Facial Expression Recognition Based on Constrained Local Models and Support Vector Machines

Nikolay Neshov[1], Ivo Draganov[2], Agata Manolova[3]

Abstract **– This paper presents a face expression recognition algorithm using Constrained Local Model (CLM). CLM is facial alignment method that is based on Active Shape Models (ASM) and Active Appearance Models (AAM). It takes the advantages of both of them and gains high accuracy. To distinguish different expression states, we use CLM model parameters that describe shape deformation in a compact form. These parameters form feature vectors for training Kernel Support Vector Machine (KSVM) classifier. The experimental results over Cohn-Kanade Extended Facial Expression (CK+) database show improvement of the recognition rate in comparison to some existing methods, suggested by other authors.**

Keywords – **Constrained Local Model (CLM), Support Vector Machines (SVM), Expression Recognition (ER), Emotion Estimation, OpenIMAJ**

I. Introduction

Facial expression analysis reflects the emotional state of people and hence provides useful information about their personality and psychopathology. The implementation of ER system can play an important role wherever humans interact with machines. This would be of help in various vital purposes, such as security, entertainment, health care, robotics, society, etc. For example, in health care, it becomes useful to integrate a facial ER module in existing surveillance system, which constantly observes patients and analyses their emotional states. Once pain presence is detected (which causes deformations in facial expression) the system would automatically notifies the doctor.

In this paper we describe a facial expression recognition system that is based on Constrained Local Model (CLM) [1] approach. The last one is built by hybridization of two parametric modelling techniques: Active Shape Models [2] and Active Appearance Models [3]. In order to develop our system we utilized the CLM implementation available in OpenIMAJ [4] (Open Intelligent Multimedia Analysis for Java) which is released as an open source under the BSD license. This model can fit a statistical shape to a facial area (detected in image) which is used to locate feature points. CLM is a form of so-called Point Distribution Model (PDM) that consists of non-rigid shape and rigid global transformation parameters. Unlike PDM, which models the appearance of the whole face, CLM takes into account local patches around landmarks of interest. This leads to more generalizability because there is no need to model the complex appearance of the whole face [5].

[1,2,3] Nikolay Neshov, Ivo Draganov and Agata Manolova are with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: nneshov@tu-sofia.bg, idraganov@tu-sofia.bg, amanolova@tu-sofia.bg,

The set of model parameters controls displacement of feature key-points according to the deformation of facial expression. When we align the shape to detected face, we construct a feature vector of these parameters which is further used to determine desired expression.

II. Algorithm Description

The main steps of suggested ER algorithm can be described using the diagram in Fig. 1.

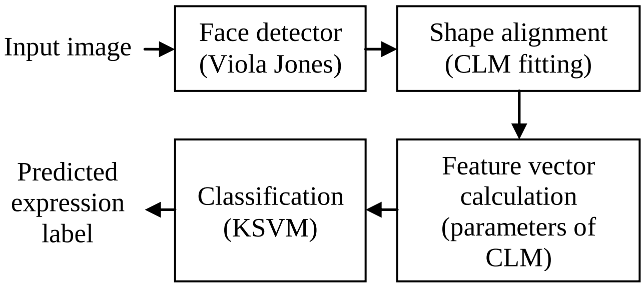

Fig.1. Overall block diagram of suggested ER alogorithm

A. Face detector

The input image is converted from color to grayscale and further examined for human faces, using Viola-Jones face detector [6]. The Viola-Jones approach provides competitive object detection rates in real-time.

B. Shape alignment

Once face location is found, the algorithm proceeds with CLM fitting strategy. In order to detect landmarks, CLM models non-rigid shape variations linearly and composes it with a global rigid transformation, placing the shape in the image frame [7]:

$$\mathbf{x}_i = s\mathbf{R}(\bar{\mathbf{x}}_i + \mathbf{\Phi}_i\mathbf{q}) + \mathbf{t}\,, \qquad (1)$$

