A Local Approach for Negative Emotion Detection

Adel Lablack, Taner Danisman, Ioan Marius Bilasco, Chaabane Djeraba

A Local Approach for Negative Emotion Detection

Adel Lablack, Taner Danisman, Ioan Marius Bilasco, Chaabane Djeraba
Laboratoire d'Informatique Fondamentale de Lille, Université de Lille 1, France
Email: {adel.lablack, taner.danisman, marius.bilasco, chabane.djeraba}@lifl.fr

***Abstract*—Recognizing human facial expression and emotion by computer is an interesting and challenging problem. In this paper, we propose a method for recognizing negative emotions through an appropriate representation of facial features from relevant face regions displayed in video streams and still images. A measure that is sensitive to facial movements is used in predefined regions of interest to detect the negative emotions. The experimentation has been performed on a standard dataset and live video streams and has showed promising results.**

I. Introduction

The automatic recognition of facial expressions has been an active research topic since the early nineties. There have been several advances in the past few years in terms of face detection and tracking, feature extraction mechanisms and the techniques used for expression classification. It can be an important component of human-computer interactions driven by high commercial values and social interests and may be used in behavioural science and in clinical practice.

The interpretation of discrete expressions differs from one application domain to another. It is a challenging task that allows to make the system react and enhance the user experience according to the way that the user is experiencing a given situation. In a shopping context, detecting instant positive feedback (a smile) might be a sign of interest, whereas negative instant feedback (a grimace) might be a sign of repulsion. Depending on the immediate reaction, the system could present more details on the product or change the product displayed to the user. In e-learning applications, the factors of reaction are to be considered at a larger temporal scale, as the user interaction with the system or distant speaker is supposed to take place continuously over a large laps of time. In this case, the measurements of actionable moods like user interest, incomprehension, frustration can be from only one global emotion. These two basic examples are empowered by different ways of representing the emotional state of the user.

In the literature, two main approaches for representing expressions can be found: the categorical discrete expression representation introduced by Ekman [1] which uses six universal facial expressions that are anger, disgust, fear, happiness, sadness and surprise. The dimensional representation is an alternative and characterize an affective state in terms of latent dimensions rather than discrete emotion categories [2]. These dimensions include evaluation, activation, control, power, etc. In particular, the evaluation dimension measures how a human feels, from positive to negative. We focus on this latter dimension as positive and negative states are the common ground for a wide variety of applications. Positive emotions are expressed in response to situations where the user is interested and enjoys its experience [3]. Negative emotions are commonly expressed in response to situations that the person finds to be irritating, frustrating, or unpleasant.

We focus on the facial expressions that convey negative emotions. Negative emotions are interesting as they may have physical correlates such as increased heart rate, blood pressure, and levels of adrenaline and noradrenaline. However, these signals needs to be measured by intrusive methods, and in this paper we focus only on a video-based non intrusive solution.

The remainder of this paper is organized as follows. Section II reviews some related work in the field of facial expression recognition in particular the ones dealing with negative emotions. In Section III, we describe the main steps of our approach that detects negative emotion from face images. Section IV presents the experimentation and reports the obtained results. Finally, we give concluding remarks and potential future work in Section V.

II. Related Work

The area of emotion recognition has been actively studied in recent years. There is an abundance of literature concerning this topic. Typically, the purpose of these systems is to recognize emotion classes or the state of an affective dimension. We focus on approaches used to detect negative emotions which are generally highly correlated with the facial expression Anger. In order to identify the negative emotions, we have used the emotion annotation and representation language (EARL) proposed by the Human-Machine Interaction Network on Emotion (HUMAINE) [4] that classifies 48 emotions.

