To Kick or Not to Kick: An Affective Computing Question

Luca Casaburi[1], Francesco Colace[2], Carmine Di Gruttola[3] and Donato Di Stasi[2]
[1]SIMASlab, Università degli Studi di Salerno, Via Giovanni Paolo II, 132, Fisciano, Italy
[2]DIIn, Università degli Studi di Salerno, Via Giovanni Paolo II, 132, Fisciano, Italy
[3]Università Degli Studi di Salerno, Via Giovanni Paolo II, 132, Fisciano, Italy

Keywords: Affective Computing, Ekman Theory, Entertainment 2.0.

Abstract: With the Web 2.0 and its services, the users are able to express their own emotions through the production and sharing of multimedia contents. In this context, we can interpret the flood of selfies that everyday invades the web as the will to overcome the limits of the textual communication. Therefore, it becomes particularly interesting the possibility to extract in an automatic way information about the emotional state of the people present in the multimedia content. The knowledge of a person's emotional state gives useful information for the personalization of many online services. Some examples of the areas where such an approach can be adopted are the services for the content recommendation, for the entertainment and for the distance training. In this paper, the techniques of sentiment extraction from multimedia contents will be applied to the sports world. In particular, the aim is to verify the possibility to predict the result of a penalty kick analyzing the player's face. The objective is to investigate the possibility to evaluate in real time the psychic conditions of an athlete during a competition. The system has been tested on an opportunely built dataset and the results are more than satisfying.

1 INTRODUCTION

Pasadena, 1994. Roberto Baggio is kicking a decisive penalty. If he failed, Italy, Baggio's team, would lose the World Cup. He takes the ball and goes towards the penalty spot. Arrigo Sacchi, Italy's coach, controls a little screen and runs to the player right away. He stops him, says something and changes him for another player who scores the penalty kick! Italy can still win its fourth World Cup! Who knows the history of football knows that things will be different: Baggio failed that penalty kick allowing the Brazil to win that world championship. However, what it would have happened if Italy's coach had a tool to evaluate the player's psychophysical state. Maybe he would have really made another player kick that penalty. Exactly from this consideration, this research paper starts with the aim to present a methodology for the determination of a player's emotional state analyzing his/her face.

At the base of this analysis, the methodology allows calculating the probability to realize a penalty kick according to the detected emotional state. The entire approach stays inside the wider and more ambitious research line named 'Affective Computing'. With the Web 2.0 and its services, the possibility to express yourself through multimedia contents is one of the mainly used modality by the network's users. Therefore, it seems clear the necessity to introduce some methodologies for the extraction of a user's emotional state by the analysis of the multimedia contents where he/she is present. The Affective Computing (Tao, 2005) is a specific branch of the artificial intelligence that realizes calculators able to recognize and express emotions. The progresses gained in the field of the HCI (Human Computer Interaction) have allowed the development of information applications that more and more take into account the users' contexts of use and permit an increasing interaction level (Amato, 2016). Nevertheless, the 'traditional' computer science chooses a kind of interaction between man and machine based on the reciprocal influence between action and reaction (Colace, 2015). The new frontiers of development of the information products aim at the implementation of affective machines that take into account the user's reaction to the system and that interact with them according to their emotional state. The development of the

Casaburi, L., Colace, F., Gruttola, C. and Stasi, D.
To Kick or Not to Kick: An Affective Computing Question.
In *Proceedings of the 18th International Conference on Enterprise Information Systems (ICEIS 2016) - Volume 2*, pages 415-420
ISBN: 978-989-758-187-8

affective computing has four possible application dimensions:

Emotional expression: it deals with the realization of interface agents able to reproduce the emotional expressions, thus communicate emotions principally through the representation of digital faces that mimic the main features of the human emotional expression. The purpose of this kind of interface is not that of equipping the machines with emotionality but allowing the machines to attribute emotions to the same interfaces.

Emotional recognition: the purpose of these products of the affective computing is that of recognizing the user's emotional state to eventually adapt to it, optimizing the tasks execution, with regard to the influence that the emotional state exercises on the human agent.

Emotional manipulation: this research line is aimed at studying the ways in which it is possible to influence the emotional state of the user in the interaction with the machine (Affect interaction).

Emotional synthesis: this is the most complex dimension of the affective computing to whom the studies about the mind are oriented, with the purpose of equipping a calculator with emotional intelligence, thus making it able to 'feel' emotions.

The detection of the emotions starting from the analysis of the facial expressions is particularly interesting.

