Journal of
Applied Sciences

science
alert

Gradient Direction Pattern: A Gray-scale Invariant Uniform Local Feature Representation for Facial Expression Recognition

Mohammad Shahidul Islam and Surapong Auwatanamongkol

Department of Computer Science, School of Applied Statistics,

National Institute of Development Administration, Bangkok, Thailand

Abstract: Local feature representations are widely used for facial expression recognition due to their simplicity and high accuracy rates achieved. However, local feature representations usually produce a long feature vector to represent a facial image and hence, require long processing time for training and recognition. To alleviate this problem, a simple gray-scale invariant local feature representation is proposed for facial expression recognition. The proposed local feature pattern at a pixel level, represented by a four-bit pattern, is derived based on the gradient directions of the gray color values of its neighboring pixels. A histogram of sixteen bins is required to count numbers of the patterns at the pixel level in a block. The histograms of all blocks in an image are concatenated to form the final local feature vector. To reduce the length of the local feature vector, a variance based feature selection method is used to select patterns that are more relevant and eight out of the sixteen possible patterns can be discarded without compromising the recognition rates. In addition, the result patterns become uniform. Experiments were performed on extended Cohn-Kanade and Japanese JAFFE datasets using Support Vector Machines as classifiers. The experimental results do show that the proposed feature representation is more effective than other local feature representations in terms of accuracy rates and processing time.

INTRODUCTION

Facial Expression is very important for daily activities, allowing someone to express feelings beyond the verbal world and understand each other from various situations. Some expressions help human to act and some give full meaning to the human communication. Mehrabian (1968) observed that the verbal part of human communication contributes only 7%, the vocal part contributes 38% and facial movement and expression yields 55% to the meaning of the communication. This means that the facial part does the major contribution in human communication as well as man-machine interaction. Due to potential applications in man-machine interactions, automatic facial expression recognition attracted much attention of the researchers in recent years (Zeng *et al.*, 2009).

The basic prototypes of facial expressions are neutral, contempt, fear, sadness, disgust, anger, surprise and happiness (Ekman and Friesen, 1978). Most of the facial expression recognition systems (FERS) built in the past are based on the Facial Action Coding System (FACS) (Tian *et al.*, 2001; Tong *et al.*, 2007; Ekman and

Friesen, 1978) etc., which involve very complex facial feature detection and extraction procedures. In these systems, the muscle movements caused by facial expressions are described with 44 different action units (AUs) (Ekman and Friesen, 1978), each of which is related to the facial muscle movements. The forty-four AUs can give up-to 7000 different combinations, with wide variations due to age, size and ethnicity. A model of AUs with multi-state face components was proposed by Tian *et al.* (2001) to represent facial expressions and neural network was used as the classifier of the facial expressions. Hidden Markov models (HMMs) were proposed by Cohen *et al.* (2003) to model human facial expressions from video sequences and the expressions were then recognized using a Naïve-Bayes classifier. (Zhang and Ji, 2005) proposed a facial expression recognition system using Dynamic Bayesian networks (DBNs) as the classification models. Facial expressions were classified based on 26 facial features around the regions of eyes, nose and mouth. A real time system for facial expression recognition was constructed by Anderson and McOwan (2006). A Multichannel Gradient

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

Model (MCGM) was introduced to capture facial motion signatures that would identify facial expressions. The Support Vector Machine Model was used for expression classification. A model considering both space and time was proposed by Yeasin *et al.* (2006) for facial expression recognition and Discrete Hidden Markov models (DHMMs) were used for classification. (Pantic and Patras, 2006) presented a method which can handle a large range of human facial behaviors by recognizing facial muscle actions that produce expressions. They applied face-profile-contour tracking and rule-based reasoning to recognize 20 AUs taking place alone or in a combination in nearly left-profile-view face image sequences. Kotsia and Pitas (2007) manually placed some of Candide grid nodes to face landmarks to create facial wire frame model for facial expressions and a Support Vector Machine (SVM) was used for classification.