where $\bar{\mathbf{x}}_i$ denotes the mean shape associated to $\mathbf{x}_i$ of the PDM's i-th landmark. $\mathbf{\Phi}_i$ are eigenvectors associated to $\mathbf{x}_i$. The set $\mathbf{p}$ = { s, $\mathbf{R}$, $\mathbf{t}$, $\mathbf{q}$ } represents PDM parameters, which contains: a global scaling s, a rotation $\mathbf{R}$, a translation $\mathbf{t}$ and a set of non-rigid parameters $\mathbf{q}$. The objective of CLM can then be interpreted as maximizing the likelihood of the model parameters such that all of its landmarks are aligned with their corresponding locations on the object in an image. Assuming conditional independence between landmark detections, the optimization problem can be represented in the following way:

$$\underset{\mathbf{p}}{\operatorname{argmax}} \prod_{i=1}^{N} p(\{l_i = aligned\} \mid \mathbf{x}_i) \,, \qquad (2)$$

where l_i denotes whether $\mathbf{x}_i$ is aligned (l_i = 1) or not (l_i = 0) and N is the number of local features (landmarks). The optimization strategy used is based on Regularized Landmark Mean-Shift [7] which has been shown as a simple and efficient technic due to the reduction in computational calculations and avoiding the local optima. The implementation of CLM in OpenIMAJ allows detection of 62 facial landmarks.

C. Feature vector calculation

Recall Eq. 1 (by disregarding global scaling, rotation and translation parameters), the weight vector $\mathbf{q}$ for the eigenvectors $\mathbf{\Phi}_i$ can describe essential variations of the mean shape. The fitting process in OpenIMAJ computes 24-dimensional parameter - $\mathbf{q}$. This parameter is used as a feature vector in our work.

D. Classification

Finally the feature vector is classified using trained Support Vector Machine (SVM) classifier with Gaussian kernel (Kernel SVM - KSVM) [8].

III. EXPERIMENTS AND RESULTS

For the experimental evaluation of the proposed ER alogorithm we used Cohn-Kanade Extended Facial Expression Database (CK+) [9]. It contains 118 subjects, annotated by 7 expressions: anger, contempt, disgust, fear, happy, sad and surprise. Expression recognition module evaluation was performed using a leave-one-person-out (LOPO) methodology to separate training and testing parts. We also applied 5-fold cross-validation method to find out the optimal kernel SVM parameters. The experimental results of classification accuracy are shown in Table 1.

TABLE I
EXPRESSION RECOGNITION CLASSIFICATION - CONFUSION MATRIX

%	anger	con-tempt	disgust	fear	happy	sad	sur-prise
anger	**51,11**	0,00	28,89	4,44	2,22	8,89	4,44
contempt	11,11	**16,67**	0,00	33,33	5,56	27,78	5,56
disgust	11,86	0,00	**77,97**	0,00	8,47	0,00	1,69
fear	4,00	4,00	4,00	**60,00**	12,00	8,00	8,00
happy	0,00	1,45	5,80	1,45	**91,30**	0,00	0,00
sad	17,86	14,29	3,57	10,71	0,00	**39,29**	14,29
surprise	0,00	2,41	1,20	2,41	0,00	3,61	**90,36**
Avg.	60,96						

It is obvious that expressions with large displacement of landmark locations (e.g. happy and surprise) resulted in more than 90,3% correct classification. For the worst case (contempt), the accuracy is 16,67%. Looking at the results reached by algorithms suggested by another authors (e.g. Active Appearance Model (AAM) that utilizes shape and texture information [9]) it can be seen that in our work the average recognition accuracy is higher (from 50,3% to 60,96%).

Following below are some typical visual examples of incorrectly recognized emotions. In Fig. 2 is shown a face expressing disgust with all feature points detected correctly but still the feature vector falls into the class representing anger. This is the most typical case for this type of detector – where the preprocessing phase proceeds normally and the classification stage produce a faulty result.

Fig.2. Disgust emotion confused with anger while all feature points are detected correctly

Another example for misclassification is shown in Fig. 3.

Fig.3. Anger emotion confused with disgust where eyes' feature points are incorrectly detected

There the detection of eyes' contours is off-center in relation to their real position. In this case the anger is confused with disgust. There are other examples of this displacement for other emotions. The reason is the low contrast of the eyes' area due to the nature of the lighting conditions. It's the second most frequently met case leading to errors.

