Facial Expression Recognition Based on Structural Changes in Facial Skin

Zeynab Shokoohi[a*], Karim Faez[b]

[a]Departement of Electrical and Computer Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran
[b]Departement of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran

Abstract

Facial expressions are the most powerful and direct means of presenting human emotions and feelings and offer a window into a persons' state of mind. In recent years, the study of facial expression and recognition has gained prominence; as industry and services are keen on expanding on the potential advantages of facial recognition technology. As machine vision and artificial intelligence advances, facial recognition has become more accessible and is now a key technique to be employed and used in creating more natural man-machine interactions, Computer vision, and health care. In this paper, we empirically evaluate facial representation based on statistical local features, Local Binary Patterns, for person-independent facial expression recognition. Different machine learning methods are systematically examined on several databases. Extensive experiments illustrate that LBP features are effective and efficient for facial expression recognition. In this paper, we proposed a face expression detection method based on the difference of a face expression and the allocated special pattern to each expression. The analysis of the image detection system locally and through a sliding window (sliding) at multiple scales, are estimated. Multiple scales are extracted aslocally binary features. Through using the change point between windows, points of face are getting a directional movement. Through using points movement of whole facial expressions and rating system that is created the superfluous points are eliminated. The classifications are taken based on the nearest neighbor. To sum up this paper, the proposed algorithms are tested on Cohn-Kanade data set and the results showed the best performance and reliability into other algorithms. We investigated LBP features for the facial skin structural changes, which is seldom addressed in the existing literature.

Keywords: Facial expression Recognition, Local Binary Patterns (LBP), Nearest Neighbor Classification, Linear Feature, Local Image Descriptor

1. Introduction

As our demand for more intelligent human-computer interaction grows, Facial expression recognition becomes an ever evolving field of interest involving such vast fields as computer aided vision, machine learning and behavioral sciences. Facial Recognition can be of use in many applications such as security [13], human-computer-interaction [16], driver safety [17] and healthcare [11]. Significant advances have been made in the field over the past decade [14, 15, and 20] with increasing interest in non-posed facial behavior in naturalistic contexts [4, 11, 18] and posed data recorded from multiple views [10, 12] and 3D imaging [19].

Common limitations in the field are:

a) Inconsistent or absent reporting of inter-observer reliability and validity of expression metadata. Emotion labels, for instance, have referred to what expressions were requested rather than what was actually performed. Unless the validity of labels can be quantified, it is not possible to calibrate algorithm performance against manual (human) standards.

b) Common performance metrics with which to evaluate the new algorithms for both AU and emotion detection. Published results for established algorithms would provide an essential benchmark against which to compare performance of new algorithms.

c) Standard protocols for common databases to make possible quantitative meta-analyses.

This paper focuses on the process of facial expression recognition in static images which is the foundation of facial expression recognition. Generally, there are two ways for static expression recognition: model based methods and appearance based methods. The difference is in the way we describe the expression information. AAM [3] is a popular model based method, while Cohn and Kanade advocate combinations of AUs [1, 2] to represent expressions. This paper will study appearance aspects and recognizing facial expression using the existing differences

* Corresponding author. Email: shokoohi64@gmail.com

in various expressions. This type of face recognition is based on human diagnosis meaning that we recognize new expressions by allocating and separating the maximum differences in facial changes with another expression. Experiments on C-K database and a hybrid database show the accuracy of the algorithm [1].

We introduce the previous work in section 2. In section 3, we present the whole algorithm frame work. In section 4, we analyze our proposed algorithm with the previous works. Finally, we conclude our findings in section5.

2. Review of Facial Expression Recognition

The origin of facial expression analysis goes back into the 19th century when Darwin originally proposed the concept of universal facial expressions in humans and animals. Since the early 1970s, Ekman and Friesen (1975) have performed extensive studies of human facial expressions, providing evidence to support this universality theory [7].