The alternative methods are the appearance-based ones which use local feature representations. The local features are much easier for extraction than those of AUs while deliver very high accuracy rates for recognition. Following are the reviews of some previously proposed local feature representations. Ahonen *et al.* (2006) proposed a 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 3×3 pixels region (Fig. 1) is computed using the gray scale color intensities of the pixel and its neighboring pixels as follows:

$$LBP = \sum_{i=1}^{P} 2^{i-1} f(g(i) - c) \tag{1}$$

Where:

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

where, 'c' denotes the gray color intensity of the center pixel, g(i) is the gray color intensity of its neighbors, P stands for the number of neighbors, i.e., 8. Figure 2 shows an example of computing the LBP value at the referenced center pixel of a given 3×3 pixels region. First, the gray color intensity values of the neighboring pixels are converted to corresponding bits in the LBP binary pattern using the f function in the Eq. 1. Then, the binary pattern is converted into the corresponding LBP value. Ojala *et al.* (2002) proposed an extension to the original LBP operator called LBP_{RIU2}. It reduces the length of the feature vector and implements a simple rotation-invariant descriptor. An LBP binary pattern is called uniform if the

Fig. 1: 3×3 pixels local region for the computation of LBP value at the center pixel of E
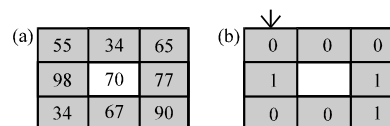
(c) $LBP = 0*2^7 + 0*2^6 + 0*2^5 + 1*2^4 + 1*2^3 + 0*2^2 + 0*2^1 + 1*2^0 = 25$

Fig. 2(a-c): Example of computing the LBP value at the center of a 3x3 block, (a) The gray color intensity values for all pixel in the block, (b) Corresponding binary pattern starting at the indicating point and (c) The Corresponding LBP value

binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa. For example, the patterns 00000000 (0 transitions), 01110000 (2 transitions) and 11001111 (2 transitions) are uniform whereas the patterns 11001001 (4 transitions) and 01010010 (6 transitions) are not. The uniform ones occur more commonly in any image textures than the non-uniform ones; therefore, the latter ones are neglected. The uniform ones yield only 59 different patterns. To create a rotation-invariant LBP descriptor, a uniform pattern can be rotated clockwise P-1 (P = no of bits in the pattern) times. If the rotated pattern matches with other pattern, then they are taken as a single pattern. Hence, instead of 59 bins, only 8 bins are needed to construct a histogram representing the local feature for a given local block. Once, the LBP local features for all blocks of a face are extracted, they are concatenated into an enhanced feature vector. This method is proven to be a growing success. It is adopted by many researchers and is successfully used for facial expression recognition (Zhao and Pietikainen, 2007; Ma and Khorasani, 2004). Although LBP features achieved high accuracy rates for facial expression recognition, LBP extraction can be time consuming. Ojansivu and Heikkila (2008) proposed a new texture descriptor LPQ (Local Phase quantization) which was blur insensitive (Ojansivu and Heikkila, 2008). The spatial blurring was represented as multiplication of the original image and a Point Spread Function (PSF) in the frequency domain. The LPQ method is based on the phase-invariant property of the original image when the PSF is centrally symmetric. Yang and Bhanu (2012) used local binary pattern and local phase quantization together

facial expression recognition system. Kabir *et al.* (2012) proposed LDPv (Local Directional Pattern-Variance) by applying weights to the feature vector using local variance and found it to be effective for facial expression recognition. He used support vector machine as a classifier. Huang *et al.* (2011) used LBP-TOP (LBP on three orthogonal planes) on eyes, nose and mouth. They also developed a method that can learn weights for multiple feature sets in facial components.

In this study, a new local feature representation is proposed to not only yield short feature vector but also effective for facial expression recognition. Proposed feature representation method can capture very rich local feature information from local 3×3 pixels region. It is also robust to monotonic gray-scale changes caused, for example, by illumination variations. When compared with LBP (Ojala *et al.*, 1996) or LBP$_{U2}$ (Ojala *et al.*, 2002), the proposed method performs better both in time and classification accuracy. The classification accuracy is also better than some recent works on facial expression recognition (Chew *et al.*, 2011; Naika *et al.*, 2012; Yang and Bhanu, 2012; Jeni *et al.*, 2012).

