Active Shape Models with 2D profiles for Stress/Anxiety recognition from face images

Martin Penev, Agata Manolova, Ognian Boumbarov

Radiocommunications and Videotechnologies Department
Faculty of Telecommunications, Technical University of Sofia
8 Kliment Ohridski Blvd., 1000 Sofia, Bulgaria
{martin, amanolova, olb}@tu-sofia.bg

Abstract

The facial expression is a visible manifestation of the emotional state, cognitive activity, intention, personality, and mental health of a person. In this paper we use an approach to distinguish between different emotions from images in particular the unhappiness related to anxiety and stress. The approach is based on Active Shape Model (ASM) with 2D profile for extracting important facial points (coordinates of edges of the mouth, eyes, eyebrows, etc.). Further the feature vectors are formed by concatenating the landmarks data from the proposed method and use the data as an input to kernel Support Vector Machine (SVM) classifier. The experimental results using Cohn-Kanade Extended Facial Expression Database show high recognition rate. 2D profile ASM significantly speeds up the fitting process comparing with 1D profile ASM by averagely over 40%.

1. Introduction

Face plays significant role in social communication. This is a 'window' to human personality, emotions and thoughts. The face we look at is a mix of both physical characteristics and emotive expressions. According to the psychological research, nonverbal part is the most informative channel in social communication. Verbal part contributes about 7% of the message, vocal – 34% and facial expression -55% [1, 2]. Facial expression recognition is a considerably challenging field to generate an intelligent system that is able to identify and understand human emotions for various vital purposes, e.g. security, society, entertainment, health care, human-computer interaction, industrial and personal robotics, surveillance and transportation. By using of the information that the face carrys, facial expressions can play an important role wherever humans interact with machines. Automatic recognition of facial expressions can act as a component of intelligent and adaptive human-machine interfaces. It is expected that introducing affective dimension can significantly enhance the users' interaction experience also can make that interaction more efficient. In general, the intelligent computer with the emotion recognition system can be used to improve the daily lives of people.

People with depression may feel sad, anxious, helpless, worthless or hopeless. They stop caring about things they used to enjoy, such as hobbies. Often they feel a loss of energy, and they may be incapable of concentrating on anything. Irregular sleep patterns and loss of appetite also indicate depression. Long-term effects of depression include chronic fatigue because of loss of energy and irregular sleep patterns. This, combined with a weakened immune system, can lead to a susceptibility to physical illness. Those with depression might also suffer chronic aches and pains. Depression also affects attention and memory. People who suffer from depression could find themselves forgetful and unable to concentrate, which can have a negative impact on their life. A history of depression may be associated with an increased risk of stroke [4]. A large body of evidence suggests that depression is associated with an increased risk of many chronic diseases, including hypertension [5, 6], diabetes [7], and particularly coronary heart disease [8]. There is also a great chance that depression and overweight will co-occur [9]. Results have shown that people with major depression who are recovering from strokes or heart attacks have a more difficult time making health care choices. They also find it more difficult to follow their doctor's instructions and to cope with the challenges their illness presents.

So for an intelligent medical system being able to recognize that the patient

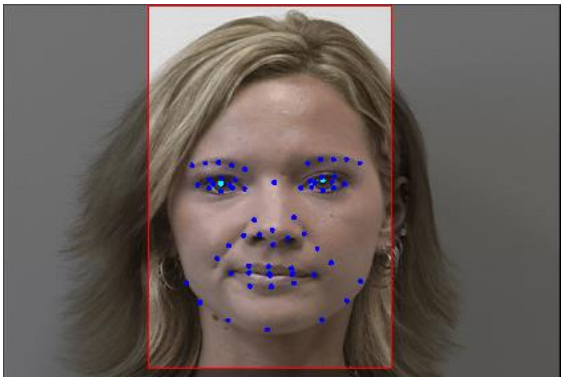
suffers from anxiety/depression is of essence.

The rest of the paper is organized as follows: In the next section we present a brief overview of ASM with 2D profile for extracting important facial points. In Section 3 we present the classification approach – kernel SVM. In Section 4 we will illustrate the experimental results for the classification of the processed data using the proposed method. Finally section 5 will conclude the paper.

2. Active Shape Model with 2D profile

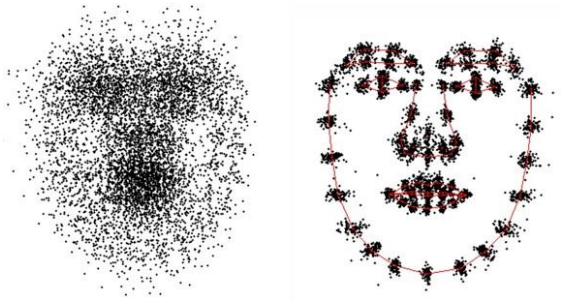
Very important for the success of an automatic facial expression detector is the face alignment/registration algorithm and the visual features derived from it. As expressions can be subtle, high accuracy is desired for the various facial features, enhancing the ability of a classifier to detect the facial expression correctly. To facilitate this, the ASMs have been widely used as they provide dense registration accuracy (i.e. 60-70 points on the face). ASM is a powerful statistical tool that has been used for the segmentation of a wide range of objects, including facial features [10].

