MIT Libraries | DSpace@MIT

MIT Open Access Articles

Detection of asymmetric eye action units in spontaneous videos

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Mikhail, M., and R. el Kaliouby. "Detection of asymmetric eye action units in spontaneous videos." Image Processing (ICIP), 2009 16th IEEE International Conference on. 2009. 3557-3560. © 2010 IEEE

As Published: http://dx.doi.org/10.1109/ICIP.2009.5414341

Publisher: Institute of Electrical and Electronics Engineers

Persistent URL: http://hdl.handle.net/1721.1/60000

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Massachusetts Institute of Technology

DETECTION OF ASYMMETRIC EYE ACTION UNITS IN SPONTANEOUS VIDEOS

Mina Mikhail

American University in Cairo
Computer Science Department
113 Kasr Al Aini Street
Cairo, Egypt
minamohebn@gmail.com

Rana el Kaliouby

Massachusetts Institute of Technology
Media Laboratory
20 Ames Street
Cambridge MA 02139 USA
kaliouby@media.mit.edu

ABSTRACT

With recent advances in machine vision, automatic detection of human expressions in video is becoming important especially because human labeling of videos is both tedious and error prone. In this paper, we present an approach for detecting facial expressions based on the Facial Action Coding System (FACS) in spontaneous videos. We present an automated system for detecting asymmetric eye open (AU41) and eye closed (AU43) actions. We use Gabor Jets to select distinctive features from the image and compare between three different classifiers—Bayesian networks, Dynamic Bayesian networks and Support Vector Machines—for classification. Experimental evaluation on a large corpus of spontaneous videos yielded an average accuracy of 98% for eye closed (AU43), and 92.75% for eye open (AU41).

Index Terms— Gabor Jets, Dynamic Bayesian Networks (DBN), Support Vector Machines (SVM), Action Units (AU), Spontaneous video

1. INTRODUCTION

Over the past decade there has been an increasing surge of interest in automated facial expression analysis. The majority of these efforts describe facial movements using the Facial Action Coding System (FACS) [1], a catalogue of 44 unique action units (AUs) that correspond to each independent motion of the face. It also includes several categories of head and eye movements. FACS enables the measurement and scoring of facial activity in an objective, reliable and quantitative way. It can also be used to discriminate between subtle differences in facial motion. For these reasons, it has become the leading method in measuring facial behavior. Human trained FACS coders are very adept at picking subtle or fleeting facial actions, which communicate a wide range of information including when a person is lying, depressed, or about to have an epileptic seizure. However, FACS-coding requires extensive training and is a labor intensive task. It takes almost 100 hours of training to become a certified coder, and between one to three hours of coding for every minute of video.

The range of potential applications of automated facial analysis systems, coupled with the labor intensive requirements of human FACS-coding, has provided the main thrust for this area of research. In this paper, we address the problem of detecting asymmetric eye open (AU41) and eye closed (AU43) action units. Applications include driver state monitoring and health application monitoring such as detection of epileptic seizures which is often accompanied with asymmetric blinking. We tested our approach on a large number of spontaneous images.

The paper advances the state-of-the-art in facial action unit classification in three ways: (1) Accurate detection of spontaneous eye movements in the presence of substantial head motion and changes in lighting conditions; (2) our approach uses Gabor Jets to generate our feature vector and compares three different classifiers in accuracy; (3) testing and training on thousands of spontaneous images that have little or no control on accompanying head motion and lighting. Our approach can be generalized and applied to detect other AUs such as nose wrinkle (AU9), mouth AUs, as well as asymmetric cheek raiser (AU6) and others.

2. RELATED WORK

Bartlett et al [2] present one of the most successful systems for detecting AUs using Gabor filters followed by support vector machines (SVMs). Faces are first localized, scaled to 96x96 pixels and then passed through a bank of Gabor filters before classification. The accuracy of AU detection decreases as the training sample decreases. Vural et al [3], improved the work done by Bartlett *et al.* [2] by retraining the system on a larger dataset. They reached an accuracy of 93% on posed images and 75% on spontaneous images. Tian *et al.* [4] detect eye state AUs for frontal images by applying Gabor filters on three points of each eye and then feed the results of the Gabor filters into a neural network.

