, 1997.

[Saïdi06] Saïdi, F. & Pradel, G. (2006). A multi-robot path planner for group navigation, Journal of Intelligent and Robotics Systems, Springer-Verlag eds, under press, 2006.

[Salvi10] J. Salvi, S. Fernandez, T. Pribanic, X. Llado, “A state of the art in structured light patterns for surface profilometry”, Pattern Recognition, vol. 43, no. 8, pp. 2666-2680, August 2010.

[Sato98] Sato, K., Shah, S., Aggarwal, L.J., 1998. Partial Face Recognition using Radial Basis Function Network, Proc. of the third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16, pp. 288-292, Nara, Japan.

[Satonaka06] Satonaka, T., Uchimura, K., 2006. A hybrid metric estimation/learning model for k- nn classifier. In: Proceedings of the 2006 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing, pp. 3337–3342.

[Scharstein03] D. Scharstein, R. Szeliski, “High-Accuracy Stereo Depth Maps Using Structured Light”, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2003, pp. 195-202.

[Schwarz10] L. A. Schwarz, D. Mateus, V. Castaneda, N. Navab, “Manifold Learning for ToF-based Human Body Tracking and Activity Recognition”, in F. Labrosse, R. Zwiggelaar, Y. Liu, and B. Tiddeman, editors, Proceedings of the British Machine Vision Conference, BMVA Press, September 2010, pages 80.1-80.11

[Schwarz11] L. A. Schwarz, A. Mkhitaryan, D. Mateus, N. Navab, “Estimating Human 3D Pose from Time-of-Flight Images Based on Geodesic Distances and Optical Flow” IEEE Conference on Automatic Face and Gesture Recognition (FG), Santa Barbara, USA, March 2011.

[Sebe02] N. Sebe, I. Cohen, A. Garg, M.S. Lew, T.S. Huang, “Emotion Recognition Using a Cauchy Naive Bayes Classifier”, International Conference on Pattern Recognition (ICPR'02), vol I, pp. 17-20, Quebec City, Canada, August 2002.

[Seiffert01] U. Seiffert, Multiple Layer Perceptron Training Using Genetic Algorithms, *European Symposium on Artificial Neural Networks* ESANN2001, Bruges (Belgium), 25-27 April 2001, pp. 159-164.

[Seo07] S.-W. Seo, K.-E. Ko, H.-C. Yang, K.-B. Sim, Behavior learning and evolution of swarm robot system using SVM, *International Conference on Control, Automation and Systems ICCAS '07*, 17-20 Oct. 2007, pp. 1238-1242.

[Shan05] C. Shan, S. Gong, and P. W. McOwan, "Robust Facial Expression Recognition Using Local Binary Patterns," Proc. IEEE International Conference on Image Processing (ICIP'05), Genova, Italy, September 2005.

[Simhon98a] Simhon, S. & Dudek, G. (1998a). A global Topological Map formed by Local Metric Maps, Proceedings of IEEE/RSJ International Conference on Intelligent Robotic Systems, Victoria, B.C., Canada, pp. 1708--1714, 1998.

[Smartkom03] SmartKom, "Smartkom database," 2003. [Online]. Available: www.smartkom.org

[Smith97] Smith, C.M. & Leonard, J.J. (1997). A multiple-hypothesis approach to concurrent mapping and localization for autonomous underwater vehicles, Proceedings of International Conference on Field and Service Robotics, pp.249-256, Canberra, Australia, 1997.

[Solis12] D. Solis, "Illustrated C# 2012", Apress, 2012.

[Stenger06] B. Stenger, A. Thayananthan , P.H.S Torr, R. Cipolla, "Model-based hand tracking using a hierarchical Bayesian filter" IEEE Transactions on Pattern Analysis and Machine Intelligence (2006)

[Stone00] P.Stone and M. Veloso, Multiagent systems: a survey from a machine learning perspective, *Autonomous Robots*, vol. 8, no. 3, July 2000, pp. 345-383.

[Tao98] H. Tao and T. S. Huang, "Connected vibrations: A modal analysis approach to non-rigid motion tracking," CVPR, pp. 735–750, 1998.

[Tedder01] Tedder, M. & Hall, L.E. (2001). Symbolic processing methods for 3D visual processing, http://www.robotics.uc.edu/papers2001/Maurice2001d.pdf.

[Thurn02] Thrun, S. (2002). Robotic mapping: A survey, Exploring Artificial Intelligence in the New Millenium, Morgan Kaufmann, http://citeseer.ist.psu.edu/thrun02robotic.html.

[Tian02] Y.-Li Tian, T. Kanade, and J. Cohn, "Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity," in Proc. Fifth IEEE Int. Conf. Automatic Face and Gesture Recognition, May, pp. 229–234, 2002.

[Tian03] Y. Tian, L. Brown, A. Hampapur, S. Pankanti, A. Senior, and R. Bolle, “Real world real-time automatic recognition of facial expression,” in IEEE PETS, Australia, March 2003.