Little work has been specifically dedicated to negative expressions as a whole in computer vision. Most of the existing approaches are using other communication channels: speech, text, bio-sensors, etc. Automatic recognition of emotions from speech has many potential applications like in call centers, a few investigations have specifically focused on the detection of angry speech [5]. Pohjalainen and Alku [6] have used auto-regressive modulation filtering to detect automatically anger in telephone speech. The rise of social media has attracted significant interest in sentiment analysis techniques such as emotion detection and opinion mining. Roberts and al. [7] analyze Micro-blogging services such as Twitter which provide researchers with a wealth of information on how individuals communicate with their social network and reflects the author's opinions and emotional states. Anger status can be determined based on the HRV spectrum, which reflects the interaction of sympathetic and parasympathetic nervous systems. A portable system for real-time reflecting the user's anger status has been proposed by Lin and al. [8] which improve the personal health and life quality by facilitating the conscious of anger and helping to control it. Recent evidence indicates that gaze direction influences the processing of facial expressions. Angry faces are judged more angry when displaying a direct (or mutual) gaze compared to an averted gaze [9]. Masnani et al. [10] have proposed an approach for anger mood detection based on face recognition and heartbeat. It combines a pulse rate detector with a combination of image filtering techniques.

In the literature, the techniques used to detect emotion in videos can be categorized in two sets [11]: geometric-based that detect and track facial feature points which are used for classification [12], [13], and Appearance-based which consider motion and texture changes [14], [15]. Gonzalez et al. [16] apply adaboost to a set that combines geometric and appearance based features extracted in facial regions that capture the AUs activation/deactivation areas to recognize facial action units. Our work relies on a local approach to detect negative emotions from face images. Hence, in order to increase the accuracy of the developed system, we are not focusing on the precise detection of specific action units in precise points of the face, but we are proposing a region-based analysis of the face.

III. Approach Description

Our approach is inspired by the facial action coding system developed by Ekman. However, rather than focusing on the recognition of the set of action units (AUs) that describe the movements of the face, we make a local analysis to find common patterns. A global analysis of each region level is performed in order to extract hints (texture deformation, movement model) reflecting the activity of the facial muscles involved in negative expressions. We do not compute AUs, but we characterize globally the regions where AUs can be originated from. With regard to the action units that convey negative expressions, we identify overlapping facial regions where the identified action units may occur.

We propose in the following an approach that indicates the presence of a negative emotion performed by a single person sitting in front of a camera. Figure 1 shows the main steps of this approach divided into two stages:

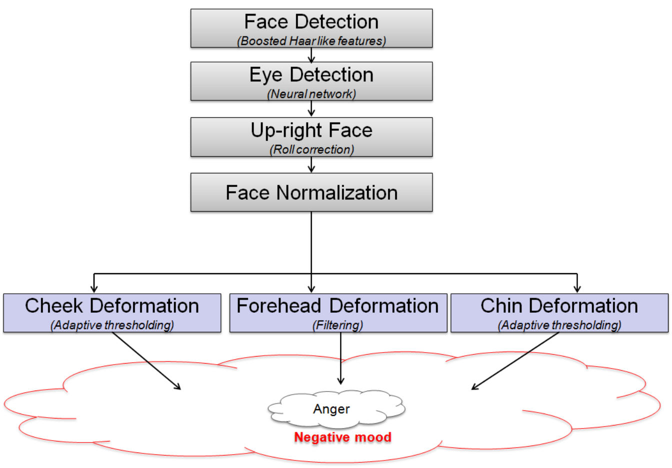

Fig. 1. Approach steps: (a) Image pre-processing stage, (b) ROI processes.

- Image pre-processing: it consists in the different processing steps that allows to extract a normalized face from the input data.
- Regions of interest processes: it consists on locating selected ROI from the face region. A specific filtering is applied to each region that indicates the presence of a negative emotion.

A. *Image pre-processing*

The image pre-processing procedure is a very important step in the facial expression recognition task. The aim of the pre-processing phase is to obtain normalized images that are uniform in size, shape and illumination, and depict only the face region.

Face detection: We use a Boosted Haar like features method. The detection of the face is performed using the Viola-Jones face detector algorithm [17] available in OpenCV library. The selected parameters achieve the best speed and performance.

Eye detection: We use a neural network based approach to locate the position of the pupils. We derive only the eye detection code from the STASM library[18] which is a variation of the implementation of the Active Shape Model proposed by Cootes. However, it works better on frontal views of upright faces with neutral expressions.

