Asian Journal of Information Technology 12 (4): 126-130, 2013
ISSN: 1682-3915

Facial Expression Recognition Using Local Arc Pattern

Mohammad Shahidul Islam and Surapong Auwatanamongkol
Department of Computer Science, School of Applied Statistics,
National Institute of Development Administration, Bangkok, Thailand

Abstract: The success of a good facial expression recognition system depends on the facial feature descriptor. This study presents a unique local facial feature descriptor, the Local Arc Pattern (LAP) for facial expression recognition. Feature is obtained from a local 5×5 pixels region by comparing the gray color intensity values surrounding the referenced pixel to formulate two separate binary patterns for the referenced pixel. Each face is divided into equal sized blocks and histograms of LAP codes from those blocks are concatenated to build the feature vector for classification. The recognition performance of proposed method was evaluated on popular Japanese Female facial expression dataset using support vector machine as the classifier. Extensive experimental results with prototype expressions show that proposed feature descriptor outperforms several popular existing appearance-based feature descriptors in terms of classification accuracy.

Key words: Facial expression recognition, image processing, local feature extraction, pattern recognition, JAFFE, computer vision

INTRODUCTION

Facial expression is the natural and immediate means of human interaction. According to Mehrabian (1968), facial expression is more meaningful than only verbal communication. Due to this reason, it catches the researchers' eye to build accurate and automatic facial expression recognition system. Many real life applications like human-computer-interaction, video indexing, driver state identification, pain assessment, patient condition observation, lie detection, etc., demand more research for fast and accurate expression recognition systems. Some applications demand high accuracy and some demand real-time recognition. A facial expression recognition system works in four phases. Detecting the face, extracting features related to facial expression from the face building a model for facial expression classification based on the extracted features and recognizing test images using the model. The vital part of a good facial expression recognition system is the second phase or feature extraction.

Mainly two types of facial feature extraction approaches are found (Tian *et al*., 2003): geometric-based approach that uses position, distance, angle and other relations between the facial components and appearance-based approach that uses texture or color combinations from the full or part of the image. Both the methods are equally popular in this field of research. In the geometric-based methods, it is necessary to find the exact location of the facial components (Shan *et al*., 2005, 2009). Most of the earlier researches on geometric-based methods were based on Facial Action Coding System where facial expressions were coded using one more Action Units (Ekman and Friesen, 1978). AUs were based on one more facial muscle movements. Kotsia and Pitas (2007) manually placed some of the Candide grid nodes to the face landmarks to create facial wire frame model for facial expressions and used a Support Vector Machine (SVM) for classification. Valstar *et al*. (2005) and Valstar and Pantic (2006) used some fiducial points on the face to create geometric features and claimed that geometric approaches are better in feature extraction than appearance-based approaches. Zhang and Ji (2005) proposed IR illumination camera for facial feature detection and tracking. To recognize the facial expressions they used Dynamic Bayesian Networks (DBNs). They marked facial expressions by detecting 26 facial features around the regions of eyes, nose and mouth (Fig. 1).

Besides geometric-based methods using AUs some local appearance-based feature representations were also proposed. The local features are much easier for extraction than those of AUs. Ahonen *et al*. (2006) proposed a new facial representation strategy for still images based on Local Binary Pattern (LBP). In this method, the LBP value at the referenced center pixel of a M×M pixel region

Corresponding Author: Mohammad Shahidul Islam, Department of Computer Science, School of Applied Statistics, National Institute of Development Administration, 10240 Bangkok, Thailand

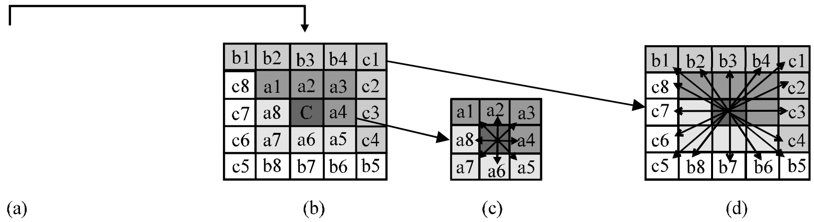

Fig. 1: Local pixels notation to formulate three different feature patterns for a single pixel, e.g., "C": a) Facial image; b) Local 5×5 pixels region; c) Pixels used for Pattern-1; d) Pixels used for Pattern-2