PROPOSED FEATURE REPRESENTATION METHOD

The proposed feature representation method is based on facial texture information of a 3×3 pixels local region. The representation operates on gray scale facial images. It is designed to capture for each pixel, the direction pattern of gradients among the gray color intensities of its neighboring pixels. Four directions of gradients are considered. The patterns represent distinctive features for the changes of gray color intensities around a particular pixel. Once the gradient direction pattern is derived for each pixel in a given block, the histogram of all possible patterns is constructed to represent the feature vector of the block. The histograms of all blocks in a given image are then concatenated to form the feature vector for the image. The representation is as follows:

Gradient direction pattern (GDP): The pattern for a pixel is computed from the 3×3 pixels region see Fig. 1. The pattern can be derived as follows:

$$gd(1) = (f-d) \tag{2}$$

$$gd(2) = (c-g) \tag{3}$$

$$gd(3) = (b-h) \tag{4}$$

$$gd(4) = (a-i) \tag{5}$$

$$D(i) = \begin{cases} 0 & gd(i)<0 \\ 1 & gd(i)\geq0 \end{cases} \tag{6}$$

where, a, b, c, d, f, g, h and i are the gray color intensities of neighboring pixels of the current pixel E, i.e. A, B, C, D, F, G, H and I, respectively. The gd(i) represents the gradient value between gray color intensities of two opposite neighboring pixels of the current pixel for the i-th direction. The D(i) corresponds to the gradient direction for the i-th direction. Thus, the binary vector of D contains 4 bits representing $2^4 = 16$ different patterns. A detailed example of the gradient direction pattern extraction at center pixel is given in Fig. 3.

Therefore, the GDP feature vector length for each block is 16. It can be shown that the GDP feature is gray-scale invariant because monotonic gray-scale changes caused, for example, by illumination variations does not affect the directions of the gradients around a pixel. Figure 4 illustrates an example of the GDP pattern being gray-scale invariant. In all the three different illumination conditions for the same local 3×3 pixels area, GDP at the center pixel sustains the same pattern which is 0101.

A feature vector for the emotional expression recognition should have all those essential features needed for classification. Unnecessary or irrelevant features can cause over-fitting due to the curse of dimensionality, as well as long learning and classification time. Hence, feature selection method is suggested as a preprocessing step to address the problem (Kumar, 2009). A feature selection method to be used for the emotional expression recognition must be a supervised one and must work with numeric values of the histograms. Linear Discriminant Analysis (LDA) had been tried but gave poor performance for the analysis. Therefore, a new

(55-90)<0   (34-67)<0   (65-34)>0   (77-98)<0
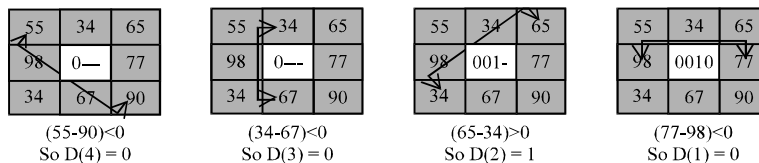So D(4) = 0   So D(3) = 0   So D(2) = 1   So D(1) = 0

Fig. 3: An example of computing GDP pattern at the center of a local 3×3 pixel region, where the gray color intensity values of its neighboring pixels are shown int the surrounding block and the derived GDP pattern is D(4) D (3) D(2) D(1) = 0010
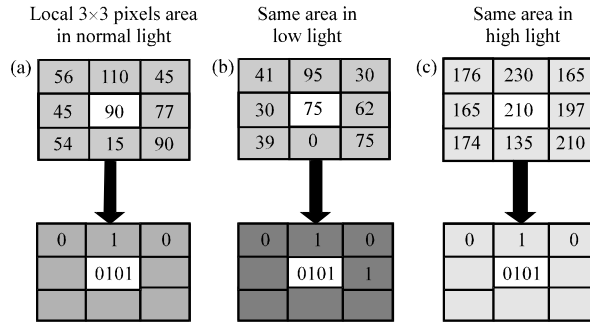
Fig. 4(a-c): Derivation of the GDP pattern at C = 0101 the center pixel of the same local 3x3 pixel area but in normal light, low light and high light conditions, respectively

method of feature selection is introduced. It selects a feature based on its power in discriminating the emotional expression classes. The discriminating power is measured by the difference between two variances of the feature value as follows. One is the variance of the feature for all given images, VAR_a and the other is the average within-class variance of the feature value, VAR_b.