These 'universal facial expressions' are those representing happiness, sadness, anger, fear, surprise, and disgust. To prove this, they provide results from studying facial expressions in different cultures, even primitive or isolated ones. These studies show that the processes of expression and recognition of emotions on the face are common enough, despite differences imposed by social rules. Ekman and Friesen used FACS to manually describe facial expressions, using still images of, usually extreme, facial expressions. This work inspired researchers to analyze facial expressions by tracking prominent facial features or measuring the amount of facial movement, usually relying on the 'universal expressions' or a defined subset of them.

In the 1990s, automatic facial expression analysis research gained much interest, mainly thanks to progress in the related fields such as image processing (face detection, tracking and recognition) and the increasing availability of relatively cheap computational power. In one of the ground-breaking and most publicized works, Mase and Pentland (1990) used measurements of optical flow to recognize facial expressions [9]. Later, Lanitis et al. used a flexible shape and appearance model for face identification, pose recovery and facial expression recognition. Black and Yacoob (1997) proposed local parameterized models of image motion to recover non-rigid facial motion, which was used as input to a rule-based basic expression classifier[5]; Yacoob and Davis (1996) also worked in the same framework, this time using optical flow as input to the rules. Local optical flow was also the basis of Rosenblum's work, utilizing a radial basis function network for expression classification. Otsuka and Ohya utilized the 2D Fourier transform coefficients of the optical flow as feature vectors for a hidden Markov model (HMM). Regarding feature-based techniques, Donato, Bartlett, Hager, Ekman, and Sejnowski (1999) tested different features for recognizing facial AUs and inferring the facial expression in the frame [6]. Oliver et al. tracked the lower face to extract mouth shape information and fed them to an HMM, recognizing again only universal expressions.

As can be seen, most facial expression analysis systems focus on facial expressions to estimate emotion-related activities. Furthermore, the introduction and correlation of multiple channels may increase robustness, as well as improve interpretation disambiguation in real-life situations. Most attempts at channel fusion evaluate speech, in addition to facial expressions. Here, expressions may be conveyed by linguistic, as well as prosodic features, such as the fundamental frequency, intensity and pause timing. Cohn and Katz (1998) as well as Chen et al [8] focused on the fundamental frequency, as it is an important voice feature for emotion recognition and can be easily extracted. It has to be noted though, that introducing speech in the expression recognition picture has to be followed by separate provisions for aural and visual information synchronization. This is essential because, in the general case, events regarding these two channels do not occur simultaneously and may affect one another (e.g. visual information from the mouth area generally deteriorates when the subject is speaking) [8].

Generally, we mean to classify 7 basic types of expressions: anger, disgust, fear, happiness, neutral, sadness and surprise. Among all the previous works, the results have presented that the accuracy rates of two expressions, i.e. happiness and surprise, are always high throughout all the reported results. Meanwhile, the samples of the other five expressions are mixed in every dimension. In such circumstances, they first distinguished the two "easy" expressions from the others. Then they could better pay attention to the other five expressions with less interference. More targeted features and some additional information is helpful to improve the classification accuracy of these five expressions [3].

Another study in Facial expression recognition is based on Neural Network approach for Facial Expression Recognition. In this method, Gabor Filtering and PCA is used for recognition[25].Another study in Facial expression recognition is based on a Neuro Fuzzy approach for Facial Expression Recognition. In this method, Local Binary feature extraction and Neuro fuzzy classification is used for recognition. In this approach, an image must be divided into 5 parts. Neuro fuzzy systems had a low Standard deviation but it is considered a difficult method for facial recognition, and thus not an ideal approach.

3. The Proposed Method

3.1- Subject Selection

First we implement the algorithm to pin point the location of the face and eyes and subsequent extraction of facial region. By means of rotation and alignment the face is normalized and cut from chin to forehead and from ear to ear. The LBP is then extracted and imported into a matrix as can be seen in Figure 1.

3.2- Normalizing the Images

Normalizing the images is done as follows:

a) Rotating the facial image so as to align it vertically
b) Scaling the images so that the distance between the eyes are the same across all samples
c) Cutting of the images to remove the background and scalp
d) Applying histogram equalization for photometric normalization

After processing, each image should be 200x150 pixels. The normalization process is presented in Figure 1.

Fig.1. Normalization processing (from left to right: Original image, rotation and normalization, final image after histogram equalization).