Pattern-1 (4-bit pattern)

Bit-1	Bit-2	Bit-3	Bit-4
$\begin{cases}1, & \text{if } a1>a5\\ 0, & \text{if } a1\le a5\end{cases}$	$\begin{matrix}1, & \text{if } a2>a6\\ 0, & \text{if } a2\le a6\end{matrix}$	$\begin{cases}1, & \text{if } a3>a7\\ 0, & \text{if } a3\le a7\end{cases}$	$\begin{cases}1, & \text{if } a4>a8\\ 0, & \text{if } a4\le a8\end{cases}$

(a)

Pattern-2 (8-bit pattern)

Bit-1	Bit-2	Bit-3	Bit-4	Bit-5	Bit-6	Bit-7	Bit-8
$\begin{cases}1, & \text{if } b1>b5\\ 0, & \text{if } b1\le b5\end{cases}$	$\begin{cases}1, & \text{if } b2>b6\\ 0, & \text{if } b2\le b6\end{cases}$	$\begin{cases}1, & \text{if } b3>b7\\ 0, & \text{if } b3\le b7\end{cases}$	$\begin{cases}1, & \text{if } b4>b8\\ 0, & \text{if } b4\le b8\end{cases}$	$\begin{cases}1, & \text{if } c1>c5\\ 0, & \text{if } c1\le c5\end{cases}$	$\begin{cases}1, & \text{if } c2>c6\\ 0, & \text{if } c2\le c6\end{cases}$	$\begin{cases}1, & \text{if } c3>c7\\ 0, & \text{if } c3\le c7\end{cases}$	$\begin{cases}1, & \text{if } c4>c8\\ 0, & \text{if } c4\le c8\end{cases}$

(b)

Fig. 2: Pattern formulation; a) 4-bit binary pattern and b) 8-bit binary pattern

is computed by thresholding the neighboring pixels gray color intensity value with the value of the center pixel as follows:

$$LBP=\sum_{i=1}^{N}2^{i-1}f\big(g(i)-C\big);\quad \text{Where, } f(x)=\begin{cases}0 & x<0\\ 1 & x\ge 0\end{cases} \tag{1}$$

Where:

C and g (i) = The gray color intensity value of the center pixel and ith neighboring pixel, respectively

N = The number of neighbors, i.e., 8 or 16

The LBP was first proposed by Ojala *et al*. (1996). He used it for texture analysis and got a very good results. Since, then it was used in many researches in many areas. Another popular holistic and appearance-based feature extraction method was Gabor Filter, named after Dennis Gabor. Facial future representations using Gabor filter is time and memory intensive (Bartlett *et al*., 2005). Lajevardi and Hussain (2012) solved some limitations of Gabor Filter using log-Gabor filter but the dimensionality of resulting feature vector was still high. Local Phase quantization (LPQ) was proposed by Ojansivu and Heikkila (2008) but like Gabor Filter, it was also very time and memory expensive. Ahsan *et al*. (2013) proposed a new feature extraction method from a local 5×5 pixels region and achieved very high accuracy but the cost of their method was very high.

Keeping all these sensitive issues in mind, this research presents a new feature extraction technique LAP (Local Arc Pattern) which overcomes almost all those cost

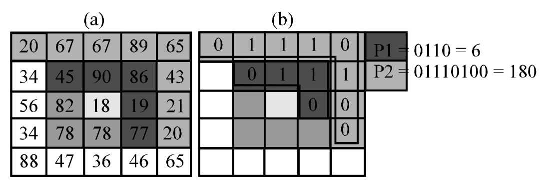

Fig. 3: Example of obtaining LAP value: a) 5×5 pixels local region; b) LAP representation of pixel 18

problems and weaknesses mentioned above. It considers local 5×5 pixel region to compute two separate patterns which together represents the local pattern at the center pixel. Unlike LTP+Gabor Filter proposed by Ahsan *et al*. (2013), it considers almost all the gray color intensity values of pixels in 5×5 pixel region. The local pattern at a pixel identifies the changes in the gray color intensities of its neighboring in all possible directions (Fig. 2).

