Facial Expression Recognition in still pictures and videos using Active Appearance Models. A comparison approach.

Dragoş Datcu Léon Rothkrantz

Abstract: *The paper highlights the performance of video sequence-oriented facial expression recognition using Active Appearance Model – AAM, in a comparison with the analysis based on still pictures. The AAM is used to extract relevant information regarding the shapes of the faces to be analyzed. Specific key points from a Facial Characteristic Point - FCP model are used to derive the set of features. These features are used for the classification of the expressions of a new face sample into the prototypic emotions. The classification method uses Support Vector Machines.*

Key words: *Facial Expression Recognition, Facial Feature Extraction, Active Appearance Models - AAM, Image Processing.*

INTRODUCTION

Computer vision has become one of the most challenging subjects nowadays. The need to extract information from images is enormous. Face detection and extraction as computer-vision tasks have many applications and have direct relevance to the face-recognition and facial expression recognition problem. According to Ekman and Friesen [5], people are born with the ability to generate and interpret only six facial expressions: happiness, anger, disgust, fear, surprise and sadness. The rest of the facial expressions have to be learned from the environment. The approach detailed in the current paper focuses on algorithms for automatic recognition of these prototypic expressions. A comparison is done on the performance of two techniques involving the analysis of facial expressions in still pictures and in videos.

RELATED WORK

The work of [10] presents a novel nonlinear generative model using conceptual manifold embedding and empirical kernel maps for facial expressions. The algorithm deals with the complex nonlinear deformations of the shape and appearance in facial expressions and provides accurate emotional based synthesis. The paper of [11] presents an approach to determine the gender and expression of faces by using Active Appearance Model. Four emotional states were employed for the analysis that was realized by using SVM classifier. The work shows an improvement of the recognition results by involving a first classification of the gender. It was found that a single cluster of Gaussian derivative responses leads to a high robustness of detection given the pose, illumination and identity [5]. The use of Active Appearance Model for extracting face shape information and the tracking of emotions in video sequences from crisis environments, based on this shape data, were recently researched in [3]. The use of AAM on still pictures for facial expression recognition in a Web-based multimodal system was investigated in [4]. The work of [12] proposes an improved fitting procedure based on a stereo AAM that uses the information about two coupled views. The recognition of expressions is improved by using a generalized discriminant analysis that combines the 3d shape and face appearance.

THE ARCHITECTURE OF THE SYSTEM

The data set used for training the facial expression recognizer was Cohn-Kanade database [7]. The database contains a set of video sequences of recordings of several subjects acting on multiple scenarios. Each video sequence includes a subject showing a specific facial expression from the neutral state to the apex of the emotion. In the original database only the last frame of each sequence is labeled using Action Units - AU codification. The process of creating the data set for training implied the selection of the

Acknowledgments. The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024.

first and the last frames from each video sequence. The final set of selected samples has the structure as illustrated in Table 1.

Table 1: The structure of the data set for facial expression recognition.

Emotion	Fear	Surprise	Sadness	Anger	Disgust	Happy
#samples	84	105	92	30	56	107

The current approach employs a face detection algorithm that is based on Viola&Jones features [14], a Facial Characteristic Point – FCP extraction method based on Active Appearance Model – AAM [2], and the classification of emotional temporal patters using Support Vector Machines – SVM [13]. The sequence of processing steps works both on still pictures and on video sequences. The video analysis is schematized in Figure 1.

