Facial Expression Recognition based on A Completed Modelling of Local Binary Pattern

Nhan Thi Cao[#1], An Hoa Ton-That[#2], Hyung-Il Choi[#3*]

[#]School of Media, Soongsil University, Seoul, Korea

[1]ctnhen@yahoo.com

[2]an_tth@yahoo.com

[3]hic@ssu.ac.kr

***Abstract*— Local binary pattern has been used as an effective feature for facial expression recognition. This feature only uses sign component of the local difference sign-magnitude transformation. However, it was also discovered that the information contained in magnitude difference can provide a significant performance improvement. This paper presents a method for facial expression recognition based on completed local binary pattern and support vector machine. The method has experimented on JAFFE facial expression database (213 images). Experimental results show the effectiveness of the proposed method for obtaining first-rate recognition result.**

Keywords— **Facial expression recognition, completed local binary pattern, support vector machine.**

I. INTRODUCTION

With the development of intelligent communication systems and data-driven animation, Facial expression recognition has been sped up in application research and obtained results better and better. There are three essential steps for an automatic facial expression recognition system: face acquisition, facial expression feature extraction and facial expression classification [1]. Many methods have been developed based on extracting appropriate features associating different classifying techniques in order to get more and more better effects of facial expression recognition. One of features investigated is Local Binary Pattern (LBP). The most important properties of LBP features are their tolerance against illumination changes and their computational simplicity [2], [3]. However, the local binary pattern only usually uses conventional sign component and ignores the magnitude component. In this paper, we tried to combine sign component and magnitude component for extracting feature called completed local binary pattern as in [4]. Experiments performed on Japanese Female Facial Expression (JAFFE) database obtained a recognition rate of above 95% for seven-class expression set.

The rest of the paper is organized as follows: In Section II, the image pre-process described for extracting face image features. Section III presents the completed local binary pattern. Section IV, the support vector machine is introduced, Section V shows experiments and the results. Finally, Section VI, the conclusions are given.

[*]Corresponding author: Hyung-Il Choi

II. THE IMAGE PRE-PROCESS

Face image pre-process is a process to attain normalized face images from input face images gotten from a camera or a database. The normalized face images are used for extracting facial expression features. This process can be divided into two steps: basic step and enhancement step. The basic step is to detect the face region of an input face image and eliminate redundant regions. This step can carry out by manual or a real-time face detector …, and enhancement step is to optimize the face region for extracting facial expression features. This step can be made by cropping methods, image normalization or image filter processes. Then the face images are rescaled and used for feature extraction. Figure 1 shows the process of face image preprocess.

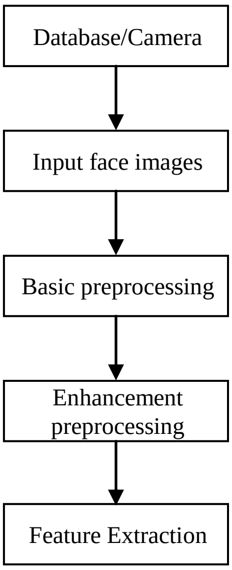

Fig. 1 The process of face image preprocess

In this paper, the image preprocess is implemented as in [5]. It included two steps of preprocess: basic process and enhancement process as following.

Human face images from a camera or a database commonly contain much redundant information e.g. background or non-face regions. Figure 2 shows human face images from database JAFFE.