These steps help us in the extraction of facial features in various facial expressions that have been popular in the past as they result in the highest level of recognition.

3.3-Local Binary Patterns (LBP)

The 3x3 per pixel neigh boring LBP function attaches a label to each pixel giving a binary result to each. The LBP function has been widely used for gender recognition.

$$s(f_p - f_c) = \begin{cases} 1 & f_p \geq f_c \\ 0 & f_p < f_c \end{cases} \quad (1)$$

$$LBP(f_c) = \sum_{p=0}^{7} s(f_p - f_c) 2^p \quad (2)$$

LBP of the center pixel is calculated using equation (2). In equation (1) and (2), f_c is the pixel center and f_p is the neigh boring of the center pixel. LBP_{PR}^{U} is the symbol used for a constant LBP [6]. Neigh boring LPB in the location of p is sampled on a circle with a circumference of R.

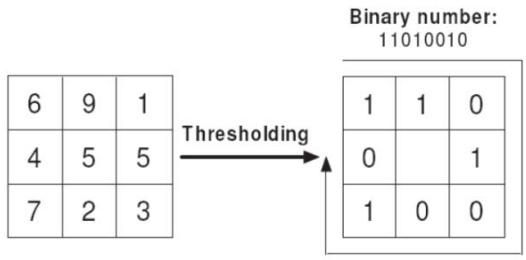


Fig. 2.Using Local Binary Patterns for neighboring pixels.

Fig. 3. LBP function for R=1 and P=8.

3.3.1- Comparing local binary patterns to other local descriptors

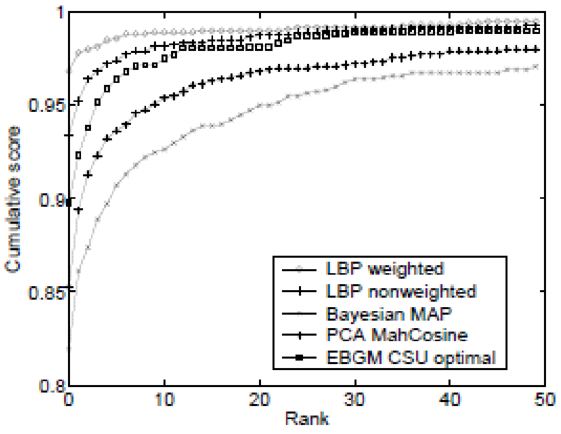
To gain a better understanding of whether the obtained recognition results are due to general idea of computing texture features from local facial regions or due to the discriminatory power of the local binary pattern operator, we compared LBP to three other texture descriptors, namely thegray-level difference histogram, homogeneous texture descriptor [21] and an improved version of the texton histogram [22]. The details of these experiments can be found in [23]. The recognition rates obtained with different descriptors are shown in Table I.

Table 1

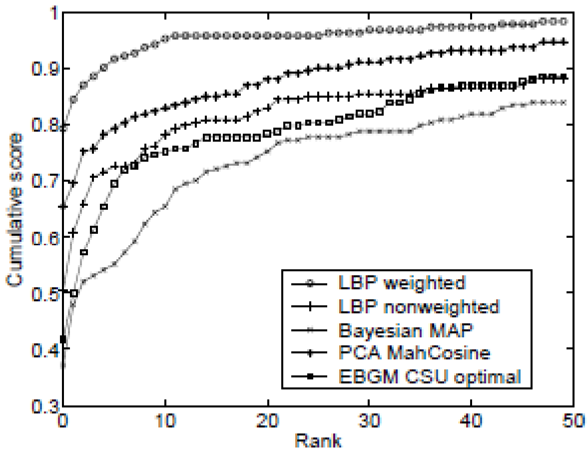
The recognition rates obtained using different texture describtors for local facial regions. The first four columns show the recognition rates for the feret test sets and the last three columns contain the meanrecognition rate of the permutation test with a 95% confidence interval.