Up-right face: We estimate the orientation of the face using the vertical positions of the two eyes. If they are not in the same position we compute the angle between these two pupil points and correct the orientation by setting the face center as the origin point and we rotate the whole frame in the opposite direction. It guaranteed a frontal upright position of the face up to 30 degrees in both sideways.

Face normalization: We use histogram equalization to normalize the image intensity by improving its contrast. It aims to eliminate light and illumination related defects from the facial area.

B. *Regions of interest processes*

Most of the basic facial expressions are negative emotions and some of them could be confused. As an example, Sebe et al. [19] shows that anger is usually confused with disgust due to the fact that these expressions share common facial movements and actions. According to the classification presented in HUMAINE [4], the negative emotions have been selected and analyzed in order to extract local patterns. We identify the AUs of the negative facial expressions like Anger which is a combination of the AUs (4+5+7+23), Fear (1+2+4+5+20+26) and Disgust (9+15+16). With the knowledge of physiological relations between the facial muscles, we present an experimentally obtained region of interest (ROI) selection for the AUs depicted in Figure 2. This scheme has been constructed with respect to AU detection accuracy and its robustness to handle inter AU correlations and partial occlusions. In each ROI a texture analysis is done in order to identify a local pattern caused by a negative emotion such as the vertical lines above the nasal root.

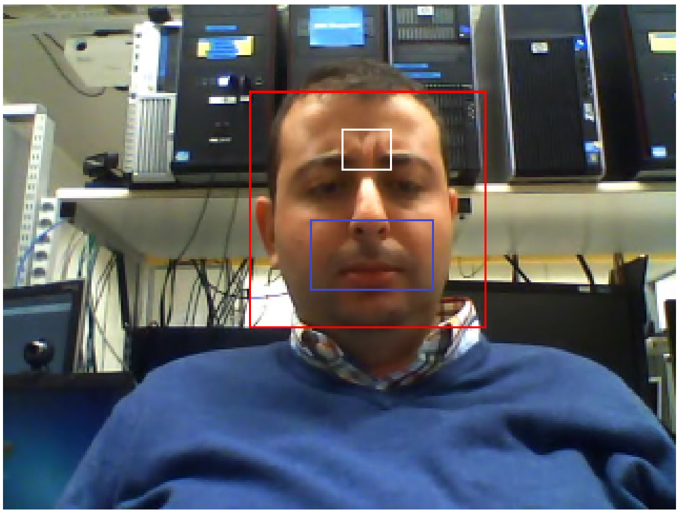

Fig. 2. Regions of Interest scheme.

1) Forehead region: This ROI is located in the upper part of the face and includes the variations of AU4 of FACS where eyebrows are lowered and drawn together. We apply Gabor filter to this region of face as shown in Figure 3. In the literature, 2D Gabor filters have been used for texture analysis and classification [20]. Gabor filters have both frequency and orientation selective properties. Therefore a 2D Gabor function is composed of a sinusoidal wave of specified radial frequency which is the spacing factor between the kernels in the frequency domain and orientation which is modulated by a 2D Gaussian. Gabor representation of a face image is computed by convolving the face image I(x, y) with the Gabor filter. Majority of AUs samples associated to negative emotion face images has vertical lines above the nasal root. So, we choose a vertical orientation for the Gabor filter with a frequency of $\sqrt{1.3}$, Mu equal to 0, Sigma equal to π and Nu equal to 3 as Gabor parameters. Then the real and imaginary responses are added together to find the magnitude response.

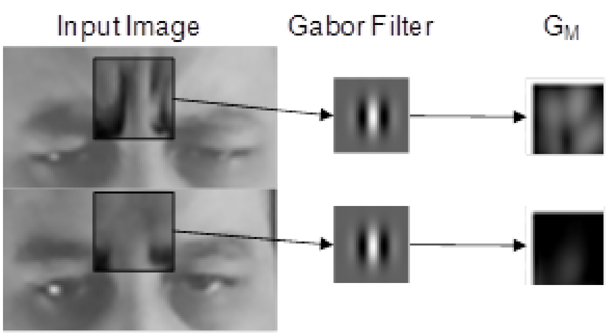

Fig. 3. Angry and neutral forehead region samples and their corresponding Gabor magnitude response.