In general, the detection of affective states from facial expressions in multimedia contents follows two main approaches: the recognition of discrete basic affects (by the adoption of a template matching approach) or the recognition of affects by the inference from movement of facial muscles according to the Facial Action Coding System (FACS) (Ekman, 1978). FACS classifies the facial movements 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 expressions. Many papers deal with this approach (Chang, 2004) 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. (Zhang, 1998) analyze the space of facial expressions to compare two classification systems: geometric-based (face is encoded by a landmark) and Gabor-based (face is encoded by Gabor features); the classification is performed with a 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 (Elad, 2003) or video sequences (Deng, 2005).

Cohen et al. (Cohen, 2003) propose a new architecture of HMM to segment and recognize facial expressions and affects from video flow, while Lee et al. (Lee, 2003) propose a method using probabilistic manifold appearances. Wang et al. (Wang, 2003) describe an automatic system that performs face recognition and affects recognition in grey-scale images of faces 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 (Deng, 2005) it is shown how to choose the Gabor features with PCA method and then LDA is used to identify the basic affects.

To extract affective states from facial expressions, the adoption of the Ekman Model is an effective approach (Ekman, 1978) (Casaburi, 2015). This model can infer six emotional states: happiness, anger, sadness, disgust, fear and surprise. This model has been enriched with the introduction of further states such as attention, fatigue and pain.

The Ekman theory can be improved with the introduction of the Facial Action Coding System (FACS), a system to name human facial movements by their appearance on the face. Movements of individual facial muscles are encoded by the FACS from slight different instant changes in facial appearance. It is a common standard to systematically categorize the physical expression of emotions and is useful to psychologists and animators. Due to subjectivity and time consumption issues, the FACS has been established as a computed automated system that detects faces in videos, extracts the geometrical features of the faces, and then produces temporal profiles of each facial movement (Hamm, 2011).

The recognition of affects by the inference from the 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 or as texture features to detect AUs. Cohn et al. (Cohn, 2004)

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) (Cohn, 2009) which analyzes videos in real-time to detect the sentiments. In this case, the face is encoded with a 2D mask that is used to interrogate a SVM to detect the associated affect. El Kaliouby et al (El Kaliouby, 2004) 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 among these points are used as features to identify different situations (open mouth, head movements, position of the eyebrows). The expressions encoded by the 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.

In this paper, we want to verify the possibility to adopt a technique based on the affective computing of FACS type to foresee the capability of an athlete to successfully complete a certain activity. The goal of this work is to try to apply the affective computing and neural networks to sporting events, especially in football. In particular, we have tried to guess the outcome of the penalty kick by analyzing the football players' face. In the next paragraph, the system architecture of the proposed approach is described. Then, the experimental campaign and the obtained results are shown and some conclusions close the paper.

2 SYSTEM ARCHITECTURE

As previously said, the purpose of this work is to apply the affective computing and neural networks to sporting events, especially in football, trying to guess the result of the penalty kick analyzing the football players' face.

The proposed framework consists of two main modules:

- Analyzer: This module is responsible for analyzing the video stream to obtain the features of interest from the face and codify them in the right format.
- Classifier: The classifier analyzes the received input feature for each frame, outputting the result of the prediction about the outcome of the penalty kick recovered in the video.

2.1 Analyzer Module

This module uses the framework of the analysis of the face presented in the paper [10] for video analysis in offline mode. This framework allows extracting the action unit and affective state for each face in the scene. The features of interest are obtained for each frame of video stream and these are inserted arrays called “Features Frame Vectors”. These features are:

- the AU on the detected face: the AU considered are AU1L / R, AU2L / R, AU4L / R, AU10, AU12L / R, AU15 L / R, AU20 L / R, AU24 L / R and AU26. The measures have a value in the range [0; 1], where 1 indicates the certainty that the AU has occurred and 0 not occurred;
- the measurement of affectiveness: the measured affective states are anger, disgust, fear, happiness, sadness, surprise. These measurements are normalized with a value in the range [0,1].

2.2 Classifier Module

The classifier module consists primarily of a SOM classifier with 22 inputs (AUs and affective state) and two outputs:

- goal: the penalty kick was scored;
- no goal: the penalty kick was wrong.

The SOM classifier has been compared with a feed-forward network and it has got the best result. The results are shown in the section of experimentation.

The SOM classifier is used to analyze the features extracted from the analyzer for each frame and the output is the label [goal / no goal]. This tag is inserted into a vector called Tag Frames Vector. The “goal” label is associated with the value 1 and the “no goal” label is associated with the value -1. In case no face is detected in the scene and so you cannot classify the frame, it is labelled with the value 0. This vector is then processed in the following steps:

1) Division of the Features Frame Vector into time windows. All time windows are organized in a vector called “Time

Windows Vector" and an example is shown in figure 1.

