Spontaneous Facial Expression Recognition Based on Histogram of Oriented Gradients Descriptor

Manar M. F. Donia[1], Aliaa A. A. Youssif[2] & Atallah Hashad[3]

[1] Modern Academy for Engineering & Technology, Computer Engineering Department, Cairo, Egypt

[2] Faculty of Computers & Information, Helwan University, Helwan, Egypt

[3] Arab Academy for Science, Technology & Maritime Transport, College of Engineering, Computer Engineering Department, Cairo, Egypt

Correspondence: Manar M. F. Donia, Modern Academy for Engineering & Technology, Computer Engineering Department, Cairo, Egypt. E-mail: manar_8884@hotmail.com

Abstract

Automatically detecting facial expressions has become an important research area. It plays a significant role in security, human-computer interaction and health-care. Yet, earlier work focuses on posed facial expression. In this paper, we propose a spontaneous facial expression recognition method based on effective feature extraction and facial expression recognition for Micro Expression analysis. In feature extraction we used histogram of oriented gradients (HOG) descriptor to extract facial expression features. Expression recognition is performed by using a Support vector machine (SVM) classifier to recognize six emotions (happiness, anger, disgust, fear, sadness and surprise). Experiments show promising results of the proposed method with recognition accuracy of 95% on static images while 80% on videos.

Keywords: Micro-Facial-Expression, emotion analysis, facial action coding system; human computer interaction, histogram of oriented gradients, face detection and tracking

1. Introduction

Decision making is a part of our day to-day life. Facial expression help humans perceive useful information; make decisions and give instant responses during social communication. It is easy to fake, and this leads to incorrect decisions. Unlike facial expressions spontaneous micro expressions are difficult to hide or fake because they are involuntary expressions shown on human faces according to emotions experienced. They are very brief, lasting only 1/25 to 1/5 of a second. Currently only highly trained individuals are able to distinguish them, but even with proper training the recognition accuracy is only 47%. Combining computer vision and behavioral sciences have large potential for developing a technology, which helps in that aspect. The major challenges are: the short duration of expressions needs a camera with high frame rates; detection of light changes in facial skin needs effective feature data extraction, representation and facial expression recognition algorithm. Facial expression analysis goes well back into the nineteenth century. Darwin published his work on expression of emotions in man and animals. He claimed that we cannot understand human emotional expression without understanding the emotional expression in animals. In 1971 Ekman and Friesen postulated six primary emotions that each has a distinctive content together with a unique facial expression. Based on the work published (Porter & Brinke, 2008; Ekman, 2009) micro expressions are the most important behavioral source of lying detection and it used for danger demeanor detection as will.

In this paper we propose a spontaneous facial expression analysis in which the face divided to specific regions, histogram oriented gradients calculated for each region to extract feature vector and apply the support vector machine to do classification. The system achieves very promising results that compare favorably with human accuracy and other methods of detection because we depend on effective feature extraction (HOG) which captures edge or gradient structure that is very characteristic of local shape, and it does so in a local representation with an easily controllable degree of invariance to local geometric and photometric transformations.

2. Related Work

Research, on spontaneous micro expressions, has progressed in two major directions psychology and computer vision.

The work done by psychologist Dr. Paul Ekman and his colleagues has become the milestone in the study of facial expression detection. After the discovery of the six universal emotions and their micro expressions

Ekman and Friesen (1978), develop a tool for measuring facial movements namely Facial Action Coding System (FACS) the most widely used method in recognizing facial expressions. FACS encodes contraction of each facial muscle (stand alone as well as combined with other muscles) in terms of action unit. Each facial expression expressed as a combination of action units.

Essa and Pentland (1995), develop two approaches (muscle and motion energy) for representing facial motion. The first approach is the peak muscle actuation template, it measure muscles actuations for many people making different expressions, and found that each expression can define a typical pattern of muscle actuation and use the result as template for recognition. The second one is a motion energy template, which generates a pattern of motion energy associated with each facial expression which represents how much movement there is at each place on the face. The overall success rate of detection was 78%.