[Tian05] Y.-L. Tian, T. Kanade, and J. F. Cohn, “Facial expression analysis,” in Handbook of Face Recognition, S. Z. Li and A. K. Jain, Eds. Springer New York, 2005, pp. 247–275.

[Tiponuț02] V. Tiponuţ, C.D. Căleanu, *Neural Networks. Architectures and algorithms*, ISBN 973-9389-66-X, “Politehnica” Publishing House, Timişoara, 2002.

[Torczon97] Torczon, V., 1997. On the convergence of pattern search algorithms. SIAM J. Optim. 7 (1), 1–25.

[Triesch] http://www.idiap.ch/resource/gestures/

[Turk91a] M. Turk and A. Pentland, “Eigenfaces for recognition”, Journal of Cognitive Neuroscience, vol.3, No.1, 1991.

[Turk91b] M. Turk and A. Pentland, “Face recognition using eigenfaces”, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1991, pp. 586-591.

[Valstar12] M.F. Valstar, M. Pantic, "Fully Automatic Recognition of the Temporal Phases of Facial Actions," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol.42, no.1, pp.28-43, Feb. 2012.

[Vandoorne11] K. Vandoorne, J. Dambre, D. Verstraeten, B. Schrauwen, P. Bienstman, “Parallel Reservoir Computing Using Optical Amplifiers”, Neural Networks, IEEE Transactions on, vol. 22, no. 9, pp.1469-1481, 2011.

[Viola04] P. Viola and M. Jones, “Robust real-time object detection,” Int. J. Comput.Vis.., vol. 57, no. 2, pp. 137–154, May 2004.

[Vukadinovici05] D. Vukadinovic and M. Pantic, “Fully automatic facial feature point detection using Gabor feature based boosted features,” in Proc. IEEE Int. Conf. Syst., Man, Cybern., 2005, pp. 1692–1698.

[Wagner74] Wagner, R.A. & M. J. Fischer, M.J. (1974). The string to string correction problem, Journal of the ACM, 21, pp. 168-173, 1974.

[Wallhoff06] F. Wallhoff, “Facial expressions and emotion database,” Technische Universität München, Tech. Rep., 2006.

[Wang05] Wang, Y., Chua, C.-S., 2005. Face recognition from 2d and 3d images using 3d Gabor filters. Image Vis. Comput. 23 (11), 1018–1028.

[Wang06a] L. Wang, K. C. Tan, C. M. Chew, *Evolutionary robotics: from algorithms to implementations*, World Scientific Series in Robotics and Intelligent Systems, Vol. 28, Jul. 2006.

[Wang06b] J. Wang, L. Yin, X. Wei, and Y. Sun, “3D facial expression recognition based on primitive surface feature distribution,” in Proc. Conf. Computer Vision and Pattern Recognition, 2006, pp. 1399–1406.

[Wang08] C C Wang, K C Wang, “Hand Posture recognition using Adaboost with SIFT for human robot interaction”, Springer Berlin, ISSN 0170-8643, Volume 370, 2008.

[Wang11] J. Wang, P. Huang, C. Chen, W. Gu, J. Chu, "Stereovision aided navigation of an Autonomous Surface Vehicle," Advanced Computer Control (ICACC), 2011 3rd International Conference on, 18-20 Jan. 2011, pp.130-133.

[Wang98] Wang, L.C., Der, S.Z., Nasrabadi, N.M., 1998. Automatic Target Recognition Using a Feature-Decomposition and Data-Decomposition Modular Neural Network. IEEE Trans. on Image Processing, vol.7, no.8, 1113-1121.

[Webb12] J. Webb, J. Ashley, Beginning Kinect Programming with Microsoft Kinect SDK, Apress, 2012.

[Weber99] Weber, K.; Venkatesh, S. & Srinivasan, M.V. (1999). Insect Inspired Robotic Homing, Adaptive Behavior, vol. 1, pp. 65-98.

[Wei04] Wei, S.D., Lai, S.H., 2004. Robust face recognition under lighting variations. In Proc. of the 17th International Conference on Pattern Recognition, vol.1, pp.354 - 357.

[Wein90] J. Wein, S. A. Zenios, Massively parallel auction algorithms for the assignment problem, *Proc. 3rd Symp. Frontiers of Massively Parallel Computation*, Nov. 1990, pp. 90–99.

[Wichert96] Wichert, G. (1996). Selforganizing Visual Perception for Mobile Robot Navigation,

[Wiskott99] Wiskott, L., Fellous, J.M., Kruger, N., Von der Malsburg, C., 1999. Face Recognition by Elastic Bunch Graph Matching. Intelligent Biometric Techniques in Fingerprint and Face Recognition, Springer-Verlag.

[Woodward11] A. Woodward, T. Ikegami, “A Reservoir Computing approach to Image Classification using Coupled Echo State and Back-Propagation Neural Networks”, IVCNZ2011, International Conference on Image and Vision Computing New Zealand, 2011.