After a binary thresholding, the sum of the total pixels in the magnitude response of the filter, just above the nasal root is examined by a threshold value to detect a negative emotion. Figure 3 shows an example of the application of Gabor filter to the nasal root for angry and neutral faces. Brighter pixels in the magnitude responses are used as an indicator of negative emotion.

2) Low face region: It corresponds to the lower part of the face region. It has been selected to check the deformation of the cheek or the chin in order to identify negative emotions. The original Local Binary Patterns (LBP) operator was introduced by Ojala et al. [21] and has been proved as a powerful means of texture description. The operator labels the pixels of an image by thresholding a 3*3 neighborhood of each pixel with the center value and considering the results as a binary number, and the 256-bin histogram of the LBP labels computed over a region is used as a texture descriptor. After labeling the low face region with the LBP operator, a histogram is constructed. It contains information about the distribution of the local patterns, such as edges, spots and flat areas, over the whole image. An adaptative thresholding is then applied in order to detect a negative emotion.

IV. Experiments and Results

We consider two types of datasets to validate our approach. It has been tested on a database that contains static images of people under different facial expressions and on a set of video streams that capture people looking toward a webcam and performing different emotions.

A. Database

We used the Karolinska Directed Emotional Faces (KDEF) database [22] which involves a set of human facial expressions of emotion to validate our measure. The set contains 70 individuals, each displaying 7 different emotional expressions (neutral, happy, angry, afraid, disgusted, sad, surprised) with each expression being captured twice. The participants were seated at a distance of approximately three meters from the camera. The absolute distance was adapted for each subject by adjusting the camera position until the subject's eyes and mouth were at specific vertical and horizontal positions on the camera's grid screen. The resolution of the images is 562x762.

We have used this dataset to validate our approach since it contains a large variety of people displaying the set of basic emotion. An emotion is correctly classified if the highest value of the measure is obtained while expressing a negative emotion. The negative emotion has been recognized successfully with a rate of 95% on this database as depicted in Figure 4.

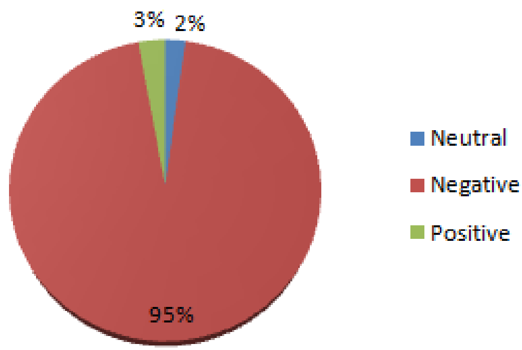

Fig. 4. The results of the negative emotion detection approach on the KDEF database.

B. Video streams

This experiment has been done in order to test the performance of the system against common environmental settings in scenarios where the user is sitting in front of a screen embedding a webcam at approximatively 50cm. The physical setting might be convenient for e-learning, video conferencing, multimedia user experiences, etc. The instructions given to the students and collaborators were to clearly express several emotions and in particular negative emotions. The screen was displaying the video captured by the webcam. Hence, the users were able to see if the expressions acted meet their own understanding of a visual representation. We made the choice to ask for the expressions to be acted and hence exaggerated as in initial tests realized on students watching for 10-minutes tutorial, negative expressions did not occurred, as students were very collaborative and, influenced by the test context and settings, showed no negative expressions. Our future plan is to extend it to more longer videos, where at some point, a complete immersion of a user will be achieved, and the user ignoring about the current constraint, will act more naturally.

The Figure 5 shows an histogram of our measure over a specific period of time for a user. The participant expressed different emotions and the peaks corresponds to negative emotions. The measure is not influenced by positive or neutral emotions where this measure has stable low values.

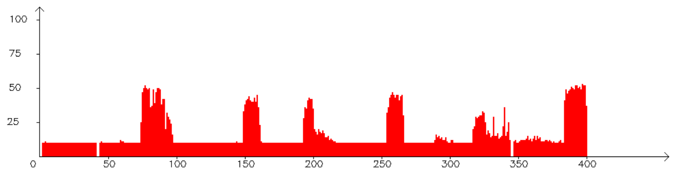

Fig. 5. The histogram of the measure that detects negative emotions over a specific period of time.

