Human Centered Interfaces for Assisted Living

Anastasios Tefas and Ioannis Pitas

Abstract Assisted living has a particular social importance in most developed societies, due to the increased life expectancy of the general population and the ensuing ageing problems. It has also importance for the provision of improved home care in cases of disabled persons or persons suffering from certain diseases that have high social impact. This paper is primarily focused on the description of the human centered interface specifications, research and implementations for systems geared towards the well-being of aged people. Two tasks will be investigated in more detail: a) nutrition support to prevent undernourishment/malnutrition and dehydration, and b) affective interfaces that can help assessing the emotional status of the elderly. Such interfaces can be supported by ambient intelligence and robotic technologies.

Key words: assisted living, automatic nutrition support, activity recognition, facial expression recognition

1 Introduction

In the last years the need for developing efficient approaches for nutrition support and well-being based on computer vision techniques has been increased. The objective of these methods is to help older persons that are in the last stages of their independent living period (e.g., people with early dementia), trying to prolong their independent living period. To this end, human centered interfaces and methods should be developed that follow an anthropocentric approach that monitors certain activities of the older persons and their behaviour in a smart home environment. These activities are considered to be related to nutrition and well-being of the older persons and

The authors are with the Department of Informatics, Aristotle University of Thessaloniki, Box 451, 54124 Thessaloniki, Greece, e-mail: {tefas,pitas}@aiia.csd.auth.gr

can be focused to the eating/drinking activity and facial expression recognition. Such interfaces can be supported by ambient intelligence and robotic technologies.

We consider as target group for our study, older persons that are in the early stages of dementia and suffer by mild memory loss. Two serious problems that the patients with early dementia face are underfeeding and dehydration. This is due to several reasons such as nerve deterioration, loss of sense of smell, apraxia (loss of the ability or will to execute or carry out learned purposeful movements), agnosia (loss of ability to recognize objects, persons, sounds, shapes, or smells), etc. A nutrition support system may be developed in order to help the older persons with early dementia. Such a system can be focused to monitor specific regions of the smart home and for pre-specified time intervals in order to respect the privacy of the older persons. We consider monitoring of the dining table where the older person uses for the daily lunches. Such a nutrition support system should have the following functionalities:

- Person appearance detection sitting on a chair in front of the eating table, in order to start monitoring.
- Face detection and/or hand detection.
- Start of eating/drinking activity detection/recognition.
- End of eating/drinking activity detection/recognition in order to measure the duration of the eating/drinking activity.
- Discrimination between eating/drinking and not eating/drinking (e.g., reading) activity.
- Analysis of the eating/drinking activity during the day.

If the monitoring system detects that the older person has not eaten/drank anything in specific time intervals (i.e., lunch time), a robotic unit may be instructed to prompt stimuli that will remind or even encourage the older person to eat and/or to drink something.

Additionally, solutions for visual monitoring of the status of the older persons using emotional status recognition (e.g., facial expressions that denote required attention by the corresponding system) have been proposed for socially intelligent robots. It is obvious that as the dementia becomes more severe, the percentage of abnormal facial expressions may increase or the older person may exhibit apathy. A facial expression recognition system may be used in order to trigger either alarms of severe deterioration of the well being or to use the robotic unit for providing more affective stimuli to the older person or prompting special exercises designed by psychologists or other dementia experts. A Cognitive Games scenario can be designed by experts (psychologists/gerontologtists/doctors). Facial expression recognition can be performed in the start or during the game. The structure or the schedule of the game can be readjusted depending on user's affective reactions to the cognitive stimuli of the game. The same module may be used in all kind of

interaction between the robot and the older person in order to give much better companionable functionalities to the robot.

The expression analysis tool should have the ability to:

- Detect a face in the camera video stream.
- Recognize if it is frontal or not.
- Classify the recognized expression to predefined classes.

First results on facial expression recognition in real users have shown that the facial expressions are rather person-dependent and that generic subspace methods cannot solve efficiently the generic problem. Indeed, the results on facial expression using different databases for training and testing indicated a dramatic drop in performance. First approaches that enhance the performance of facial expression recognition algorithms include enrichment of the training database with geometrically distorted training samples. Moreover, the performance is radically improved if the test person is included in the training dataset. That is, person specific algorithms are more appropriate for the expression recognition task [8].