[Wright10] Wright, J., Yang, A.Y., Ganesh, A., Sastry, S., Ma, Y., 2010. Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Machine Intell. 31, 210–227.

[Xu09a]Xu, C., Li, S., Tan, T., Quan, L., 2009a. Automatic 3d face recognition from depth and intensity Gabor features. Pattern Recognition 42 (9), 1895–1905.

[Xu09b] Xu, Y., Yao, L., Zhang, D., Yang, J.-Y., 2009b. Improving the interest operator for face

[Yang11] J. Yang, L. Nanni (Eds.), State of the art in Biometrics, InTech, July, 2011.

[Yao93] X. Yao, A review of evolutionary artificial neural networks, *International Journal of Intelligent Systems*, Vol. 8, No. 4, 1993, pp. 539-567.

[Yin06] L. Yin, X. Wei, Y. Sun, J. Wang, M. Rosato, "A 3D facial expression database for facial behavior research," in 7th International Conference on Automatic Face and Gesture Recognition (FG2006), IEEE Computer Society TC PAMI, 2006, pp. 211-216.

[Yin08] L. Yin, X. Chen, Y. Sun, T. Worm, M. Reale, "A High-Resolution 3D Dynamic Facial Expression Database," The 8th International Conference on Automatic Face and Gesture Recognition (FGR08), 17-19 September 2008.

[Zeng06] Z. Zeng, Y. Fu, G. I. Roisman, Z. Wen, Y. Hu, and T. S. Huang, "Spontaneous Emotional Facial Expression Detection," Journal of Multimedia, vol. 1, no. 5, pp. 1-8, 2006.

[Zhang03] N. Zhang, D.C. Wunsch, Fuzzy logic in collective robotic search, *The 12th IEEE International Conference on Fuzzy Systems FUZZ '03*, Vol. 2, 25-28 May 2003, pp. 1471 – 1475.

[Zhang04] Zhang, J., Li, S., Wang, J., 2004. Manifold learning and applications in recognition.

[Zhao02] Zhao, W.Y., Chellappa, R., 2002. Image-based Face Recognition: Issues and Methods. In: Javidi, B. (Ed.), Image Recognition and Classification: Algorithms, Systems, and Applications. M. Dekker, pp. 375–402.

[Zhao03]Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A., 2003. Face recognition: A literature survey. ACM Comput. Surveys 35 (4), 399–458.

[Zhao04] Zhao, Z.Q., Huang, D.S., Sun, B.Y., 2004. Human face recognition based on multi-features using neural networks committee. Pattern Recognition Letters 25 (12), 1351-1358.

[Zhao05] W. Zhao, R. Chellappa, Face Processing: Advanced Modelling and Methods, Academic Press, 2005.

[Zhao08] Zhao, T., Liang, Z., Zhang, D., Zou, Q., 2008. Interest filter vs. interest operator: Face recognition using fisher linear discriminant based on interest filter representation. Pattern Recognition Lett. 29 (13), 1849–1857.

[Zhu06] A. Zhu, S.X. Yang, A Neural Network Approach to Dynamic Task Assignment of Multirobots, *IEEE Transactions on Neural Networks*, Vol. 17, Issue 5, Sept. 2006, pp. 1278 – 1287.

[Zinck08] A. Zinck, A. Newen, "Classifying emotion: a developmental account", Synthese, Vol.161, No.1, pp.1-25, 2008.

[Emgu13]Emgu CV, http://www.emgu.com/wiki/index.php/Main_Page

3.2 List of publications 2001 - 2014

2013

1. [Bio2013-1]Simion G., David C., Gui V., **C.-D. Căleanu**, Otesteanu M." *Fingertip-based Real Time Tracking and Gesture Recognition for Natural User Interfaces*", Advances in Electrical and Computer Engineering - in review
2. [Bio2013-2]Simion G., **C.-D. Căleanu**."*Sparse Feature for Hand Gesture Recognition: A Comparative Study*", 36th International Conference on Telecommunications and Signal Processing (TSP), July 2-4, Rome, Italy, pag. 858-861, 978-1-4799-0403-7/13/ IEEE, 2013.
3. **[Expr2013-1] C.-D. Căleanu,** "*Face Expression Recognition: a Brief Overview of the Last Decade*", 8th IEEE International Symposium on Applied Computational Intelligence and Informatics, May 23–25, pp. 157- 161, Timişoara, Romania, IEEE Catalog Number: CFP1345C-USB (pendrive proceedings); CFP1345C-PRT (printed proceedings), ISBN Number: 978-1-4673-6398-3 (pendrive); 978-1-4673-6397-6 (printed), Timişoara, Romania, 2013.
4. [Bio2013-3] Ana-Andreea Nagâţ, **C.-D. Căleanu**, "*Face Identification Using Kinect Technology*", 8th IEEE International Symposium on Applied Computational Intelligence and Informatics, May 23–25, pp. 169- 172, IEEE Catalog Number: CFP1345C-USB (pendrive proceedings); CFP1345C-PRT (printed proceedings), ISBN Number: 978-1-4673-6398-3 (pendrive); 978-1-4673-6397-6 (printed), Timişoara, Romania, 2013.