Method	fb	fc	dup I	dup II	lower	mean	upper
Difference histogram	0.87	0.12	0.39	0.25	0.58	0.63	0.68
Homogeneous texture	0.86	0.04	0.37	0.21	0.58	0.62	0.68
Texton Histogram	0.97	0.28	0.59	0.42	0.71	0.76	0.80
LBP (nonweighted)	0.93	0.51	0.61	0.50	0.71	0.76	0.81

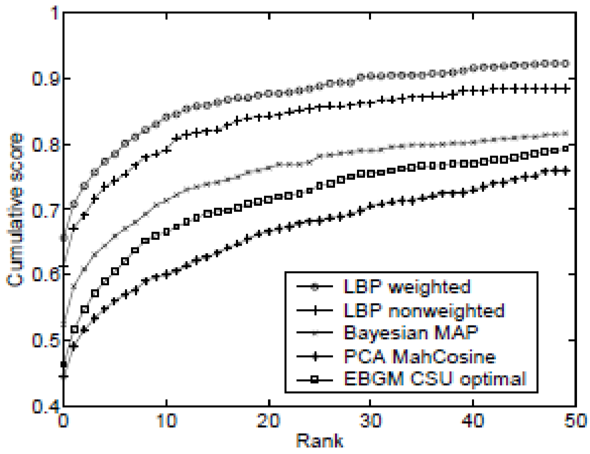
It should be noted that no weighting for local regions was used in this experiment. The results show that the tested methods work well with the easiest fb probe set, which means that they are robust with respect to variations of facial expressions, whereas the results with the fc probe set show that other methods than LBP do not survive changes in illumination. The LBP and text on give the best results in the dup I and dup II test sets. We believe that the main explanation for the better performance of the local binary pattern operator over other texture descriptors is its tolerance to monotonic gray-scale changes. Additional advantages are the computational efficiency of the LBP operator and that no gray-scale normalization is needed prior to applying the LBP operator to the face image. The cumulative scores of the LBP and other local descriptors are addressed in Figure 4.



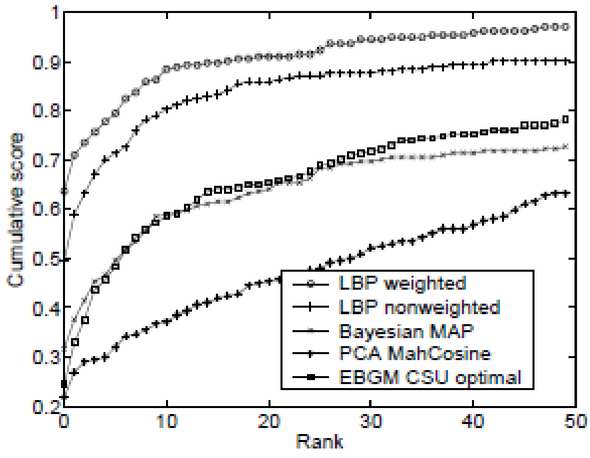
(a)



(b)



(c)



(d)

Fig. 4. The cumulative scores of the LBP and other local descriptors on the (a) fb, (b) fc, (c) dup I and (d) dup II probe sets

3.4-Creating Movement Patterns for Facial Elements

In order to calculate facial changes in expression, we divide the face into neighbor finding partitions and convert the main facial features achieved in step 1-3 into Figure 5.

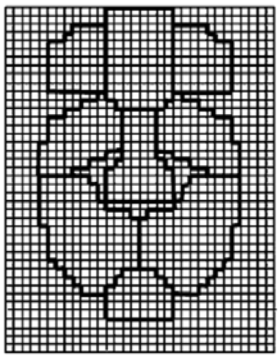


Fig. 5.Partitioning of the face.

By partitioning the human face as above, each facial movement can be clearly analyzed and detected. Figure 6 shows facial template incursion[24].

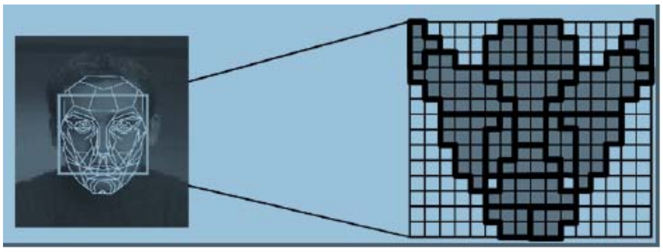


Fig. 6.Putting template (on the face).