2 Eating/Drinking Activity Recognition

The main objective is to develop and use up-to-date technology to support independent living of older persons as long as possible in their own homes. A system that automatically recognizes eating and drinking activity, using video processing techniques, would greatly contribute to prolonging independent living of older persons in a non-invasive way. For this purpose, video processing techniques have been devised, which are related to the nutrition support use case scenarios. These methods are based either on primitive human body configurations, the so-called dynemes, or primitive action sequences, the so-called action volumes, and utilize Artificial Neural Networks (ANNs), Fuzzy Vector Quantization (FVQ), Linear Discriminant Analysis (LDA) techniques. Several pre-processing steps are needed in these methods.

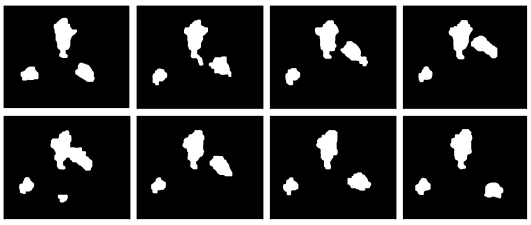

Fig. 1 Sequence of frames depicting eating activity of an older person .

2.1 Preprocessing

For the face detection task, an algorithm that uses Haar-like features has been applied. A cascade of classifiers is employed in order to perform face detection. This approach was firstly introduced in [13]. A simple tracker based on face's previous positions is used to follow the possible movement (transposition) of the face and to ensure that the location of the face is known for each frame. The skin region can be detected using analysis of the HSV histogram inside the facial region. Additionally, support vector machines can be used to learn the person specific skin regions.

Another pre-processing method developed to extract the person's ROIs required for the classification process (binary masks) is background removal (background subtraction). Both static and dynamic background extraction approaches were performed. Optimal results obtained by using the static case, namely the subtraction of the first frame of the video. Morphological operators were then utilized in order to optimize the final result of the mask. The output of the function is a single binary object. An example of a pre-processed video sequence is shown in Fig. 1.

2.2 Activity recognition based on dynemes and fuzzy distances

Eating/drinking activity recognition has been performed by using a method that is based on Fuzzy Vector Quantization (FVQ) and Linear Discriminant Analysis (LDA). Elementary action videos were preprocessed in order to produce binary posture masks of fixed size. In the training phase, the training binary posture masks were clustered in a number of clusters using a Fuzzy C-Means (FCM) algorithm and basic action units, the dynemes, were obtained. Fuzzy distances between all the binary posture masks corresponding to elementary action videos were used to represent them in the dyneme space. LDA was exploited in order to specify an optimal subspace in which action representations of different action classes are linearly separable. Dyneme representations of elementary action videos were mapped to this space and discriminant action representations were obtained. In the classification phase, the unknown elementary action video was classified using a Nearest Centroid (NC) algorithm. More details can be found in [3]. The action recognition method for the continuous recognition problem is shown in Fig. 2.

Resistance to video observation by the older persons that have privacy concerns, should be taken into consideration by any developed algorithm. That is, Privacy Preserving Technologies should be used in a certain extent. Additionally, an on/off functionality should be foreseen for the use of video cameras. Thus, the user will be fully responsible for using the camera if he wants to. Alternative user interfaces using touch screens, or voice responses

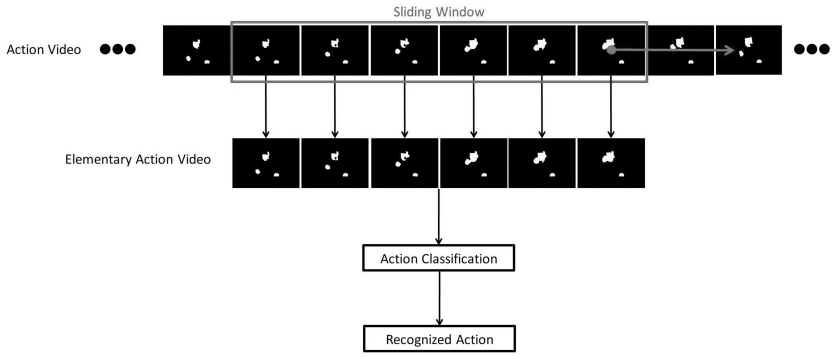

Fig. 2 Description of the action recognition algorithm applied to continuous action videos.

may be forseen. In our case, the visual representations used by the described technologies are not using the visual information as it comes from the camera but instead they transform it in what we call privacy preserving representations. For example Fig. 1 presents an eating activity as it is used internally before recognition. It is obvious that no identity information is kept. Moreover, these representations are only used on the fly, during recognition and they are not stored or transmitted.