Shreve and Godavarthy (2011), proposed a method for the automatic spotting of facial expressions in long videos comprising of macro- and micro-expressions using spatio-temporal strain where they compute motion and strain separately. They compute motion using robust optical flow but to get a reliable solution they must take in consider that intensity of a point on a moving object remains constant across a pair of frames, and pixels in a small image window move with a similar velocity; for strain computation they used finite difference method. They achieve accuracy of 85% for macro-expressions and 74% for micro-expressions detection.

3. Proposed Algorithm

We proposed a method to recognize spontaneous micro expression. The primary input is a video sequence of the subject; we expect only one face at time.

The proposed algorithm consists of three main phases as shown in Figure 1: face detection and tracking; feature extraction; facial expression classification.
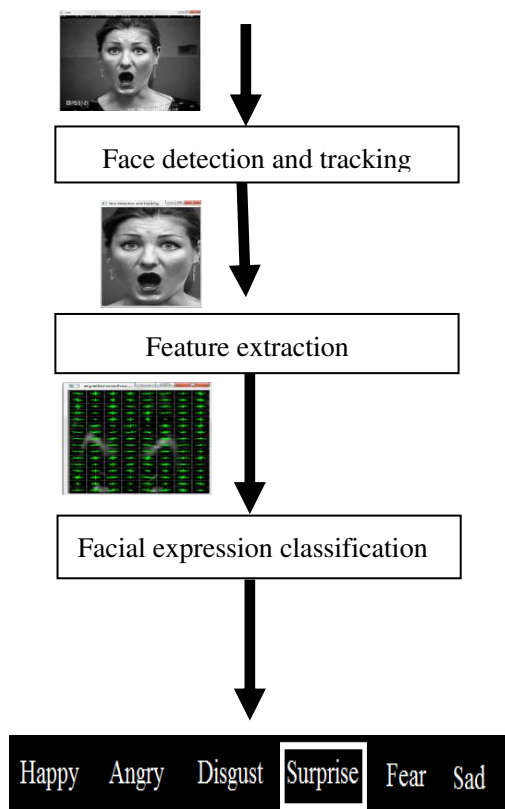

Figure 1. Block diagram of the proposed algorithm

3.1 Face Detection and Tracking

Face detection comes with its own challenges such as camera quality, illumination, facial hair, pose and rotation of the face. We detect and track the face as shown in Figure 2 by using Continuously Adaptive Mean shift (CAMshift) which based on the Mean shift algorithm. Camshift algorithm one of the promising methods of face tracking.

CAMshift algorithm consists of the following steps:

1)   Select an initial location of the search window.

2)   Create the color histogram within the search window to represent the face under test.

3)   Calculate the color probability distribution of the region centered at the Mean Shift search window.

4)   Mean shift algorithm is performed on the area until suitable convergence. The zeroth moment and centroid coordinates are computed and stored.

5)   For the next frame, center the search window at the mean location found in Step4 and set the window size to a function of the zeroth moment. Go to Step 3.

Mean shift algorithm calculates the area and the center of mass under the window using the following equations.

Zeroth moment:

$$M_{00} = \sum_x \sum_y I(x, y) \tag{1}$$

First order moment:

$$M_{01} = \sum_x \sum_y yI(x, y) \tag{2}$$

$$M_{10} = \sum_x \sum_y xI(x, y) \tag{3}$$

The center of mass (x_c, y_c) are then calculated using the following equation:

$$x_c = \frac{M10}{M00} \qquad\qquad y_c = \frac{M01}{M00} \tag{4}$$

Where $I(x,y)$ is the pixel value at position (x,y) in the image, and x and y range over the search window.