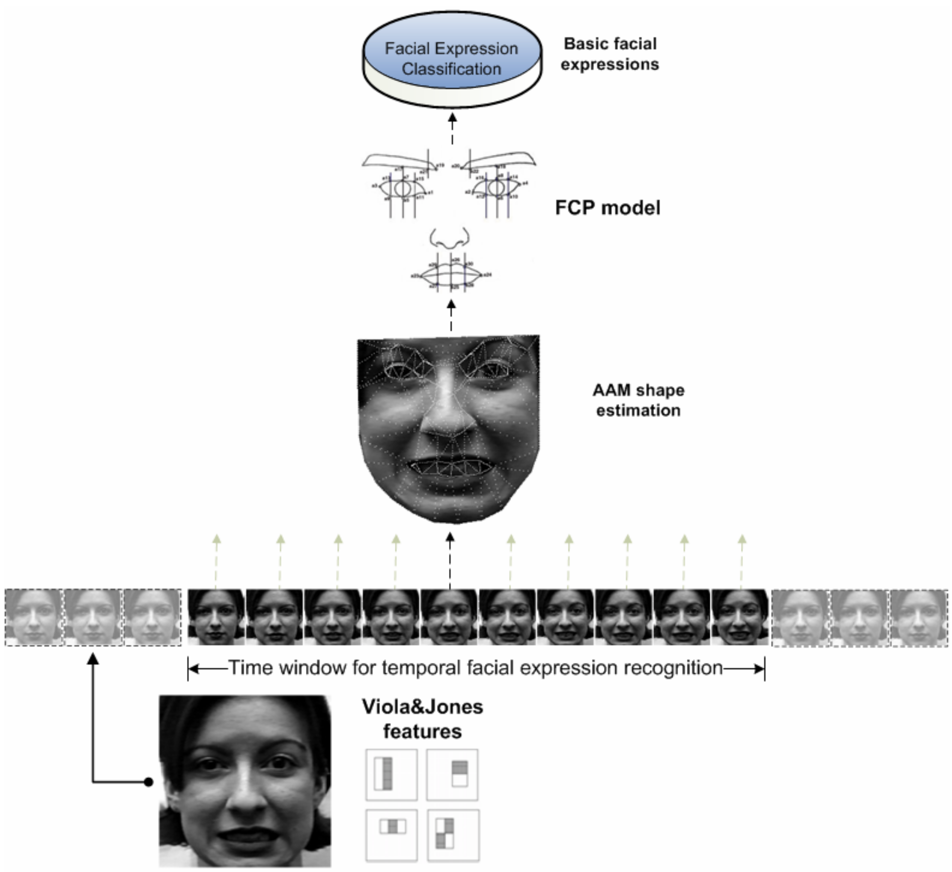

Figure 1: The steps involved in the automatic facial expression recognition.

Facial Characteristic Point - FCP model

The shape information extracted for the Active Appearance Model – AAM from each face sample, is used to compute a set of suitable parameters that describe well the appearance of the facial features. The AAM models the shape and texture data from the faces contained in the training data set. The variability of the shapes and textures from the training data set is determined by taking into account the mean face shape and face

texture (Figure 2). The AAM model is built using 19 shape modes, 24 texture modes and 22 appearance modes retaining 95.49% of the combined shape and texture variation. The shape vector contains 58 face shape points.
During the test step, a search on all modes of variation corresponding to the projection vector according to the Principal Component Analysis - helps at fitting optimally the model to the new, unseen face sample. The model robustly handles a certain degree of rotation, scaling and translation. The effect of illumination variation is minimized by scaling the texture data of the face samples during the training of AAM. Although AAM is capable of overcoming to some degree the influence of occlusion, new samples containing partially occluded faces were introduced to increase the performance of the model fitting.

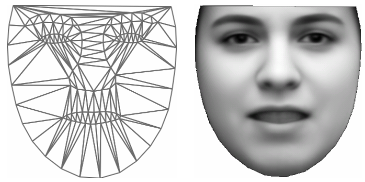

Figure 2: The mean face shape (left) and the mean face texture aligned to the mean shape (right).

The first step in identifying specific features for the facial expression recognizer is the selection of the optimal key points on the face area from the shape data. The key points are defined as Facial Characteristic Points (FCPs) and the FCP-set (Figure 3) is derived from Kobayashi & Hara model [8].