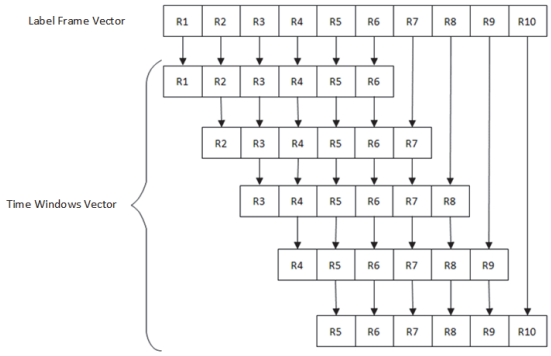

Figure 1: Example of Time Windows Vector.

2) Calculation of the arithmetic averages for each time window. The arithmetic averages are stored in a vector called Arithmetic Vector.
3) Calculation of the weighted average from the arithmetic vector. The weighted average is calculated using the arithmetic means of each time window and the weights are the order of succession of the windows. The result is given as input to the evaluator.
4) Evaluation of the weighted average. This step returns:
   - goal if the average is greater than α;
   - no goal if the average is less than $-\alpha$;
   - uncertain if the average is between α and $-\alpha$.

The value of α has been obtained considering it maximizing the results. So, the best value has been fixed as $\alpha = 0,2$. The approach is depicted in figure 2.

3 EXPERIMENTAL CAMPAIGN AND OBTAINED RESULTS

For the testing phase, we have created a dataset. The matches have been examined in a period ranging from 1994 to 2014. The videos of the matches have been taken from YouTube in different formats and resolutions, preferring HD quality (figure 2).

Then the videos have been cut to take the face of the player trying to find the frames from the moment he places the ball to the point when he starts to run. The videos have been cut in order to have only the face of the player even if they are very short. In fact,

Figure 2: Frames extracted by the videos in the Dataset.

their duration is between 1 second up to a maximum of 5 seconds.

The dataset consists of 80 videos with goal and 80 videos without goal, for a total of 160 videos. It has been divided into two parts: the training-set includes a total of 100 videos of which 50 are with goal and the other 50 are without goal; the remaining videos are used as test-set. In terms of percentage, the 62.5% of videos is part of the training-set and the remaining 37.5% is part of the test set.

In figure 3 the analys conducted frame by frame and the final results given by the system.

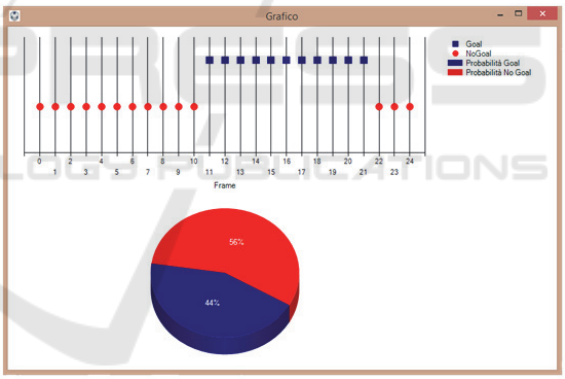

Figure 3: The obtained results frame by frame.

Figure 4: The obtained result.

The results obtained from the test phase, using a feed-forwarding network and a SOM classifier, are

reported in the table below. Videos are considered waste if the result is uncertain.

Table 1: Results test of system with feed-forward and SOM network.

Feed-forwarding Network		SOM Classifier	
Windows size = 6		**Windows size = 6**	
Recall	**0,194**	**Recall**	**0,647**
Precision	**0,406**	**Precision**	**0,543**
Waste	**0,143**	**Waste**	**0,131**
Windows size = 8		**Windows size = 8**	
Recall	**0,206**	**Recall**	**0,647**
Precision	**0,412**	**Precision**	**0,537**
Waste	**0,131**	**Waste**	**0,119**
Windows size = 10		**Windows size = 10**	
Recall	**0,214**	**Recall**	**0,652**
Precision	**0,441**	**Precision**	**0,536**
Waste	**0,125**	**Waste**	**0,106**

The best results are obtained with a SOM network with a time window of 10 frames, reporting a recall of 65.2% and a precision of 53.6%.

If we consider discarded even the videos in which the number of frames for the analysis of the face is less than the time window, we obtain the results shown in the following table, where we consider only the SOM classifier.

Table 2: Results test with SOM network on the video with the number of frames acceptable for the analysis of the face.

SOM Classifier	
Windows size = 6	
Recall	**0,726**
Precision	**0,584**
Waste	**0,263**
Windows size = 8	
Recall	**0,722**
Precision	**0,549**
Waste	**0,325**
Windows size = 10	
Recall	**0,731**
Precision	**0,567**
Waste	**0,363**

In this case, the best result is obtained with an acceptable number of rejects, with the classifier SOM and the time window equal to 6 frames. The recall is of 72.6% and the precision of 58.9%.

