

Robust Facial Expression Recognition Based on Local Directional Pattern

Taskeed Jabit, Md. Hasanul Kabir, and Oksam Chae

Automatic facial expression recognition has many potential applications in different areas of human computer interaction. However, they are not yet fully realized due to the lack of an effective facial feature descriptor. In this paper, we present a new appearance-based feature descriptor, the local directional pattern (LDP), to represent facial geometry and analyze its performance in expression recognition. An LDP feature is obtained by computing the edge response values in 8 directions at each pixel and encoding them into an 8 bit binary number using the relative strength of these edge responses. The LDP descriptor, a distribution of LDP codes within an image or image patch, is used to describe each expression image. The effectiveness of dimensionality reduction techniques, such as principal component analysis and AdaBoost, is also analyzed in terms of computational cost saving and classification accuracy. Two well-known machine learning methods, template matching and support vector machine, are used for classification using the Cohn-Kanade and Japanese female facial expression databases. Better classification accuracy shows the superiority of LDP descriptor against other appearance-based feature descriptors.

Keywords: Image representation, facial expression recognition, local directional pattern, features extraction, principal component analysis, support vector machine.

Manuscript received Mar. 15, 2010; revised July 15, 2010; accepted Aug. 2, 2010.

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2010-0015908).

Taskeed Jabit (phone: +82 31 201 2948, email: taskeed@khu.ac.kr), Md. Hasanul Kabir (email: hasanul@khu.ac.kr), and Oksam Chae (email: oschae@khu.ac.kr) are with the Department of Computer Engineering, Kyung Hee University, Yongin, Rep. of Korea.
doi:10.4218/etrij.10.1510.0132

I. Introduction

Facial expression provides the most natural and immediate indication about a person's emotions and intentions [1], [2]. Therefore, automatic facial expression analysis is an important and challenging task that has had great impact in such areas as human-computer interaction and data-driven animation. Furthermore, video cameras have recently become an integral part of many consumer devices [3] and can be used for capturing facial images for recognition of people and their emotions. This ability to recognize emotions can enable customized applications [4], [5]. Even though much work has already been done on automatic facial expression recognition [6], [7], higher accuracy with reasonable speed still remains a great challenge [8]. Consequently, a fast but robust facial expression recognition system is very much needed to support these applications.

The most critical aspect for any successful facial expression recognition system is to find an efficient facial feature representation [9]. An extracted facial feature can be considered an efficient representation if it can fulfill three criteria: first, it minimizes within-class variations of expressions while maximizes between-class variations; second, it can be easily extracted from the raw face image; and third, it can be described in a low-dimensional feature space to ensure computational speed during the classification step [10], [11]. The goal of the facial feature extraction is thus to find an efficient and effective representation of the facial images which would provide robustness during recognition process. Two types of approaches have been proposed to extract facial features for expression recognition: a geometric feature-based system and an appearance-based system [12].

In the geometric feature extraction system, the shape and

location of facial components are considered, and geometric relationships between these components are used to form a feature vector. These geometric relationships may be example positions, distances, and angles. For instance, Zhang and others [13] used the geometric positions of 34 fiducial points as facial features to represent facial images. Another widely-used facial description is the Facial Action Coding System, where facial expressions are represented by one or more action units (AUs) [14]. Valstar and others [15], [16] presented detection by classifying features calculated from tracked fiducial facial points and urged that geometric approaches have similar or better performance than appearance-based approaches in facial expression analysis. However, geometric representation of facial geometry requires accurate and reliable facial component detection and tracking, which are difficult to accommodate in many situations [9].

The appearance-based system models the face images by applying an image filter or filter banks on the whole face or some specific regions of the face to extract changes in facial appearance. Principal component analysis (PCA) has been widely applied to extract features for face recognition [17], [18]. PCA is primarily used in a holistic manner. More recently, independent component analysis (ICA) [19], [20], enhanced ICA [3], and Gabor wavelet [21] have been utilized to extract facial feature either from whole-face or specific face regions for modeling facial changes. Donato and others [22] performed a comprehensive analysis of different techniques, including PCA, ICA, local feature analysis, and Gabor wavelet, to represent images of faces for facial action recognition and demonstrate that the best performance can be achieved by ICA and Gabor wavelet. However, convoluting a facial image with multiple Gabor filters of multiple scales and orientations makes the Gabor representation very intensive as regards time and memory.

