

FACIAL LANDMARKING LOCALIZATION FOR EMOTION RECOGNITION USING BAYESIAN SHAPE MODELS

Hernan F. Garcia¹, Alejandro T. Valencia¹ and Alvaro A. Orozco¹

Technological University of Pereira, La Julita Periera, Risaralda, Colombia,
hernan.garcia@utp.edu.co, cristian.torres@utp.edu.co,
aaog@utp.edu.co

ABSTRACT

This work presents a framework for emotion recognition, based in facial expression analysis using Bayesian Shape Models (BSM) for facial landmarking localization. The Facial Action Coding System (FACS) compliant facial feature tracking based on Bayesian Shape Model. The BSM estimate the parameters of the model with an implementation of the EM algorithm. We describe the characterization methodology from parametric model and evaluated the accuracy for feature detection and estimation of the parameters associated with facial expressions, analyzing its robustness in pose and local variations. Then, a methodology for emotion characterization is introduced to perform the recognition. The experimental results show that the proposed model can effectively detect the different facial expressions. Outperforming conventional approaches for emotion recognition obtaining high performance results in the estimation of emotion present in a determined subject. The model used and characterization methodology showed efficient to detect the emotion type in 95.6% of the cases.

KEYWORDS

Facial landmarking, Bayesian framework, Fitting, Action Unit

1. INTRODUCTION

This work is focus to the problem of emotion recognition in images sequences with varying facial expression, illumination and change pose in a fully automatic way. Although some progress has already been made in emotion recognition, several unsolved issues still exist. For example, it is still an open problem which features are the most important for emotion recognition. It is a subject that was seldom studied in computer science. In order to define which facial features are most important to recognize the emotion type, we rely on Ekman study to develop a methodology which be able to perform a emotion recognition trough a robust facial expression analysis [1].

Although there have been a variety of works in the emotion recognition field, the results are not yet optimal, because the techniques used for facial expressions analysis are not robust enough. Busso et al. [2], presents a system for recognizing emotions through facial expressions displayed in live video streams and video sequences. However, works such as those presented in [3] suggest

that developing a good methodology of emotional states characterization based on facial expressions, leads to more robust recognition systems. Facial Action Coding System (FACS) proposed by Ekman et al. [1], is a comprehensive and anatomically based system that is used to measure all visually discernible facial movements in terms of atomic facial actions called Action Units (AUs). As AUs are independent of interpretation, they can be used for any high-level decision-making process, including the recognition of basic emotions according to Emotional FACS (EM-FACS), the recognition of various affective states according to the FACS Affect Interpretation Database (FACSAID) introduced by Ekman et al. [4], [5], [6], [7]. From the detected features, it is possible to estimate the emotion present in a particular subject, based on an analysis of estimated facial expression shape in comparison to a set of facial expressions of each emotion [3], [8], [9].

In this work, we develop a novel technique for emotion recognition based on computer vision techniques by analysing the visual answers of the facial expression, from the variations that present the facial features, specially those in regions of FACS. Those features are detected by using bayesian estimation techniques as Bayesian Shape Model (BSM) proposed on [10], which from the a-priori object knowledge (face to be analyzed) and being help by a parametric model (Candide3 in this case) [11], allow to estimate the object shape with a high accuracy level. The motivation for its conception, is an extension of the Active Shape Model (ASM) and Active Appearance Model (AAM) which predominantly considers the shape of an object class [12] [13]. Although facial feature detection is widely used in facial recognition tasks, databases searching, image restoration and other models based on coding sequences of images containing human faces. The accuracy which these features are detected does not reach high percentages of performance due some issues present at the scene as the lighting changes, variations in pose of the face in the scene or elements that produce some kind of occlusion. Therefore, it is important to analyze how useful would be the image processing techniques developed for this work. The rest of the paper is arranged as follows. Section 2 provides a detailed discussion of model-based facial feature extraction. Section 3 presents our emotion characterization method. Sections 4 and 5 discuss the experimental setup and results respectively. The paper concludes in Section 6, with a summary and discussion for future research.

2. A BAYESIAN FORMULATION TO SHAPE REGISTRATION

The probabilistic formulation of shape registration problem contains two models: one denotes the prior shape distribution in tangent shape space and the other is a likelihood model in image shape space. Based on these two models we derive the posterior distribution of model parameters.

2.1 Tangent Space Approximation

Assuming that a planar shape is described by N landmark points in the image, we can represent it by a $2N$ -dimensional vector s_i . The difference between two planar shapes is usually measured by their Procrustes distance [14]. Furthermore, given a set of training shape vectors $\{s_i\}_{i=1}^L$ the most popular way to align them into a common co-ordinate frame is Generalized Procrustes Analysis (GPA) [15]. The procedure essentially equals to minimize a quadratic loss function defined by $L(\mu) = \sum_{\text{all } i} \|T_i(s_i) - \mu\|^2$, where $T_i(s_i)$ is a $2D$ similarity transform of s_i .

The tangent space is a linear approximation of the general shape space in the vicinity of the mean shape vector. More specifically, the tangent space C_μ is defined as the space normal to $T(\mu)$ and passing through μ . The Euclidean distance in the tangent space is a good approximation to the Procrustes distance, if most of shape instances are close. s_i can be transformed onto C_μ by aligning s_i with μ as $\{x_i \in C_\mu : x_i = T_i(s_i)\}_{i=1}^L$. x_i is often referred as tangent shape vector and represented as a $2N$ -dimensional vector. The residuals are computed as $\{t_i = x_i - \mu\}_{i=1}^L$ in tangent space instead of image space, to remove the difference introduced by similarity transform T_i .

Note that the dimension of C_u is $2N - 4$, where the degenerated dimensionality is corresponding to the degree of freedom of similarity transformation in a $2D$ Euclidean space. Furthermore, since any transformed shape vector from μ can be represented by a linear combination of $\{e, e^*, \mu, \mu^*\}^1$, the complement space of C_μ is spanned by $\{e, e^*, \mu, \mu^*\}$. Therefore, the covariance matrix of tangent shape,

$$Var(X) = \frac{1}{L-1} \sum_{i=1}^L (x_i - \mu)(x_i - \mu)^T \quad (1)$$

will has at least four zero eigenvalues with corresponding eigenvectors $\{e, e^*, \mu, \mu^*\}$. In other words, the tangent shape variances in this complement space must be zero. Figure 1 shows the align process of the landmark points for the training set using *Procrustes Analysis* [15].

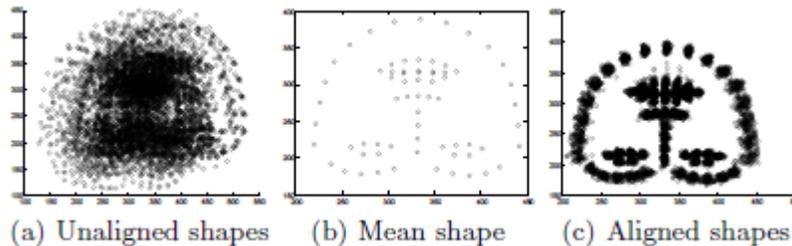


Fig. 1. The mean face model and all the training sets normalized face models [16]

3. EMOTION CHARACTERIZATION

3.1 Facial Expression Analysis

Facial expressions are generated by contraction or relaxation of facial muscles or by other physiological processes such as coloring of the skin, tears in the eyes or sweat on the skin. We restrict ourselves to the contraction or relaxation of the muscles category. The contraction or relaxation of muscles can be described by a system of 43 Action Units (AUs) [1]

Labeling of the Emotion Type After choosing that AUs are used to represent the facial expressions that correspond to each emotion, we chose Candide3 vertexes used to define the

configuration of the face and the calculation of the FACS [11]. Figure 2 shows the shapes of the Candide3 models playing each emotion.

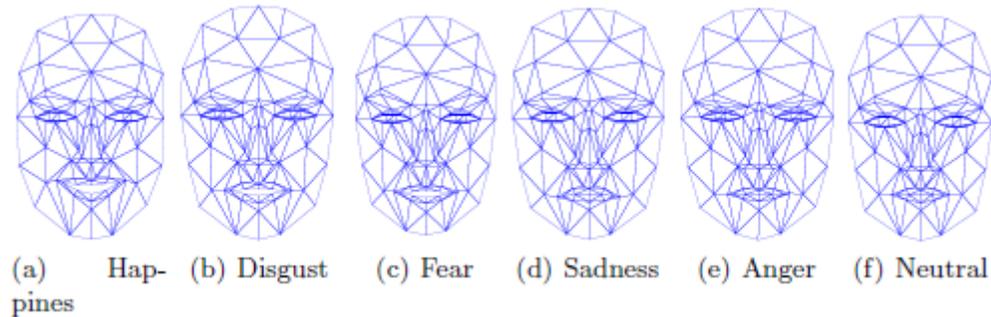


Fig. 2. Shape of each emotion type

4. EXPERIMENTAL SETUP

In this work we used the MUCT database which consists of 3755 faces from 276 subjects with 76 manual landmarks [16].

4.1 Evaluation Metrics

The shape model will be estimated for every subject, contemplating 5 changes of pose exposure (-35° and 35° in the axis X and $(-25^{\circ}$ and $25^{\circ})$ in the axis Y to simulate the nods) and 1 frontal, with which 1124 images will be had for changes of pose exposure and 276 frontal images. The first step is to compute the average error of the distance between the manually labeled landmark points p_i and points estimated by the model \hat{p}_i for all training and test images. We compute the relative error between the manually labeled landmark points and landmark points estimated by the model for the eyelids region [17].

4.2 Landmarking The Training Set

The initial step is the selection of images to incorporate into the training set. In order to decide which images will be include in the training set, the desired variations must be considered. In order to include the whole region of the labeled face, there will be used a set of landmarks, which consist in a group of 76 landmark points that depict the whole face regions in detail (eyes, nose, mouth, chin, etc.) [16].

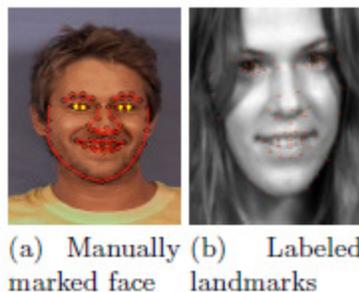


Fig. 3. Samples of the landmarking process

After estimate the landmark points of BSM, procrustes analysis is performed to align an estimated facial expression with respect to standard shape (Neutral emotion). Then, we extract all the combinations of the estimated vertexes model and perform combinations of these sets (complete set, set of eyes, eyebrows and mouth).

5. RESULTS

5.1 Bayesian shape Model Estimation Error

In Table 5.1, It can be seen that although the accuracy in the estimation of the points is greater for images of the training set, the average error is also small for the test images. This is due to a rigorous procedure in the training and model building in which there were considered to be the biggest quantity of possible forms to estimate. Moreover, it is noted that although the average error for images with changes in pose is a bit higher than in the case of frontal images, this indicates that the accuracy in estimating the model is so high for images that show changes in the pose of the face. Also, it is of highlighting that the times average of estimation of the model BSM, they are relatively small which would help in applications on line.

5.2 Distribution on the Relative Error

Figure 4(a) shows the distribution function of the relative error against successful detection rate, on which it is observed that for a relative error of 0:091 in the case of the adjustment of the right eye, 0:084 for the left eye and 0:125 for the mouth region in frontal images, the detection rate is 100%, indicating that the accuracy in the BSM model fit is high. In addition to images with pose variations in BSM is achieved 100% for adjusting the eye region for relative errors to 0:120, 0:123 and 0:13 for the mouth region; being this much lower than the established criterion of 0:25. It is considered that the criterion $Rerr < 0:25$ is not adapted to establish a detection like correct and can be not very suitable when it is desirable to realize facial features detection in images with minor scale. That's why it is considered to be a successful detection for errors $Rerr < 0:15$ [18].

Table 1. Average estimation error for BSM

N	Front		Pose Changes		
	Error [pix]	T [ms]	Error [pix]	T [ms]	
Training	1120	3.27 ± 0.276	8.9	3.53 ± 0.277	9.1
Test	280	3.31 ± 0.280	8.3	3.44 ± 0.254	8.8

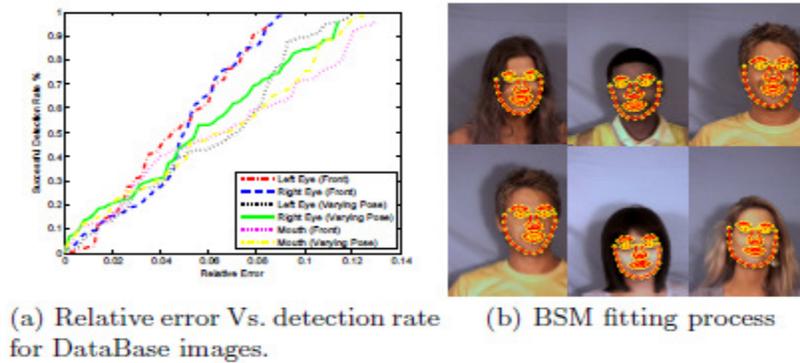


Fig. 4. Shape of each emotion type.

Finally a set of fitting images by BSM model is shown in Figure 4(b) evidencing the robustness of the proposed model.

5.3 Emotion Recognition

Table 2 shows the average success rates in the emotion detection obtained for all sets used in this analysis using the MUCT database proving the high accuracy of the RMSE metric. Another important factor in this parsing is the fact that the sets which provide more evidence on the emotional estimation state are the sets of B;C (mouth and eyebrow separated) with 95:6% of average success rate, B;C;O (mouth, eyebrows and eyes separate set) with 92:2%, CO;B (Eyebrow, Eyes together and Mouth separately) with 91:1%, and where the first of these provides more discriminant results for the estimation of the emotional detection rate.

Table 2. Successful rate emotion recognition for MUCT database

Set	Metric performed		
	Average RMSE	Average Correlation	Average Euclidean
Total	86.7%	72.2%	71.1%
COB	83.3%	80.0%	78.9%
CO	75.6%	71.1%	71.1%
B	75.6%	71.1%	63.3%
C	83.3%	83.3%	78.9%
O	63.3%	54.4%	54.4%
CB	84.4%	83.3%	81.1%
OB	75.6%	63.3%	54.4%
B,C,O	92.2%	91.1%	85.6%
BC,O	83.3%	83.3%	77.8%
BO,C	90.0%	86.7%	71.1%
CO,B	91.1%	84.4%	85.6%
B,C	95.6%	92.2%	85.6%
B,O	77.8%	74.4%	71.1%
C,O	77.8%	78.9%	78.9%

6. CONCLUSION AND FUTURE WORKS

This paper presents a bayesian shape model for emotion recognition using facial expression analysis. By projecting shape to tangent shape, we have built the model describing the prior distribution of face shapes and their likelihood. We have developed the BSM algorithm to uncover the shape parameters and transformation parameters of an arbitrary figure. The analyzed regions correspond to the eyes, eyebrows and mouth region on which detection and tracking of features is done using BSM. The results show that the estimation of the points is exact and complies with the requests for this type of systems. Through quantitative analysis, the robustness

of the BSM model in feature detection is evaluated, and it is maintained in nominal pose for a range between $[-40^0$ to $40^0]$ on X and $[-20^0$ to $20^0]$ for Y . The used model and the methodology of characterization showed efficiency to detect the emotion type in 95:55% of the evaluated cases, showing a high performance in the sets analysis for the emotion recognition. In addition, due to high accuracy in detecting and characterizing features proposed to estimate parameters associated with the emotion type, the designed system has great potential for emotion recognition, being of great interest in human/computer research systems.