As the results show, the overall performances of the system are not excellent. The reasons have to be find above all in the scarce length of the considered videos and in the difficulty to find and analyze the faces. Even the typology of the shooting, typical of the penalty kick, is particularly difficult to analyze with the proposed approach. In any case, when the system analyzes videos at least three seconds long, the performances of the system improve with a value for the recall of 0,878 and for the precision of 0,726. Moreover, the choice of analyzing only the face excludes other features, obtainable for example with techniques of BLP, which can be used to determine the outcome of the penalty kick.

4 CONCLUSIONS

In this paper, we have presented a methodology based on the determination of the emotional state of a person and the use of a neural network for the prevision of the outcome of a penalty kick. The technique, tested on a specifically built dataset, gives enough satisfying results and indicates a possible approach for the determination of the emotional state of an athlete and a more efficient use of his/her potentials. Possible future developments are the application of the methodology to other sports in addition to football and the introduction of BLP techniques for determining the athlete's emotional state.

ACKNOWLEDGEMENTS

The research reported in this paper has been supported by the Project Cultural Heritage Information System (CHIS) PON03PE_00099_1 CUP E66J140000 70007 – D46J1400000 0007 and the Databenc District.

REFERENCES

Tao, J., Tieniu, T., 2005, Affective Computing: A Review, *Affective Computing and Intelligent Interaction, LNCS 3784*, Springer, pp. 981–995.

Ekman, P., Friesen, W., 1978, *Facial Action Coding*

System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press.

Chang, Y., Changbo, Hu, Turk, M., 2004, *Probabilistic expression analysis on manifolds*, In Proceedings of the 2004 IEEE computer society conference on Computer vision and pattern recognition, CVPR'04, pp. 520-527.

Zhang, Z., Lyons, M., Schuster, M., Akamatsu, S., 1998, *Comparison Between Geometry-based and Gabor-wavelets-based Facial Expression Recognition Using Multi-layer Perceptron*, Third IEEE Intel. Conf. On Automatic Face and Gesture Recognition, pp. 454-459.

Elad, A., Kimmel, R., 2003, *On bending invariant signatures for surfaces*, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 25, Issue: 10, pp. 1285–1295.

Deng, H., Jin, L., Zhen, L., Huang, J., 2005, *A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA*, International Journal of Information Technology, Vol. 11 No. 11, pp. 86-96.

Cohen, I., Sebe, N., Garg, A., Chen, L.S., Huang, T.S., 2003, *Facial Expression Recognition From Video Sequences: Temporal and Static Modeling*, Computer Vision and Image Understanding, pp. 160-187.

Lee, K., Ho, J., Yang, M.H., Kriegman, D., 2003, *Videobased Face Recognition Using Probabilistic Appearance Manifolds*, Conference on Computer Vision and Pattern Recognition, pp. 313-320.

Wang, H., Ahuja, N., 2003, *Facial Expression Decomposition*, Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03), vol.2, pp. 958-965.

Casaburi, L., Colace, F., De Santo, M., Greco, L., 2015, *Magic mirror in my hand, what is the sentiment in the lens?": An action unit based approach for mining sentiments from multimedia contents*, Journal of Visual Languages & Computing, 27, pp. 19-28.

Hamm, J., Kohler, C. G., Gur, R. C., Verma, R., 2011, *Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders*, Journal of Neuroscience Methods 200 (2), pp. 237–256.

Cohn, J. F., Reed, L. I., Ambadar, Z., Xiao, J., Moriyama T., 2004, *Automatic Analysis and Recognition of Brow Actions and Head Motion in Spontaneous Facial Behavior*, IEEE International Conference on Systems, Man and Cybernetics, pp. 610-616.

J. F. Cohn, J. F., Lucey, S., Saragih, J., Lucey, P., De la Torre, F., 2009, *Automated Facial Expression Recognition System*, IEEE International Carnahan Conference on Security Technology, pp. 172-177.

El Kaliouby, R., Robinson P., 2004, *Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures*, In Proc. Int'l Conf. Computer Vision & Pattern Recognition, pp. 181-200.

Amato, F., Colace, F., Greco, L., Moscato, V., Picariello, A., 2016, *Semantic processing of multimedia data for e-government applications*, J. Vis. Lang. Comput. 32: pp. 35-41.

Colace, F., Casaburi, L., De Santo, M, Greco, L., 2015, *Sentiment detection in social networks and in collaborative learning environments*, Computers in Human Behavior 51: pp. 1061-1067.