2012

5. [Bio2012-1] G. Simion, **C.-D. Căleanu**, *"A ToF 3D Database for Hand Gesture Recognition",* Proceedings of International Symposium on Electronics and Telecommunications, ISETC'12, Timişoara, România, Noiembrie 15-16, pag. 363-366, IEEE Catalog Number: CFP1203L-PRT, ISBN: 978-1-4673-1174-8 http://dx.doi.org/10.1109/ISETC.2012.6408145, 2012.
6. [Rob2012-1] R. Mîrşu, S. Micuţ, **C.-D. Căleanu**, D. B. Mîrşu, "*Optimized Simulation Framework for Spiking Neural Networks using GPU's*", Advances in Electrical and Computer Engineering, vol. 12, nr. 2, pag. 61- 68, ISSN 1582-7445, http://dx.doi.org/10.4316/AECE.2012.02011, Mai, 2012.
7. [Expr2012-1] X. Wang, X. Mao, **C.-D. Căleanu**, M. Ishizuka, "*Statistical shape analysis for face movement manifold modeling*", Optical Engineering, vol. 51, nr. 3, ISSN,0091-3286, pag. 0370040-0370048, http://dx.doi.org/10.1117/1.OE.51.3.037004, 29 Mar. 2012.

2011

8. **[FaceID2011-1] C.-D. Căleanu**, X. Mao, G. Pradel, S. Moga, Y. Xue, "*Combined pattern search optimization of feature extraction and classification parameters in facial recognition*", Pattern Recognition Letters, Volume 32, Issue 9, 1 July 2011, Pages 1250-1255, ISSN 0167-8655, DOI: http://dx.doi.org/10.1016/j.patrec.2011.03.019.

9. [Rob2011-1] R. Mîrşu, **C.-D. Căleanu**, V. Tiponuţ, "*GPU Accelerated Model for Liquid State Machine Based on spiking Neurons"*, MENDEL 2011, 17th International Conference on Soft Computing, ISSN 1803-3814, pp. 275-281, June 15-17, Brno, Czech Republic, 2011.

2010

10. [Expr2010-1] YuLi Xue, Xia Mao, **C.D. Căleanu**, ShanWei Lv., "*Layered Fuzzy Facial Expression Generation of Virtual Agent*", Chinese Journal of Electronics, ISSN: 1022-4653, Vol. 19, No.1, pag. 69-74, 2010.
11. [Expr2010-2] YuLi Xue, Xia Mao, **C.D. Căleanu**, ShanWei Lv., "*Robust Facial Expression Recognition Under Occlusion Condition*", Journal of Beijing University of Aeronautics and Astronautics, (in Chinese), vol. 4, no. 36, ISSN: 1001-5965, pag. 429-433, 2010.
12. [Expr2010-3] X.K. Wang, X. Mao, **C.D. Căleanu**,"*Nonlinear Shape-Texture Manifold Learning*", IEICE Transactions on Information and Systems, vol. E93-D, no. 7, pag. 2016-2019, DOI: http://dx.doi.org/10.1587/transinf.E93.D.2016, ONLINE ISSN : 1745-1361, PRINT ISSN : 0916-8532, Iulie 2010.
13. **[Rob2010-1] C.D. Căleanu**, X. Mao, V. Tiponuţ, Y. Xue, "*Direct Search as Unsupervised Training Algorithm for Neural Networks*", 14th WSEAS International Conference on SYSTEMS , Latest Trends on Systems, Vol. I, pag. 575-579, Corfu Island, Greece ISSN: 1792-4235, ISBN: 978-960-474-214-1, Iulie 22-24, 2010.
14. [Expr2010-4] YuLi Xue, Xia Mao, **C.D. Căleanu**, Q. Chang, "*Layered fuzzy facial expression generation based on sociality, emotion and physiology*", Pattern Recognition and Artificial Intelligence, Volume 23, Issue 5, Pages 663-670, ISSN: 1003-6059, http://mssbyrgzn.periodicals.net.cn/default.html, October 2010.
15. **[Book2010-1] C.D. Căleanu,** A. Filip, V. Tiponuţ – " *Electronic Devices and Circuits. Experiments and Simulations* ", Second Edition, ISBN 978-606-554-160-3, 195 pagini, Ed. Politehnica, Timişoara, 2010.
16. **[Book2010-2] C.D. Căleanu**, V. Tiponuţ, V. Maranescu, A. Filip – "*Electronic Devices*", ISBN 978-606-554-042-2, 180 pagini, Ed. Politehnica, 2010.