After implementing the changes in the subject's facial expressions, we can record the changes in all the facial areas. At this point, every small movement can be categorized into an overall facial pattern as can be seen in Figure 7.

Fig. 7.Stepwise motion for the facial expression of a smile.

Using the method above, we can calculate the general movements for a particular facial expression and generate a reference pattern for that expression as in Figure 8.

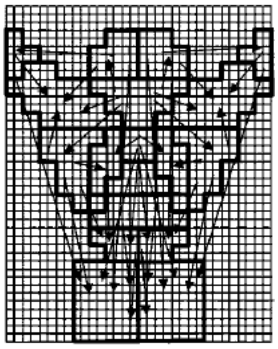


Fig. 8.Movement of facial regions in the facial expression of a smile.

3.5-Removing Redundant Facial Movements

After extracting the common rules and patterns in a specific facial expression, we then remove the movements that are not common or are subject specific according to the following method:

a) Selecting an expression from the 7 major facial expressions.

b)Movement patterns of all samples are compared with regard to the angle of motion (a max tolerance of 10% angle variation is acceptable).

c) Redundant images within the facial partitions that have not been selected or are absent in 20 percent of samples are removed.

d)Remaining patterns are normalized as to remove further complexities.

3.6- Image Classification

Among the various classification approaches (Neural network, SVM …), the Nearest Neighbor method has had the best classification results with the purposed method. Based on the tests conducted, the nearest neighbor considered with K=1 has given the best results as can be seen in Figure 9.

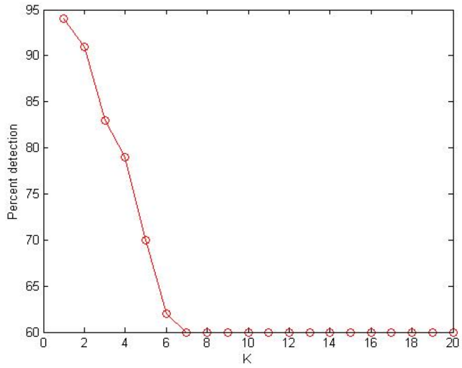


Fig. 9.rate of K for best results in the Nearest Neighbor classification

4- Assessment of the Proposed Method

For the assessment of the proposed method, the Cahn-Kanade dataset was used in which 123 subjects were sampled in 7 facial expressions. Each expression was subsequently admitted. Table 2 shows the advantages of the proposed method compared to other methods. The comparisons were carried out based on a 60% of image and 40% of data.

Table 2
Comparison of the method with others

Methods	Recognition Rate
A Hierarchical Algorithm with Multi-Feature Fusion for Facial Expression Recognition(LBP- Gabor)	84.5%
Neuro Fuzzy Model for Human Face Expression Recognition	81.5%
Proposed Method	91.6%

5- Conclusion

In this method a different approach has been presented in which various facial regions are partitioned and movements are assessed. The facial movements are then classified and the patterns are extracted. The results confirmed that the purposed method is more accurate than the conventional methods. The experimental results also demonstrated significant performance improvements due to the consideration of facial movement features.

References

[1] P. Baldi, K. Hornik. Neural networks and principal, “The AR face database,”CVC Technical Report #24, Tech. Rep., 1998.

[2] Z. Zhang, etc. A Hierarchical Algorithm with Multi-FeatureFusion for Facial Expression Recognition . 21st InternationalConference on Pattern Recognition (ICPR 2012) November 11-15,2012.

[3] P. Lucey, etc. The Extended Cohn-Kanade Dataset(CK+): A complete dataset for action unit and emotionspecifiedexpression. CVPR 2010.

[4] M.Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, J. Movellan. Automatic Recognition of Facial Actions in SpontaneousExpressions. Journal of Multimedia, 2006

[5] M.Black, Y.Yacoob (1997). Recognizing facial expressions inimage sequences using local parameterized models of image motion.International Journal of Computer Vision, 25(1), 23–48.