$$VAR_a = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{N_i}(a_i^j - \overline{a})^2 \qquad (7)$$

$$VAR_a = \frac{1}{N}\sum_{i=1}^{C}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}(a_i^j - \mu_i)^2\right)*N_i \qquad (8)$$

$$\Delta VAR = \Delta VAR_a - \Delta VAR_b \qquad (9)$$

Where:

$$\overline{a} = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{N_i}a_i^j$$

and:

$$\mu_i = \frac{1}{N_i}\sum_{j=1}^{N_i}a_i^j$$

a_i^j denotes the feature value of j-th training sample of the ith emotional expression class, μ_i stands for the mean of the feature value of the ith emotional expression class $\overline{a}$ and represents the mean of the feature value of all the training samples. N_i is the number of training samples in the ith class, N is the total number of training samples and ΔVAR represents the difference between the two variances. The high value of the variance difference for a feature means that the average within-class variance of the feature values is quite smaller than the total variance

of the feature value of all the training samples regardless of their classes. Hence, the feature should be suitable for distinguishing samples from one class to the others and so possesses high discriminating power. Features can then be ranked for selection based on values of their ΔVAR.

The feature selection method was performed on the two datasets, the Extended Cohn-Kanade Dataset (CK+) (Lucey *et al.*, 2010) and JAFFE Dataset (Lyons *et al.*, 1997). The results of feature selection on both datasets show that the features with binary patterns consisting of at most one 0-1 or 1-0 transition, i.e., 0000, 0001, 0011, 0111, 1111, 1000, 1100 and 1110, have high values of the variance differences while the others have very low values. Therefore, the features with those eight GDP are selected. This helps to reduce the length of the feature vector by half as well as makes the feature patterns uniform. The effects of the selection on the recognition accuracy will be discussed in the next section.

Figure 5 shows the framework for the proposed facial expression recognition system.

The framework consists of the following steps:

- For each of training images, convert it to gray scale if in different format
- Detect the face in the image, resize using bilinear interpolation and divide it into equal size blocks
- Compute GDP feature for each pixel of the image
- Construct the histogram for each block
- Concatenate the histograms to get the feature vector
- Build a Multiclass Support Vector Machine for face expression recognition using feature vectors of the training images
- Do step 1 to 5 for each of testing images and use the Multiclass Support Vector Machine from step 6 to identify the face expression of the given testing image
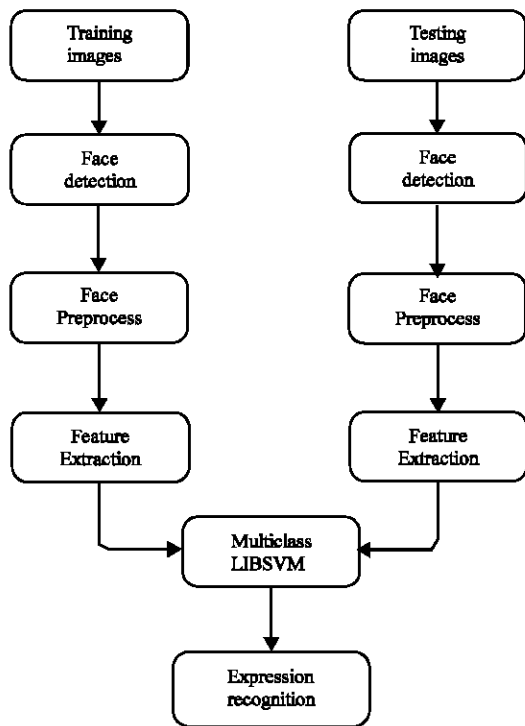
Fig. 5: Framework of the proposed facial expression recognition system

EXPERIMENTS AND RESULTS

Two datasets were used in the experiments, CK+ (Lucey *et al.*, 2010) and JAFFE (Lyons *et al.*, 1997). There are 326 peak facial expressions from 123 subjects in CK+ divided into seven emotion categories. They are 'Anger', 'Contempt', 'Disgust', 'Fear', 'Happy', 'Sadness' and 'Surprise'. No subject with the same emotion is collected more than once. All the facial images in the dataset are posed. Figure 6 shows some face samples from the dataset.

JAFFE dataset contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. The photos were taken at the Psychology Department in Kyushu University. Every expression is taken more than once from each subject. Some sample faces from the JAFFE dataset are shown in Fig. 7.

Table 1 shows the numbers of instances of expressions of both datasets.

The steps of face detection, preprocessing and feature extraction are illustrated in Fig. 8.