2009

17. I. Lie, V. Tiponut, I. Bogdanov, S. Ionel, **C. D. Căleanu**, "*Automated Meter Reading System for Heat Costs Allocation*", WSEAS Transactions on Circuits and Systems, Issue 2, Volume 8, ISSN 1109-2734, pag. 177-186, February 2009.
18. I. Lie, V. Tiponut, **C. D. Căleanu**, "*Rotation Detector Using FM Principles*", Recent Advances in Circuits, Proceedings of the 13th WSEAS International Conference on CIRCUITS, WSEAS International Conferences, Rodos, Greece, ISBN 978-960-474-096-3, ISSN 1790-5117, pag. 119-122, July 22-25, 2009.
19. [Rob2009-1] V. Gui, **C. D. Căleanu**, "*On the Effectiveness of Multiscale Mode Filters in Edge Preserving Image Filtering*", Recent Advances in Systems, Proceedings of the 13th WSEAS International Conference on SYSTEMS, WSEAS International Conferences, Rodos, Greece, ISSN 1790-2769, ISBN 978-960-474-097-0, pag. 190-195, July 22-25, 2009.

20. [Rob2009-2] F. Alexa, V. Gui**, C. D. Căleanu**, C. David, "*Fast contur extraction*", Proceedings of the 8th WSEAS International Conference on CIRCUITS, SYSTEMS, ELECTRONICS, CONTROL & SIGNAL PROCESSING (CSECS '09), ISSN: 1790-5117, ISBN: 978-960-474-139-7, pag. 232-235, Puerto De La Cruz, Tenerife, Canary Islands, Spain, December 14-16, 2009.

2008

21. [Rob2008-1] S. Ionel, V. Tiponuţ, **C.D. Căleanu,** I. Lie**, "***A Unified Treatment of Deterministic and Random Transients in Electrical Circuits*", WSEAS TRANSACTIONS on CIRCUITS and SYSTEMS, Vol. 7, Issue 3, ISSN: 1109-2734, pag. 87-95, Martie 2008.
22. I. Lie, V. Tiponuţ, I. Bogdanov, S. Ionel**, C. D. Căleanu**, "*A Low Cost CPLD-Based Ultrasonic Flowmeter*", WSEAS TRANSACTIONS on CIRCUITS and SYSTEMS, Vol. 7, Issue 3, ISSN: 1109-2734, pag. 102-108, Martie 2008.
23. [Rob2008-2] V. Tiponuţ, S. Popescu, I. Bogdanov, **C.D. Căleanu**, "*Obstacles Detection System For Visually Impaired Guidance*", WSEAS International Conference on SYSTEMS, Heraklion, Crete Island, Greece, ISBN: 978-960-6766-83-1, ISSN: 1790-2769, pag.350-354, Iulie 22-24, 2008.
24. I. Lie, V. Tiponuţ, I. Bogdanov, S. Ionel, **C.D. Căleanu**, "*Data Acquisition System for Heat Costs Allocation*", Proceedings of the 12th WSEAS International Conference on CIRCUITS, Heraklion, Crete Island, Greece, ISBN: 978-960-6766-82-4, ISSN: 1790-5117, pag. 163-168, Iulie 22-24, 2008.
25. **[Rob2008-3] C. D. Căleanu**, V. Tiponuţ, I. Bogdanov, I. Lie, "*Emergent Behaviour Evolution in Collective Autonomous Mobile Robots*", WSEAS International Conference on SYSTEMS, Heraklion, Crete Island, Greece, ISBN 978-960-6766-83-1, ISSN 1790-2769, pag. 428-433, Iulie 22-24, 2008.
26. [Rob2008-4] F. Alexa, V. Gui, **C. D. Căleanu**, C. Botoca, "*Lossless Data Compression Using Neural Networks*", Proceedings of the 7th WSEAS International Conference on CIRCUITS, SYSTEMS, ELECTRONICS, CONTROL and SIGNAL PROCESSING (CSECS'08), Puerto De La Cruz, Tenerife, Canary Islands, Spain, ISSN: 1790-5117, ISBN: 978-960-474-035-2, pag. 128-132, Dec. 15-17, 2008.

2007

27. **[FaceID2007-1] C.D. Căleanu**, V. Gui, F. Alexa, "*Direct Search Optimized Feature Extraction*", WSEAS Transactions on Systems and Control, Issue 2, Vol.1, pag. 113 – 120, ISSN 1991-8763, Dec., 2006.
28. [FaceID2007-2] V. Gui, F. Alexa, **C. D. Căleanu**, D. Fuiorea, "*Motion Segmentation and Analysis in Video Sequences*", WSEAS Transactions on Circuits and Systems, Issue 1, Vol. 6, pag. 142-148, ISSN 1109-2734, Ian., 2007.
29. **[FaceID2007-3] C.D. Căleanu**, D.S. Huang, V. Gui, V. Tiponuţ, V. Maranescu, "*Interest Operator vs. Gabor Filtering for Facial Imagery Classification*", Pattern Recognition Letters, Vol. 28, nr. 8, pag. 950-956, Elsevier, ISSN 0167-8655, 2007, DOI: http://dx.doi.org/10.1016/j.patrec.2006.12.013.
30. **[FaceID2007-4] C.D. Căleanu**, C. Botoca – "*C# Solutions for a Face Detection and Recognition System*", FACTA UNIVERSITATIS, Ser. Elec. Energ., Vol. 20, no. 1, ISSN 0353-3670, pag. 93-105, Nis, Serbia, 2007.