This paper extends eye state detection by accurately detecting eye states in thousands of spontaneous images with substantial degrees of head motion and changes in the light-

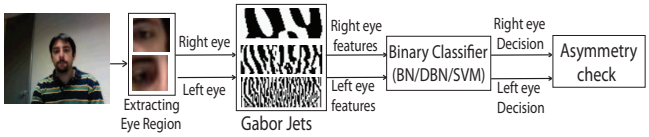

Fig. 1. Multi-level approach for asymmetric AU detection.

ing. We also propose the use of Bayesian Networks and SVMs instead of using Neural Networks, resulting in a huge boost in accuracy that reaches up to 98% for AU43 and 93% for AU41, instead of 83% reported in Tian *et al.* [4].

3. APPROACH

As shown in Fig. 1, we present a multi-level approach to detect asymmetric eye open or eye closed AUs in video. Since the purpose of this research is to differentiate between eye open and eye close, there was no need to extract all the facial features and train the eye open/close classifier on features that are not related to the eye. For every incoming frame of the video, we first locate the left and right eye regions. The regions are then passed through a bank of Gabor Jets, which are then fed into a classifier. We used three different classifiers for comparison: static Bayesian network, Dynamic Bayesian Network (DBN) and support vector machines (SVM). Congruency between classification results of the left eye and right eye determines the presence of asymmetry or not.

4. EYE REGION LOCALIZATION

In order to detect faces in an image, we used Google's face-tracker. The tracker uses a generic face template to bootstrap the tracking process, initially locating the position of 22 facial land-marks including the eyes, mouth, eyebrows and nose. We use the eye brow, inner and outer eye corner feature points to locate the eye region as shown in Fig. 2. From each frame, we extract two images representing the left and right eyes. The eye brow feature point represents the maximum Y coordinate for each eye image. The minimum Y coordinate is the reflection of the eye brow feature point on right pupil feature point. The inner and outer eye corner feature points represent the maximum and minimum X for the eye rectangle.

5. FEATURE EXTRACTION

After generating the eye images from the video frames, we wanted to extract features from these images to be used in our classification. Since the videos have substantial degrees of head motions which give different scales and orientations for the face, we decided to use Gabor filters. Gabor filters convolves the image with a Gaussian function multiplied by a sinusoidal function. The Gabor filters are considered to be orientation and scale tunable edge detector. The statistics of these features can be used to characterize the underlying texture information [5].

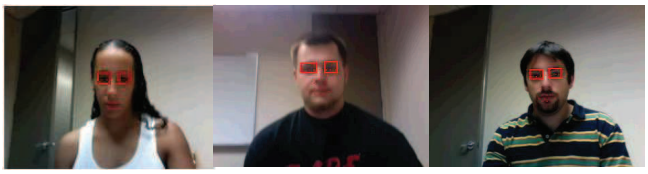

Fig. 2. The result of locating the right eye and left eye from the frame using facial feature points.

One major disadvantage of Gabor filters is that it is computationally expensive, making it difficult to be applied in real-time applications [6]. To detect video images in real-time, we decided to use Gabor Jets which describe the local image contrast around a given pixel in angular and radial directions [6]. Gabor Jets are characterized by the radius of the ring around which the Gabor computation will be applied. We chose the center of our Gabor Jets to be the center of pupil. So an image of 3x3, as shown in Fig. 3, is passed to the Gabor filters with 4 scales and 6 orientations to generate 216 features representing the magnitude of the Gabor filters.

6. CLASSIFICATION

In order to train our classifiers, we chose 80 eye left images captured from spontaneous videos. 40 out of the 80 images were a representative set of eye open and the other 40 were representative set of the eye closed. We experimented our approach with three different classifiers.

6.1. Static Bayesian Network

We created a Bayesian Network for detecting the eye open or eye closed AU. The Bayesian Network is defined by number of hidden states (N) and number of observed states (M) and number of parameters λ_j = (π,A):

- N, the number of states in the model $S = \{S_1, ..., S_N\}$; S_1 is a discrete hidden node representing whether the eye is open or closed. $S_2, S_3, ..., S_N$ are continuous observed states representing the Gabor Jets generated features;
- A= $\{a_{ij}\}$, is an N x N matrix to represent the topology of the network where aij = 1 indicates an edge from node i to node j. The structure of our Bayesian Network is shown in Fig. 4;
- π_i is the probability of state i in case of a discrete node or the mean and variance of state i in case of a continuous node.