(a)   (b)

Figure 2. Face detection and tracking: (a) Example of input frame. (b) Face image

3.2 Feature Extraction

The face has many muscles which are known as "the muscles of facial expression". Unlike other muscles, the facial muscles insert directly into the skin. Contraction of the muscles causes the skin to move which forms expression. To recognize all facial muscles movement at once is a difficult task. So after the face image obtained we divide the face manually into six main regions of interest as shown in Figure 3 (eyebrows, left eye, right eye, between eyes, nose and mouth) since they are the most representative regions in terms of facial motion according to (FACS). The regions extracted by marking 10 points in the first frame of the video then the locations of the

facial regions calculated. After extracting facial regions we apply the histogram orientation gradient (HOG) descriptor on each region to extract features.
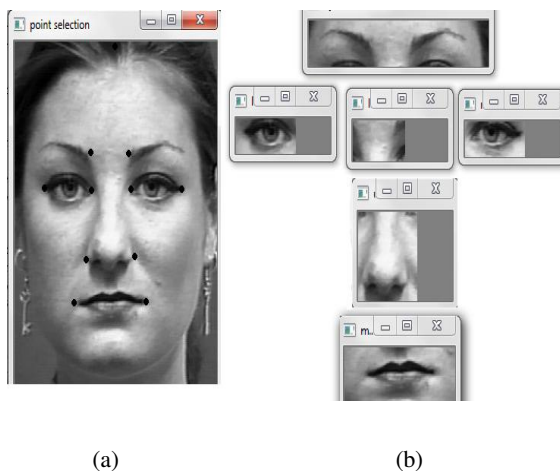

(a)   (b)

Figure 3. Facial region extraction (a) Marked points (b) Facial regions

HOG algorithm consists of three steps:

1)   Gradient Computation.

2)   Orientation Binning.

3)   Block Normalization.

3.2.1 Gradient Computation

HOG is a feature descriptor based on gradient orientation or edge direction so we apply sobel filter which computes the gradient magnitude and orientation to detect edges in the facial region image as shown in Figure 4 (b).

The gradient magnitude and orientation of each pixel calculated using Equations (5, 6):

$$m(x,y) = \sqrt{d_x(x,y)^2 + d_y(x,y)^2} \tag{5}$$

$$\theta = arctan\left(\frac{d_y(x,y)}{d_x(x,y)}\right) \tag{6}$$

3.2.2 Orientation Binning

The facial region image divided into a no of cells, and for each cell a histogram of gradient directions for the pixels within the cell compiled. Each pixel casts a weighted vote for an orientation based histogram channel based on the values found in the gradient computation. The histogram channels are evenly spread over 0 to 180 degrees as shown in Figure 4 ©.

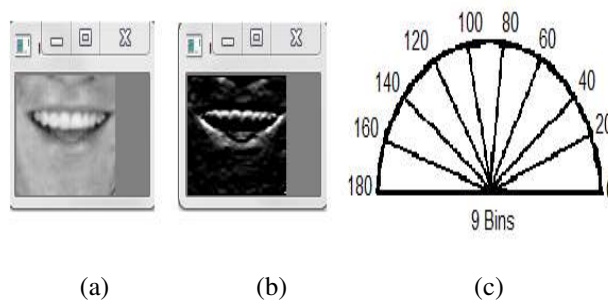(a)   (b)   (c)

Figure 4. Feature extraction (a) Extracted mouth region (b) Magnitude value (c) Orientation histogram bins

3.2.3 Block Normalization

Cells grouped together into larger blocks as shown in Figure 5. These blocks overlapped, which means that each cell contributes more than once to the final descriptor. The local histograms are contrast normalized. This normalization results in better invariance to changes in illumination or shadowing.

The normalization process is done by Equation 7 as in:

$$v \rightarrow \sqrt{v/(\|v\|_1 + \epsilon)} \tag{7}$$

Let v be the un-normalized descriptor vector and ϵ a small normalization constant to avoid division by zero.
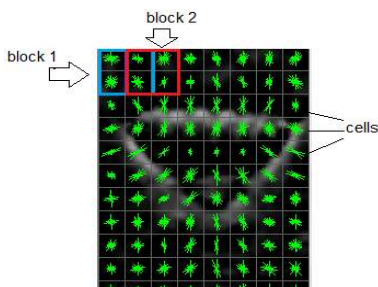

Figure 5. Overlapped blocks with HOG visualization

3.3 Facial Expression Classification

The classification based on the supervised learning technique. We have used a set of training images of different expressions. After we apply our algorithm on training images we concatenated descriptor values in one vector and feed it to SVM with n-class classification and a linear kernel.