31. **[Rob2007-1] C.D. Căleanu**, V. Tiponuţ, I. Bogdanov, S. Ionel, I. Lie, "*C# and .NET Framework for uC communication protocol implementation*", Proceedings of the 11th WSEAS International Conference on COMPUTERS, pag. 583-586, Agios Nikolaos, Crete Island, Greece, ISSN: 1790-5117, ISBN: 978-960-8457-92-8, July 26-28, 2007.
32. [Rob2007-2] S. Ionel, V. Tiponuţ, **C.D. Căleanu**, I. Lie, "*Continuity Relations for Random Transients in Electrical Circuits*", Proceedings of the 11th WSEAS International Conference on CIRCUITS, pag. 147-151, Crete Island, Greece, ISSN: 1790-5117, ISBN: 978-960-8457-89-8, July 23-25, 2007.
33. [Rob2007-3] V. Tiponuţ, S. Ionel, **C.D. Căleanu**, I. Lie, "*Improved Version of an Integrated Environment for Assisted Movement of Visually Impaired*", Proceedings of the 11th WSEAS International Conference on SYSTEMS, pag. 87-91, Agios Nikolaos, Crete Island, Greece, ISSN: 1790-5117, ISBN: 978-960-8457-89-8, July 23-25, 2007.
34. I. Lie, V. Tiponuţ, I. Bogdanov, S. Ionel, **C.D. Căleanu**, "*The Development of CPLD-Based Ultrasonic Flowmeter*", Proceedings of the 11th WSEAS International Conference on CIRCUITS, pag. 190- 193, Crete Island, Greece, ISSN: 1790-5117, ISBN: 978-960-8457-89-8, Iulie 23-25, 2007.
35. [Rob2007-4] V. Gui, F. Alexa, **C.D. Căleanu**, "*Barley Seeds Classification with Genetically Optimized Kernel Density Estimator*", Proceedings of the 6th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics (CIMMACS '07), pag. 131-134, Puerto De La Cruz, Tenerife, Canary Islands, Spain, ISBN: 978-960-6766-21-3, ISSN: 1790-5117, Decembrie 14-16, 2007.
36. **[FaceID2007-5] C. D. Căleanu**, F. Alexa, M. Dăneţi, "*Modalităţi de implementare pentru un sistem de detecţie şi recunoaştere facială*", Studia Univ. "Vasile Goldiş" nr. 17, Ser. Şt. Exacte, pag. 70-76, ISSN: 1584 - 2355, „Vasile Goldis" University Press, Arad, 2007.
37. [Book2007-1] G. Pradel, **C. D. Căleanu** – "Symbolic trajectory description in mobile robotics" în Sascha Kolski (editor), "Mobile Robots, Perception & Navigation", ISBN 3-86611-283-1, Pro Literatur Verlag, Germany, March 2007.

2006

38. [Rob2006-1] G. Pradel, I. Bogdanov, **C.-D. Căleanu**, "*Methods for Salient Frescoes Selection Based Mobile Robots Navigation*", Proceedings of the 2006 IEEE International Conference on Control Applications, Munich, Germany, October 4-6, pag. 3259 – 3264, IEEE Catalog nr. 06CH37824C, ISBN 0-7803-9796-7, 2006.
39. [Rob2006-2] V. Tiponut, A. Gacsady, **C.-D. Căleanu**, I. Gavriluţ, "*Neural Network Guided Robot Collectivity – An Experimental Setup*", 7th WSEAS International Conference on NEURAL NETWORKS (NN'06), pag. 41 – 46, Cavtat, Croatia, Iunie 12-14, 2006.
40. [Rob2006-3] V. Tiponut, I. Gavrilut, **C. D. Căleanu** , A. Gacsadi, "*Development of a Neural Network Guided Mobile Robot Collectivity*", WSEAS TRANSACTIONS on CIRCUITS and SYSTEMS Issue 6, Volume 5, ISSN 1109-2734, pag. 805-812, June 2006.
41. V. Maranescu, Pleşu N., **C. D. Căleanu**, C. Toma, "*Accurate Modeling of Carbon-Polymer Composites Thermal Path in High Power Density Electronic Devices*", Annals of West University of Timişoara – Series of Chemistry, 15(1), pag. 65 – 74, ISSN 1224-9513, 2006.