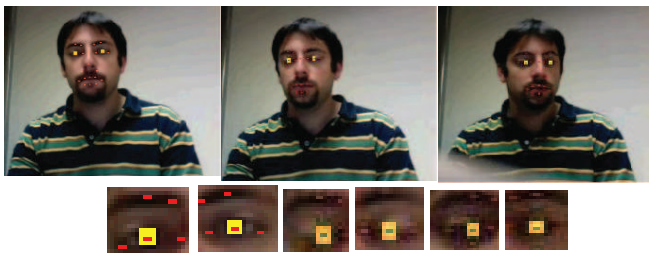

Fig. 3. The figure represents the selected (Gabor Jets) where we are applying Gabor filter on.

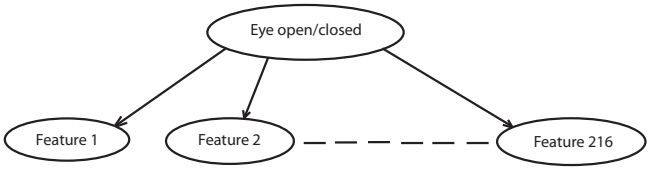

Fig. 4. The structure of the Bayesian Network used to infer the presence of eye open/closed AU.

The probability of eye open is defined by

$$P(X_1, X_2, X_3, ..., X_n) = \prod_{i=1}^{n} P(X_i|parents(X_i))$$

6.2. Dynamic Bayesian Network (DBN)

In order to make use of the temporal relations between AUs, we experimented with Dynamic Bayesian Networks instead of static Bayesian networks. DBNs are defined in the same way like static Bayesian network with some extra parameters. First, we have to define the inter relation between the different time slices. In our case, we found that our hidden node at time t is dependent only on the hidden node at t-1. The model is given by the joint probability distribution:

$$P(X_i, Y, \Theta) = P(Y|X_i, B_\Phi)P(X_i|A, \Pi)$$

Where Θ represents the number of time slice and the observation function B_Φ is parameterized by the conditional probability distribution that model the dependency between the two nodes. We detected the desired AU based on 5 previous time slices.

6.3. Support Vector Machines (SVMs)

Another classifier that was experimented with is an SVM classifier. SVMs view the input data, 216 Gabor Jet features, as two sets of vectors in a 216-dimensional space. SVM will construct a separating hyperplane in that space that maximizes the margin between the two data sets. A good hyperplane will be the one that has the highest distance to different points in different classes [7]. We trained one SVM for detecting eye open or eye closed. In our research we experimented with linear kernels and compared the results of applying SVM with the results obtained from static Bayesian Networks and Dynamic Bayesian Networks.

Table 1. Results of applying the three classifiers to the eye images.

AU	# images	BN		DBN		SVM	
		True	%	True	%	True	%
Open	1919	1780	92.7	1780	92.7	1809	94.3
Closed	150	147	98	147	98	140	93.3
	2069	1927	93.1	1927	93.1	1949	94.2

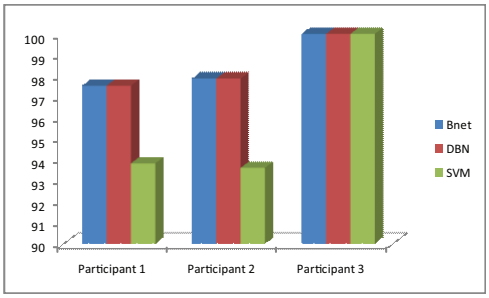

Fig. 5. Accuracy of applying the three classifiers to three different participants

7. EXPERIMENTAL EVALUATION

7.1. Classification Results

We have created a large database of images taken from spontaneous videos with an average duration of thirty minutes. The videos used for testing are from a sip study that Affective Computing at MIT Media Laboratory conducted in collaboration with major beverage company. Each participant is seated in front of a laptop (with a built-in webcam) and given a choice of two beverages that were located on the left and right of the laptop.

We used different images for testing than those used for training. The chosen images have head pitches which range from -9.2 to 12.23 degrees, head yaws which range from -

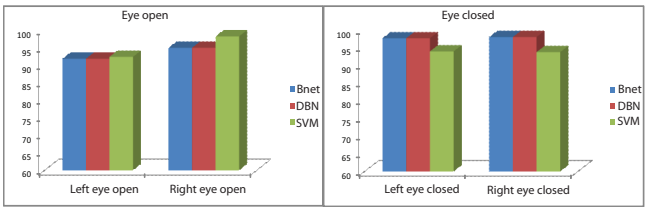

Fig. 6. Accuracy of applying the three classifiers, trained on the left eye images, to the left and right eye of the same participant.