ASMs are based on the combination of a Point Distribution Model (PDM) plus a set of local image texture models. The PDM describes the shape variability of the template and the texture models describe the image variability around each of its points. In order to correctly model the shape of human faces, it is first represented by a set of landmarks. Each landmark describes a particular part of the face such as the tip of nose and the eye pupils as shown in Figure 1. In order to build a model that is flexible enough to cover the most typical variations of facial shapes, a sufficiently large training set has to be used. The landmarks are manually labelled for each training image to generate a set of training shapes. An object shape is represented by a set of labelled points or landmarks. The number of landmarks should be large enough to show the overall shape. Our landmarking scheme uses 68 model points.



"Figure 1. The landmarks labelled on a facial image"

The coordinates of all landmarks for each image are stored as a vector, called a shape, in the form $\mathbf{x} = (x_1, y_1, ..., x_N, y_N)^T$, where x_i and y_i are the coordinates of the i-th landmark and N is the number of landmarks used. The shapes in the training set are aligned with each other using Generalized Procrustes Analysis (GPA) which scales, rotates and translates each training shape so that the corresponding landmarks in the various images of the training set are as close as possible [11]. An example is illustrated in Figure 2.



"Figure 2. Training set before and after alignment"

The PDM is constructed by applying Principal Component Analysis (PCA) to the set of shapes in the training set. PCA is computed on these shapes and the eigenvalues of the covariance matrix contributing to 97% of the variation are sorted and used to store their

corresponding eigenvectors in a matrix P. The mean shape is the mean of the aligned training shapes (which in our case are manually landmarked faces). The mean of the *N* training shapes is calculated as follows:

$$\bar{x} = 1/N \sum_{i=1}^{N} x_i \,, \qquad (1)$$

where x_i is the *i*-th shape in the set of the *N* training images.

By subtracting the mean shape calculated above from each training shape, the covariance matrix of the deviations from the mean is calculated as follows:

$$S_s = 1/N \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T \qquad (2)$$

The eigenvectors and eigenvalues of the covariance matrix are subsequently obtained by using the equation:

$$S_s \mathbf{v}_k = \lambda_k \mathbf{v}_k \,, \qquad (3)$$

where $\mathbf{v}_k$ and λ_k are, respectively, the *k*-th eigenvector and eigenvalue of S_s. Any shape in the training set can be approximated by the mean shape and a weighted sum of *t* eigenvectors. This is mathematically represented as:

$$x = \bar{x} + P\mathbf{b} \,, \qquad (4)$$

where $P = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_t)$ is a matrix whose columns are the *t* eigenvectors corresponding to the first *t* largest eigenvalues, and $\mathbf{b} = (b_1, b_2, ..., b_t)^T$ is a vector of weights applied to the eigenvectors. These weights (b_k) are the model parameters which control the shape generated by the PDM; thus, new facial shapes can be generated by varying these parameters. At the same time, these parameters are constrained in order for the model shape to be consistent with those in the training set. Since most of the population lies within three standard deviations of the mean, the model parameters are chosen to lie within the range given by:

$$-3\sqrt{\lambda_k} \le b_k \le 3\sqrt{\lambda_k} \qquad (5)$$

where λ_k is the eigenvalue that corresponds to the *k*-th eigenvector of the *t* eigenvectors.

The next stage involves profiling to generate statistical models of the grey level intensities of the region around each landmark to build a subspace that spans the variations of the exemplar training images. The classical ASM employs 1D profile, which is generated by sampling grey-level intensities of *m* pixels centered at each landmark. The direction of the profile is chosen to be along a line orthogonal to the shape boundary at each landmark. The local grey-level gradient pattern is then utilized for searching a new position of each landmark along a line centered at the landmark. The model shape is iteratively changed for its better fit to the target face in the image. The classical ASM performs well if the landmarks of the initial model shape are placed close to their targets. However, the initial model shape could be placed only roughly whereby the landmarks often get placed away from their targets. Thus resulting in long search lines and consequently making the target search computationally expensive. It could also distract the landmarks by local structures in the image. In order to overcome the above drawbacks of the classical ASM, a multi-resolution approach of ASM, known as multi-resolution ASM (MRASM), is proposed in [12]. In this scheme, an ASM is first applied to a coarse image to roughly place the model shape near the target object, and then applied to finer images to refine the shape fitting. An image pyramid containing a set of images with different resolutions is used in order to build the LGGM for multiple image resolutions. For each level of the image pyramid, the corresponding LGGM is utilized for searching a new position of each landmark. Unlike the classical ASM, the multi-resolution approach requires only short search lines; thus reducing the risk of distraction of the landmarks by local structures in the image, and decreasing the computational complexity [12]. However,