LOCAL ARC PATTERN

Methodology: A 5×5 pixels local region is used to calculate LAP pattern for the center pixel of the region, C (Fig. 2). The gray color intensity values of a1, a2, a3, a4, b1, b2, b3, b4, c1, c2, c3 and c4 are used to formulate the LAP binary pattern. LAP pattern consists of one 4-bit binary pattern and one 8-bit binary pattern (Fig. 3), say Pattern-1 (P1) and Pattern-2 (P2). P1 is computed using the gray color intensity values of a1, a2, a3, a4, a5, a6, a7 and a8 and P2 is computed using the gray color intensity values of b1, b2, b3, b4, b5, b6, b7, b8, c1, c2, c3, c4, c5, c6, c7 and c8 (Fig. 2).

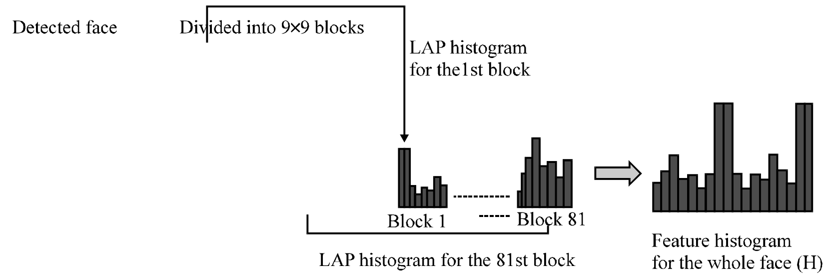

Fig. 4: Facial feature extraction

P1 can have at most $2^4 = 16$ bit combinations and P2 can have at most $2^8 = 256$ bit combinations. For each combination, a bin is created to count the number of occurrences of the combination within a given block. Sixteen bins for P1 and 256 bins for P2 are concatenated to build the LAP histogram for a block.

Therefore, the feature vector length for the proposed method is 16+256 (272) per block. A detailed example of obtaining LAP patterns from a 5×5 pixels region in shown in Fig. 4. Once histograms of all blocks in an image have been computed they are concatenated to form a final feature vector of the image (Fig. 4).

Feature dimensionality reduction using variance: For the LAP representation, feature vector dimension is 272 per block before feature selection. Not all the features in this feature vector are necessary for the classification if they can not quite differentiate faces of different facial expression classes.

The features having higher variances values would have higher power to differentiate faces of different facial expression classes than ones with lower varaince values. So, variance values of the features can be used as indicators for feature selection. The variance value for each feature t of the feature vector can be calculated using Eq. 2:

$$VAR_t = \frac{1}{N}\sum_{j=1}^{N}\left(\alpha_t^j - \mu_N\right)^2 \tag{2}$$

Where:

α^j_t = The value of the feature t of the jth training sample
μ_N = The mean value of the feature t
N = The number of total training samples

The features are then sorted in descending order according to their variance values. The top M features with the highest variances values are then selected as the most contributing features to be used for the classification.

Expression classifier: For differentiating facial expression, the researchers used several classification techniques. Shan *et al*. (2005) did a comparative analysis on four machine learning techniques, namely Template Matching, Linear Discriminant Analysis (LDA), Linear Programming and Support Vector Machine. He showed that SVM was the best in terms of classification accuracy. In this study, SVM is adopted as the classifier for facial expression.

EXPERIMENTS AND RESULT

In general, six or seven type of facial expressions are used to evaluate the facial expression recognition system (Shan *et al*., 2005). The performance of the proposed local descriptor was evaluated on the well-known Japanese Female Facial Expression (JAFFE) Dataset (Lyons *et al*., 1997). The dataset contains 213 images of 7 facial expressions (6 basic facial expressions+1 neutral) posed by 10 Japanese female models. The images in the dataset were taken at the Psychology Department in Kyushu University. An unpublished matlab code fdlibmex was used to detect the face from an image and re-dimension it to 99×99 pixels as a part of preprocessing. The image was then divided into 9×9 blocks. Each contains 11×11 pixels. No further alignment or no attempt was made to remove illumination changes. Linear, polynomial and Radial Basis Function (RBF) kernels were used in LIBSVM to classify the testing images. A ten fold none overlapping cross validation was performed. The 90% of the images from each expression were used for training LIBSVM. The remaining 10% of the images were used for testing. For each fold, different 10% of the images were chosen for testing and it is user-dependent. Ten rounds of training and testing were performed and the average confusion matrix for proposed method was reported. The kernel parameters for the classifier were set to: s = 0 for SVM type C-Svc, t = 0/1/2 for linear, polynomial and RBF kernel, respectively, c = 100 is the cost of the SVM, g = 1/(length of feature vector dimension), b = 1 is for probability