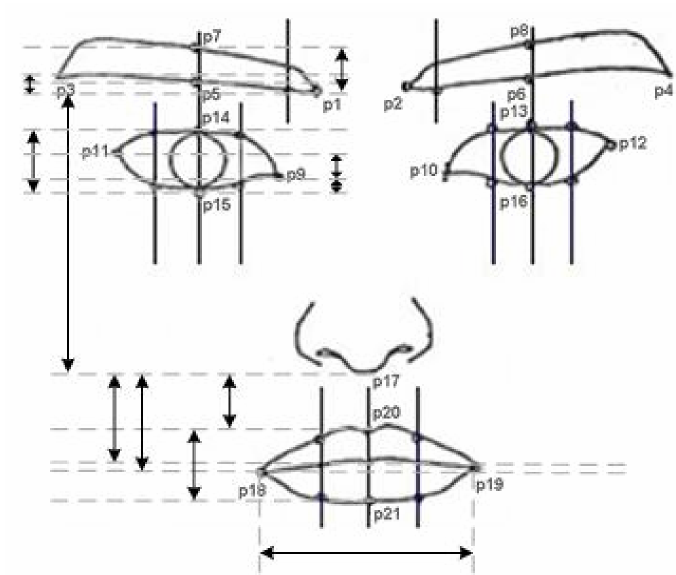

Figure 3: The Facial Characteristic Point FCP model.

In the next step, a transform converts the FCP-set to some parameters v_i of an intermediate model. The parameterization has the advantage of providing the classifier with data that encode the most important aspects of the typical variations induced by the facial expressions.
Furthermore, it acts as a dimensionality reduction procedure, since the dimension of the feature space is lower than the dimension of the image space. An advantage of the model is that it also can handle certain degree of asymmetry by using some parameters for both left and right sides of the face.

The symmetry of the model is assumed to make the recognition process of facial expressions robust to occlusion or poor illumination i.e. if the left eye area is not directly visible do not use related information. The robustness assumption is based on the supposition that the face detection procedure is also efficient enough in such working conditions so as to be able to correctly detect the faces.

CLASSIFICATION OF FACIAL EXPRESSIONS

The method used to encode the emotional patterns takes into account the subtle changes of the face shapes in different emotional postures. The algorithm for recognizing emotions is based on an approach that uses a FCPs distance oriented model. The complete list of such parameters is given in Table 2. A second approach employs a prior technique of detecting the Action Units – AUs [5] followed by the determination of facial expressions by taking into account the AU sequences.

Table 2: The set of visual feature parameters.

		Visual feature			Visual feature			Visual feature
v_1	$(P_1, P_7)_y$	Left eyebrow	v_7	$(P_{14}, P_{15})_y$	Left eye	v_{13}	$(P_{17}, P_{20})_y$	Mouth
v_2	$(P_1, P_3)_y$	Left eyebrow	v_8	$(P_9, P_{11})_y$	Left eye	v_{14}	$(P_{20}, P_{21})_y$	Mouth
v_3	$(P_2, P_8)_y$	Right eyebrow	v_9	$(P_9, P_{15})_y$	Left eye	v_{15}	$(P_{18}, P_{19})_y$	Mouth
v_4	$(P_2, P_4)_y$	Right Eyebrow	v_{10}	$(P_{13}, P_{16})_y$	Right eye	v_{16}	$(P_{17}, P_{18})_y$	Mouth
v_5	$(P_1, P_{17})_y$	Left Eyebrow	v_{11}	$(P_{10}, P_{12})_y$	Right eye	v_{17}	$(P_{17}, P_{19})_x$	Mouth
v_6	$(P_2, P_{17})_y$	Right eyebrow	v_{12}	$(P_{10}, P_{16})_y$	Right eye			

Different selection methods for creating the input feature set of the SVM classifier were defined. In the case of still picture analysis, the features are chosen from the values of the parameters v_i where $i = 1..17$, that were recorded from the frames showing the apex of specific emotions, for all the prototypic facial expressions.

In the case of video analysis, the variances in the parametric model describe emotional patterns for each facial expression. These capture the subtle changes in the shapes of facial features during the video sequence. Figure 4 depicts the temporal deformation templates for the face shape for different emotions. The method used to encode the emotional patterns take into account the total variance for each parameter of the same face in all the frames in the temporal analysis window. If $V_T = (v_1, v_2, ..., v_m)$ is the vector of parameters extracted from the video frame at time T, then $PV_{T'} = \Delta V$ is the variance of the parameters according to the last and the first frame in the sequence starting at time T'.