V. Conclusion

In this paper, we have proposed a method that detects the negative emotion of a person. We have defined a measure that is sensitive to common patterns in local regions of interest. The method doesn't require prior specific training processes or to recognize a set of action units and it has been applied to detect negative emotions. The results are promising in term of robustness and have been extended with the estimation of negative emotion in the context of video input. The contextualization of the system is also important towards reliable estimations. Introducing context information to optimize the system configuration allows to use the same underlying detection algorithms in different locations and in an efficient way.

Acknowledgment

This work was conducted in the context of the ITEA2 "Empathic Products" project, ITEA2 1105, and is supported by funding from the French Services, Industry and Competitivity General Direction.

References

[1] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press, 1978.

[2] R. Calvo and S. DMello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Transactions on Affective Computing*, no. 1(1), pp. 18–37, 2010.

[3] T. Danisman, I. M. Bilasco, N. Ihaddadene, and C. Djeraba, "Automatic facial feature detection for facial expression recognition," in *5th International Conference on Computer Vision Theory and Applications (VISAPP)*, 2010.

[4] "Humaine emotion annotation and representation language (earl)." [Online]. Available: http://emotion-research.net/earl

[5] W. Kim and J. H. L. Hansen, "Angry emotion detection from real-life conversational speech by leveraging content structure," in *International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010.

[6] J. Pohjalainen and P. Alku, "Automatic detection of anger in telephone speech with robust autoregressive modulation filtering," in *International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2013.

[7] K. Roberts, M. A. Roach, J. Johnson, J. Guthrie, and S. M. Harabagiu, "Empatweet: Annotating and detecting emotions on twitter," in *International Conference on Language Ressources and Evaluation*, 2012.

[8] A.-C. Lin, F.-Y. Yen, M.-H. Sun, F.-S. Shang, S.-T. Tang, and J.-H. Lin, "A real-time portable analyzer for anger emotion," in *International Conference on Electronics and Information Engineering (ICEIE)*, 2010.

[9] M. P. Ewbank, C. Jennings, and A. J. Calder, "Why are you angry with me? facial expressions of threat influence perception of gaze direction," *Journal of Vision (JOV)*, no. 9(12):16, pp. 1–7, November 2009.

[10] M. Mohamed, N. S. Jusoh, and I. L. Ahmad, "Anger mood detection based on face recognition and heartbeat," *International Journal of Engineering Research and Technology (IJERT)*, no. 2, February 2013.

[11] T. Senechal, V. Rapp, H. Salam, R. Séguier, K. Bailly, and L. Prevost, "Combining aam coefficients with lgbp histograms in the multi-kernel svm framework to detect facial action units," in *9th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2011.

[12] M. Khademi and L.-P. Morency, "Relative facial action unit detection," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2014.

[13] Z. Wang, S. Wang, and Q. Ji, "Capturing complex spatio-temporal relations among facial muscles for facial expression recognition," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[14] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. N. Metaxas, "Learning active facial patches for expression analysis," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[15] S. Wan and J. Aggarwal, "Spontaneous facial expression recognition: A robust metric learning approach," *Pattern Recognition (PR)*, no. 47(5), pp. 1859–1868, 2014.

[16] I. Gonzalez, H. Sahli, V. Enescu, and W. Verhelst, "Context-independent facial action unit recognition using shape and gabor phase information," in *4th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2011.

[17] P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[18] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *10th European Conference on Computer Vision (ECCV)*, 2008.

[19] N. Sebe, I. Cohen, A. Garg, and T. S. Huang, *Machine Learning in Computer Vision*, S. Verlag, Ed., 2005.

[20] M. Zhou and H. Wei, "Face verification using gaborwavelets and adaboost," in *18th International Conference on Pattern Recognition (ICPR)*, 2006.

[21] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition (PR)*, no. 29(1), pp. 51–59, 1996.

[22] D. Lundqvist, A. Flykt, and A. hman, *The Karolinska Directed Emotional Faces (KDEF)*, P. s. CD ROM from Department of Clinical Neuroscience, Ed. Karolinska Institutet, 1998.