Journal of Visual Languages and Computing 27 (2015) 19–28

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

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

"Magic mirror in my hand, what is the sentiment in the lens?": An action unit based approach for mining sentiments from multimedia contents ☆

Luca Casaburi [a], Francesco Colace [a,*], Massimo De Santo [a], Luca Greco [b]

[a] DIEM – Department of Information Engineering, Electrical Engineering and Applied Mathematics, Università degli Studi di Salerno, Fisciano, Salerno, Italy
[b] DIIN – Department of Industrial Engineering, Università degli Studi di Salerno, Fisciano, Salerno, Italy

A R T I C L E   I N F O

Article history:
Received 19 September 2014
Received in revised form
14 January 2015
Accepted 15 January 2015
Available online 3 February 2015

Keywords:
Affective computing
Ekman theory
Emotional intelligence
Human computer interaction

A B S T R A C T

In psychology and philosophy, emotion is a subjective, conscious experience characterized primarily by psychophysiological expressions, biological reactions, and mental states. Emotion could be also considered as a "positive or negative experience" that is associated with a particular pattern of physiological activity. So, the extraction and recognition of emotions from multimedia contents is becoming one of the most challenging research topics in human–computer interaction. Facial expressions, posture, gestures, speech, emotive changes of physical parameters (e.g. body temperature, blush and changes in the tone of the voice) can reflect changes in the user's emotional state and all this kind of parameters can be detected and interpreted by a computer leading to the so-called "affective computing". In this paper an approach for the extraction of emotions from images and videos will be introduced. In particular, it involves the adoption of action units' extraction from facial expression according to the Ekman theory. The proposed approach has been tested on standard and real datasets with interesting and promising results.

1. Introduction

Affective computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects. It is an interdisciplinary field spanning computer science, psychology, and cognitive science [2].

Affective computing gets its name from the field of psychology (where "affect" is, basically, a synonym for "emotion") and could offer benefits in an almost limitless range of applications: e-learning, e-health, e-therapy, entertainment, marketing [1,4,5,11,14,3,48].

In this scenario, the face and the facial expressions can be a powerful communication channel to convey emotions and opinions [7]. A facial expression can be defined as a visible manifestation of the emotional state, cognitive activity, intention, personality and psychology of a person [1]; it has been observed that facial expressions contribute strongly to a multimedia message, more than the vocal and the verbal components [11].

In general, the detection of affective states from facial expression in multimedia contents follows two main approaches: the recognition of discrete basic affects (by the adoption of a template matching approach); the

☆ This paper has been recommended for acceptance by Shi Kho Chang.

* Corresponding author.
E-mail addresses: lcasaburi@unisa.it (L. Casaburi), fcolace@unisa.it (F. Colace), desanto@unisa.it (M. De Santo), lgreco@unisa.it (L. Greco).

http://dx.doi.org/10.1016/j.jvlc.2015.01.001
1045-926X/

recognition of affects by the inference from movement of facial muscles according to the Facial Action Coding System (FACS) [8]. FACS classifies the facial movement as Action Units (AUs) and describes the facial expressions as a combination of AUs.

The first approach requires the execution of two main steps: the face's representation through features (landmarks or filtered images) and the classification of facial expression. Many papers deal with this approach such as [18] that shows how to represent facial expressions in a space of faces. In this case the face is encoded as a landmark (58 points) and the classification is performed through a probabilistic recognition algorithm based on the manifold subspace of aligned face appearances.

Zhang et al. [21] analyze the space of facial expressions to compare two classification systems: *geometric-based* (face is encoded by a landmark) and *Gabor-based* [9] (face is encoded by Gabor features); the classification is performed with two-layer preceptor network. They show that the best results are obtained with a network of 5–7 hidden preceptors to represent the space of expression. In this way, the facial expression analysis can be performed on static images [21,23] or video sequence [22,24,6].

