Facial Expression Analysis by k-means Clustering on Fiducial points of face

Previte Modesto, Saul Reitano
University of Messina

***Abstract*—Human beings produce thousands of facial actions and emotions in a single day. These come up while communicating with someone and at times even when alone. These expressions vary in complexity, intensity, and meaning. This paper proposes a novel method to predict what emotion is being expressed by analyzing the face. The algorithm, because of the speed of execution, could also be used for micro expression analysis. 11 fiducial points are taken on the image after a face recognition algorithm is used. 7 classes of images are formed. These classes are the main expressions: sadness, happiness, anger, fear, disgust, surprise and neutral. Training is done by studying the relationship between the fiducial points for each class of image. Using this relationship a new image is classified by making use of the k-means algorithm.**

Keywords—expression analysis, k-means, fiducial points;

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 a single feature vector, that represents the whole face image. Examples of holistic approaches are the fiducial points as proposed by Gowda et al [3], the linear discriminant analysis as proposed by Martinez et al [4], using LS-SVM as proposed by Gowda 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 on in, facial expression analysis refers to computer systems that make an attempt to automatically analyze and recognize facial motions and facial feature changes from visual information. Sometimes facial expression analysis has been confused with emotion analysis. For emotion analysis, a higher level knowledge is required. For example, although facial expressions do convey emotion, they can also express intention, cognitive processes, physical effort etc. Interpretation is further aided by context, voice, body gestures, individual differences. Computer facial expression analysis systems need to analyze the facial actions regardless of context, culture, gender, and so on.

The accomplishments in the related areas for example psychological studies, human movement analysis, face detection, face tracking and recognition make the automatic facial expression analysis possible to a very high degree. Facial expression analysis includes both measurement of facial motion and recognition of expression. The general approach to automatic facial expression analysis (AFEA) consists of three steps: face acquisition from image, facial data extraction and representation of that data, and last facial expression recognition.

Face acquisition is a pre-processing stage to automatically determine the region containing the face for the input images or sequences. It can be a detector to detect face for each frame, or just detect the face in the first frame and then track the face in the case of a video sequence. After the face is located, the next step is to extract and represent facial changes caused by facial expressions.

In facial feature extraction for expression analysis, there are mainly two different types of approaches namely geometric feature-based methods and appearance-based methods. The geometric facial features represent the shape and locations of facial components (example eyes, brows, nose, etc.). The facial components or feature points are extracted to form a feature vector that represents the face geometry. With appearance-based methods, however, image filters, such as Gabor filters are applied to either the whole-face or part of a face image 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

Gowda et al. [8] worked on color images instead of gray scale and found that the color components in color images provides more information than gray scale images. They introduced a tensor perceptual color framework (TPCF) for the facial expression recognition (FER), which is based on information contained in color facial images. The TPCF enabled multi-linear image analysis in different colour spaces, and demonstrates that colour components provide additional information for robust FER. Saeid Fazli et al. [9] found that in their study if the number of samples is less in comparison to the dimensionality of the image then Linear Discrimination Analysis (LDA) alone is insufficient for feature reduction. To increase the performance Principal Component Analysis should be used before LDA.

Li et all [10] proposed a cascade of fixed filters and trainable non-linear 2-D filters, which were based on the biological mechanism of shunting inhibition. The fixed filters were used to extract the primitive features, whereas the adaptive filters were trained to extract the more complex facial features for classification by SVMs.

Zhang et al [11] proposed a unified framework for a comparative study on the widely used texture (LBP, Gabor) and geometric features using Adaboost, mRMR and SVM feature selection algorithms. Their experiments on the demonstrated the benefits of fusing geometric and texture features.

Gao et al [12] presented a methodology for facial expression recognition from a single static image by using line-based caricatures. The recognition process was completely automatic. It also addressed the computational expensive problem and was thus said to be suitable for real-time applications. The proposed approach used structural and geometrical features of a user sketched expression model to match the line edge map (LEM) descriptor of the input face image.