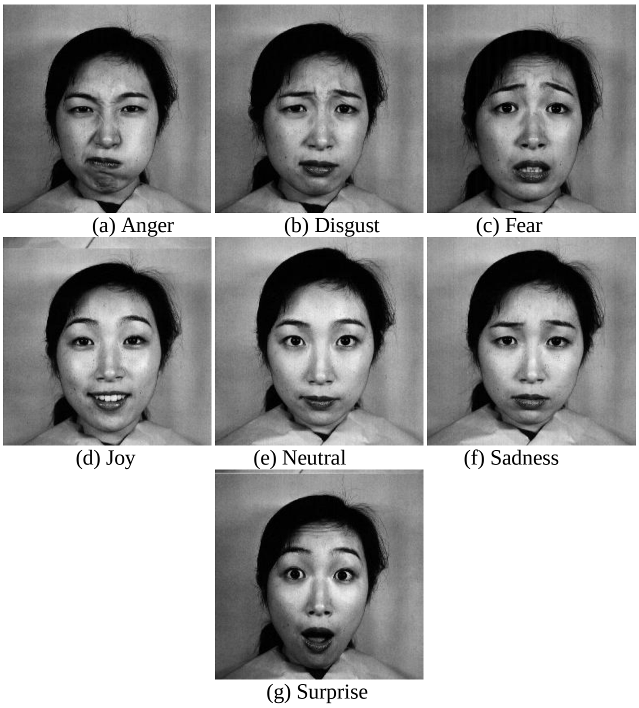

Fig. 2 Face images from database JAFFE

In the basic step, the robust real-time face detection algorithm developed by Viola and Jones [6] is applied. Figure 3 shows the robust real-time face detector applied for a face image and Fig. 4 shows human face images are obtained from Fig. 2 by the algorithm.

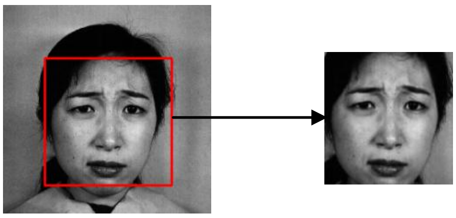

Fig. 3 Face region cropped by the robust real-time face detector (red square)

However, face images still contain some redundant areas that can impact accurate recognition result and processing speed, so in the enhancement step, we used a cropping technique as in Fig. 5.

First determining size of square *S* used for cropping the human face in images. The side w_2 of square S will be equal to the widthwise of the human face. The size of square S depends on each database even each image. However, based on tested results of some databases by the image preprocess method as in [5], the widthwise of the human face accounts for from 75% to 82% of the widthwise of face images obtained from the robust real-time face detector. These percentages are calculated on images of each database to select the side w_2 of the square in pixels. Values of w_2 are counted as experimental parameters.