[6] G.Donato, M.Bartlett, J.Hager, P.Ekman, T.Sejnowski(1999).Classifying facial actions. IEEE Transactions on Pattern Analysis andMachine Intelligence, 21(10).

[7] P.Ekman, W.Friesen (1975). Unmasking the face. Prentice-Hall.FP5 IST ERMIS project. Emotionally rich man–machineintelligentsystem, http://www.image.ntua.gr/ermis.

[8] J. F.Cohn, G. S.Katz (1998). Bimodal expressions ofemotion by faceand voice Workshop on face/gesture.recognition and their pplications, the sixth ACMinternational multimedia conference, Bristol, England

[9] K.Mase, A.Pentland (1990). Lip reading by optical.flow. IEICE of Japan, J73-D-II, 6, 796–803.

[10] R. Gross, I. Matthews, S. Baker, T. Kanade. The CMU MultiplePose,Illumination, and Expression (MultiPIE). Technical report,Robotics Institute, Carnegie Mellon University, 2007.

[11] P. Lucey, J. Cohn, S. Lucey, I. Matthews, S. Sridharan, K. Prkachin. Automatically Detecting Pain Using Facial Actions. InProceedings of the International Conference on Affective Computingand Intelligent Interaction, pages 1–8, 2009.

[12] M. Pantic, M. Valstar, R. Rademaker, L. Maat. Fully automaticFacial Action Recognition in Spontaneous Behavior. In Proceedingsof the International Conference on Multimedia and Expo, pages 317–321, 2005.

[13] A. Ryan, J. Cohn, S. Lucey, J. Saragih, P. Lucey, F. D. la Torre,A. Rossi.Automated Facial Expression Recognition System. InProceedings of the International Carnahan Conference on SecurityTechnology, pages 172–177, 2009.

[14] Y. Tian, J. Cohn, T. Kanade. Facial expression analysis. InS. Li and A. Jain, editors, The handbook of emotion elicitation andassessment, pages 247–276. Springer, New York, NY, USA.

[15] Y. Tong, W. Liao, and Q. Ji. Facial Action Unit Recognition byExploiting Their Dynamic and Semantic Relationships.IEEETransactions on Pattern Analysis and Machine Intelligence,29(10):1683–1699, 2007.

[16] A. Vinciarelli, M. Pantic, H. Bourlard. Social Signal Processing:Survey of an Emerging Domain. Image and Vision Computing,31(1):1743–1759, 2009.

[17] E.Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan.Automated Drowsiness Detection for Improved Driver Saftey-Comprehensive Databases for Facial Expression Analysis. InProceedings of the International Conference on AutomotiveTechnologies,2008.

[18] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan.Towards Practical Smile Detection. IEEE Transactions on PatternAnalysis and Machine Intelligence, 31(11):2106–2111, 2009.

[19] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A high-resolution3D dynamic facial expression database. In Proceedings of theInternational Conference on Automatic Face And Gesture Recognition,2008.

[20] Z. Zeng, M. Pantic, G. Roisman, and T. Huang. A Survey of AffectRecognition Methods: Audio, Visual and Spontaneous Expressions.IEEE Transactions on Pattern Analysis and Machine Intelligence,31(1):39–58, 2009.

[21] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, “Color and texture descriptors,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6, pp. 703–715, Jun 2001.

[22] M. Varma, A. Zisserman, “Texture classification: Are filter banks necessary?” in Proc.IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, pp. 691–698.

[23] T. Ahonen, M. Pietik¨ainen, A. Hadid, T. M¨aenp¨a¨a, “Face recognition based on appearance of local regions,” in Proc. 17th International Conference on Pattern Recognition, 2004, pp. III: 153–156.

[24] I. A. Essa, “Coding, Analysis, Interpretation, Recognition of Facial Expressions,” Georgia Institute of Technology,Graphics, Visualization and Usability Center, TechnicalReport # GIT-GVU-98-23, August 1998.

[25] M. N. Dailey, Garrison W. Cottrell, Curtis Padgett,Ralph Adolphs, “EMPATH: A Neural Network that CategorizesFacial Expressions” in Journal of Cognitive Neuroscience 14:8, pp. 1158–1173.