the 1D profile used in the classical ASM has a quite obvious shortcoming: every model point can only search its optimal location along its normal to the boundary and consequently, the search can converge to a local minimum and produces poor fitting results. In order to extend its flexibility, 2D profile ASM is proposed by Pei, J. [13]. Instead of sampling a single vector of gradients just along the profile normal in 1D profile ASM, an additional vector of gradients along the profile tangential is further used in 2D profile ASM. By sampling two vectors of gradients along these two directions, two sequential feature sets g_i^1 and g_i^2 can be obtained. The optimization process of each single model point can be realized by searching for the optimum along two directions in turn.

3. Support Vector Machines for Expression Classification

SVM is very popular and powerful method for binary and multi-class classification as well as for regression problems. For two class separation, SVM applies a maximum margin manner that estimates the optimal separating hyper plane. In our investigation we used the LibSVM library (http://www.csie.ntu.edu.tw/~cjlin/libsvm/) and Matlab functions. In general SVMs can only solve binary classification problems. For multi-class classification, LibSVM computes decision surfaces for all class pairs (one-against-one technique) and then find the correct class by a voting mechanism.

Let us given sample and label pairs $(x^{(i)}, y^{(i)})$, where $x^{(i)} \in R^m$, $y^{(i)} \in \{-1;1\}$ and $i = 1,...,K$. Here, for class "1" and for class "2", $y^{(i)} = 1$ and $y^{(i)} = -1$, respectively. We also define a feature map - $\phi: R^m \to \mathrm{H}$, where H denotes Hilbert space. The kernel implicitly performs the dot product calculations between mapped points: $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathrm{H}}$. Now, the SVM can be formulated as following optimization problem:

$$\min_{\mathrm{w},b,\xi} \frac{1}{2} \mathrm{w}^T \mathrm{w} + C \sum_{i=1}^{K} \xi_i, \quad (6)$$

$$\text{s.t. } y^{(i)}(\mathrm{w}^{\mathrm{T}} \phi(x^{(i)}) + b) \geq 1 = \xi_i, \ \xi_i \geq 0, \quad (7)$$

where $\xi_i (i = 1,...,K)$ are slack variables which measure the degree of misclassification of their associated training data points with respect to the current decision boundary and margin.

In our experiments, we applied kernel SVM classifier with the pure distance substitution linear kernel (k^{lin}) and the Gaussian radial basis kernel (k^{rbf}). The values of the penalty error C and the parameter for the Gaussian radial basis kernel are logarithmically varying along a suitable grid and only the best recognition rates are presented in the next section.

4. Experimental results

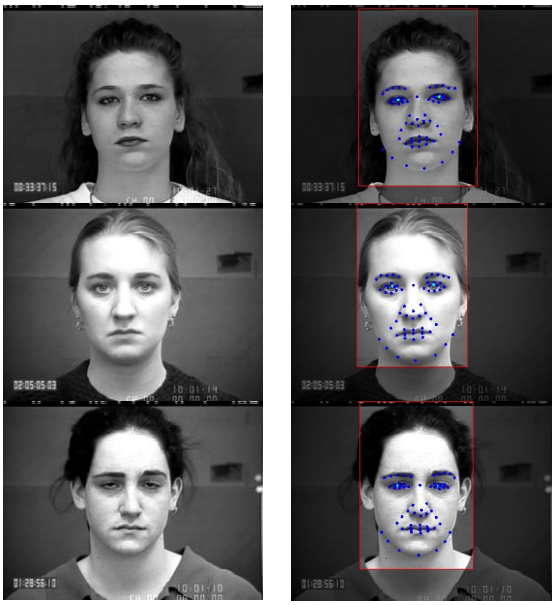
In our experiments we used the extended Cohn-Kanade (CK+) database [3], which contains 593 sequences from 123 subjects. From these, 118 subjects are annotated with 7 universal emotions: anger, contempt, disgust, fear, happy, sad and surprise (Table 1). The corresponding labels are used as the "ground truth" data.

"Table 1. Distribution of emotion labels for 118 subjects from CK+ database".

