FACIAL ACTIVITY DETECTION TO MONITOR ATTENTION AND FATIGUE

Óscar Cobos, Jorge Munilla, Ana M. Barbancho, Isabel Barbancho, Lorenzo J. Tardón
Universidad de Málaga, Andalucía Tech, E.T.S.I. Telecomunicación,
Dpt. Ingeniería de Comunicaciones, Campus Teatinos s/n, 29071 Málaga
ocm@ic.uma.es, munilla@ic.uma.es, abp@ic.uma.es, ibp@ic.uma.es, lorenzo@ic.uma.es

ABSTRACT

In this contribution, we present a facial activity detection system using image processing and machine learning techniques. Facial activity detection allows monitoring people emotional states, attention, fatigue, reactions to different situations, etc., in a non-intrusive way. The designed system can be used in many fields such as education and musical perception. Monitoring the facial activity of a person can help us to know if it is necessary to take a break, change the type of music that is being listened to or modify the way of teaching the class.

1. INTRODUCTION

Human-machine interaction systems have improved with facial recognition. Pioneering works emerged between the 70s and 80s [1], when Facial Action Coding System (FACS) was developed. Facial activity detection allows monitoring people emotional states, attention, fatigue, reactions to different situations, etc., in a non-intrusive way. These detection can be used in many fields such as education and musical perception.

In this contribution we present a multi-purpose facial activity detection system to monitor attention and fatigue. The system has been made entirely in Matlab, making use of two specialized toolbox: Computer Vision System and USB Webcam. The operation of the system needs a computer and a Webcam USB2.0 with a frame rate of at least 60 fps.

2. SYSTEM DESCRIPTION

The general structure of the facial activity detection system developed in this work to monitor attention and fatigue is shown in Fig. 1. In this figure, it can be seen that the developed system consists on different stages that, although arranged in a row, are strongly related to each other with feedback information. The first state is intended to detect a face in each frame of the video as well as the Regions Of Interest (ROIs) on it: eyes and mouth. Then, the ROIs are tracked, so that even if the person moves, the regions of interest are always located in the scene. Once they ROIs are determined in any frame, it must be decided the status of each face part: eyes open or close and mouth open or close. Further, the movement and head turns are determined. Finally, a temporary analysis of the results is made to determine drowsiness using the frequency and duration of blinking and yawning.

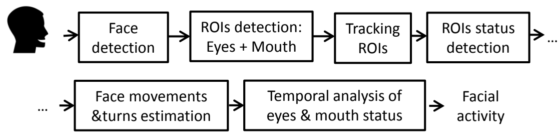

Figure 1. General structure of facial activity detection system.

2.1 Detection and tracking of face and ROIs (eyes and mouth)

The general state diagram of the detection and tracking of face and ROIs (eyes and mouth) are represented in Fig. 2.

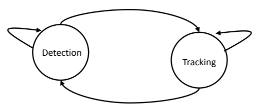

Figure 2. State diagram of the detection and tracking of face and ROIs.

The detection of face and ROIs is performing using the Matlab functionality *vision.CascadeObjectDetector*, which is based on detection by sliding window. The Viola & Jones algorithm [2] is the selected processing, including certain improvements such as rotated Haar characteristics and the use of weak classifiers CART (Classification and Regression Tree) with up to four Haar features. The performing of tracking are made using the KLT algorithm. Fig. 3 shows the result of the detection and tracking subsystem.

Figure 3. Result of detection and tracking subsytem.

2.2 Eyes and mouth state detection

The state detection is performed in each frame, that is, in a static picture.

2.2.1 Eyes state detection

Eyes status detection, open or close, has been performed using the model of Bag of Words. This model needs the creation of a visual dictionary, based on the analysis of labelled pictures with open or close eyes. A visual dictionary has been created, while a SVM classifier has been trained and used to classify (tested with cross-correlation).

2.2.2 Mouth state detection

Segmentation method [3] has been selected to perform the mouth status detection. The segmentation method is based on colour, and the YIQ model provides good results in mouth detection [4]. Specifically, channel Q is enough for this task, because the lips and the inside of the mouth (except the teeth) have a high purple content.