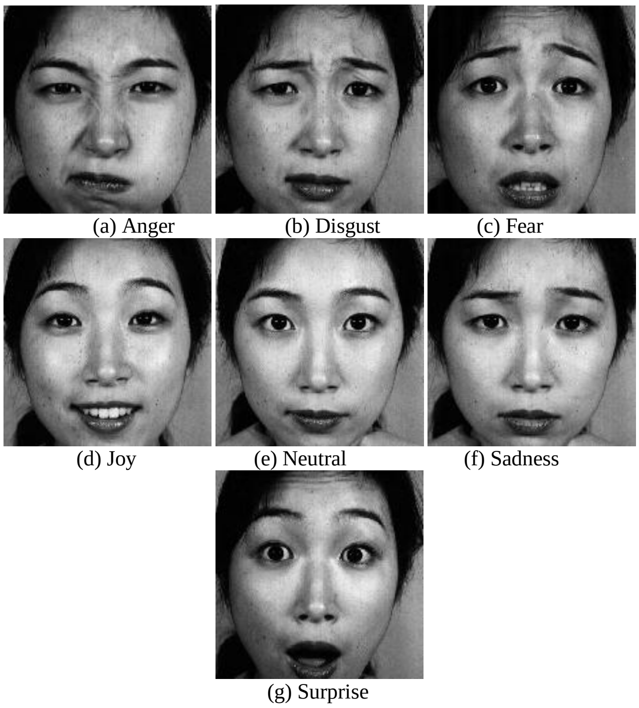

Fig. 4 Scaled face images obtained from the robust real-time face detector

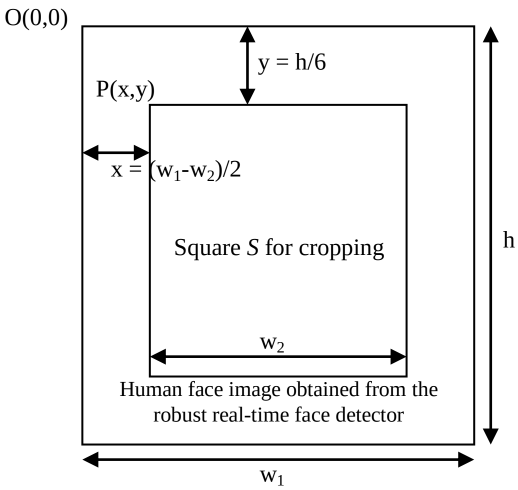

Fig. 5 Shows face region cropped by the cropping method

Then determine co-ordinate *P(x, y)* from left-up corner of the image applied the square *S* to crop the human face. Let *O(0, 0)* is co-ordinate at left-up corner of human face image obtained from the the robust real-time face detector, h is the height of the face image, w_1 is the width of the face image and w_2 is the width of the square. So, co-ordinates y = h/6 and x = $(w_1-w_2)/2$. Expression y = h/6 based on neutral expression face image. Normally, forehead region occupies one-fourth of human face height. Thus forehead region occupies a not small region on human face region but it does not contain much essential information of face expressions. For this reason, two-third (2/3) of upper forehead region is trimmed and one-third (1/3) of lower forehead region from eyebrows is retained.

Finally, the human face image obtained from the robust real-time face detector is cropped by square S at co-ordinate P(x,y). Figure 6 shows the cropping technique applied for a face image and Fig. 7 shows the human face images were cropped by the technique.

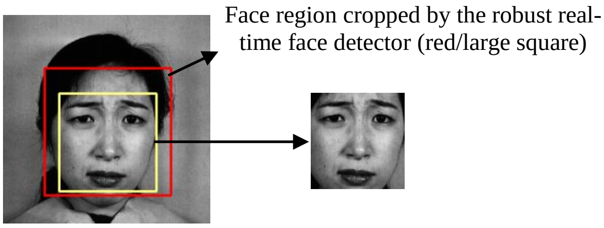

Fig. 6 Face region cropped by the cropping technique (the small square)

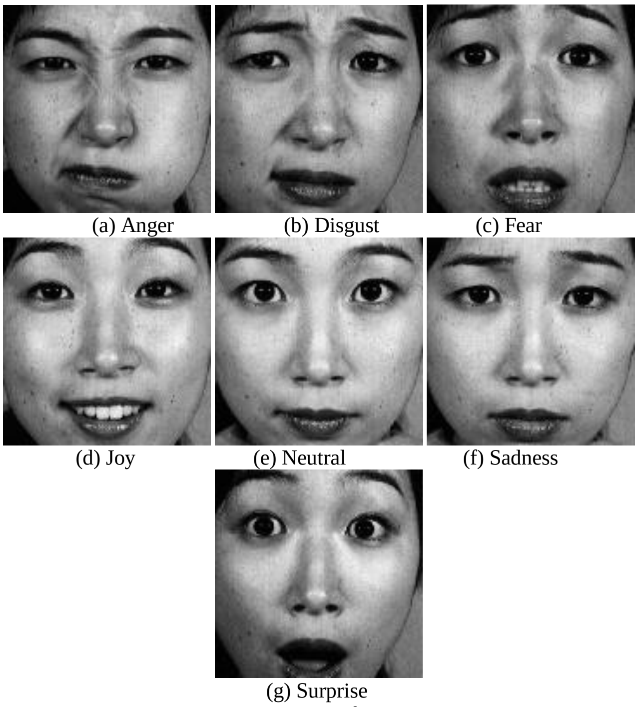

Fig. 7 Face images after cropping

This method aims at reducing processing time in steps of feature extraction and facial expression recognition, and most important being to improve the rate of facial expression recognition.

III. The Completed Local Binary Pattern for Facial Expression Recognition

A. Review of Local Binary Pattern

The LBP (Local Binary Pattern) operator was first introduced as a complementary measure for local image contrast [2], [3]. A local binary pattern code is computed for a pixel in an image by comparing it with its neighbors as in (1):

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, s(x) = \begin{cases} 1, x \geq 1 \\ 0, x < 1 \end{cases} \quad (1)$$

where g_c is gray value of the central pixel, g_p is the gray value of its neighbors, P is the total number of involved neighbors and R is the radius of the neighborhood. Based on the operator, each pixel of image is labeled by a LBP code.