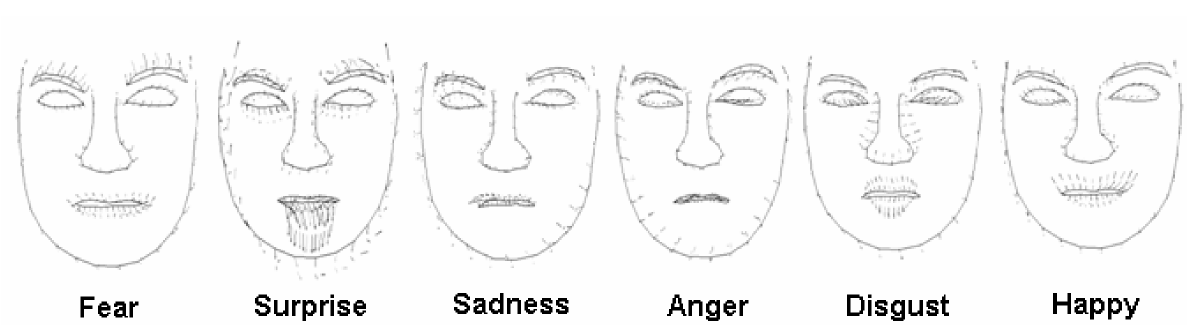

Figure 4: Shape deformation templates for the prototypic emotions.

Table 3 presents the results in the case of the recognition of facial expressions by using Support Vector Machines – SVM as classifier. The left table illustrates the recognition results for the analysis of facial expressions in still pictures and the right table shows the analysis for video data. From the algorithmic point of view, the difference consists in the type of feature used for classification. The method used for presenting the results is 2-fold Cross Validation.
As expected, it can be easily seen from the tables that the performance in the case of video sequences is higher than that in the case of still pictures. This comes from the fact that more information is employed to distinguish between the emotional classes for the video analysis. In fact, each feature actually describes the variance of the same feature used in still picture analysis.
Figure 5 shows the Receiver Operating Characteristic - ROC curves in both cases, for still picture and video data. The facial expression recognition results can be compared with the performance achieved by other methods that use AAM face appearance vectors, such as 3-layer feed-forward neural network [9] or with Fisher's discriminant analysis [1].

Table 3: The confusion matrix for the facial expression recognition using SVM (polynomial kernel of degree 3) – still picture (left) and video sequence (right) analysis.

(%)	Fear	Surprise	Sadness	Anger	Disgust	Happy
Fear	**84.70**	3.52	3.52	4.70	1.17	2.35
Surprise	12.38	**83.80**	0.95	0	0	2.85
Sadness	6.45	3.22	**82.79**	1.07	3.22	3.22
Anger	3.44	6.89	6.89	**75.86**	6.89	0
Disgust	0	0	7.14	10.71	**80.35**	1.78
Happy	7.54	8.49	2.83	3.77	4.71	**72.64**

(%)	Fear	Surprise	Sadness	Anger	Disgust	Happy
Fear	**88.09**	2.38	4.76	3.57	1.19	0
Surprise	0	**88.67**	2.83	8.49	0	0
Sadness	5.43	2.17	**85.86**	2.17	1.08	3.26
Anger	10.71	0	3.57	**85.71**	0	0
Disgust	5.35	5.35	3.57	1.78	**82.14**	1.78
Happy	4.62	0	7.40	2.77	5.55	**79.62**