Face detection was done using fdlibmex library from Matlab. The library consists of single mex file with a single function that takes an image as input and returns the frontal face. The face was then masked using an elliptical shape. According to the past research on facial expression recognition, higher accuracy can be achieved if the input face is divided into several blocks (Ojala *et al.*, 2002). Thus GDP features were extracted from each block using proposed method. Gathering feature histograms of all the blocks produces a unique feature vector for a given image. A ten-fold none overlapping cross validation was used in the experiments. Using LIBSVM (Chang and Lin, 2011) a multiclass support vector machine with randomly chosen 90% from each expression as the training images is constructed. The remaining 10% of images was used as testing images. There was no overlap between the folds and it was user-dependent. Ten rounds of training and testing were conducted and the average confusion matrix for each proposed method was reported and compared against the others. The kernel parameters for the classifier are set to: s = 0 for SVM typeC-Svc, t = 1 for polynomial kernel function, c = 1 is the cost of SVM, g = 1/(length of feature vector dimension), b = 1 for probability estimation. Other kernel and parameter settings were also tried but the above setting is found to be suitable for both datasets with seven classes of facial expressions.

Several experiments were conducted to find the suitable face dimension and number of blocks. For CK+ dataset, 180×180 pixels face dimension and 9×9 = 81 blocks yields the best performance in terms of accuracy. For JAFFE dataset, 99×99 pixels face dimension and 9×9 = 81 blocks yields the best performance in terms of accuracy. To prove GDP method to be gray-scale invariant, the gray color intensity of random instances from the datasets were manually changed according to different illumination conditions, see the example instances in Fig. 9.

The accuracy results are found to be consistent even after changing the illumination, therefore, the proposed method is bound to be gray-scale invariant.

The classification accuracy before and after feature selection for both datasets are shown in Table 2. It can be seen from the two results that the feature selection does not affect the accuracy achieved on both datasets while helps reduce the number of features by half.

The experimental results for face expression recognition using the proposed feature representation method on both datasets are shown using confusion matrices in Table 3.

It can be seen from the two confusion matrices that some particular expression classes, e.g., contempt and fear, are consistently more difficult to classify than the others. Some instances of these expressions are difficult to distinguish even by a human.
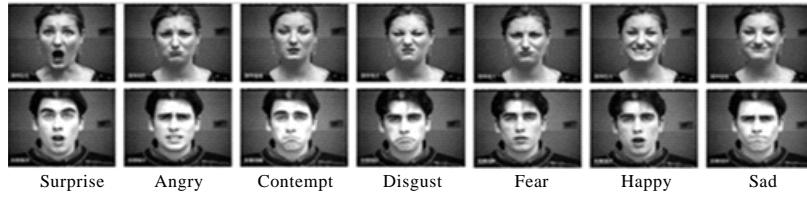
Fig. 6: Some image samples from Extended Cohn-Kanade (CK+) dataset for the 7 facial expressions, two in each column for each

Fig. 7: Some image samples from JAFFE dataset for the 7 facial expressions, two in each column for each
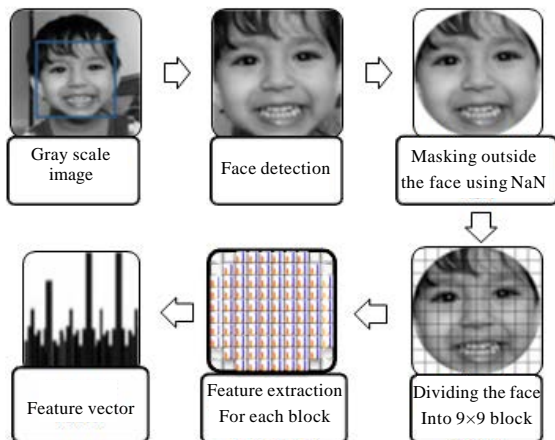

Fig. 8: Steps of facial feature extraction

Table 1: Expression instances from each dataset

Expression class	CK+	JAFFE
Anger	45	30
Disgust	59	29
Fear	25	32
Happy	69	31
Sadness	28	31
Surprise	82	30
Contempt	18	-
Neutral	-	30

Table 2: Classification accuracy before and after feature dimension reduction

	Classification accuracy	
	Before feature selection (%)	After feature selection (%)
CK+	91.96	91.96
JAFFE	92.04	92.11

Table 3: Confusion Matrices results for (a) CK+ and (b) JAFFE dataset

CK+		Angry	Contempt	Disgust	Fear	Happy	Sad	Surprise
Prediction	Angry	82.2	4.4	6.7	0.0	0.0	6.7	0.0
	Contempt'	5.6	83.3	0.0	5.6	0.0	5.6	0.0
	Disgust	1.7	0.0	96.6	0.0	1.7	0.0	0.0
	Fear	4.0	4.0	4.0	88.0	0.0	0.0	0.0
	Happy	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	Sad	14.3	3.6	0.0	3.6	0.0	75.0	3.6
	Surprise	0.0	1.2	0.0	0.0	0.0	0.0	98.8