For facial expression recognition, the uniform LBP code is usually used. A LBP code is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular [3]. For example, 00000000, 001110000 and 11100001 are uniform patterns.

An uniform LBP operator is denoted $LBP_{P,R}^{u2}$. The subscript describes the operator using a *(P,R)* neighborhood; the superscript u2 indicates using only uniform patterns and labeling all remaining patterns with a single label. A histogram of a labeled image $f_k(x, y)$ can be defined as following:

$$H_i = \sum_{x,y} I(f_k(x,y) = i), \quad i = 0, \dots, n-1 \quad (2)$$

where n is the number of different labels produced by the LBP operator and

$$I(A) = \begin{cases} 1 & A \text{ is true} \\ 0 & A \text{ is false} \end{cases} \quad (3)$$

This histogram contains information about the distribution of the local micro-patterns, e.g. spots, edges, corners or flat areas etc., over the whole image. To represent the face efficiently, features extracted should retain spatial information. For this reason, the face image can be divided into m small regions R_0, R_1,…, R_m as shown in Fig. 9. So a spatially enhanced histogram is expressed as

$$H_{i,j} = \sum_{x,y} I(f_k(x,y) = i) \quad I\left((x,y) \in R_j\right) \quad (4)$$

where $i = 0,\dots, n-1$, $j = 0,\dots, m-1$.

B. Local Difference Sign-Magnitude Transform

According to [4], based on a central pixel g_c and its P circularly and evenly spaced neighbors g_p, $p = 0, 1, \dots, P-1$, the difference between g_c and g_p can be calculated as $d_p = g_p - g_c$.

The local difference vector $[d_0,\dots, d_{p-1}]$ describes the image local structure at g_c and can be decomposed into two components:

$$d_p = s_p * m_p \quad with \begin{cases} s_p = sign(d_p) \\ m_p = |d_p| \end{cases} \quad (5)$$

where $s_p = \begin{cases} 1, d_p \geq 0 \\ 0, \ d_p < 0 \end{cases}$ is sign of d_p and m_p is the magnitude of d_p. The Eq. (5) is called the local difference sign-magnitude transform and it transforms the local difference vector $[d_0,\dots, d_{p-1}]$ into a sign vector $[s_0,\dots, s_{p-1}]$ and a magnitude vector $[m_0,\dots, m_{p-1}]$. Fig. 8 shows an example of the transformation.

25	48	76
19	**32**	41
36	87	9

(a)

-7	16	44
-13		9
4	55	-23

(b)

0	1	1
0		1
1	1	0

(c)

7	16	44
13		9
4	55	23

(d)

Fig. 8 (a) A 3x3 sample block, (b) local difference, (c) sign component and (d) magnitude component

C. Completed LBP with CLBP_S and CLBP_M operators

The transformation shows that the original LBP uses only the sign vector to code the local pattern because it is proved that d_p can be more accurately approximated by using the sign component s_p than the magnitude component m_p. However, it is also found that the magnitude component may contribute additional discriminative information for pattern recognition if it is properly used.

The sign component is the same as the original LBP operator defined in (1). In completed LBP, this component is denoted CLBP_S operator, whereas the magnitude component is continuous values as a replacement for the binary ‘1’ and ‘0’ values. To code this component in a consistent format with that of sign component to exploit their additional information, the magnitude component is denoted CLBP_M operator and defined as following:

$$CLBP\_M_{P,R} = \sum_{p=0}^{P-1} t(m_p, c)2^p , t(x,c) = \begin{cases} 1, x \geq c \\ 0, x < c \end{cases} \quad (6)$$

where the threshold c is to be determined adaptively and set as the mean value of m_p from the whole image.

Two CLBP_S and CLBP_M operators have same binary string format, so they can be used together for pattern recognition. To form a CLBP descriptor, histograms of CLBP_S and CLBP_M codes of the image are made then they can be combined by two ways: in concatenation or jointly. In the first way, the histograms of the CLBP_S and CLBP_M codes are calculated separately, and then concatenate the two histograms together. This CLBP scheme can be represented as “CLBP_S_M”. In the second way, a joint 2D histogram of the CLBP_S and CLBP_M codes are calculated. This CLBP scheme is represented as “CLBP_S/M”.

D. Extracting CLBP feature

The algorithm of extracting CLBP features for facial expression recognition can be summarized as following:

- Face image registration for extracting LBP features
- Apply the robust real-time face detector to the face image
- Crop the face image as Fig. 5 (section II)
- Resize the face image to appropriate resolution (square with 8-multiple side)
- Divide the face image into square regions being 8x8 pixels as in Fig. 9

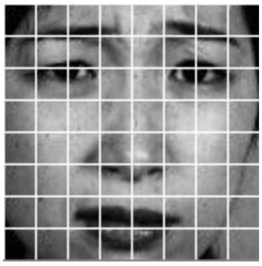

Fig. 9 A face image resized in 64x64 pixels and divided 8x8 regions of 8x8 pixels

- Calculate CLBP_S and CLBP_M code of each pixel in each region
- Build up uniform histogram for CLBP_S and CLBP_M of each region.
- Concatenate histograms of CLBP_S and CLBP_M of each region into CLBP_S_M
- Concatenate histograms CLBP_S_M of regions from left to right, up to down to obtain CLBP features of the face image or its feature vector.

IV. Support Vector Machine

Support Vector Machine (SVM) proposed by Vladimir N. Vapnik [7]-[9] is a powerful, effective and popular machine learning technique for data classification. This method based on statistical learning theory with close foundation of mathematics to ensure that the results are optimal. SVM performs an implicit mapping of data into higher (perhaps infinite) dimensional feature space and then constructs a separating hyperplane with the maximal margin separate data in this higher dimension space. Many applications have confirmed SVM obtaining high results for classifying facial expression [5], [10]-[12].

Given a training set of labeled examples $\{(x_i, y_i), i = 1,\ldots, l\}$ where $x_i \in R^n$ and $y_i \in \{-1, 1\}$, a new test example x is classified by the following function:

$$f(x) = sign(\textstyle\sum_{i=1}^{l} u_i y_i K(x_i, x) + b) \quad (7)$$

where u_i are Lagrange multipliers of a dual optimization problem that describe the separating hyperplane, $K(. , .)$ is a Kernel function, and b is the threshold parameter of the hyperplane. The training sample x_i with $u_i > 0$ is support vectors. $K(x_i, x_j)$ is kernel based on a non-linear mapping Φ that mapped the input data into higher dimensional space and in the form of $\Phi(x_i).\Phi(x_j)$. Some frequently used kernel functions being used in SVM are the linear, polynomial, and Radial Basis Function (RBF) kernels.

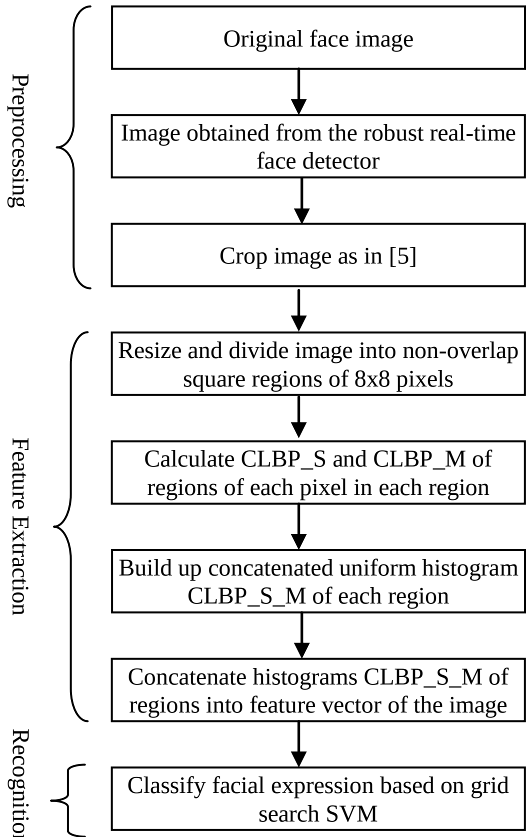

Fig. 10 Experimental process

SVM makes binary decisions, so the multi-class classification here is accomplished by using the one-against-rest technique, which trains binary classifiers to discriminate one expression from all others, and outputs the class with the largest output of binary classification [13].

In this work, we used the SVM functions with Radial Basis Functions kernel. In order to choose optimal parameters, we implement grid-search approach as in [14].

V. Experiments and Results

A. Database

We have applied our proposed method on Japanese Female Facial Expression (JAFFE) database. JAFFE database [15] includes 213 gray images of ten Japanese female facial expression. Each person represents seven different facial expressions: anger, disgust, fear, joy, neutral, sadness and surprise. Most each facial expression of each subject has 3 different images, but there are three cases having two images and six cases having four images. Original images from the database have a resolution of 256x256 pixels. In our experiments, we selected all 213 images as experiment samples. A few examples of facial expression images from the JAFFE database are shown in Fig. 11.

B. Experimental process

We used 3-fold cross-validation method for experiments on platform C++. Experimental process is shown in Fig. 10 and can be summarized as following:

- Extracting CLBP feature as in section III. D
- Classify facial expression based on grid search SVM

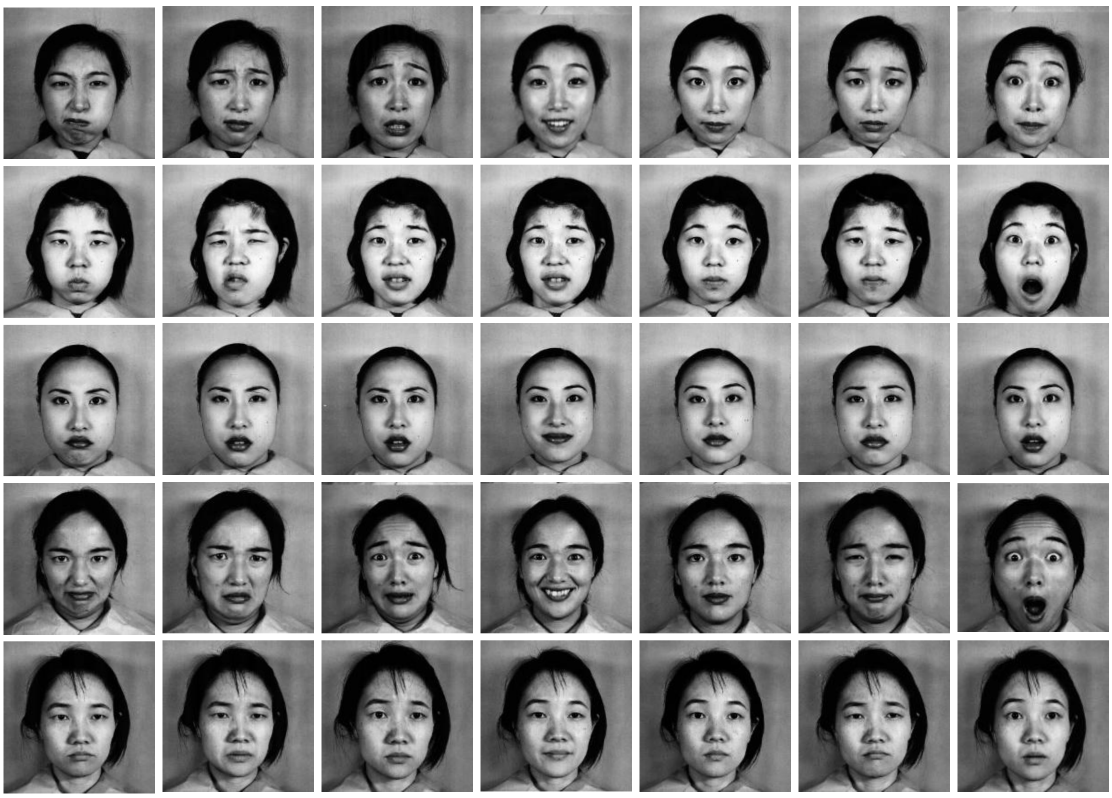
Fig. 11 The sample facial expression images from JAFFE database

[14]

TABLE I
CONFUSION MATRIX OF JAFFE DATABASE AT 80% OF PERCENTAGE W_2/W_1

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Neutral (%)	Sadness (%)	Surprise (%)
Anger	**100.00**	0.00	0.00	0.00	0.00	0.00	0.00
Disgust	3.33	**93.34**	3.33	0.00	0.00	0.00	0.00
Fear	0.00	6.36	**90.61**	0.00	0.00	3.03	0.00
Joy	0.00	0.00	0.00	**100.00**	0.00	0.00	0.00
Neutral	0.00	0.00	0.00	0.00	**100.00**	0.00	0.00
Sadness	0.00	0.00	0.00	0.00	6.67	**93.33**	0.00
Surprise	0.00	0.00	3.33	3.33	3.33	0.00	**90.10**
						Average:	**95.32**

Experimental results of the JAFFE database are shown in Table 1.

VI. CONCLUSIONS

We presented a novel experimental method of facial expression recognition, based on completed local binary pattern. Experimental results show that this method can obtain remarkably more accurate recognition rate in comparison with other methods even with small scaled images. The completed local binary pattern feature can be combined with Gabor filter to may get better results.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2013R1A1A2012012)

We would like to thank Professor Michael J. Lyons for the use of the JAFFE database.

REFERENCES

[1] Y. L. Tian, T. Kanade and J. F. Cohn, *Handbook of face recognition*, chapter 11, Facial expression analysis Springer, Heidelberg, 2005.

[2] Timo Ojala, Matti Pietikainen and David Harwood, *A comparative study of texture measures with classification based on featured distribution*, Pattern Recognition, vol. 29, no. 1, pp. 51~59, 1996.

[3] Timo Ojala, Matti Pietikainen and Topi Maenpaa, *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7), pp. 971~987, 2002.

[4] Zhenhua Guo, Lei Zhang and David Zhang, *A Completed Modeling of Local Binary Pattern Operator for Texture Classification*, IEEE Transactions on Image Processing, vol. 19, no. 6, pp. 1657~1663, 2010.

[5] Nhan Thi Cao, An Hoa Ton-That and Hyung-Il Choi, *An efficient Method of Face Image Preprocess for Facial Expression Recognition*, International Journal of Engineering Associates, vol. 2, no. 4, pp. 10~16, 2013.

[6] Paul Viola and Michael J. Jones, *Robust real-time face detection*, International Journal of Computer Vision, 57(2), pp. 137~154, 2004.

[7] Corina Cortes and Vladimir Vapnik, *Support-Vector Networks*, Machine Learning, 20, pp. 273-297, 1995.

[8] Vladimir N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.

[9] Vladimir N. Vapnik, *An Overview of Statistical Learning Theory*, IEEE Transactions on Neural Network, Vol. 10, pp. 988~999, 1999.

[10] Irene Kotsia and Ioannis Pitas, *Facial expression recognition in image sequences using geometric deformation features and support vector machines*. IEEE Trans on Image Process, vol. 16, no. 1, pp. 172~187, 2007.

[11] Daw-Tung Lina and De-Cheng. Pan, *Integrating a mixed-feature model and multiclass support vector machine for facial expression recognition*, Integrated Computer-Aided Engineering 16, pp. 61–74, IOS Press, 2009.

[12] Chen-Chiung Hsieh and Meng-Kai. Jiang, *A Facial Expression Classification System based on Active Shape Model and Support Vector Machine*, IEEE International Symposium on Computer Science and Society, pp. 311~314, 2011.

[13] Caifeng Shan, Shaogang Gong and Peter W. McOwan, *Facial expression recognition based on Local Binary Patterns: A comprehensive study*, Image and Vision Computing 27, pp. 803~816, 2009.

[14] Chih-Wei Hsu, Chih-Chung Chang and Chih-Jen Lin, *A Practical Guide to Support Vector Classification*, Tech. Rep., Taipei, 2010.

[15] Michael J. Lyons, Julien Budynek and Shigeru Akamatsu, *Automatic classification of single facial images*, IEEE Transaction on Pattern Analysis Machine Intelligent, vol. 21, no. 12, pp. 1357~1362, 1999.