Tian et al [13] developed an Automatic Face Analysis (AFA) system to analyze facial expressions based on both the permanent facial features such as brows, eyes, mouth and the transient facial features such as deepening of facial furrows in a near frontal-view face image sequence. The AFA system recognized fine-grained changes in facial expression into action units (AUs) of the Facial Action Coding System (FACS), instead of a few prototypic expressions. This was shown to be better than the compared algorithms in terms of accuracy.

Gowda et al [14] proposed using an ensemble of deep learning models for obtaining best accuracy. They used it in an action recognition model. Another important part of this experiment is determining the fiducial points of an image. 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: local and global. In local methods individual fiducial points are detected and then processed and no additional information is needed or utilised. The global methods are characterized by detecting more fiducial points in comparison, for this they use deformable models, less susceptible to pose and illumination variations than local methods.

The design of a classifier is probably the most important stage of a fiducial point detection algorithm. In this stage, several different machine learning algorithms could be used as shown by Gowda et al [15]. Particularly, some studies use a classification method called Support Vector Machine. The mathematical formulation of the SVM is obtained by optimization problem with restrictions.

There are many recent papers with regards to SVM. 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 coefficients 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 same problem. The authors 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 (Support Vector Classification) [18], for fiducial point detection.

III. Proposed Algorithm

Before The detection of Fiducian points follows after two steps: training and testing. First the SVM is trained using a set of 497 images taken from the JAFFE database[21] and the AR database [23] .

Next a new image is taken as a testing image and then the testing is done. In any case the first step is the pre-processing step. The pre-processing step is composed of two different stages. First the face is detected using the Viola Jones algorithm as proposed by Viola et al in [20]. Next the face is made to scale to a 320x240 image to obtain a clearer image consisting only of the facial features.

In the training stage first, the image is pre-processed. To determine the position of Fiducian points we find areas of high probability of its occurrence by using a Gaussian Mixture Model inspired by the mathematical derivations in [14].

This Gaussian Mixture model(GMM) consists of 4 models developed around the Gaussian prior model(GPM) by changing the parameters of the prior model to accommodate a larger set of points into the classification with higher accuracy.

A candidate to any fiducial point with a label q in the image, having coordinates equal to y, is considered to be inside the ellipse that is defined by the Mahalanobis distance to the average of all the fiducial points in the training set, given by (3):

$$\max_{\text{label(w)=q}} \{1.05\|w-\mu\|^2\} \geq (y-\mu)^T \Sigma_X^{-1} (y-\mu) \qquad (1)$$

X is the random vector whose realisation is equal to the ficudial points, is the vector mean and $\sum_X$ is the covariance matrix of *X*.

Next we use a Wiener filter to remove some noise and focus on each facial feature to obtain the fiducial points of that feature. Each block is represented as A_z and is usually of the size 13x13.

Further in Figure 1 we can see a sample picture being taken and the resulting output of each phase can also be seen.

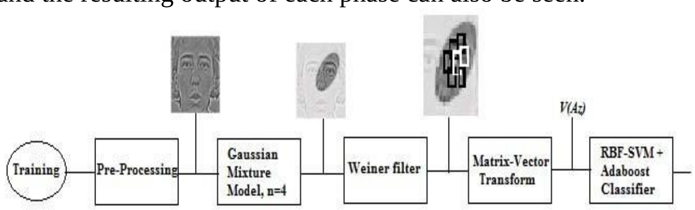

Fig 1. Training block diagram

The end output of determining the fiducial points is as shown in Figure 2.

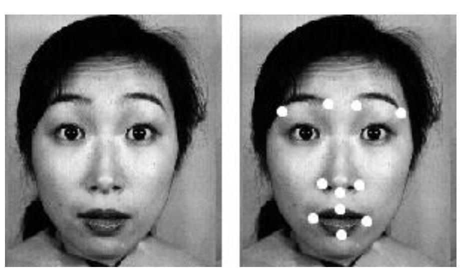

Fig 2 (a) original face (b) face with fiducial points

7 clusters of images are formed manually. This is done by taking 25 images corresponding to each expression: sadness, happiness, anger, fear, disgust, surprise and neutral. So 175 images in total are used for the training purposes. It is easy to obtain a geometric relationship form the fiducial points obtained. Figure 3 shows a sample on the way the fiducial points were treated with regards to the algorithm.