42. V. Maranescu, N. Pleşu, C. Toma, **C.D. Căleanu**, M. Tănase, "*Nonlinear Thermal Impedance Electrical Model for Polymers Used in Electronic Industry*", Chem. Abs. Materiale plastice, Vol. 43, Nr.2, MPLAAM 43(2), pag. 165-170, ISSN 0025-5289, 2006.
43. [Rob2006-4] V. Maranescu, **C. D. Căleanu**, C. Toma, "*High input, wide output voltage range linear regulators*", Scientific Bulletin of the "POLITEHICA" University of Timişoara, Transactions on Electronics and Communications, Tomul 51 (65), Fasc. 1, pag. 205-208, ISSN 1583-3380, 2006.
44. **[FaceID2006-1] C.D. Căleanu**, V. Gui, F. Alexa, "*Face Recognition via Direct Search Optimized Gabor Filters*", The 5th WSEAS International Conference on System Science and Simulation in Engineering (ICOSSE'06), Tenerife, Canary Islands, Spain, December 16-18, pag. 209-214, ISBN 960-8457-57-2, 2006.
45. [Bio2006-1]V. Gui, F. Alexa, **C.D. Căleanu**, D. Fuiorea, "*Robust 2D Moving Object Segmentation and Tracking in Video Sequences*", The 5th WSEAS International Conference on System Science and Simulation in Engineering, (ICOSSE'06) Tenerife, Canary Islands, Spain, December 16-18, pag.232-236, ISBN 960-8457-57-2, 2006.

2005

46. V.-I. Maranescu, **C.D. Căleanu**, M. Mocofan – "*Stimulus Generator for Spice Simulations Using Graphical Programming Language*", Analele Universităţii "Eftimie Murgu" Reşiţa, An XII, nr.1. pag. 200-204, ISSN 1453-7394, 2005.

2004

47. **[Bio2004-1] C.D. Căleanu**, M. Mocofan, V. Maranescu - "*Benchmarking Feedforward Neural Networks Training Algorithms*", ACTA TEHNICA NAPOCENSIS. ELECTRONICS AND TELECOMMUNICATIONS, Nr. 2, Vol. 45, pag. 18 - 23, ISSN 1221-6542, 2004.
48. [Expr2004-1] M. Mocofan, **C.D. Căleanu**, V. Maranescu - "*Tree content based search algorithm for an image databases*", ACTA TEHNICA NAPOCENSIS. ELECTRONICS AND TELECOMMUNICATIONS, Nr. 2, Vol. 45, pag. 24 - 29, ISSN 1221-6542, 2004.

2003

49. V.Tiponuţ, **C.D. Căleanu**, S.Tiponuţ, C.Gal – "*A Distance Lab Experiment in Electronic Devices and Circuits Discipline Study*", Proceedings of International Scientific Conference *microCAD 2003*, University of Miskolc, ISBN 963 661 547 0, pag.129-134, 6-7 Martie, Hungary, 2003.
50. [Rob2003-1] P. Hoppenot, G. Pradel, **C.D. Căleanu**, Nicolas Perrin, Vincent Sommeilly – "*Towards a symbolic representation of an indoor environment*", IEEE-SEE-CESA2003 - Computing Engineering in Systems Applications, CD ROM paper no. S1-R-00-0048, ISBN 2-9512309-5-8, 9-11 Iulie, Lille, Franţa, 2003.
51. **[Book2003-1] C.D. Căleanu** – "*Electronic Devices and Circuits. Experiments and Simulations*", ISBN 973-625-086-5, "Politehnica" Publishing House, Timişoara, 2003.

2002

52. A. Popa, **C.D. Căleanu** – "*Predicţia coeficientului de variaţie la fineţe a firelor tip lână pieptănată cu ajutorul reţelelor neuronale artificiale*", Analele Universităţii din Oradea, Fascicula Textile-Pielărie, Sesiunea de comunicări ştiinţifice, ISSN 1582-5590, pag.87-90, Oradea, 29-31 Mai, 2002.
53. A. Popa, **C.D. Căleanu** – "*Reţele neuronale utilizate pentru predicţia caracteristicilor firelor tip lână pieptănată*", Proceedings of the Scientific Communications Meeting, Analele Universităţii Aurel Vlaicu, Seria Textile, Fascicola Tehnologii Textile, ISSN1582-3431, pag. 93-96, 2002.
54. **C.D. Căleanu**, C. Gal, V. Tiponuţ – "*A proposal regarding "Electronic Devices and Circuits" discipline distance learning study*", Buletinul Ştiinţific al Universităţii "POLITEHNICA" din Timişoara, Proceedings of the Symposium on Electronics and Telecommunications, Etc.2002, Tom 47(61), Fascicola 1,2 vol.I, ISSN1224-6034, Timişoara, pag.57-60, Sept.19-20, 2002.
55. **[Expr2003-1] C.D. Căleanu**, M. Mocofan – "*Neural Network based Algorithm for Content Multimedia Databases Searching*", Buletinul Institutului Politehnic Iaşi, Tomul XL VIII (LII), Fasc. 5A, Seria Electrotehnică, Energetică, Electronică, ISSN 0258-9109, pag.69-74, 2002.
56. **[Rob2003-1] C.D Căleanu** – "*Fuzzy versus Neural Techniques for Prediction*", Proceedings of the International Conference COMMUNICATIONS 2002, Military Technical Academy, "POLITEHNICA" University of BUCHAREST and IEEE Romanian Section, ISBN 973-8290-67-8, pag. 288-293, 5 – 7 Decembrie, Bucureşti, 2002.
57. **[FaceID2003-1] C.D Căleanu** – "*Interest Operator and Neural Networks for Facial Detection*", Proceedings of the International Conference COMMUNICATIONS 2002, Military Technical Academy, "POLITEHNICA" University of BUCHAREST and IEEE Romanian Section, ISBN 973-8290-67-8, pag. 294-297, 5 – 7 Decembrie, Bucureşti, 2002.
58. [Book2002-1] V. Tiponuţ, **C.D. Căleanu** – "*Neural Networks. Architectures and algorithms*", ISBN 973-9389-66-X, "Politehnica" Publishing House, Timişoara, 2002.