JAFFE		Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Prediction	Angry	93.3	0.0	0.0	0.0	0.0	6.7	0.0
	Disgust	6.9	86.2	3.4	0.0	0.0	3.4	0.0
	Fear	0.0	3.1	81.3	3.1	3.1	6.3	3.1
	Happy	0.0	0.0	0.0	96.8	0.0	3.2	0.0
	Neutral	0.0	0.0	0.0	0.0	96.7	3.3	0.0
	Sad	3.2	0.0	3.2	3.2	0.0	90.3	0.0
	Surprise	0.0	0.0	3.3	3.3	0.0	0.0	93.3

Experiments were also performed on the two datasets using other appearance-based methods, i.e., LPQ, LBP and LBPu2, using the same experimental setup

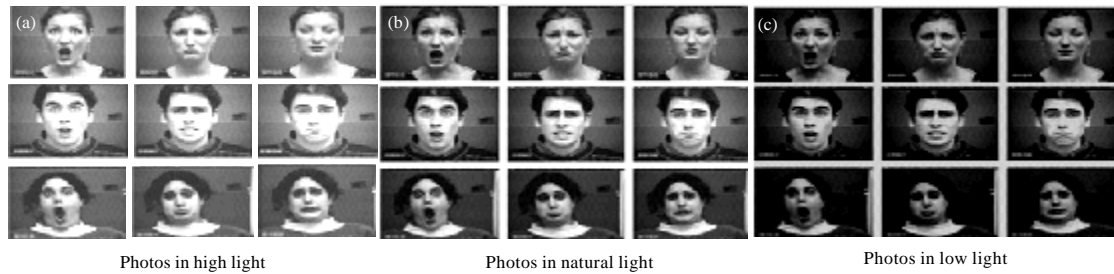Photos in high light Photos in natural light Photos in low light

Fig. 9(a-c): Samples from CK+ dataset in different illumination conditions (different gray-scale intensity)

Table 4: Classification accuracy and processing time comparison CK+ dataset and JAFFE dataset

Method	Classification accuracy (%)	Feature extraction time (sec)	Learning time (sec)	Classification time (sec)
CK+				
Proposed method	91.96	0.019	1.66	0.002
LPQ	80.21	0.29	66.00	0.070
LBP	91.64	0.07	66.00	0.070
LBP$_{u2}$	90.12	0.06	37.00	0.025
JAFFE				
Proposed method	92.11	0.006	1.16	0.002
LPQ	79.60	0.095	46.00	0.070
LBP	91.14	0.026	46.00	0.070
LBP$_{u2}$	90.98	0.025	25.00	0.025

Table 5: Comparison on facial expression recognition accuracy obtained by the proposed method and by some other recent works

Author	Method	No. of subjects	No. of Images	Classification accuracy (%)
Chew *et al.* (2011)	Appearance-based (PDM)	123	327	80+
Naika *et al.* (2012)	Appearance-based (EAR-LBP)	123	327	82
Yang and Bhanu (2012)	Appearance-based (LBP + LPQ)	123	316	83
Jeni *et al.* (2012)	Shape-based (68 Landmarks)	123	593	87
Proposed	GDP	123	326	92

and execution environments. The achieved classification accuracy, feature extraction time for a single facial image, learning time for a single fold and classification time of a single image are shown in comparison with those of the proposed method in Table 4. It is clear that GDP outperforms all the methods in terms of the processing time and classification accuracy rate. Due to the limitations to implement some other recent works on facial expression recognition, only classification accuracy on CK+ of the proposed method is compared with the results of some other recent works in Table 5. It can be seen that the proposed method outperforms the other works as well.

CONCLUSION

A gray-scale invariant uniform and local feature representation method is proposed for facial expression recognition. For each pixel in a gray scale image, the method extracts the local binary pattern using four possible gradient directions between two opposite neighboring pixels in 3×3 region. A variance-based feature selection is proposed and used to reduce the number of features by half. The resultant features become uniform with no more than one e.g., 0-1 or 1-0 transition in the binary patterns. It can also be shown that the features are gray scale invariant and not affected by different illumination conditions. The method is very effective for facial expression recognition and outperformed LPQ, LBP and LBP$_{U2}$, in terms of both classification accuracy and processing time. The classification accuracy is also better than those achieved by some other recent works on facial expression recognition. Future research work may include the incorporating AdaBoost or SimpleBoost algorithm into the facial expression recognition process. It is expected they can boost the accuracy even further.