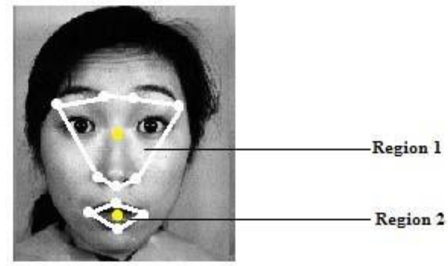

Fig 3. Relationship between ficudial points

The yellow dots in the image represent the centroid of that particular region. 2 variables are needed for the classification: one that keeps note of centroid of region 1 and another for region 2. 25 images of each expression are subject to the same procedure. So 7 clusters are formed in this way. Each cluster having 2 variables representing it say C_{i1} and C_{i2} where i stands for the number of any 1 particular cluster. Hence when a new image is going to be tested, first the fiducial points are obtained and the values of centroids are calculated. Using these values the image is classified to one of the 7 clusters using k-means or in this case 7-means. The 2 relationship between the centroids relative to the coordinate of each fiducial point will in turn give us an expression. The relation we use is the average distance of each point in a region from the centroid.

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. Table 1 represents the confusion matrix for [22].

Table 1. Confusion matrix for template matching with 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**

Table 2 represents matrix for SVM with RBF.

Table 2. Confusion matrix for SVM with RBF

	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**

Table 3 represents confusion matrix for Gabor filter related algorithm.

Table 3. Confusion matrix for Gabor filter related [11]

	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**

Table 4 represents the confusion matrix for LDA and PCA algorithm proposed in [9].

Table 4. Confusion matrix for LDA+PCA [9]

	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**

As can be seen from the confusion matrices, the gabor filter related method and the LDA and PCA combined algorithm perform best among the selected algorithms. Now table 5 corresponds to the confusion matrix for the 7 emotions by the proposed algorithm.

Table 5. Confusion matrix for proposed algorithm

	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	**89.8**	4.1	1.8	0	4.3
Joy(%)	0	0	0	**97.7**	0	0	2.3
Sadness(%)	4.1	0	0	0	**89.8**	3.9	2.2
Surprise(%)	0	0	2.7	0	1.4	**94.1**	1.8
Neutral(%)	0	0	0.5	3.1	1.1	0	**95.3**

Table 6 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.

Table 6. Different types of model Results comparison

Name of algorithm	Accuracy of classification(%)
Template matching with LBP	81.18
SVM with RBF	89.31
Gabor Filter Related	90.84
LDA+PCA	91.74
Proposed	**93.71**

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., 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.

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

[5] 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.

[6] Moghaddam, B., Nastar, C. and Pentland, A., 1996, August. A Bayesian similarity measure for direct image matching. In Pattern Recognition, 1996., Proceedings of the 13th International Conference on (Vol. 2, pp. 350-358). 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] 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.

[9] Fazli, S., Afrouzian, R. and Seyedarabi, H., 2009, November. High-performance 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] Zhang, L., Tjondronegoro, D. and Chandran, V., 2012, July. Discovering the best feature extraction and selection algorithms for spontaneous facial expression recognition. In 2012 IEEE International Conference on Multimedia and Expo (pp. 1027-1032). IEEE.

[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] 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).

[15] Gowda, S.N., 2016, October. Face verification across age progression using facial feature extraction. In 2016 International Conference on Signal and Information Processing (IConSIP) (pp. 1-5). IEEE.

[16] Araujo, G.M., Júnior, 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., Júnior, T., D., P., Santos, V., K. and Junior, W., W., "Fiducial Points Detection using SVM Linear Classifiers", 2014, In CS & IT-CSCP pp 23-31.

[18] Gowda, S.N., 2016, October. Using Blowfish encryption to enhance security feature of an image. In 2016 6th International Conference on Information Communication and Management (ICICM) (pp. 126-129). IEEE.

[19] Vapnik, V., 2013. The nature of statistical learning theory. Springer Science & Business Media.

[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] 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.

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