4. Experimental Results and Dissections

We used two sets of dataset. One is well-known The Cohn Kanade dataset which used in many researches and become a golden standard dataset, the other one collected from the internet randomly.

4.1 Datasets

Our database comprises of two datasets.

4.1.1 The Cohn Kanade Dataset

It is one of the famous public datasets. It used for research in automatic facial image analysis and synthesis and for perceptual studies. It consists of 486 FACS-coded sequences from 97 subjects. Subjects range in age from 18 to 30 years 65% was female; 15% were African-American and 20% Asian or Latin. Cohn Kanade is available in two versions CK and CK+. In our research we depend on version 2 CK+ which augmented the dataset further to include 593 sequences from 123 subjects. It includes both posed and non-posed (spontaneous) expressions, which is fully FACS coded.

4.1.2 Other Videos

Dr. Paul Ekman one of the famous researchers in micro-expression has mentioned several examples of micro-expressions. We have obtained some of those from the internet.

We prepared a dataset from the available portion to examine our recognition system. The proposed system trained on 10 different subjects each of which performed different facial expressions. It also tested in two phases, the first phase was on posed expression of another 10 subjects performing different expressions, and the second phase of testing was on the found videos that are not included in training.

4.2 Result

A summary of the results are given in Tables (1, 2). The proposed system can classify the spontaneous facial expressions (happy, sad, fear, angry, disgust and surprise in the two phases of testing. The results prove that the proposed system could classify happy, angry, disgust, sad and surprise in maximum rates but fear in minimum rate. It's due to the shorter distance between brows, eyes, nose and mouth; during fear expression the distance between them is almost equal to that in natural face but in other expression it has significant space changes which make them more detectable than fear. The system implemented in MS VC++ 2010 with opencv2.3.1 on

core i5 windows 7 workstation.

Table 1. Experiential results of confusion matrix for facial expression recognition in videos

%	Happy	Disgust	Sad	Angry	Surprise	Fear
Happy	**83.83**	**10.29**	0	**5.88**	0	0
Disgust	0	**82.08**	0	**17.92**	0	0
Sad	0	0	**75.71**	**2.81**	**7.16**	**14.32**
Angry	0	**16.42**	0	**83.58**	0	0
Surprise	0	0	0	0	**80.3**	**19.7**
Fear	0	**18.57**	0	**9,71**	0	**71.01**

Table 2. Experiential results of confusion matrix for facial expression recognition in posed expression

%	Happy	Disgust	Sad	Angry	Surprise	Fear
Happy	**100**	0	0	0	0	0
Disgust	0	**100**	0	0	0	0
Sad	0	0	**81.8**	**9.09**	**9.09**	0
Angry	0	0	0	**100**	0	0
Surprise	0	0	0	0	**100**	0
Fear	0	0	0	**9.09**	**18.18**	**63.64**

5. Conclusion

Facial micro-expressions is a useful behavior source for danger demeanor detection, lie sign and hostile intent. In this paper we proposed a spontaneous facial expression recognition system for micro-expression analysis. We address the problems of detection and tracking a human face in videos and how to recognize spontaneous expressions presented in those faces .A histogram orientation gradient descriptor used for feature extraction and a support vector machine for classification. The proposed system tested on both static images and videos; acquired results of 95% recognition rate on static images and 80% recognition rate on videos.

6. Challenges and Future Work

In this work we proposed a system that deals with one frontal face with no rotation at a time. In future work we will handle more than one face and deals with rotated faces by adding an alignment step. We will also deal with other types of image sequences such as 3D images.

Acknowledgements

References

Allen, J. G., Xu, R. Y. D., & Jen, J. S. (2004). *Object Tracking Using CamShift Algorithm and Multiple Quantized Feature Spaces*. Pan-Sydney Area Workshop on Visual Information Processing Conferences in Research and Practice in Information Technology.