ACKNOWLEDGMENT

REFERENCES

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

Anderson, K. and P.W. McOwan, 2006. A real-time automated system for the recognition of human facial expressions. IEEE Trans. Syst. Man Cybern. B Cybern., 36: 96-105.

Chang, C.C. and C.J. Lin, 2011. Libsvm: A library for support vector machines. ACM Trans. Intell. Syst. Technol., Vol. 2, No. 3.

Chew, S.W., P. Lucey, S. Lucey, J. Saragih, J.F. Cohn and S. Sridharan, 2011. Person-independent facial expression detection using constrained local models. Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, March 21-25, 2011, Santa Barbara, CA., USA., pp: 915-920.

Cohen, I., N. Sebe, A. Garg, L.S. Chen and T.S. Huang, 2003. Facial expression recognition from video sequences: Temporal and static modeling. Comput. Vision Image Understanding, 91: 160-187.

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.

Huang, X., G. Zhao, M. Pietikainen and W. Zheng, 2011. Expression recognition in videos using a weighted component-based feature descriptor. Proceedings of the 17th Scandinavian Conference on Image Analysis, May 2011, Ystad, Sweden, pp: 569-578.

Jeni, L.A., A. Lorincz, T. Nagy, Z. Palotai, J. Sebok, Z. Szabo and D. Takacs, 2012. 3D shape estimation in video sequences provides high precision evaluation of facial expressions. Image Vision Comput., 30: 785-795.

Kabir, H., T. Jabid and O. Chae, 2012. Local directional pattern variance (LDPv): A robust feature descriptor for facial expression recognition. Int. Arab J. Inf. Tecnol., 9: 382-391.

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.

Kumar, A.C., 2009. Analysis of unsupervised dimensionality reduction techniques. Comput. Sci. Inform. Syst., 6: 217-227.

Lucey, P., J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, 2010. The extended cohn-kande dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression. Proceedings of the 3rd IEEE Workshop on CVPR for Human Communicative Behavior Analysis, June 18, 2010, San Francisco, CA., USA.

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

Ma, L. and K. Khorasani, 2004. Facial expression recognition using constructive feedforward neural networks. IEEE Trans. Syst. Man Cybernet. Part B: Cybernet., 34: 1588-1595.

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

Naika, C.L.S., S.S. Jha, P.K. Das and S.B. Nair, 2012. Automatic facial expression recognition using extended AR-LBP. Proceedings of the 6th International Conference on Information Processing: Wireless Networks and Computational Intelligence, August 10-012, 2012, Bangalore, India, pp: 244-252.

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.

Ojala, T., M. Pietikainen and T. Maenpaa, 2002. Multiresolution Gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Analysis Machine Intellig., 24: 971-987.

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.

Pantic, M. and I. Patras, 2006. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. IEEE Trans. Syst. Man Cybernet. Part B: Cybernet., 36: 433-449.

Tian, Y.L., T. Kanade and J.F. Cohn, 2001. Recognizing action units for facial expressions analysis. IEEE Trans. Pattern Analysis Machine Intell., 23: 97-115.

Tong, Y., W. Liao and Q. Ji, 2007. Facial action unit recognition by exploiting their dynamic and semantic relationships. IEEE Trans. Pattern Anal. Machine Intelli., 29: 1683-1699.

Yang, S. and B. Bhanu, 2012. Understanding discrete facial expressions in video using an emotion avatar image. IEEE Trans. Syst. Man Cybern. B Cybern., 42: 980-992.

Yeasin, M., B. Bullot and R. Sharma, 2006. Recognition of facial expressions and measurement of levels of interest from video. IEEE Trans. Multimedia, 8: 500-508.

Zeng, Z., M. Pantic, G.I. Roisman and T.S. Huang, 2009. A survey of affect recognition methods: Audio, visual and spontaneous expressions. IEEE Trans. Pattern Anal. Machine Intelli., 31: 39-58.

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

Zhao, G. and M. Pietikainen, 2007. Dynamic texture recognition using localbinary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell., 29: 915-928.