In Fig. 4 is shown considerable displacement of the mouth's feature points covering only about a half of the targeted area. The reason is the misleading effect from the lighter area of the teeth right below the darker upper lip and followed by the darker area of the tongue which easily is mistaken with the lower lip. This case appears with almost the same frequency as the latter.

Fig.4. Surprise emotion confused with fear where lips' and partially face countor around the chin feature points are detected incorrectly

In Fig.5 is shown similar effect but here the feature points outline an area higher than expected.

Fig.5. Sadness emotion confused with fear where lips' feature points oclude an area higher than the mouth and right part of the face contour is slightly translated to left

The eyebrows' feature points form curves sometimes laying below (Fig.6) or above (Fig.7) the eyebrows themselves. It causes misclassification of emotions of different types but more rarely than the previous misdetections. In some cases (Fig. 8) there is no error introduced by this effect.

A more rare case is when the mouth's feature points fall below almost the entire lower lip or partially covers it (Fig.9).

Fig.6. Surprise emotion confused with sadness where eyebrows' feature points fall below the upper boundaries of the eyebrows themselves. Left part of the face contour is slightly displaced to left

Fig.7. Disgust emotion confused with anger where eyebrows' feature points arised above the upper boundaries of the eyebrows themselves. The other features are correctly detected

Fig.8. Correctly recognized disgust emotion while eyebrows' feature points are well above the upper line of the eyebrows themselves. Eyes' contours are slightly displaced below as well

Fig.9. Sadness emotion confused with surprise where mouth's feature points are considerably shifted belowc overing only portion of the lower lip

The opposite case when mouth is detected high above its original position (Fig. 10) is also not that frequently taking place.

Fig.10. Sadness emotion confused with surprise where mouth's feature points are considerably shifted belowcovering only portion of the lower lip

Fig.11. Sadness emotion confused with contempt – eyes' contours are slightly folded where the lower part is lifted up by around1/3 of the eye's height

In this case it's possible to have some of the nose points also incorrectly located which additionally minister to misclassification of the emotions, e.g. of sadness with surprise.

In Fig. 11 is given probably the rarest case where eyes' feature points do not extend enough to the full height of the eyes themselves. Again some emotions may not be recognized correctly.

The close examination of all these cases (Fig.2-Fig.11) may very well help in the further development of the tested approach to get more accurate results.

IV. CONCLUSION

In this paper we presented an algorithm for expression classification that can distinguish seven expressions. Comparing the results over CK+ database with some other state-of-art methods we got improvement of accuracy rate more than 10%.

ACKNOWLEDGEMENT

This paper is under the support of the research project № 151ПР0013-07, which is funded by RDS, TU-Sofia, Bulgaria.

REFERENCES

[1] J. Saragih, S. Lucey, J. Cohn, Face Alignment through Subspace Constrained Mean-Shifts, IEEE International Conference on Computer Vision (ICCV), 2009.

[2] B. T. Cootes, C. J. Taylor, D. H. Cooper, J. Graham. Active shape models their training and application. Comput. Vis. Image Underst., vol. 61 no. 1, pp. 38–59, Jan. 1995.

[3] T. Cootes, G. Edwards, C. Taylor. Active appearance models. IEEE Transactions On Pattern Analysis And Machine Intelligence, vol. 23, no. 6, pp. 681–685, June 2001.

[4] J. Hare, S. Samangooei, D. Dupplaw. OpenIMAJ and ImageTerrier: Java libraries and tools for scalable multimedia analysis and indexing of images. In Proceedings of ACM Multimedia 2011, MM '11, pp. 691–694. ACM, 2011.

[5] T. Baltrusaitis, P. Robinson, L. P. Morency, 3D constrained local model for rigid and nonrigid facial tracking. In: CVPR, 2013.

[6] P. Viola, M. J. Jones, Robust real-time face detection, Int. J. Comput. Vis., vol. 57, no. 2, pp. 137–154, 2004.

[7] J. M. Saragih, S. Lucey, J. F. Cohn. Face alignment through subspace constrained mean-shifts. IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 - October 4, 2009. IEEE. pp. 1034-1041, 2009.

[8] K. Crammer, Y. Singer, On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines, J. of Machine Learning Research 2, pp. 265–292, 2001.

[9] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, pp. 94-101, 2010.