Cohen et al. [22] propose a new architecture of HMM to segment and recognize facial expression and affects from video flow, while Lee et al. [24] propose a method using probabilistic manifold appearance. Wang et al. [27] describe an automatic system that performs face recognition and affect recognition in grey-scale images of face by making a classification on a space of faces and facial expressions. This system can learn and recognize if a new face is in the image and which facial expression is represented among basic affects. In [25] a methodology to choose the Gabor features with the PCA method is shown and then LDA is used to identify the basic affects. Bartlett et al. [26] propose a system for facial expressions' extraction from video that chooses Gabor features with AdaBoost algorithm and then affects are classified by a SVM. Garbas et al. [31] extract features from the face through a LBP filter and choose the most representative ones using Real-AdaBoost algorithm. Finally the faces are classified as positive or negative by a binary classifier.

Although the template matching approach typically obtains the best performance in terms of accuracy when used on standard datasets, it also shows some drawbacks on the computational side. In fact, a large number of parameters are involved when encoding face as an image: much more complex algorithms need to be used, requiring higher computation times and memory allocation. There are also some issues with the classification process: more features require a longer training phase.

A different approach aims at mining affective states from facial expressions by the use of the Ekman Model [7,8]. This model can infer six emotional states: happiness, anger, sadness, disgust, fear and surprise; recently, it has also been enriched by the introduction of states such as attention [15], fatigue [16] and pain [17].

The Ekman theory can be improved by the introduction of the Facial Action Coding System (FACS). As previously said, Facial Action Coding System is a system to taxonomize human facial movements by their appearance on the face. The process of categorizing physical expressions of emotions has proven useful to psychologists and to animators. Anyway, due to subjectivity and time consumption issues, FACS is actually employed in automated systems that detect faces in videos, extract the geometrical features of the faces, and then produce temporal profiles of each facial movement [47].

The recognition of affects by the inference from movement of facial muscles according to the FACS requires three steps: feature extraction, AUs recognition and basic affect classification. Parts of the face, such as eyebrows, eyes, nose and lips, are analyzed and encoded in sets of points [28],[29] or as texture features [17],[30] to detect AUs. Valstar et al. [28] introduce a method to detect the AUs starting from eyebrows, classifying their movements as spontaneous or voluntary by the use of a Relevance Vector Machine approach. After the detection of the AUs, it classifies their affective class by the adoption of a probabilistic decision function.

The FACS approach has been adopted in the automated Facial Image System (AFA) [32] which analyzes video in real-time to detect the sentiments. In this case, the face is encoded with a 2D mask which is used to interrogate a SVM to detect the associated affect. Robinson et al. [15] have developed a system that analyzes real-time video streams to detect the presence of one of the following moods: concordant, discordant, focused, interested, thinking and unsure. The face is encoded by 24 points and the distances between these points are used as features to identify different situations (open mouth, head movements, position of the eyebrows); the expressions encoded by FACS are recognized by a chain of HMM for each possible action and the computation of the probability of each state is obtained by the use of a Bayesian Network. Anyway, most Ekman based systems in literature show limited performance when compared to template matching approach.

In this paper we propose a novel approach for analyzing facial expressions and recognizing emotion from multimedia contents. This method uses the AUs approach for recognizing basic emotions [18] and implements a new technique for extracting feature points from the face and measuring emotion. The classic prototypes have been extended introducing the concept of combinations of AUs: when a combination occurs, a bonus or a penalty is assigned to the measure of emotions. In this way, a more detailed model can be obtained. Our method has proven to obtain better results than the classic Ekman based approaches and also outperforms template matching based systems in some cases. In particular, it provides better performance not only in terms of precision but also for execution time and this makes it suitable for real time video applications.

The paper is organized as follows: the proposed approach is discussed in the next section. In Section 2 results of test on CK+ dataset [35], for image analysis and video analysis, and on MMI Facial Expression [34], eNTERFACE'05 [12] and Cam3D [19] datasets for video analysis are presented. The obtained results are discussed in last section.

2. The proposed framework

As previously said, in this paper a framework for mining the emotive states of people appearing in multimedia