The performance of the described method on activity recognition for three classes (i.e., apraxia, eat, drink) has been found to be in the range 80%-90%. The dataset that was used contained multiple recording of different persons in several sessions (days).

3 Facial Expression Recognition based on subspace learning

Facial expressions and gestures complement verbal communication in everyday life, conveying information about emotion, mood and ideas [15]. The facial expressions play central role in an everyday conversation. Even the voice intonation present lower impact on efficient communication than the facial expressions do [10]. Consequently, a successful automatic facial expression recognition system is expected to significantly facilitate the human-computer interaction. Furthermore, it could be integrated in many technologies of this kind, bordering behavioral science and medicine, (e.g., assisted living) [11].

A transparent way of monitoring the human emotional state is by using a video camera, which automatically detects human face and captures the facial expressions. Following this approach, the data used as input to the expression analysis tool would be a video stream, namely successive luminance or color image frames. Many techniques have been proposed in the literature for im-

plementing this tool. Some of them use static images, while others work with image sequences. Furthermore, the image representations used for expression recognition are local or global ones. Local (or landmark-based) techniques employ fiducial image points or point grids (e.g., the candide model) and their deformations for facial expression recognition [5]. Global techniques use image features derived from the entire facial image region of interest (ROI) [6]. The classification techniques operating on these image representations, have been categorized into template-based, also known as appearance-based methods, (fuzzy) rule-based, ANN-based, HMM-based and Bayesian [12].

Subspace learning methods are based on principles originally used for statistical pattern recognition and have been successfully implemented in many computer vision problems, such as, facial expression classification [9]. The problem that emerges, when it comes to appearance-based methods, is that usually initial images lie on a high dimensional space. The main goal of subspace learning methods is to reduce the data dimensionality, maintaining the meaningful information.

In subspace learning techniques, the high dimensionality of the initial image space is reduced into a lower one. Several criteria have been employed in order to find the bases of the low dimensional spaces. Some of them have been defined in order to find projections that represent the data in an optimal way, without using the information about the way the data are separated to different classes, e.g., Principal Component Analysis (PCA) [4] and Non-Negative Matrix Factorization (NMF) [7]. Other criteria deal directly with the discrimination between classes, e.g., Discriminant Non-Negative Matrix Factorization (DNMF) [14], Linear Discriminant Analysis (LDA) [1], and Clustering Discriminant Analysis (CDA) [2]. Subspace learning methods are usually combined with a classifier, like k-Nearest Neighbor (KNN), Nearest Centroid (NC), Nearest Cluster Centroid (NCC) or Support Vector Machine (SVM) in order to classify the data in the new low-dimensional space.

Among the various subspace learning methods LDA is the most popular when the objective is classification. However, LDA confronts some fundamental problems such as, the small sample size problem, where the number of samples is smaller than their dimensionality. Clustering Discriminant Analysis (CDA) [2] is a subspace learning method that has been developed in order to handle cases where the data within a class are not normally distributed. For instance, a class might consist of a mixture of Gaussians. Specifically, CDA attempts to exploit the potential subclass structure of the classes of the data.

As it has mentioned, the first crucial step towards automatic facial expression recognition is face detection. The output of this procedure is a bounding box (facial region of interest, facial ROI), which is ideally placed around the facial area. The image information within this bounding box is subsequently used as input to the classification algorithm. In general, the preprocessing steps are usually not clearly described and the bounding box, used for recognition, is arbitrarily selected, implying that only small displacements of the

bounding box may occur. However, when it comes to automatic real-world applications, inaccuracies regarding the face detection are expected and a systematic preprocessing is needed.

An additional major source of inaccuracies could be attributed to the difficulty of creating a single model that could operate optimally in cases of different people. It is common-knowledge that there is a great variation in the way several facial expressions are performed by distinct persons, due to personality or cultural background variations. This fact creates difficulties in developing a generic facial expression recognition algorithm. However, there are cases, where the users are, a priori, known. For instance, in cognitive robotics for assisted living, the persons that interact with the robot are typically known, are few (in many cases just one person) and do not change over long period of time. In this case, attempting to model the way that the facial expressions are performed by the specific persons is more reasonable rather than using a generic approach.

The motivation of our work was to create a facial expression recognition system that would be fast and would operate in realistic assisted living environments involving few persons (e.g., one elderly person living independently). The solution we followed was the one based on subspace techniques. The sensitivity of subspace learning methods when the registration of the facial ROI prior to recognition fails, even slightly ($\approx 6\%$ on the distance between the eyes) has been discussed in [9, 8]. Thus we have proposed a training set enrichment approach for improving significantly the performance of subspace learning techniques in the facial expression recognition problem. The contribution of enriching the training set with images of a tested person, in order to create person specific recognizers, thus, improving the subspace learning and the recognition performance has been also indicated.

We have performed a systematic experimental study in order to measure robustness of various subspace techniques against geometrical transformations. After the enrichment with transformed images, a clear improvement in the performance is observed in the vast majority of the cases for the enriched versions of the applied databases that ranges in the interval 4%-20%. Robustness when enriching the training set is systematically observed in our experiments. Additionally, it is observed that the more transformations are used the greater the improvement of the accuracy becomes. Moreover, it can be noted that when the facial expression recognition system is meant to be used for a specific person, very high performance can be achieved using person-dependent training.

4 Conclusions

In this paper several solutions for Human Centered Interfaces for Assisted Living have been discussed. Assisted living has a particular social importance for the provision of improved home care in cases of disabled persons or

persons suffering from certain diseases that have high social impact. Human centered interface specifications, research and implementations for systems geared towards the well-being of aged people have been presented. Two tasks have been investigated in more detail: a) nutrition support to prevent undernourishment/malnutrition and dehydration, and b) affective interfaces that can help assessing the emotional status of the elderly. Such interfaces can be supported by ambient intelligence and robotic technologies.

Acknowledgements This work has been funded by the Collaborative European Project MOBISERV FP7-248434 (http://www.mobiserv.eu), An Integrated Intelligent Home Environment for the Provision of Health, Nutrition and Mobility Services to the Elderly.

References

1. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(7), 711–720 (1997)
2. Chen, X.W., Huang, T.S.: Facial expression recognition: A clustering-based approach. Pattern Recognition Letters **24**(9-10), 1295–1302 (2003)
3. Gkalelis, N., Tefas, A., Pitas, I.: Combining fuzzy vector quantization with linear discriminant analysis for continuous human movement recognition. IEEE Transactions on Circuits and Systems for Video Technology **18**(11), 1511–1521 (2008)
4. Jolliffe, I.: Principal Component Analysis. Springer Verlag (1986)
5. Kotsia, I., Zafeiriou, S., Pitas, I.: A novel discriminant non-negative matrix factorization algorithm with applications to facial image characterization problems. IEEE Transactions on Information Forensics and Security **2**(3-2), 588–595 (2007)
6. Kyperountas, M., Tefas, A., Pitas, I.: Salient feature and reliable classifier selection for facial expression classification. Pattern Recognition **43**(3), 972–986 (2010)
7. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**, 788–791 (1999)
8. Maronidis, A., Bolis, D., Tefas, A., Pitas, I.: Improving Subspace Learning for Facial Expression Recognition Using Person Dependent and Geometrically Enriched Training Sets. Neural Networks, accepted for publication (2011)
9. Maronidis, A., Tefas, A., Pitas, I.: Frontal view recognition using spectral clustering and subspace learning methods. In: Int. Conf. Artificial Neural Networks, *Lecture Notes in Computer Science*, vol. 6352, pp. 460–469. Springer (2010)
10. Mehrabian, A.: Communication without words. Psychology Today **2**(4), 53–56 (1968)
11. Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: The state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **22**(12), 1424–1445 (2000)
12. Pantic, M., Rothkrantz, L.J.M.: Toward an affect-sensitive multimodal human-computer interaction. Proceedings of the IEEE **91**(9), 1370–1390 (2003)
13. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 511–518 (2001)
14. Zafeiriou, S., Tefas, A., Buciu, I., Pitas, I.: Exploiting discriminant information in non negative matrix factorization with application to frontal face verification. IEEE Transactions on Neural Networks **17**(3), 683–695 (2006)
15. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **31**(1), 39–58 (2009)