Table 1: Confusion matrix for LAP, classification accuracy achieved 94.41%

	Actual						
C	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Prediction							
Angry	100.0	00.0	00.0	00.0	000.0	00.0	00.0
Disgust	000.0	93.1	06.9	00.0	000.0	00.0	00.0
Fear	000.0	03.1	84.4	00.0	003.1	06.3	03.1
Happy	000.0	00.0	00.0	96.8	000.0	03.2	00.0
Neutral	000.0	00.0	00.0	00.0	100.0	00.0	00.0
Sad	003.2	00.0	06.5	03.2	000.0	87.1	00.0
Surprise	000.0	00.0	00.0	03.3	000.0	00.0	96.7

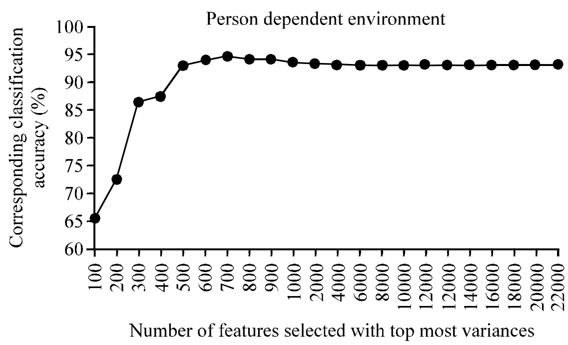

Fig. 5: Number of features selected with top most variances vs. Corresponding Classification Accuracy (%)

estimation. This setting of LIBSVM was found to be suitable for JAFFE dataset with seven classes of data. The RBF kernel normally achieves slightly better recognition accuracy than linear or polynomial kernels (Chang and Lin, 2011). However, the experiments achieved better accuracy using polynomial kernel. The accuracies achieved using linear, polynomial and RBF kernel are 94.11, 94.41 and 93.96%, respectively. No fine-tuning of the C and g parameter had been performed. Further tuning may increase the RBF kernel performance substantially.

Figure 5 shows the plot between accuracy rate and the number of the top most contributing features selected for the classification. It is found that the number of selected features at 700 gives the highest accuracy of 94.41%. Therefore, these 700 features are used to build the classification model and to validate the test images.

To get a better picture of the experimental results of individual facial expression types, the confusion matrix of 7-class expression recognition is given in Table 1. The results are compared with those of some earlier researches on JAFFE dataset (Table 2). All of earlier researches the are based on appearance methods. The feature extraction running time of all the methods can not be directly comparable due to different experimental setup and execution environments.

Table 2: Comparison of classification accuracy of LAP+LIBSVM with some other systems on JAFFE dataset (NN: Neural Network, LDA: Local Discriminant Analysis)

Researchers	Methods	Classifier	Classification accuracy (%)
Proposed	LAP	Multi class SVM (Poly)	94.41
Subramanian *et al.* (2012)	LBP	SVM	88.09
Lyons *et al.* (1999)*	Gabor Filter	LDA-based classification	92.00
Zhang *et al.* (1998)	Gabor Filter	NN	90.10
Guo and Dyer (2003)	Gabor Filter	Linear Programming	91.00

*Used a subset of the dataset

Table 3: Accuarcy and Feature Extraction Time Comparisons

Diameter	LAP (Proposed)	LBP	LBP_{U2}
Classification accuracy	94.41%	91.14%	90.12%
Feature dimension	700	20736	4779
Feature extraction time	0.035 sec	0.068 sec	0.065 sec

Experiments were also performed on JAFFE dataset using other well known local feature methods, i.e., LBP, LBU_{u2}. Table 3 compares the feature dimension feature extraction time and accuracy achieved from using the LAP and the other methods as the feature representations. It can be seen from the experimental results that the LAP outperforms the other methods on both accuracy and feature extraction time.