Bradski, G., & Kaehler, A. (2008). *Learning OpenCV* (1st ed.). In USA September.

Dalal, N., & Triggs, B. (2005). *Histograms of Oriented Gradients for Human Detection*. In CVPR.

Darwin, C., & Darwin, F. (Eds.). *The Expression of the Emotions in Man and Animals* (2nd ed.). J. Murray, London.

Ekman, P. (2001). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. W.W. Norton and Company.

Ekman, P. (2009). *Lie catching and micro-expressions.* The Philosoph of Deception, Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195327939.003.0008

Ekman, P., & Friesen, W. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto.

Ekman, P., Rolls, E. T., Perrett, D. I., & Ellis, H. D. (1992). Facial Expressions of Emotion: An Old Controversy and New Findings and Discussio. *Philosophical Transactions: Biological Sciences.* http://dx.doi.org/10.1098/rstb.1992.0008

Endres, J., & Laidlaw, A. (2009). Micro-expression recognition training in medical students: A pilot Study. *BMC Medical Education.* http://dx.doi.org/10.1186/1472-6920-9-47

Essa, I. A., & Pentland, A. P. (1995). Facial Expression Recognition using a Dynamic Model and Motion Energy. International Conference on Computer Vision. http://dx.doi.org/10.1109/ICCV.1995.466916

Fasel, B., & Luettin. J. (2002) Automatic facial expression analysis: A survey. *Journal of Pattern Recognition Society*, February 2002.

Frank, M. G., Herbasz, M., Sinuk, K., Keller, A., & Nolan, C. (2009). I see how you feel: Training laypeople and professionals to recognize fleeting emotions. In international Communication Association, Sheraton New York City.

Gomathi, V., Ramar, K., & Jeevakumar, A. S. (2010). A Neuro Fuzzy approach for Facial Expression Recognition using LBP Histograms. *International Journal of Computer Theory and Engineering.* http://dx.doi.org/10.7763/IJCTE.2010.V2.147

Huang X., Li, X., & Zhao, G. (2013). A Spontaneous Micro-expression Database: Inducement, collection and baseline. Automatic Face and Gesture Recognition (FG), 10th IEEE International Conference and Workshops on, 2013.

Khandait, S. P., Thool, R. C., & Khandait, P. D. (2011). Automatic Facial Feature Extraction and Expression Recognition based on Neural Network. *International Journal of Advanced Computer Science and Applications*.

Kobayashi, T., Hidaka, A., & Kurita, T. (2008). Selection of Histograms of Oriented Gradients Features for Pedestrian Detection, M. Ishikawa et al. (Eds.), 2008.

Oliver, N., Pentland, A., & BeHrard, F. (2000). LAFTER: A real-time face and lips tracker with facial expression recognition. *Journal of Pattern Recognition Society*, February 2000. http://dx.doi.org/10.1016/S0031-3203(99)00113-2

Polikovsky, S., Kameda, Y., & Ohta, Y. (2010). Detection and Measerment of Facial Micro-Expression Characteristics for Psychological Analysis, In IEICE technical report, 2010.

Shreve, M., & Godavarthy, S. (2011). *Macro- and Micro-expression spotting in long videos using spatio-temporal strain, Automatic Face & Gesture Recognition and Workshops*. IEEE International Conference, 2011.

Wang, S. S., Chen, H. L. & Yan, W. J. (2013). Face recognition and Micro-expression recognition based on discriminant tensor subspace and extreme learning machine. springer science+business media new york, 2013.

Youssif, A. A., & Asker, W. A. (2011). Automatic Facial Expression Recognition System Based on Geometric and Appearance Features. *Computer and Information Science*, 2011. http://dx.doi.org/10.5539/cis.v4n2p115

Zeng, Z., Fu, Y., Roisman, G. I., Wen, Z., Hu, Y., & Huang, T. S. (2006). Spontaneous Emotional Facial Expression Detection . *Journal of multimedia,* 2006.

Copyrights