Fig. 7. Lower action units: chin raiser (AU17), mouth open (AU27), cheek puffer (AU13). Upper action units: nose wrinkler (AU9), cheek raiser (AU6), frown (AU2)

16.8 to 26.9 and head rolls which range from -5.0 to 7.0. We tested our methodology on 2100 images and Table 1 shows the accuracy of applying the three different classifiers on eye open and eye closed images. It is obvious that the accuracy of the DBN is the same like that of the BN. This is because the hidden node is dependent on 216 observed node in case of BN and 221 observed nodes in case of the DBN which includes the Gabor features and the five temporal nodes of the previous time slices. The effect of the extra five nodes, in case of the DBN, will have a minor effect on probability of the hidden node compared to the other 216 nodes.

To ensure that our approach is general and can be used on participants that the classifier is not trained on, we tested on three participants whose images are not used in the training. Fig. 5 shows the accuracy of the three classifiers for each participant. The images that were selected for training were extracted from the video of the first participant only. However, we used different images from the video of the first participant for testing.

We also, trained our classifiers on left eye images only and used 400 right eye images for testing on the same classifier to make sure that we do not need a separate classifier for the right eye. The results of applying the classifier on the right eye images shown in Fig. 6 shows that our classifiers work well even if they are trained on left eye images only.

7.2. Discussion

Our methodology depends mainly on the accuracy of the tracker. Since the center of the Gabor Jet is determined by one of the feature points generated by the tracker, any substantial drift in this feature point will result in misclassification of the eye images.

Our approach can be easily generalized to detect other AUs as shown in Fig. 7. For instance, we can easily detect mouth open (AU27) by making the center of the Gabor Jets in the center of the mouth. We can also, detect the presence of a frown (AU2) by making the center of the Gabor Jets at the center of the forehead.

8. CONCLUSION

This paper describes a methodology for differentiating between eye open (AU41) and eye closed (AU43). Detecting such AUs is important for different applications such as driver state monitoring and health application monitoring. We presented the results of applying three different machine learning classifiers on Gabor Jets features. We reached an average accuracy of 98% for eye closed and 93% for eye open. Our next step is to test our methodology on different AUs such as mouth closed and mouth stretch. And in order to account for the inaccuracy of the tracker feature point, we will work on increasing the size of Gabor Jets to 5x5 or 7x7 or 9x9. This will require using a feature extraction algorithm such as Adaboost in order to reduce the training and inference time and to be able to apply our approach in real-time.

9. ACKNOWLEDGMENTS

The authors would like to thank Hyungil Ahn and Rosalind W. Picard for making the corpus available for this work. The authors would also like to thank Ahmed Sameh, Joshua Gluckman for their help in this research and Google for making the face tracker available to our research.

10. REFERENCES

[1] P. Ekman, W.V. Friesen, J.C. Hager, and A.H. Face, "Facial Action Coding System," 1978.

[2] M.S. Bartlett, G.C. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, and J.R. Movellan, "Automatic recognition of facial actions in spontaneous expressions," *Journal of Multimedia*, vol. 1, no. 6, pp. 22–35, 2006.

[3] E. Vural, M. Çetin, A. Erçil, G. Littlewort, M. Bartlett, and J. Movellan, "Machine Learning Systems For Detecting Driver Drowsiness," in *Proceedings of the Biennial Conference on Digital Signal Processing for in-Vehicle and Mobile Systems*, 2007.

[4] Y. Tian, T. Kanade, and J.F. Cohn, "Eye-state action unit detection by gabor wavelets," *Lecture notes in computer science*, pp. 143–150, 2000.

[5] W. Y. Ma and B. S. Manjunath, "Texture features and learning similarity," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 1996, pp. 425–430.

[6] U. Hoffmann, J. Naruniec, A. Yazdani, and T. Ebrahimi, "Face Detection Using Discrete Gabor Jets and Color Information," in *International Conference on Signal Processing and Multimedia Applications*, 2008, pp. 76–83.

[7] S.R. Gunn, "Support Vector Machines for Classification and Regression," *ISIS Technical Report*, vol. 14, 1998.