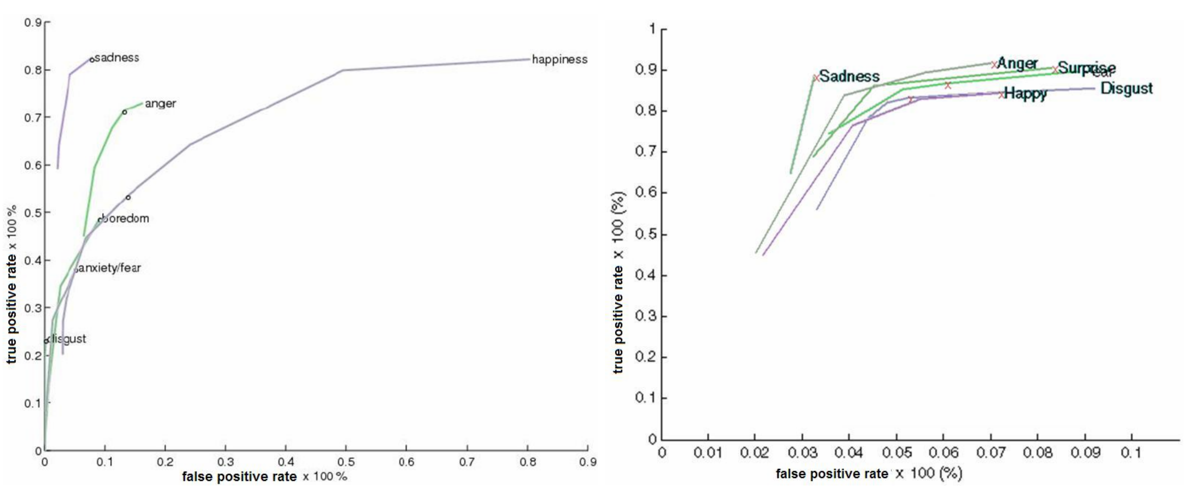

Figure 5: ROC graph that show the committees with the highest true positive rates for each emotion class – still picture (left) and video sequence (right) analysis.

CONCLUSIONS AND FUTURE WORK

In the current paper we described two methods for automatic facial expression recognition using face shape information extracted using an Active Appearance - AAM model. The methods were used on static pictures and video data analysis. One differs from the other in the type of selected features. The two classifiers consisted in Support Vector Machines – SVM that were trained using the same data set. The results clearly show the advantage of using information from video frames over the processing on one picture only. Further work is employed to improve the performance of the expression recognition algorithm by identifying another set of more relevant features. In parallel, the performance of other classifiers is also analyzed.

REFERENCES

[1] Abboud, B, F. Davoine and M. Dang. 2004. Facial expression recognition and synthesis based on an appearance model. Signal Processing: Image Communication, 19: 723–740.

[2] Cootes, T. F. and G. J. Edwards, and C. J. Taylor. 1998. Active appearance models. Lecture Notes in Computer Science, 1407:484–502.

[3] Datcu, D. and L.J.M. Rothkrantz. 2007. Facial expression recognition using Active Appearance Model in crisis environments, ISCRAM'07, ISBN 9789054874171, Delft, The Netherlands, 515-524.

[4] Datcu, D. and L.J.M. Rothkrantz. 2007. Multimodal Web based system for human emotion recognition, ISC'07, ISBN 978-90-77381-34-2, Delft, The Netherlands, 91-98.

[5] Ekman, P. and W. Friesen. 1978. Facial Action Coding System. Consulting Psychologists Press, Inc., Palo Alto California, USA.

[6] Gourier, N., D. Hall and J. L. Crowley. 2004. Estimating face orientation from robust detection of salient facial features, in Proc of Pointing 2004, International Workshop on Visual Observation of Deictic Gestures.

[7] Kanade, T., J. Cohn; and Y. Tian. 2000. Comprehensive database for facial expression analysis. Proc. IEEE Int'l Conf. Face and Gesture Recognition. 46–53.

[8] Kobayashi, H. and F. Hara. 1997. Facial Interaction Between Animated 3D Face Robot and Human Beings. IEEE Computer Society Press, 3732-3737.

[9] Kuilenburg, Hans van, M. Wiering and M. den Uyl. 2005. A Model Based Method for Automatic Facial Expression Recognition. ECML'05, 194-205.

[10] Lee, C.S. and A. Elgammal, Nonlinear Shape and Appearance Models for Facial Expression Analysis and Synthesis. Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), 1: 497 – 502.

[11] Saatci, Y., and C. Town. 2006. Cascaded Classification of Gender and Facial Expression using Active Appearance Models. The 7th Conference on Automatic Face and Gesture Recognition FGR'06.

[12] Sung, J., S. Lee, D. Kim. 2006. A Real-Time Facial Expression Recognition using the STAAM. ICPR'06, 18: 275-278.

[13] Vapnik, V. 1995. The nature of statistical learning.Springer, New York.

[14] Viola, P. and M. Jones. 2001. Robust Real-time Object Detection. Second International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling.