Using Bezier Curve analysis in context of Expression Analysis

Previste Modeste

***Abstract*—Affective computing is an area of research under increasing demand in the field of computer vision. Expression analysis, in particular, is a topic that has been undergoing research for many years. In this paper, an algorithm for expression determination and analysis is performed for the detection of seven expressions: sadness, anger, happiness, neutral, fear, disgust and surprise. First, the 68 landmarks of the face are detected and the face is realigned and warped to obtain a new image. Next, feature extraction is performed using LPQ. We then use a dimensionality reduction algorithm followed by a dual RBF-SVM and Adaboost classification algorithm to find the interest points in the features extracted. We then plot bezier curves on the regions of interest obtained. The curves are then used as the input to a CNN and this determines the facial expression. The results showed the algorithm to be extremely successful.**

I. Introduction

Face recognition and analysis is a study that has been going on for many years. Applications for face recognition range from commercial ones to security. However, a system must be trained to recognise a face. Even if trained there are a number of factors that limit the opportunity to recognise a face as shown in [1]. These include: image quality, lighting conditions of the image etc. There are different methods that are available for face recognition as mentioned by Zhao et al [2]. The two main approaches in these are: the holistic approach and the feature extraction approach. In case of the holistic approaches, face recognition is done by making use of single feature vector that then represents the whole face image. Examples of holistic approaches are the eigenfaces as proposed by Turker et al [3], the linear discriminant analysis as proposed by Martinez et al [4], the discriminative common vector as proposed by Cevikalp et al [5], the bayesian intrapersonal classifier as proposed by Moghaddam et al [6], and the classifiers trained by Neural networks as proposed by Rowley et al [7].

From here onwards, facial expression analysis means that the computer system makes an attempt to automatically recognize and analyze facial motion and feature changes from simple visual information. Computer facial expression analysis systems need to analyze the facial actions regardless of context, culture, gender, and so on. The achievements in the related areas for example human movement analysis, psychological studies, face tracking, face detection and recognition, make automatic facial expression analysis plausible to a very high degree.

Facial expression analysis includes measurement of facial motion as well as recognition of expression. A general approach to automatic facial expression analysis consists of three steps: face acquisition from the image, facial data extraction and representation of it, and finally facial expression recognition. Face acquisition is the pre-processing stage used to automatically obtain the region containing the face for the input images. It could be a detector to detect the face for each frame or just detect the face in the first frame and further track the face as in the case of a video sequence.

After the face is located, the next step is to extract and simultaneously represent changes caused by expressions. In case of facial feature extraction for expression analysis there are mainly two different types of approaches: appearance based methods and geometric feature-based methods. The geometric facial features represent the locations and shapes of facial components. The facial components also called as feature points are extracted to form a feature vector that is then used to represent the face geometry.

With appearance based methods, however, filters, such as the Gabor filters are applied to the whole-face or a part of a face image used to extract a feature vector. Chapter 2 presents an outline about recent trends in both fiducial points detection and also facial expression analysis. Chapter 3 talks about the proposed algorithm. Chapter 4 shows us the results of the algorithm and compare it with some of the recent, most efficient ones. Chapter 5 talks about the conclusion of the entire experiment and what possibilities remain to make the algorithm possibly even more efficient than it currently is.

II. Related work

S M Lajevardi et al. [8] worked on color images instead of the usual gray scale and they found out that the color components in color images provided more information than gray scale images and this helped to achieve an improved and robust facial expression recognition system. They introduced a tensor perceptual color framework abbreviated as TPCF for facial expression recognition (FER), which is based on the information contained in the color facial images. The TPCF enables multi-linear image study in different colour spaces and proves their study that colour components provides additional information for a robust FER algorithm.

S Fazli et al. [9] found that in their study, if the number of samples taken is less in comparison to dimensionality of the image, then the Linear Discrimination Analysis (LDA) alone is insufficient for feature-reduction. To enhance performance, Principal Component Analysis should be used before LDA.

Li et al [10] proposed a method which included a cascade of fixed filters and a set of trainable non-linear 2-D filters, which were based on a biological structure of avoiding inhibition. The fixed filters were utilized to obtain the primitive or basic features, whereas the adaptive filters were made to train to extract the more complicated facial features for classification by the SVMs.

Zhang et al [11] proposed a consolidated framework for a study to compare some of the widely used textures and geometric features making use of mRMR, Adaboost and SVM feature selection algorithm. Their experiments demonstrated the positives of fusing texture and geometric features.

Gao et al [12] presented a method for FER from a single static image by utilizing line-based caricatures. The recognition process was automatic completely. The proposed approach used geometrical and structural features of user sketched expression model to match the line edge map descriptor of input face image.

Tian et al [13] worked on an Automatic Face Analysis (AFA) system to study the facial expressions based on both the permanent facial features such as eyes, brows, mouth and the ephemeral facial features such as deepening of facial furrows in a near frontal view face-image sequence. The Automatic Face Analysis system recognized fine-grained alterations in facial expression into action units of the Facial Action Coding System instead of a few prototypic expressions. This was shown to be better than the compared algorithms in terms of accuracy. This work also proposed multistate facial component models for modeling and tracking the various facial features such as eyes, lips, brows, furrows and cheeks. During tracking, precise parametric descriptions of the facial features were obtained. Using this, the expressions could be easily extracted from the image. Another important part of this experiment is determining the fiducial points of an image. In this case, the fiducial points refer to the points of interest in the facial features for drawing the bezier curve.

The amount of research work being done to detect fiducial points on the face is constantly increasing as shown by Waldir et al [14]. In recent times, the approaches for this purpose could be divided into two main categories namely: local and global. In local methods individual fiducial points are discovered and then processed and no additional information is utilised or needed.

The global methods are described by detecting more fiducial points in comparison, for this they use deformable models which are much lesser susceptible to pose and illumination variations than the local methods. The design of a classifier is probably the pivotal stage of a fiducial point detection algorithm.

In this stage several different machine learning algorithms could be utilized as shown by Jahanbin et all [15]. There are many recent research papers with regards to Support Vector Machines. For example, Silva et al [14] propose a face recognition subsystem framework that makes use of fiducial points detection. The detection of the fiducial points is a combination of two different techniques. The first is by using Gabor filters for local detection and then following is the use of a human face anthropometric measurement.

The system proposed by Araujo et al [16] also explores the exact problem. They used classifiers based on Inner Detector Product correlation filters. These filters are designed by making use of principal components. In [17] Eduardo et al proposed the use of a SVM mathematical formulation called C-SVC [18], for fiducial point detection. In [19] S N Gowda et al extended the idea proposed by [17] and used a dual classification scheme to increase accuracy of the detection of fiducial points. Adaboost was used along RBF-SVM for the approach.

III. PROPOSED APPROACH

A. *Image Normalization*

Not all images are correctly aligned. Keeping in mind the end goal to take out the shape varieties, for example, head posture, scale etc image normalization is applied on facial pictures. Image normalization is done by warping the images obtained from the training set into the mean shape obtained. 68 fiducial points are marked on the image using the algorithm proposed in [24]. All the images are warped into their mean shape by utilizing Delaunay's method [25] (Fig. 1-b) along with affine transformations. As the head poses of facial images are not correctly aligned, the warped images need to be rotated as the last part of the image normalization process (Fig.1-c).

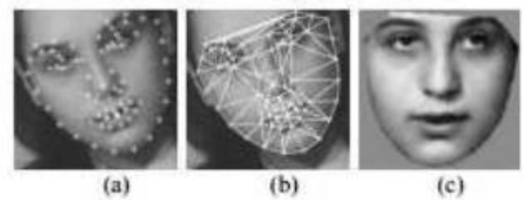

Figure 1. a) 68 landmark points on facial image b) Delaunay's method output c) image warped into mean shape d) face aligned

B. *Facial Feature Extraction*

We use LPQ (Local Phase Quanitization) as our descriptor for feature extraction. LPQ is a texture descriptor that was developed on the blur invariance property of the Fourier spectrum [26]. The LPQ codes are obtained by computing discrete fourier transforms in local image windows and represented as a histogram. Consider B(u), O(u) and P(u) as the discrete fourier transforms of the blurred image, original image and Point Spread Function, then we can obtain a relation between the three as shown in (1),

$$B(u) = O(u).F(u) \tag{1}$$

On consideration of the phase of the spectrum we obtain (2),

$$\angle B(u) = \angle O(u) + \angle F(u) \tag{2}$$

Let us consider the blur Point Spread Function as p(x) and assume that p(x) is centrally symmetric, this implies p(x) = p(x), and that the Fourier transform is always real valued. If the NxN neighborhood around a pixel x is denoted as Nx, the two dimensional (2-D) DFT of Nx is defined by

$$F(u,x) = \sum_{y \epsilon N_x} f(x-y)e^{-2\pi u^T y} = w_u^T f_x \tag{3}$$

where wu is the basis vector of the 2-D DFT at some frequency 'u' and fx is the vector containing all NxN pixels in Nx. The LPQ feature descriptor describes the input image as a histogram containing 256 bins, without considering the spatial information regarding the pixels. To enhance the discriminative power of the descriptor, spatial histograms are produced.

C. Kernel PCA for dimensionality reduction

Dimensionality reduction is an important phase in any machine learning algorithm, especially in terms of increasing the speed of the algorithm by reducing the number of random variables. To aid with this, we utilise the kernel principal component analysis method. This is a non linear dimensionality reduction technique and was chosen for convenience.

D. Interest point detection on Facial features

For this step, we use the method proposed in [19]. 11 fiducial points were obtained as can be seen in figure 2(a). The points were divided into regions as seen in Figure 2(b).

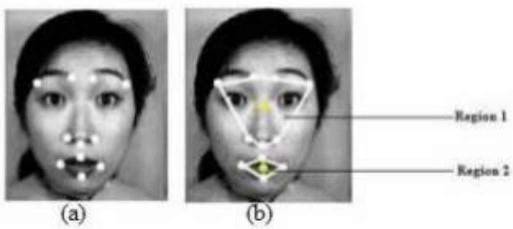

Figure 2. a) 11 fiducial points on facial image b) divided regions depicted

The interest points obtained, in turn are used for developing the bezier curve. The algorithm for choosing of the fiducial points was decided on the basis of speed and also that 11 points were sufficient for the bezier curve to provide a good rate of accuracy.

E. Bezier Curve

The bezier curve is a parametric curve used to model smooth curves that can be scaled indefinitely. A bezier curve can be drawn over any number of points. We draw two bezier curves over regions 1 and 2 specified in figure 2(b). The Bezier curve generates a smooth curve considering global shape information.

The curve passes through the first and last control points and since we want a closed curve, we use the same point as first and last control point. If there are L+1 control points, the position is defined as Pk : (xk, yk),[k in range (0,L)] considering 2-D shapes. The coordinate points are taken and later blended to form P(t), which describes the path of Bezier polynomial function between P0 and PL as seen in (4):

$$P(t) = \sum_{k=0}^{L} P_k BEZ_{k,L}(t) \tag{4}$$

The Bezier blending function is represented as BEZ and shown in (5),

$$BEZ_{k,L}(t) = (\frac{L}{k}) t^k (1-t)^{L-k} \tag{5}$$

$$BEZ_{k,k}(t) = (\frac{k}{k}) t^k (1-t)^{k-k} = t^k \tag{6}$$

F. Prediction Phase

Convolutional Neural Networks (CNNs) have had great successes with regards to image classification, face detection etc. We use a fully convolutional network. The input to the proposed CNN is an image that contains only the bezier curve. The CNN gives us a score of similarity between the input image and all the 7 expressions in consideration.

The 7 expressions in consideration are: happiness, surprise, sadness, neutral, anger, fear and disgust.

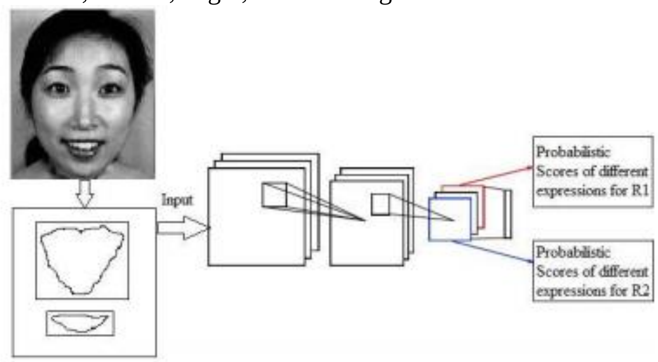

Figure 3. Architecture of proposed system

The CNN is trained with using a set of 497 images taken from JAFFE [21] and AR [23]. The different expressions are grouped before hand and training is done for each expression. For each group, one individual Bezier curve image is developed by taking the mean of all pixel positions of the curve for individual images. As an example, consider the happiness group. In this group assume there are ’z’ images. This implies we will have ’z’ bezier curve images as ouput.

To obtain 1 standard image for the testing phase we create 1 image by taking the average of all pixels for each individual output. If we consider Pi as the pixel value for the ith image and consider only pixels within the bezier curve, then Pmain = average (Pi) for i in range [0,z]. We repeat this process until each expression has one standard bezier curve representing it.

We use ordinary procrustes analysis to obtain the similarity score from the CNN. Since, we have 2 curves that are not of the exact dimensions and cannot be directly compared we try and fit them together for the comparison. This means we perform translation, rotation, scaling and then compare the shapes.

4 steps are involved in the ordinary procrustes analysis,

1) Translation: Translational components could be terminated from an object by translating the object. This would make the centroid of all the points lie at the origin.

2) Scaling: Similarly, the scale component could be removed by scaling the object such that the root mean square distance from the points to the translated origin is 1.

3) Rotating: The rotating component is much more complicated than the translational and scaling component. Assume 2 objects whose scale and translational components have been removed.

One of the objects is used as a reference orientation. The second object is rotated around that.

4) Shape Comparison: The difference between the shapes of the objects is calculated can after ”superimposing” the two objects by scaling, translating and optimally rotating. The square root of the sum of squared distances between corresponding points can be used as a statistical measure of this difference in shape as seen in (7):

$$d^2 = \sum P_{i_{std}} - P_{i_{test}} \tag{7}$$

Here, Pistd is any point on the standard bezier curve developed and Pitest is the corresponding point on the test bezier curve. ’d’ is calculated for both regions and summed up tp give ’D’. ’D’ is the score assigned by the CNN to the test image regarding each possible expression.

So for a test image there will be 14 values for ’D’, 2 per expression and 1 for each region. The classification to a particular expression is based on the highest value of ’D’ for an expression.

IV. EXPERIMENTAL ANALYSIS

We compare the algorithm for the same JAFFE database and AR database with 4 other algorithms namely: SVM with a Radial Basis Function (RBF), Template matching with LBP features, LDA with PCA [9] and Gabor Filter based [11]. We construct the confusion matrix for each process to determine the strength and accuracy of the algorithm. Figure 4 represents the confusion matrix for LBP Features[22].

	Anger	Disgust	Fear	Joy	Sadness	Surprise	Neutral
Anger	66.8	3.9	1.9	0	7.3	0	20.1
Disgust	0	92.5	2.5	0	0	0	5
Fear	0	0	70	17	3.0	0	10
Joy	0	0	2.5	90.1	0	0	7.4
Sadness	6.4	0	0	0	61.2	0.8	31.6
Surprise	0	0	1.3	0	0.5	92.5	5.7
Neutral	0	0	0.8	0.4	3.6	0	95.2

Figure 4. Confusion Matrix for LBP Features [22]

Figure 5 represents matrix for SVM with RBF. Figure 6 represents confusion matrix for Gabor filter related algorithm. This model was proposed and executed in [11] and showed very good accuracy. Figure 7 represents the confusion matrix for LDA and PCA algorithm proposed in [9]. Now Figure 8 corresponds to the confusion matrix for the 7 emotions by the proposed algorithm.

	Anger	Disgust	Fear	Joy	Sadness	Surprise	Neutral
Anger	85.1	3.9	0	0	8.6	0	2.4
Disgust	0	97.5	2.5	0	0	0	0
Fear	0	0	80.7	7.4	3.0	0	8.9
Joy	0	0	0	97.1	0	0.6	2.3
Sadness	12.4	0	0	0	75.1	4.2	8.3
Surprise	0	0	0.5	0	0.5	97.5	1.5
Neutral	0	1	0.8	3.4	2.6	0	92.2

Figure 5. Confusion Matrix for SVM and RBF

	Anger	Disgust	Fear	Joy	Sadness	Surprise	Neutral
Anger	87.1	2.8	0	0	7.7	0	2.4
Disgust	0	97.5	1.3	0	1.2	0	0
Fear	0	0	85.7	4.4	1.1	0	8.8
Joy	0	0	0	97.1	0	0.4	2.5
Sadness	9.1	0	0	0	79.1	2.4	9.4
Surprise	0	0	1.2	0	1	96.3	1.5
Neutral	0	1	0.5	3.5	1.9	0	93.1

Figure 6. Gabor Filter Related Confusion Matrix

	Anger	Disgust	Fear	Joy	Sadness	Surprise	Neutral
Anger	89.4	1.8	0	0	5.9	0	2.9
Disgust	0	96.3	1.4	0	1.4	0	0.9
Fear	0	0	87.2	3.8	1.3	0	7.7
Joy	0	0	0	97.3	0	0.3	2.4
Sadness	9.1	0	0	0	84.1	3.4	3.4
Surprise	0	0	2.5	0	1.2	94.8	1.5
Neutral	0	1.2	0.8	3.2	1.7	0	93.1

Figure 7. LDA + PCA Confusion Matrix

	Anger	Disgust	Fear	Joy	Sadness	Surprise	Neutral
Anger	93.4	2.8	0	0	1.9	0	1.9
Disgust	0	95.9	1.6	0	1.5	0	1
Fear	0	0	85.8	8.1	1.8	0	4.3
Joy	0	0	0	94.7	0	3	2.3
Sadness	7.1	0	0	0	84.8	4.9	3.2
Surprise	0	0	2.7	0	1.4	94.1	1.8
Neutral	0	0	0.5	3.1	1.1	0	95.3

Figure 8. Confusion Matrix for proposed algorithm

Method	Accuracy
Template Matching with LBP	81.18
SVM with RBF	89.31
Gabor Filter Related	90.84
LDA + PCA	91.74
Proposed	92.00

Table 1. Accuracy comparison

Table I provides a comparison of all the algorithms in terms of their accuracy. This accuracy is calculated by considering true positive percentage for each expression and adding this percentage for all 7 expressions and dividing this sum by 7. This accuracy comparison provides a good comparison tool as an average and is much simpler to understand.

CONCLUSION

The proposed algorithm first detects fiducial points on the face. Then bezier curves are generated out of these points and a CNN develops a score comparison for the different expressions by comparing the test bezier curve with the bezier curve generated by the points generated as a standard for the expression by the CNN. The expression with the highest score is taken and assigned to the test image.

Using these bezier curves, we can obtain a relationship between fiducial points and using this relationship we can determine the expression detected. It was seen that on average the proposed approach gave a 92.00 percent accuracy, greater than the next best algorithm tested, LDA+PCA. It performed much better than some of the older algorithms, like template matching and SVM with RBF.

REFERENCES

[1] Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Mobahi, H., Ma, Y.. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012 Feb;34(2):372-386.

[2] Zhao, W., Chellappa, R., Phillips, P.J. and Rosenfeld, A., 2003. Face recognition: A literature survey. ACM computing surveys (CSUR), 35(4), pp.399-458.

[3] Gowda, S.N., 2017. Human activity recognition using combinatorial Deep Belief Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 1-6).

[4] Martnez, A.M. and Kak, A.C., 2001. Pca versus lda. IEEE transactions on pattern analysis and machine intelligence, 23(2), pp.228-233.

[5] Cevikalp, H., Neamtu, M., Wilkes, M. and Barkana, A., 2005. Discriminative common vectors for face recognition. IEEE Transactions on pattern analysis and machine intelligence, 27(1), pp.4-13.

[6] Gowda, S.N., 2016, September. Innovative enhancement of the Caesar cipher algorithm for cryptography. In 2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA)(Fall) (pp. 1-4). IEEE.

[7] Rowley, H.A., Baluja, S. and Kanade, T., 1998. Neural network-based face detection. IEEE Transactions on pattern analysis and machine intelligence,20(1), pp.23-38.

[8] Lajevardi, S.M. and Wu, H.R., 2012. Facial expression recognition in perceptual color space. IEEE Transactions on Image Processing, 21(8), pp.3721-3733.

[9] Fazli, S., Afrouzian, R. and Seyedarabi, H., 2009, November. Highperformance facial expression recognition using Gabor filter and probabilistic neural network. In Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on (Vol. 4, pp. 93-96). IEEE.

[10] Li, P., Phung, S.L., Bouzerdom, A. and Tivive, F.H.C., 2010, July. Feature selection for facial expression recognition. In Visual Information Processing (EUVIP), 2010 2nd European Workshop on (pp. 35-40). IEEE

[11] Gowda, S.N. and Yuan, C., 2018, December. ColorNet: Investigating the importance of color spaces for image classification. In *Asian Conference on Computer Vision* (pp. 581-596). Springer, Cham.

[12] Gao, Y., Leung, M.K., Hui, S.C. and Tananda, M.W., 2003. Facial expression recognition from line-based caricatures. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 33(3), pp.407- 412.

[13] Tian, Y.I., Kanade, T. and Cohn, J.F., 2001. Recognizing action units for facial expression analysis. IEEE Transactions on pattern analysis and machine intelligence, 23(2), pp.97-115.

[14] Waldir, S.S., Arajo, G.M., da Silva, E.A. and Goldenstein, S.K., 2010, September. Facial fiducial points detection using discriminative filtering on principal components. In 2010 IEEE International Conference on Image Processing (pp. 2681- 2684). IEEE.

[15] Gowda, S.N., 2016, December. Age estimation by LS-SVM regression on facial images. In International Symposium on Visual Computing (pp. 370-379). Springer, Cham.

[16] Araujo, G.M., Jnior, W.S., Silva, E.A. and Goldenstein, S.K., 2010, October. Facial landmarks detection based on correlation filters. In Proceedings of the IEEE International Telecommunication Symposium.

[17] Silva, S., E., L., Jnior, T., D., P., Santos, V., K. and Junior, W., W., ”Fiducial Points Detection using SVM Linear Classifiers”, 2014, In CS and ITCSCP pp 23-31.

[18] Chang, C.C. and Lin, C.J., 2012. LIBSVM: A library for support vector machine, 2001.

[19] Gowda, S.N., 2016, November. Fiducial Points Detection of a Face Using RBF-SVM and Adaboost Classification. In Asian Conference on Computer Vision (pp. 590- 598). Springer, Cham.

[20] Viola, P. and Jones, M., 2001. Robust real-time object detection.International Journal of Computer Vision, 4 pp 137-154.

[21] Lyons, M., Akamatsu, S., Kamachi, M. and Gyoba, J., 1998, April. Coding facial expressions with gabor wavelets. In Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on (pp. 200-205). IEEE.

[22] Cohen, I., Sebe, N., Garg, A., Chen, L.S. and Huang, T.S., 2003. Facial expression recognition from video sequences: temporal and static modeling.Computer Vision and image understanding, 91(1), pp.160- 187.

[23] Martinez, A.M., 1998. The AR face database. CVC Technical Report, 24.

[24] Milborrow, S. and Nicolls, F., 2014, January. Active shape models with SIFT descriptors and MARS. In Computer Vision Theory and Applications (VISAPP), 2014 International Conference on (Vol. 2, pp. 380-387). IEEE.

[25] Gowda, S.N. and Yuan, C., 2020. StegColNet: Steganalysis based on an ensemble colorspace approach. arXiv preprint arXiv:2002.02413.

[26] Ojansivu, V. and Heikkil, J., 2008, July. Blur insensitive texture classification using local phase quantization. In International conference on image and signal processing (pp. 236-243). Springer Berlin Heidelberg.