CONCLUSION

A novel feature representation for facial expression recognition is proposed in this study. Facial feature pattern at a pixel is extracted from pixels gray color intensity values of its neighboring pixels in 5×5 pixels region. To reduce the dimension of the feature vector, features are then selected based on their variance values. The experiments, conducted on JAFFE dataset, demonstrate the superiority of the proposed method over several other appearance-based methods. The proposed method successfully classifies nearly 95% of facial expressions accurately. Extensive experiments illustrate that the proposed method is more effective and takes less extraction time for facial expression recognition than the others. Future works include working with motion pictures where face registration is necessary.

ACKNOWLEDGEMENT

This research was supported by the 2012-2013 research fund of National Institute of Development Administration (NIDA), Bangkok.

REFERENCES

Ahonen, T., A. Hadid and M. Pietikainen, 2006. Face description with local binary patterns: Application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell., 28: 2037-2041.

Ahsan, T., T. Jabid and U.P. Chong, 2013. Facial expression recognition using local transitional pattern on gabor filtered facial images. IETE Tech. Rev., 30: 47-52.

Bartlett, M.S., G. Littlewort, M. Frank, C. Lainscsek, I. Fasel and J. Movellan, 2005. Recognizing facial expression: Machine learning and application to spontaneous Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 20-25, 2005, San Diego, CA, USA., pp: 568-573.

Chang, C.C. and C.J. Lin, 2011. LIBSVM: A library for support vector machines. ACM Trans. Intellig. Syst. Technol., 2: 1-39.

Ekman, P. and W. Friesen, 1978. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, CA., USA.

Guo, G.D. and C.R. Dyer, 2003. Simultaneous feature selection and classifier training via linear programming: A case study for face expression recognition. IEEE Conf. Comput. Vision Pattern Recogn., 1: I-346-I-352.

Kotsia, I. and I. Pitas, 2007. Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Trans. Image Process., 16: 172-187.

Lajevardi, S.M. and Z.M. Hussain, 2012. Feature extraction for facial expression recognition based on hybrid face Adv. Electr. Comput. Eng., 9: 63-67.

Lyons, M.J., J. Budynek and S. Akamatsu, 1999. Automatic classification of single facial images. IEEE Trans. Pattern Anal. Mach. Intell., 21: 1357-1362.

Lyons, M.J., M. Kamachi and J. Gyoba, 1997. The japanese female facial expression (JAFFE) database. http://www.kasrl.org/jaffe_info.html.

Mehrabian, A., 1968. Communication without Words. Psychol. Today, 2: 53-56.

Ojala, T., M. Pietikainen and D. Harwood, 1996. A comparative study of texture measures with classification based on featured distributions. Pattern Recognit., 29: 51-59.

Ojansivu, V. and J. Heikkila, 2008. Blur insensitive texture classification using local phase quantization. Proceedings of the 3rd International Conference on Image and Signal Processing, July 1-3, 2008, Cherbourg-Octeville, France, pp: 236-243.

Shan, C., S. Gong and P.W. McOwan, 2005. Robust facial expression recognition using local binary patterns Image Process., 2: 370-373.

Shan, C., S. Gong and P.W. McOwan, 2009. Facial expression recognition based on local binary patterns: A comprehensive study. Image Vision Comput., 27: 803-816.

Subramanian, K., S. Suresh and R.V. Babu, 2012. Meta-cognitive neuro-fuzzy inference system for human emotion Proceedings of the International Joint Conference on Neural Networks, June 10-15, 2012, Brisbane, QLD, pp: 1-7.

Tian, Y.L., T. Kanade and J.F. Cohn, 2003. Facial Expression Analysis. In: Handbook of Face Recognition, Li, S.Z. and A.K. Jain (Eds.). Springer, New York, pp: 247-274.

Valstar, M. and M. Pantic, 2006. Fully automatic facial action unit detection and temporal. Proceedings of the Conference on Computer Vision and Pattern Recognition, June 17-22, 2006, New York, pp: 149-156.

Valstar, M.F., I. Patras and M. Pantic, 2005. Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern, June 25, 2005, San Diego, CA, USA., pp: 76-84.

Zhang, Y. and Q. Ji, 2005. Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Trans. Pattern Anal. Mach. Intell., 27: 699-714.

Zhang, Z., M. Lyons, M. Schuster and S. Akamatsu, 1998. Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron. Proceedings of the International Conference Automatic Face and Gesture Recognition, IEEE Computer Society Washington, DC., USA., April 14-16, pp: 454-459.