Among the appearance-based feature extraction methods, the local binary pattern (LBP) method which was originally introduced for the purpose of texture analysis [23] and its variants [24], [25] were used as a feature descriptor for facial expression representation [9]. The LBP method is computationally efficient and robust to monotonic illumination changes. However, it is sensitive to non-monotonic illumination variation and also shows poor performance in the presence of random noise [26], [27]. The local directional pattern (LDP) method, a more robust facial feature proposed by Jabid and others [27], demonstrated better performance for face recognition compared to LBP. In this work, we have analyzed the performance of the proposed LDP feature in characterizing different facial expression. We empirically study the effectiveness of facial image representation based on LDP for recognizing human expression. The performance of this

representation is evaluated using template matching and support vector machine (SVM). Extensive experiments with two widely-used expression databases, namely, the Cohn-Kanade (CK) facial expression database [28] and the Japanese female facial expression (JAFPE) database [21], demonstrate that the LDP feature is more robust in extracting the facial features, and it is also superior in classifying expressions compared to LBP and Gabor wavelet features. We also find that the LDP method performs stably and robustly over a useful range of lower resolution face images.

II. LBP

The LBP operator, a gray-scale invariant texture primitive, has gained significant popularity for describing the texture of an image [26]. It labels each pixel of an image by thresholding its P -neighbor values with the center value and converts the result into a binary number by using

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

where g_c denotes the gray value of the center pixel (x_c, y_c) and g_p corresponds to the gray values of equally spaced pixels P on the circumference of a circle with radius R .

The values of neighbors which do not fall exactly on pixel position are estimated by bilinear interpolation. In practice, (1) means that the signs of the differences in a neighborhood are interpreted as a P -bit binary number, resulting into 2^P distinct values for the binary pattern. This individual pattern value is capable of describing the texture information at the center pixel g_c . The process of generating this P -bit pattern is shown in Fig. 1.

One variation of the original LBP, known as uniform LBP, is proposed from the observation that certain LBPs appear more frequently in a significant image area. These patterns are considered uniform because they contain very few transitions from 0 to 1 or 1 to 0 in a circular bit sequence. For example, the patterns 00000000 and 11111111 have zero transitions, 00011000 has two transitions, and 10001101 has four transitions. Shan and others [2] used this variant of the LBP, which has at most two transitions (LBP^{u2}), for their facial

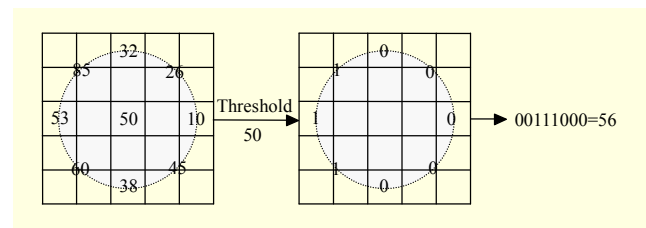


Fig. 1. Basic LBP operator.

expression recognition task. Though the LBP shows good recognition accuracy in a constraint environment, it is sensitive to random noise and non-monotonic illumination variation.

III. LDP

An LBP operator encodes the micro-level information of edges, spots, and other local features in an image using information of intensity changes around pixels. Some researchers apply the LBP operator on gradient image to encode the texture [29], [30]. These variations simply replace the intensity value with the gradient magnitude value of that pixel. Then the LBP code is calculated trivially. Lack of robustness of those methods can be alleviated by encoding the edge response in a different direction from a pixel. Being motivated by this, we propose LDP that computes the edge response values in different directions and uses these to encode the image texture. Since the edge responses are less sensitive to illumination and noise than intensity values, the resultant LDP feature describes the local primitives, including different types

of curves, corners, and junctions, in a more stable manner and also retains more information.

The proposed LDP method assigns an 8 bit binary code to each pixel of an input image. This pattern is then calculated by comparing the relative edge response values of a pixel in different directions. The Kirsch, Prewitt, and Sobel edge detectors are some of the different representative edge detectors which can be used in this regard. The Kirsch edge detector [31] detects different directional edge responses more accurately than the others because it considers all 8 neighbors [32]. Given a central pixel in the image, the eight-directional edge response values $\{m_i\}$, $i=0, 1, \dots, 7$ are computed by Kirsch masks, M_i , in eight different orientations centered on the pixel's position. These masks are shown in the Fig. 2.

The response values are not equally important in all directions. The presence of a corner or an edge shows high response values in some particular directions. Therefore, we need to know the most prominent k directions to generate the LDP. Here, the top- k directional bit responses, b_i , are set to 1. The remaining $8-k$ bits of the 8 bit LDP pattern are set to 0. Finally, the LDP code is derived by

$$LDP_k = \sum_{i=0}^7 b_i(m_i - m_k) \times 2^i, \quad b_i(a) = \begin{cases} 1, & a \geq 0, \\ 0, & a < 0, \end{cases} \quad (2)$$

where m_k is the k -th most significant directional response. Figure 3 shows the mask response and LDP bit positions, and Fig. 4 shows an exemplary LDP code with $k=3$.

1. Robustness of LDP

Since edge responses are more stable than intensity values, LDP provides the same pattern value even if there is some presence of noise and non-monotonic illumination changes. For instance, Fig. 5 shows a small image patch, before and after adding Gaussian white noise. After the addition of noise, the 5th bit of the LBP has changed from 1 to 0. Thus, the LBP pattern is changed from uniform to non-uniform code. Since edge response values are more stable than gray values, LDP provides the same pattern value under the same noise and non-monotonic illumination changes. In addition to this, an extensive demonstration has been reported [33] where the LDP

$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}$	$\begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}$	$\begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$	$\begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$
East M_0	Northeast M_1	North M_2	North west M_3
$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}$	$\begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}$	$\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix}$	$\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$
West M_4	South west M_5	South M_6	South east M_7

Fig. 2. Kirsch edge masks in all eight directions.

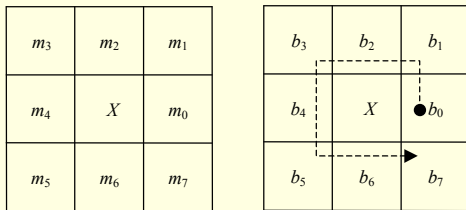


Fig. 3. (a) 8-directional edge response positions and (b) LDP binary bit positions.

85	32	26		313	97	503		0	0	1
53	50	10	$\{M_i\}$	537	X	393	m_k	1	X	1
60	38	45		161	97	161		0	0	0

LDP binary code = 00010011
LDP decimal code = 19

Fig. 4. LDP code with $k=3$.

85	32	26	LBP = 00111000 LDP = 00010011	81	29	32	LBP = 00101000 LDP = 00010011
53	50	10		38	58	15	
60	38	45		65	43	47	

Fig. 5. Stability of LDP vs. LBP: (a) original LBP and LDP values of image patch and (b) LBP and LDP values of image patch with noise.

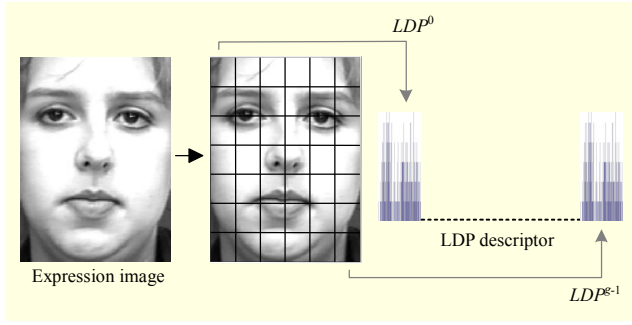


Fig. 6. Expression image is divided into small regions from which LDP histograms are extracted and concatenated into LDP descriptor.

robustness is proved by analyzing with a set of image patches.

2. LDP Descriptor

After computing all the LDP code for each pixel (r, c) , the input image I of size $M \times N$ is represented by an LDP histogram, H , using

$$H(i) = \sum_{r=1}^M \sum_{c=1}^N f(LDP_k(r, c), i), \quad f(a, i) = \begin{cases} 1, & a = i, \\ 0, & a \neq i, \end{cases} \quad (3)$$

where i is the LDP code value. The resulting histogram H is the LDP descriptor of that image. For a particular value of k , the histogram H has C_k^8 number of bins. The resultant LDP histogram describes a local region similar to that of scale invariant feature transform (SIFT) feature [34]. SIFT is a histogram of gradient orientations, whereas the proposed LDP descriptor is a histogram of encoded gradient values.

LDP descriptor contains detail information of an image, such as edges, spot, corner, and other local texture features. However, a descriptor computed over the whole face image encodes only the occurrences of the micro-patterns without any knowledge about their locations. However, for face images, some degree of locations and spatial relationship represents the image content better [35], [36]. Consequently, we modified the histogram to an extended histogram, where the image is divided into g regions R_0, R_1, \dots, R_{g-1} as shown in Fig. 6, and the LDP^i histogram is built for each region R_i . Finally, concatenating all the basic LDP^i distributions yields the LDP descriptor.

IV. Feature Dimensionality Reduction

An effective feature vector should contain only the essential information which carries higher discriminating capacity to formulate the classification task easily. Though inadequate features normally lead to a failure with a good classifier, having too many features may again increase time and space

complexities with no guaranteed advantage in the classification process. Therefore, dimensionality reduction (DR) is an important step in solving the problem of dimensionality in an efficient manner [37]. DR techniques can be broadly clustered into two groups: techniques which transform the existing features to a newly reduced set of features and techniques which select a subset of existing features. In this paper, PCA and AdaBoost techniques are employed which fall into first and second category, respectively.

In PCA, eigenvectors or principal components (PCs) are computed from the covariance data matrix. Then, each input feature is approximated by a linear combination of the top-most few eigenvectors. These weight coefficients form a new representation of the feature vector. The matrix represents the eigenspace defined by all the eigenvectors, and each eigenvalue defines its corresponding axis of variance. Usually, some eigenvalues are close to zero and can be discarded as they do not contain much information. The selected eigenvectors associated with the top eigenvalues define the newly reduced subspace. The LDP feature vector, projected onto the new subspace defined by the top eigenvectors, found from PCA that few dimensions defined by eigenvalues contain significant amount of discriminative information. Thus, the principal component representation of facial expression image can be obtained with a lesser dimension LDP representation.

AdaBoost [38] provides a simple yet effective approach for stage-wise learning of a nonlinear classification function. AdaBoost learns a small number of weak classifiers whose performance is just better than random guessing and boosts them iteratively into a strong classifier of higher accuracy. In our proposed LDP descriptor, classification capability of each bin is considered as a weak classifier. In each iteration, a weak classifier which minimizes the weighted error rate is selected, and the distribution is updated to increase the weights of the misclassified samples and reduce other weights. The basic form of AdaBoost is for two-class problems. A set of N labeled training examples is given as $(x_1, y_1) \dots (x_N, y_N)$, where $y_i \in \{+1, -1\}$ is the class label for the example $x_i \in R_n$. Both 6-class and 7-class expression recognition problems are multiclass problem; hence, we used the generalized multi-class multi-label AdaBoost algorithm proposed in [39].

V. Facial Expression Recognition Using LDP

Template matching, linear discriminant analysis, linear programming, and SVM are machine learning techniques available to classify facial expressions. A comparative analysis was carried out [9] with these techniques, and SVM perform the best. Accordingly, we verify the effectiveness of proposed

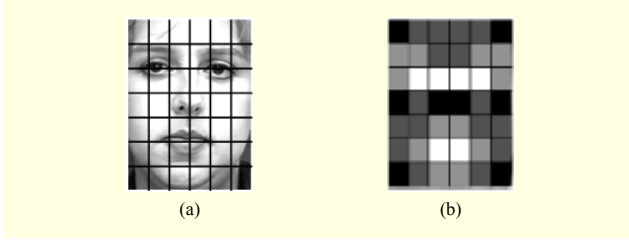


Fig. 7. (a) Facial image divided into 7×6 sub-regions and (b) weights assigned for the weighted χ^2 measure. Black indicates weight of 0.0, dark gray indicates 0.5, light gray indicates 1.0, and white indicates 1.5.

facial feature in classifying expression using SVM. Besides this, we also employ template matching technique due to its simplicity.

1. Template Matching

A template for each class of expression images is formed to model that particular expression. During the training phase, the LDP histograms of expression images in a given class are averaged to generate the template model M . For recognition, a dissimilarity measure is evaluated against each template, and the class with the smallest dissimilarity value announces the match for the test expression, S . Chi-square statistic, χ^2 , is frequently used as the dissimilarity measure, but sometimes weighted χ^2 statistics are used to give more or less importance to particular regions such as eye, nose, and mouth areas. In our case, we opted to use the weighted χ^2 statistic for template matching, and adopted weights are shown in Fig. 7.

$$\chi_w^2 = \sum_{i,j} w_i \frac{(S_i(j) - M_i(j))^2}{S_i(j) + M_i(j)}, \quad (4)$$

where w_i is the weight of region R_i .

2. SVM

SVM theory is a well-established statistical learning theory that has been successfully applied in various classification tasks in computer vision [40]. SVM performs an implicit mapping of data into a higher dimensional feature space and finds a linear separating hyper-plane with maximal margin to separate the data. Given a training set of labeled examples $T = \{(s_i, l_i), i = 1, 2, \dots, L\}$, where $s_i \in \mathbb{R}^p$, and $l_i \in \{-1, 1\}$, a new test data x is classified by

$$f(x) = \text{sign} \left(\sum_{i=1}^L \alpha_i l_i K(x_i, x) + b \right), \quad (5)$$

where α_i are Lagrange multipliers of dual optimization problem, b is a bias or threshold parameter, and K is a kernel function.

The training samples x_i with $\alpha_i > 0$ are called the support vectors, and the separating hyper-plane maximizes the margin with respect to these support vectors. Among the various kernels found in the literature, linear, polynomial, and radial basis function (RBF) kernels are the most frequently used ones.

SVM makes binary decisions, and multi-class classification can be achieved by adopting the one-against-rest or several two-class problems. In our work, we adopt the one-against-rest technique, which trains a binary classifier for each expression to discriminate one expression from all others and outputs the class with the largest output. We carried out a grid-search on the hyper-parameters in a cross-validation approach for selecting the parameters, as suggested in [41]. The parameter setting producing the best cross-validation accuracy was picked.

VI. Experimental Setup and Dataset Description

Most facial expression recognition systems attempt to recognize a set of prototypic emotional expressions like anger, disgust, fear, joy, sadness, and surprise [9]. This 6-class expression set can also be extended as a 7-class expression set by including a neutral expression. In this work, our effort is devoted to recognize both 6-class and 7-class prototypic expressions. The performance of our proposed system is evaluated with the two well-known image datasets; namely, the CK facial expression database [28] and the JAFFE database [21].

The CK database consists of 100 university students who at the time of their inclusion were between 18 to 30 years old; 65% were female, 15% were African-American, and 3% were Asian or Latino. Subjects were instructed to perform a series of facial expression displays starting from neutral or nearly neutral to one of six target prototypic emotions. Image sequences from neutral to target display were digitized into 640×480 or 640×690 pixel arrays of gray scale frames. In our setup, we selected 408 image sequences from 96 subjects, each of which was labeled as one of the six basic emotions. For 6-class prototypic expression recognition, the three most expressive image frames were taken from each sequence that resulted into 1,224 expression images. In order to build the neutral expression set, the first frame (neutral expression) from all 408 sequences was selected to make the 7-class expression dataset (1,632 images). Seven expression images, one from each prototypic expression from CK database, are displayed in Fig. 8(a).

The JAFFE database contains only 213 images of female facial expressions expressed by 10 subjects. Each image has a resolution of 256×256 pixels with almost the same number of images for each categories of expression. The head in each

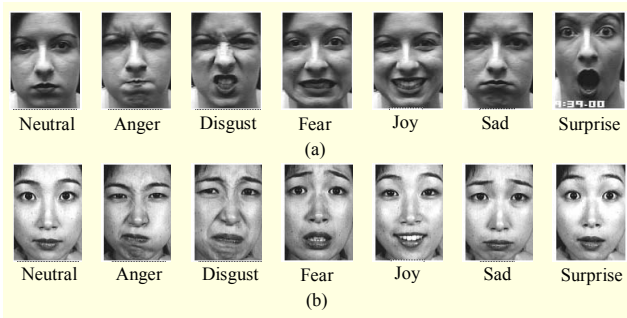


Fig. 8. Sample expression images of each prototypic expression from (a) CK database and (b) JAFFE database.



Fig. 9. Original face and cropped region as an expression image.

image is usually in frontal pose, and the subject's hair was tied back to expose all the expressive zones of her face. Tungsten lights were positioned to create an even illumination on the face. The actual names of the subjects are not revealed, but they are referred with their initials: KA, KL, KM, KR, MK, NA, NM, TM, UY, and YM. Figure 8(b) refers to the seven prototypic expressions of the person with initial KA.

After choosing the images, they were cropped from the original one using the positions of two eyes and resized into 150×110 pixels. For the CK database, the ground-truth of eye position data is provided. For other image databases, an existing eye detection technique that provides good detection accuracy [42] was used. Automatic face cropping and resizing have been done with the position of both eyes in such a way that they are a distance, D , apart. A distance of $0.5D$ between the boundaries of both eyes has been maintained. The height of the image is $2.7D$ with level of eye located $2D$ apart from bottom boundary as shown in Fig. 9. No further alignment of facial features such as alignment of mouth was performed in our algorithm. Since LDP is robust in illumination change, no attempt was made to remove illumination changes. In our experiment, we carried out a 7-fold cross-validation scheme where each dataset is randomly partitioned into seven groups separately. Six groups were used as a training dataset to train the classifiers or model their templates, while the remaining groups were used as testing datasets. The above process was repeated seven times, and the average recognition rate was calculated.

VII. Result and Discussion

In this section, we show how we first tried to find the optimal parameter settings for LDP-based facial image representation. These optimal parameter settings are employed for the facial image representation, and the classification performance is analyzed with images from the CK and JAFFE databases. The effects of the DR technique are also discussed. Finally, the robustness of proposed method is presented for recognizing a variety of lower resolution expression images.

1. Determining Optimal LDP Parameters

The recognition accuracy of the proposed method can be influenced by adjusting two parameters: the number of prominent directions used to encode in the LDP pattern and the number of regions into which the image is divided. In order to determine the optimal values of these two parameters, we first fixed the number of regions, g , and found the optimal value for k . It may be noted that $k=1$ gives the symmetric descriptor as $k=7$ because $C_1^8 = C_7^8$. Therefore, the parameter k is verified with the value from $\{1, 2, 3, 4\}$. Next, with the determined k value, we searched for the optimal value for g , that is, the number of divided regions. In our experiment, we evaluate four different cases: 3×3 , 5×5 , 7×6 , and 9×8 . All these experiments are carried out with template matching using images from the CK database.

Table 1 shows the performance for different k values with the facial images divided into 42 (7×6) regions. It can be observed that the best recognition rate is achieved when $k=3$. Though the

Table 1. Recognition performance for different k .

	6-class expression (%)	7-class expression (%)	Vector length of LDP feature
$k = 1$	80.1	78.1	336
$k = 2$	86.4	84.0	1,176
$k = 3$	89.2	86.9	2,352
$k = 4$	87.7	86.0	2,940

Table 2. Recognition performance for different number of regions.

	6-class expression (%)	7-class expression (%)	Vector length of LDP feature
$g = 3 \times 3$	85.1	82.5	504
$g = 5 \times 5$	87.4	86.0	1,400
$g = 7 \times 6$	89.2	86.9	2,352
$g = 9 \times 8$	89.1	85.4	4,032

LDP descriptor's dimension becomes higher with $k=4$, the recognition rate does not improve. This observation is in accordance with the fact that larger descriptor does not always contain more discriminative information. In order to determine the optimal number image division, images are sub-divided into 3×3 , 5×5 , 7×6 , and 9×8 blocks. Table 2 lists the effect of different number of regions on the recognition performance. Having a small number of regions leads to a lower recognition rate (below 83%). While increasing the number of regions, the recognition performance starts to increase as the descriptor feature incorporates more local and spatial relationship information. However, after a certain point, too many sub-regions incorporate unnecessary local information that might degrade the performance. From our observation, 7×6 regions provide optimal recognition performance. Therefore, we concluded that $k=3$ and $g=7 \times 6$ are the optimal parameter values for the proposed LDP-based facial expression image representation.

2. Recognition Performance Using the Optimal Parameters

The optimal parameter values are employed in recognizing facial expression images, which were collected from the CK and JAFFE databases beforehand, and a better recognition rate conforms the efficiency of the proposed LDP-based method. The basic template matching technique provides a recognition accuracy of 89.2% and 86.9% in a 6-class and 7-class expression recognition problem, respectively, with images from the CK database, whereas with images from the JAFFE database, the same technique achieved an accuracy of 87.4% and 82.6%, respectively. Tables 3 and 4 provide results

Table 3. Recognition performance with template matching using CK database.

Feature descriptor	6-class recognition (%)	7-class recognition (%)
Gabor [43]	83.7 ± 4.5	78.9 ± 4.8
LBP [9]	84.5 ± 5.2	79.1 ± 4.6
LDP	89.2 ± 2.5	86.9 ± 2.8

Table 4. Recognition performance with template matching using JAFFE database.

Feature descriptor	6-class recognition (%)	7-class recognition (%)
Gabor [43]	81.9 ± 6.4	75.5 ± 5.8
LBP [9]	83.7 ± 6.7	77.2 ± 7.6
LDP	87.4 ± 5.6	82.6 ± 4.1

Table 5. Expression recognition performance with different methods using SVM on CK database.

6-class expression			
Feature descriptor	Liner kernels (%)	Polynomial kernels (%)	RBF kernels (%)
Gabor [43]	89.4 ± 3.0	89.4 ± 3.0	89.8 ± 3.1
LBP [9]	91.5 ± 3.1	91.5 ± 3.1	92.6 ± 2.9
LDP	94.9 ± 1.2	94.9 ± 1.2	96.4 ± 0.9
7-class expression			
Feature descriptor	Liner kernels (%)	Polynomial kernels (%)	RBF kernels (%)
Gabor [43]	86.6 ± 4.1	86.6 ± 4.1	86.8 ± 3.6
LBP [9]	88.1 ± 3.8	88.1 ± 3.8	88.9 ± 3.5
LDP	92.8 ± 1.7	92.8 ± 1.7	93.4 ± 1.5

Table 6. Expression recognition performance with different methods using SVM on JAFFE database.

6-class expression			
Feature descriptor	Liner kernels (%)	Polynomial kernels (%)	RBF kernels (%)
Gabor [43]	85.1 ± 5.0	85.1 ± 5.0	85.8 ± 4.1
LBP [9]	86.7 ± 4.1	86.7 ± 4.1	87.5 ± 5.1
LDP	89.9 ± 5.2	89.9 ± 5.2	90.1 ± 4.9
7-class expression			
Feature descriptor	Liner kernels (%)	Polynomial kernels (%)	RBF kernels (%)
Gabor [43]	79.7 ± 4.2	79.7 ± 4.2	80.8 ± 3.7
LBP [9]	80.7 ± 5.5	80.7 ± 5.5	81.9 ± 5.2
LDP	84.9 ± 4.7	84.9 ± 4.7	85.4 ± 4.0

comparing LBP and Gabor features with the CK and JAFFE databases which clearly exhibit the superiority of the proposed LDP-based expression recognition system.

SVM is a well-devised machine learning technique that provides excellent classification accuracy in pattern recognition. Therefore, we conducted the recognition using SVM with different kernels to classify the facial expressions. The comparative generalized performances with the SVM classifier based on different features are shown in Tables 5 and 6. It is observed that despite LDP representation having less feature dimensionality than LBP or Gabor representation, it performs more stably and robustly than both representations.

So far, we have discussed the average recognition accuracy of several prototypic expressions. To get a better picture of the recognition accuracy of individual expression types, the confusion matrices (CMs) for 6-class and 7-class expression

Table 7. CM of 6-class expression recognition (%) using SVM on CK database.

	Anger	Disgust	Fear	Joy	Sad	Surprise
Anger	95.6	2.5	0.0	0.0	1.5	1.5
Disgust	0.0	96.5	3.5	0.0	0.0	0.0
Fear	1.5	0.0	96.0	2.5	0.0	0.0
Joy	0.0	0.0	0.0	98.0	0.0	2.0
Sad	0.5	1.5	0.0	0.0	98.0	0.0
Surprise	0.0	0.0	0.0	3.0	0.0	97.0

Table 8. CM of 7-class expression recognition (%) using SVM on CK database.

	Anger	Disgust	Fear	Joy	Sad	Surprise	Neutral
Anger	86.9	0.9	0.9	0.0	0.0	0.9	10.4
Disgust	2.0	94.2	0.0	0.0	0.0	0.0	3.8
Fear	1.5	0.0	94.4	0.0	0.0	0.0	4.1
Joy	0.0	0.0	0.7	98.9	0.0	0.0	0.4
Sad	1.1	0.5	0.0	0.0	92.6	0.0	5.8
Surprise	0.0	0.0	0.0	0.0	0.0	99.0	1.0
Neutral	5.9	1.2	0.7	0.0	2.7	0.2	89.3

Table 9. CM of 6-class facial expression recognition (%) using SVM on JAFFE database.

	Anger	Disgust	Fear	Joy	Sad	Surprise
Anger	92.6	7.4	0.0	0.0	0.0	0.0
Disgust	4.9	85.3	0.0	0.0	9.8	0.0
Fear	0.0	0.0	90.4	0.0	4.8	4.8
Joy	0.0	0.0	0.0	95.8	2.1	2.1
Sad	4.5	10	0.0	0.0	83.2	0.0
Surprise	0.0	0.0	2.4	2.4	0.0	95.2

Table 10. CM of 7-class facial expression recognition (%) using SVM on JAFFE database.

	Anger	Disgust	Fear	Joy	Sad	Surprise	Neutral
Anger	94.3	5.7	0.0	0.0	0.0	0.0	0.0
Disgust	5.9	80.1	4.0	0.0	10.0	0.0	0.0
Fear	0.0	2.7	86.3	0.0	4.2	0.0	6.8
Joy	0.0	0.0	0.0	95.2	2.4	2.4	0.0
Sad	6.1	10.8	0.0	2.8	77.5	0.0	2.8
Surprise	0.0	0.0	3.5	3.5	0.0	89.6	3.4
Neutral	2.8	0.0	0.0	0.0	5.1	7.6	84.5

recognition with template matching using the CK database are given in Tables 7 and 8, respectively. As we include the neutral



Fig. 10. Example images of disagreement. In the JAFFE database, the expressions are labeled (from left to right): sadness, surprise, and joy. However, the recognition results are joy, joy, and neutral, respectively.

expression in the 7-class recognition problem, the accuracy of the other six expressions gets lower because some facial expression samples are confused with a neutral expression. The same recognition task is also carried with images from the JAFFE database, and results shown in Tables 9 and 10 further validate the strength of the LDP.

We observed that recognition accuracy in JAFFE database is relatively lower than that of the CK database. One of the main reasons behind this is that some expressions in the JAFFE database had been labeled incorrectly or expressed inaccurately. Thus, depending on whether these expressional images are used for training or testing, the recognition result is influenced. Figure 10 shows examples of the labeled expressions and our recognition results which clarify this finding.

3. Effect of Dimensionality Reduction

In this subsection, we show that the feature dimension is reduced through PCA and AdaBoost. Then, the effect of this reduced feature on the recognition rate is analyzed. At first, the LDP descriptor is projected onto the subspace for DR as defined by the significant PCs from PCA. The dimension of the subspace determines the new feature vector's dimension. As discussed before, only those dimensions which contain the most information are desired, and unnecessary elements should be discarded. In this subsection, the optimal number of PCs is determined, and the new feature space is found from those PCs. Figure 11 shows the recognition rate for a different number of PCs varying from 60 to 260. With 240 PCs, the projected features achieved a recognition performance of above 96% and 93% for 6-class and 7-class facial expression recognitions, respectively. With a higher number of transformed features, the recognition rate shows almost a constant performance.

We also analyzed the effect of feature dimension using AdaBoost. The LDP histogram generated from every sub-region makes the feature vector long. A subset of features which has more discriminating capability to classify expression is selected using AdaBoost. During AdaBoost, training for each expression classifier continued until the distributions for the positive and negative samples were completely separated. The

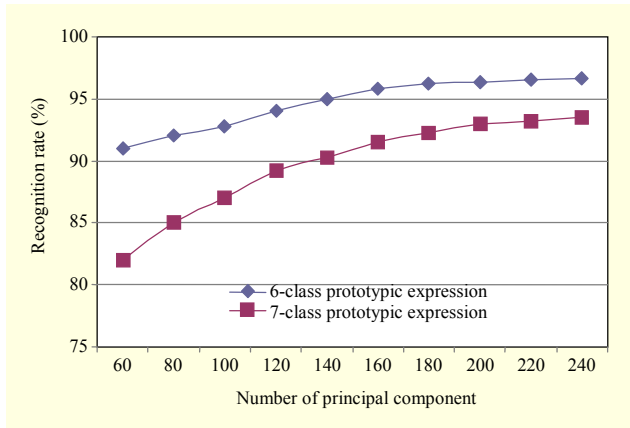


Fig. 11. Recognition rate of prototypic facial expressions by varying the number of features from PC subspace in PCA.