2001

59. **[Book2001-1] C.D. Căleanu**, V. Tiponuţ – "*Neural Networks. Aplications*", ISBN 973-9389-67-8, "Politehnica" Publishing House, 2001.

3.3 List of grants - selection

[GCD2013]
Project titles: “Generic Central Display”
Participant as: Contract director
Funding Institution: Continental Automotive
Contract No.: 13/24.02.2012
Value: 33889 lei (2012: 32491 lei, 2013: 1398 lei)

[EXPR2011]
Poject title: New Methods for Facial Expression Analisys and Recognition
Participant as: Grant director
Program: PN II, IDEAS Exploratory Research, code 945/2008.
Funding institution: Executive Unit for Financing High Education and University Research (UEFISCSU)
Contract No.: 599/19.01.2009
Value: 277416,15 lei (2009: 102416,15 lei, 2010: 0 lei, 2011: 175000 lei)

[VIDEO2011]
Project title: Statistic and Semantic Modelling in Video Sequences Analisys
Participant as: Team member
Program: PN II, IDEAS Exploratory Research, code 931/2008.
Funding institution: Executive Unit for Financing High Education and University Research (UEFISCSU)
Contract No.: 651/19.01.2009
Value: 2009: 89285,87 lei; 2010: 150000 lei; 2011: 186479 lei.

[EMO2010]
Project title: “Research on Emotional Facial Expression recognition in Complicated Environment”
Participant as: Grant director
Program: PN II, CAPACITIES, Module III, Bilateral Research Projects, România-China, 39-5/2008.
Funding Institution: National Authority for Scientific Research (ANCS)
Contract No.: 222/15.04.2009
Value: 29 304,45 lei lei (2009: 10549,43 lei, 2010: 18755,02 lei)

[RENEW2008]
Project title: Research Institute for Renewable Energy
Participant as: Team member
Program: MedCt 129/14.03.2008
Value: 1841350 lei

[NEURAL2006]
Project title: C#/.NET Implementation for a Facial Detection and Recognition Neural System.
Participant as: Grant director.
Program: CNCSIS AT MEdC, AT41, 58 GR/19.05.2006.

Value: 200 mil. ROL
Funding institution: CNCSIS - Education Minister.

[AI2005]
Project title: Artificial Intelligence in building a face detection and recognition system.
Participant as: Grant director.
Program: CNCSIS AT MEdC, AT69, 27688/14.03.2005.
Value: 100 mil. ROL
Funding institution: CNCSIS - Education Minister.

[SIARAS2005]
Project title: SIARAS, SKILL-BASED INSPECTION AND ASSEMBLY FOR RECONFIGURABLE AUTOMATION SYSTEMS
Participant as: Team member
Program: EU Sixth Framework Programe Priority, FP6- 017146, 2005.
Value: 43,567 Euro
Funding institution: European Community.

[TRACK2005]
Project title: Object tracking estimation in video sequences
Participant as: Team member
Program: Grant CNCSIS MEdC, A type, cod 600, no. 27688/14.03.2005.
Value: 43,000. ROL
Funding institution: CNCSIS - Education Minister.

[ROBOTS2004]
Project title: The mobile robots navigation trough symbolic representation of the environment.
Participant as: Grant director.
Program: CNCSIS AT MEdC 32940/22 June 2004.
Value: 65 mil ROL
Funding institution: CNCSIS - Education Minister.

[SYMBOLIC2003]
Project title: Environment Symbolic Representation for Mobile Robots Navigation
Participant as: Team member
Program: Framework co-operation between “POLITEHNICA” University Timisoara and - Laboratoire Systemes Complexes, Universite D`Evry Val D`Essonne, France
Value: 55000 Euro
Funding institution: Education Minister, France.

[AI2003]
Project title: Artificial Intelligence-based Universal Prediction Kernel
Participant as: Grant director
Program: Romanian Academy, 2003.
Value: 35 mil. ROL
Funding institution: Romanian Academy.