2.3 Face movements and turns estimation

Determining the attention of a person includes the detection of movements and turns made by the face. This point on, the analysis must be performed comparing images along time, not just individual images.

2.3.1 Face movements estimation

The method used compares the detected face regions (eyes and mouth) of the current image with the detected face regions of the closest image in which the face detection stage was performed. To this end, a simple algorithm has been developed based on calculating and comparing the central points of the ROIs that delimit the face.

2.3.2 Face turns estimation

The estimation of face turns is based on determining three angles: pitch, roll and yaw. Since the designed system only has one camera, it is necessary to perform 3D orientation from a 2D image. The selected method [5], is based on estimations of the head orientation.

2.4 Temporal analysis of eyes and mouth status

Determining activities such as blinking, prolonged eyes closing, yawning, surprise, etc. need to analyze several consecutive images.

2.4.1 Temporal analysis of eyes state

Temporal analysis of eyes state has to take into account that the average duration of a human blink is between 50 and 500ms. To this end, the video camera is configured with a frame-rate of at least 60fps. Fig. 4 shows the finite state machine to determine the state of eyes between staying awake, blinking and sleeping. Associated to this state machine is the estimation of the frequency and duration of blinkings. The condition to change to the state sleeping is that the eyes were closed more than 500ms.

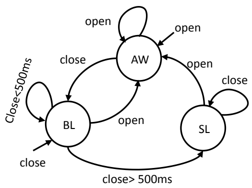

Figure 4. Finite state machine to determine the state of eyes between awaking (AW), blinking (BL) and sleeping (SL).

2.4.2 Temporal analysis of mouth status

Since the objectives of the described system are monitoring attention and fatigue, it is crucial to determine if the person has the mouth closed, is talking or yawning. Fig. 5 shows the finite state machine to determine the state of mouth between close, talking and yawning. Associated to this state machine is the estimation of the frequency and the duration of the yawns. The condition of the yawing is mouth open for over 3 seconds.

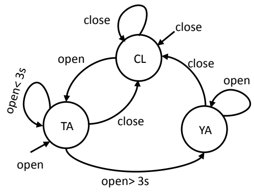

Figure 5. Finite state machine to determine the state of mouth between close (CL), talking (TA) and yawning (YA).

3. RESULTS

The designed system has been tested with 30 different people (15 males and 15 females). We have tested the system both together and ins isolated stages. * Detection of face and ROIs. Success rate is near 100%, being the mouth detection the worst case with a success probability of 85%. * Tracking of face and ROIs. Traking is lost in 1 out of 16 tests. * Eyes state detection. Classification accuracy is close to 95%. * Mouth state detection. Classification accuracy is close to 90%. * Temporal analysis of eyes and mouth state. They are always detected correctly, but the detection delay is 500ms for eyes and 3s for mouth.

4. CONCLUSIONS

In this contribution, we present a facial activity detection system using image processing and machine learning techniques. This system allows detecting if the eyes are open, blinking or closed by being asleep, as well as detecting the mouth closed, talking or open for yawning. This facial activity detection allows monitoring people emotional states, attention, fatigue, reactions to different situations, etc., in a non-intrusive way. Monitoring the facial activity of a person allows us to know if it is necessary to take a break, change the type of music that is being listened to or modify the way of teaching the class.

Acknowledgments

This work has been funded by Campus de Excelencia Internacional Andalucía Tech-Universidad de Málaga under educational innovation project PIE17-021. This activity was done in collaboration with Servicio de Publicaciones y Divulgación Científica and the department Ingeniería de Comunicaciones, both at Universidad de Málaga.

5. REFERENCES

[1] P. Ekman and W. Friesen, “Measuring facial movement,” in *Journal of Environmental Psychology*, 1976.

[2] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Computer Vision and Pattern Recognition CVPR2001*, 2001.

[3] C. Solomon and T. Breckon, *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*. John Wiley & Sons, 2011.

[4] U. Saeed and J. Dugelay, “Combining edge detection and region segmentation for lip contour extraction,” in *International Conference on Articulated Motion and Deformable Objects*, 2010, pp. 11–20.

[5] G. Mateos, *Procesamiento de caras humanas mediante integrales proyectivas*. Tesis Doctoral. Universidad de Murcia., 2007.