Emotion	Number images
Anger	45
Contempt	18
Disgust	59
Fear	25
Happy	69
Sad	28
Surprise	83

The dense facial landmarking annotation that we use allows us to achieve excellent fitting accuracy. We use 68 landmark points which are dense enough and optimally placed to not only perform accurate annotation based on the

underlying facial geometry of typical human faces but to also model varied facial expressions. Initialization of an ASM is vital to ensuring high fitting accuracy. This initialization is usually provided by the bounding box around a face that is returned by a face detector. However, face detectors, such as Viola-Jones [14], are prone to errors that result in scale and translation effects in the bounding box. We use OpenCV's face detection implementation which performs well only for frontal faces without in-plane rotation. The 2D profile extension of ASM that we used made the landmark extraction process less sensitive to model initialization errors. Examples from the dataset with the proposed algorithm are illustrated in Figure 3.



"Figure 3. Examples for ASM with 2D profile".

We trained a multi-class SVM using leave-one-subject-out cross validation method in which all images of the test subject were excluded from the training data. The conducted results of the classification accuracy are shown in Table 2. Our goal is to correctly classify the negative emotions related to stress and anxiety: fear, sadness, anger and contempt.

"Table 2. Emotion classification confusion matrix for CK+ database using the proposed method".

%	Anger	contempt	fear	sad	all others
Anger	**80,22**	6,67	0,00	2,22	10,89
Contempt	16,67	**73,78**	5,56	0,00	4,00
Fear	4,00	4,00	**76,00**	4,00	12,00
Sad	10,29	0,00	0,00	**82,14**	7,57
all others	8,89	1,20	0,00	10,11	**79,80**

5. Conclusion

In this paper we proposed a method for automatic facial expression classification in the context of stress/anxiety recognition task. The system is able to detect a human face from still image, extract feature vectors (the coordinates of specific important facial key-points) and then classify expression presented in the face using trained SVM. From the conducted experiments on CK+ database, the classification rate vary between 76% and 82,14%. The system is capable to distinguish between the negative emotions "angry", "sad", "fear" and "contempt" expressions and all the rest.

The facial landmark information is very efficient for facial expression recognition provided that details of coordinates' changes are determined precisely. The current system can be extended by texture based algorithms to improve performance.

6. Acknowledgments

This paper was supported by Contract № 142ПД0028-07 of Technical University-Sofia, Research Sector. Research project: «Algorithms for image segmentation of face components» - 2014.

References

[1] S. Li, A. Jain, *Handbook of Face Recognition*, 2nd ed. Springer, 2011.

[2] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, T.J. Sejnowski, "Classifying Facial Actions", *IEEE Trans. Pattern Analysis and*

Machine Intelligence, Vol. 21, No. 10, pp. 974-989, 1999

[3] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis", *Proc. 4th IEEE Int.Conf. on Automatic Face and Gesture Recognition*, pp. 46–53, 2000

[4] Jia-Yi Dong,Yong-Hong Zhang, Jian Tong; Li-Qiang Qin, Depression and Risk of Stroke: A Meta-Analysis of Prospective Studies, Stroke 1 January 2012: 32-37.

[5] Jonas BS, Franks P, Ingram DD. Are symptoms of anxiety and depression risk factors for hypertension? Longitudinal evidence from the National Health and Nutrition Examination Survey I Epidemiologic Follow-up Study. Arch Fam Med. 1997; 6: 43–49.

[6] Davidson K, Jonas BS, Dixon KE, Markovitz JH. Do depression symptoms predict early hypertension incidence in young adults in the CARDIA study? Coronary Artery Risk Development in Young Adults. Arch Intern Med. 2000;160: 1495–1500.

[7] Mezuk B, Eaton WW, Albrecht S, Golden SH. Depression and type 2 diabetes over the lifespan: A meta-analysis. Diabetes Care. 2008;31: 2383–2390.

[8] Nicholson A, Kuper H, Hemingway H. Depression as an aetiologic and prognostic factor in coronary heart disease: A meta-analysis of 6362 events among 146 538 participants in 54 observational studies. Eur Heart J. 2006; 27:2763–2774.

[9] Albert J, Myles S Faitha, Kelly C Allisona, Depression and obesity, *Biological Psychiatry*, Volume 54, Issue 3, 1 August 2003, Pages 330–337.

[10] A. Lanitis, C.J. Taylor, and T.F. Cootes, "Automatic Interpretation and Coding of Face Images Using Flexible Models", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 743-756, July 1997.

[11] J. C. Gower, "Generalized Procrustes Analysis," *Psychometrika*, vol. 40, no. 1, pp. 33-51, March 1975.

[12] T. F. Cootes, C. J. Taylor, A. Lanitis, "Active Shape Models: Evaluation of a Multi-Resolution Method for Improving Image Search," in *Proc. 5th British Mach. Vision Conf.*, York, 1994, pp. 327-336.

[13] Pei, J. I. A. "2D Statistical Models." (2010).

[14] Viola, P., Michael Jones, J.: Robust real-time object detection. In: *Second International Workshop on Statistical and Computational Theories of Vision-Modeling Learning, Computing, and Sampling* (2001)