Knowledge-Based Systems 23 (2010) 350–356

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

On assisting a visual-facial affect recognition system with keyboard-stroke pattern information

I.-O. Stathopoulou *, E. Alepis, G.A. Tsihrintzis, M. Virvou

Department of Informatics, University of Piraeus, Piraeus 185 34, Greece

ARTICLE INFO

Article history:
Available online 22 November 2009

Keywords:
Human–computer interaction
Multimodal affective systems
Emotion recognition
Emotion perception
Pattern recognition
Facial expression

ABSTRACT

Towards realizing a multimodal affect recognition system, we are considering the advantages of assisting a visual-facial expression recognition system with keyboard-stroke pattern information. Our work is based on the assumption that the visual-facial and keyboard modalities are complementary to each other and that their combination can significantly improve the accuracy in affective user models. Specifically, we present and discuss the development and evaluation process of two corresponding affect recognition subsystems, with emphasis on the recognition of six basic emotional states, namely happiness, sadness, surprise, anger and disgust as well as the emotion-less state which we refer to as neutral. We find that emotion recognition by the visual-facial modality can be aided greatly by keyboard-stroke pattern information and the combination of the two modalities can lead to better results towards building a multimodal affect recognition system.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Recently, the recognition of emotions of users while they interact with software applications has been acknowledged as an important research topic. How people feel may play an important role on their cognitive processes as well [12]. Thus the whole issue of human-computer interaction has to take into account users' feelings. Picard [22] points out that one of the major challenges in affective computing is to try to improve the accuracy of recognizing people's emotions. Improving the accuracy of emotion recognition may imply the combination of many modalities in user interfaces. Indeed, human emotions are usually expressed in many ways. For example, as we articulate speech we usually move the head and exhibit various facial emotions [13].

There is an increasing interest within the human–computer interaction (HCI) community in designing affective engagement with interfaces [15]. This is especially the case of computer-based educational applications that are targeted to students who are in the process of learning. Learning is a complex cognitive process and it is argued that how people feel may play an important role on their cognitive processes as well [12]. A way of improving interaction and, thus, learning is recognizing the users' emotions by observing them during their engagement with the educational application and then adapting its interaction to their emotional state. Indeed, research in psychology and neurology shows that both body and mind are involved in emotional experiences [7,8,10] and emotions influence people's bodily movements [9]. Therefore, observing users may provide a system with adequate information for recognizing users' emotions. Picard [22], on the other hand, argues that people's expression of emotion is so idiosyncratic and variable, that there is little hope of accurately recognizing an individual's emotional state from the available data. Therefore, many researchers have pointed out that there is a need for combining evidence from many modes of interaction so that a computer system can generate as valid hypotheses as possible about users' emotions (e.g. [18,20]).

Towards this task, a shortage of empirical studies have appeared in the literature. Indeed, after an extensive search of the literature, we found that there is a shortage of empirical evidence concerning the strengths and weaknesses of these modalities. The most relevant research work is that of De Silva et al. [11] who performed an empirical study and reported results on human subjects' ability to recognize emotions. However, De Silva et al. focus on the audio signals of voice concentrating on the pitch and volume of voice rather than lingual keywords that convey affective information. On the other hand, in our research we have included the lingual aspect of users' spoken words on top of the pitch and volume of voice and have compared the keyboard-stroke patterns results with the results from the other two modes so that we can see which modality conveys more information for human observers.

Ideally evidence from many modes of interaction should be combined by a computer system so that it can generate as valid hypotheses as possible about users' emotions. This view has been

* Corresponding author. Tel.: +30 2107273973.
E-mail addresses: iostath@unipi.gr (I.-O. Stathopoulou), talepis@unipi.gr (E. Alepis), geoatsi@unipi.gr (G.A. Tsihrintzis), mvirvou@unipi.gr (M. Virvou).

0950-7051/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2009.11.007

supported by many researchers in the field of human–computer interaction [5,21,22]. However, progress in emotion recognition based on multiple modalities has been quite slow. Although several approaches have been proposed to recognize human emotions based on facial expressions or speech, relatively limited work has been done to fuse these two and other modalities to improve the accuracy and robustness of the emotion recognition system [3]. Specifically, in the area of unimodal emotion recognition, there have been many studies using different, but single, modalities. Facial expressions [19,24], vocal features [6,25], body movements and postures [4,2], physiological signals [23] have been used as inputs during these attempts, while multimodal emotion recognition is currently gaining ground [21,3].

Nevertheless, most of the works consider the integration of information from facial expressions and speech and there are only a few attempts to combine information from body movement and gestures in a multimodal framework. Gunes and Piccardi [14], for example, fused at different levels facial expressions and body gestures information for bimodal emotion recognition. Further, el Kaliouby and Robinson [16] proposed a vision-based computational model to infer acted mental states from head movements and facial expressions. So far the problem of emotion recognition through multiple modalities in human–computer interaction has been approached by other mathematical methods. A lot of them have been described in a comprehensive review of the field made in [17]. Such methods include rule-based systems, discriminate analysis, fuzzy rules, case-based and instance-based learning, linear and nonlinear regression, neural networks, Bayesian learning, Hidden Markov Models, Bayesian networks, etc. However, multi-criteria decision making methods have not been used yet in the problem of affect recognition through multiple modalities.

In view of the above, it is our aim to improve the accuracy of visual-facial emotion recognition by assisting it with information from other modalities. In this paper, we are considering assisting the visual-facial modality with keyboard-stroke pattern information. Currently, a system that combines two modalities, namely keyboard-stroke pattern information and audio-lingual information, has been already constructed and is described briefly in [1]. Towards combining keyboard-stroke patterns and the visual-facial modality, we had to determine the extent to which these two different modalities can provide emotion recognition independently. Moreover, we had to specify the strengths and weaknesses of each modality. In this way, we could determine the weights of the criteria that correspond to the respective modalities from the perspective of a human observer. In previous work of ours, we conducted empirical studies involving human subjects and human observers concerning the recognition of emotions from keyboard-stroke patterns and visual-facial modalities and presented the results from their combination [35]. In this paper, we present the results from combining the visual-facial modality with keyboard-stroke pattern information and discuss the advantages that derive from their combination. Specifically, in Section 2, we briefly present our facial expression recognition system, which constitutes the visual-facial modality and present evaluation results regarding its recognition accuracy. In Section 3, we present our keyboard-stroke pattern information-based emotion recognition system and evaluate its performance. In Section 4, we combine the two emotion recognition modalities. Finally we draw conclusions and point to future work, in Section 5.

2. Visual-facial affect recognition modality

2.1. Facial expression database

Since our search in the literature and World Wide Web did not result to a complete facial expression database we built our own facial expression database. The process of acquiring image data and building this database is described extensively in [32]. The final dataset consisted of 250 different persons, each forming the seven expressions: "neutral", "happy", "sad", "surprised", "angry", "disgusted" and "bored-sleepy".

2.2. Feature description

From the collected dataset, we identified differences between the "neutral" expression of a model and its deformation into other expressions. This led us to the identification of the some important facial features [32], that can represent these changes in mathematical terms, so as to form the feature vector. These facial points are widely used in facial processing systems and they can help us in the computation of the facial features which will be used as an input to the classifiers. The aim of feature extraction process is to convert pixel data into a higher-level representation of shape, motion, color, texture and spatial configuration of the face and its components. Specifically, we locate and extract the corner points of specific regions of the face, such as the eyes, the mouth and the eyebrows, and compute variations in size or orientation from the "neutral" expression to another one. Also, we extract specific regions of the face, such us the forehead or the region between the eyebrows, so as to compute variations in texture. Namely, the extracted features are:

- Mouth ratio.
- Left Eye ratio.
- Right eye ratio.
- Head ratio.
- Texture of the chin: measurement of the changes of the texture of the chin compared to "neutral" expression.
- Texture of the region between the eyebrows: measurement of the changes of the texture f the region between the eyebrows compared to "neutral" expression.
- Texture of the left cheek: measurement of the changes of the texture of the left cheek compared to "neutral" expression.
- Texture of the right cheek: measurement of the changes of the texture of the right cheek compared to "neutral" expression.
- Texture of the forehead: measurement of the changes of the texture of the forehead compared to "neutral" expression.
- Mouth orientation: measurement of the changes of the orientation of the mouth compared to "neutral" expression.
- Left brow orientation: measurement of the changes of the orientation of the left brow compared to "neutral" expression.
- Right brow orientation: measurement of the changes of the orientation of the right brow compared to "neutral" expression.

The above features form the resulting feature vector which is fed to the classifiers for training and testing as we describe in the next section. The feature extraction process and system's results are analyzed and presented for various stages of the development of our system in [26,27,30,29,28,31,33].

2.3. Neural network architecture

In order to classify facial expressions, we developed a two layer artificial neural network which is fed with the input data: (1) mouth dimension ratio, (2) mouth orientation, (3) left eye dimension ratio, (4) right eye dimension ratio, (5) measurement of the texture of the left cheek, (6) measurement of the texture of the right cheek, (7) left eye brow direction, (8) right eye brow direction, (9) face dimension ratio, (10) measurement of the texture of the forehead, (11) measurement of the texture of the region between the brows, and, (12) measurement of the texture of the chin. The network produces a 7-dimensional output vector which can be

Download English Version:

https://daneshyari.com/en/article/403087

Download Persian Version:

https://daneshyari.com/article/403087

Daneshyari.com