Due to the current problem of facial pose changes in the scene, it would be of great interest to design a 3D model that will be robust to these variations. Also, Taking advantage of sets analysis demonstrated in this work, it would be interesting to develop a system that in addition to use the sets analysis, add an adaptive methodology to estimate the status of each facial action unit to improve performance of the emotion recognition system.

REFERENCES

- [1] Ekman, P.: Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. 2nd edn. Owl Books, 175 Fifth Avenue, New York (2007)
- [2] Carlos, B., Zhigang, D., Serdar, Y., Murtaza, B., Chul Min, L., Abe, K., Sungbok, L., Ulrich, N., Shrikanth, N.: Analysis of emotion recognition using facial expressions, speech and multimodal information. In: in Sixth International Conference on Multimodal Interfaces ICMI 2004, ACM Press (2004) 205{211
- [3] Song, M., You, M., Li, N., Chen, C.: A robust multimodal approach for emotion recognition. Neurocomput. 71 (2008) 1913{1920
- [4] Ekman, P., Rosenberg, E.: What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford Univ. Press (2005)
- [5] Griesser, R.T., Cunningham, D.W., Wallraven, C., B ultho, H.H.: Psychophysical investigation of facial expressions using computer animated faces. In: Proceedings of the 4th symposium on Applied perception in graphics and visualization. APGV '07, ACM (2007) 11{18
- [6] Cheon, Y., Kim, D.: A natural facial expression recognition using differential aam and knns. In: Proceedings of the 2008 Tenth IEEE International Symposium on Multimedia. ISM '08, Washington, DC, USA, IEEE Computer Society (2008) 220{227
- [7] Cheon, Y., Kim, D.: Natural facial expression recognition using differential-aam and manifold learning. Pattern Recogn. 42 (2009) 1340{1350
- [8] Ioannou, S., Raouzaïou, A., Tzouvaras, V., Mailis, T., Karpouzis, K., Kollias, S.: Emotion recognition through facial expression analysis based on a neurofuzzy network. Elsevier 18 (2005) 423{435
- [9] Cohen, I., Garg, A., Huang, T.S.: Emotion recognition from facial expressions using multilevel hmm. In: in In Neural Information Processing Systems. (2000)
- [10] Xue, Z., Li, S.Z., Teoh, E.K.: Bayesian shape model for facial feature extraction and recognition. Pattern Recognition 36 (2004) 2819{2833
- [11] Ahlberg, J.: Candide-3 - an updated parameterised face. Technical report, Report No. LiTH-ISY-R-2326, Dept. of Electrical Engineering, Linköping University, Sweden (2001)
- [12] Cootes, T., Taylor, C., Pt, M.M.: Statistical models of appearance for computer vision (2004)
- [13] Matthews, I., Baker, S.: Active appearance models revisited. Int. J. Comput. Vision 60 (2004) 135{164
- [14] Gower, J.C.: Generalised Procrustes analysis. Psychometrika 40 (1975) 33{51
- [15] Larsen, R.: Functional 2D procrustes shape analysis. In: 14th Scandinavian Conference on Image Analysis. Volume 3540 of LNCS., Berlin Heidelberg, Springer Verlag (2005) 205{213
- [16] Milborrow, S., Morkel, J., Nicolls, F.: The MUCT Landmarked Face Database. Pattern Recognition Association of South Africa (2010) <http://www.milbo.org/muct>.
- [17] Hassaballah M., I.S.: Eye detection using intensity and appearance information. In: IAPR Conference on Machine Vision Applications. (2009)
- [18] M., H., Ido, S.: Eye detection using intensity and appearance information. (2009) 801809