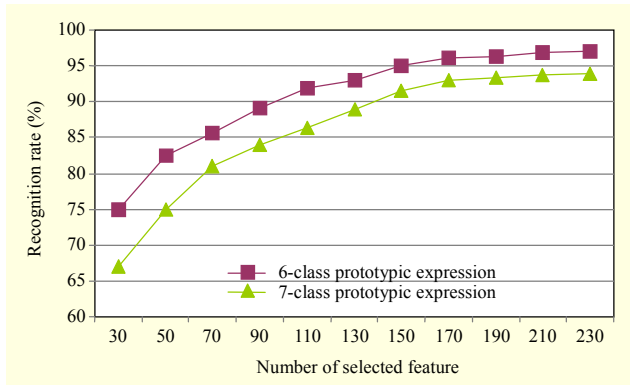


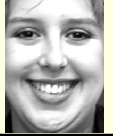



Fig. 12. Recognition rate of prototypic facial expressions by varying the number of features selected from boosting technique.

total number of features selected using this procedure was 230, and these selected features are used to further classify the expressions using SVM. The generalization performances in 6-class and 7-class recognitions as a function of the number of selected features are shown in Fig. 12.

4. Evaluation at Different Resolution

In environments like smart meeting, visual surveillance, and old-home monitoring, only low-resolution video input is available [44]. Deriving AUs from such facial images are critical problems. In this subsection, we explore the recognition performance on low-resolution images with the LDP descriptor. Four different resolutions of face images were studied: 150×110 , 75×55 , 48×36 , and 37×27 . Low-resolution images were formed by down-sampling the original images. All face images were divided into 42 (7×6) regions for building the LDP descriptor. To compare with the methods based on LBP and Gabor wavelet features, we conducted similar experiments

Table 11. Recognition performance in low-resolution facial (CK database) images.

	150×110	75×55	48×36	37×27
Feature				
Gabor [43]	89.8 ± 3.1	89.2 ± 3.0	86.4 ± 3.3	83.0 ± 4.3
LBP [9]	92.6 ± 2.9	89.9 ± 3.1	87.3 ± 3.4	84.3 ± 4.1
LDP	96.4 ± 0.9	95.5 ± 1.6	93.1 ± 2.2	90.6 ± 2.7

on the 6-class prototypic expression recognition using SVM with RBF kernel. Table 11 lists the recognition results with LBP, Gabor, and the proposed LDP feature. As with low-resolution images, it is difficult to extract geometric features [45]; therefore, appearance-based methods seem to be a good alternative. Our analysis with the LDP feature demonstrates that the proposed descriptor performs robustly and stably over a range of expressions, even with low-resolution facial image.

Our experimental results validate that the proposed LDP performs better than LBP in expression recognition. Nevertheless, it is relatively more expensive than that of LBP because it needs to compute different edge responses with a compass mask. Instead of convoluting the image pixels with a 3×3 mask, the edge responses can easily be generated with the help of integral images. This enables computation of each edge response with only eight additive operations, which in turns allows the proposed method, which is suitable for real time application.

VIII. Conclusion

This paper describes a new local facial descriptor based on LDP codes for facial expression recognition. The LDP code contains local information encoding the texture, and the descriptor contains the global information. Extensive experiments illustrate that the LDP features are effective and efficient for expression recognition. The discriminative power of the LDP descriptor mainly lies in the integration of the local edge response pattern. Furthermore, with dimensionality reduction techniques, like PCA or Adaboost, the newly transformed LDP features also maintain a high recognition rate with lower computational cost. Once trained, our system can be used in consumer products for human-computer interaction which require recognition of facial expressions. Psychological experiments by Bassili [46] have suggested that facial expressions can be recognized more accurately from sequence images than from a single image. In future, we plan to explore

the sequence images and incorporate temporal information with the LDP descriptor.

References

- [1] Y.L. Tian et al., "Real World Real-time Automatic Recognition of Facial Expressions," *Proc. IEEE Workshop Performance Evaluation of Tracking and Surveillance*, 2003.
- [2] C. Shan, S. Gong, and P.W. McOwan, "Robust Facial Expression Recognition using Local Binary Patterns," *Proc. IEEE Int. Conf. Image Process.*, 2005, pp. 914-917.
- [3] M.Z. Uddin, J.J. Lee, and T.S. Kim, "An Enhanced Independent Component-Based Human Facial Expression Recognition from Video," *IEEE Trans. Consum. Electron.*, vol. 55, no. 4, 2009, pp. 2216-2224.
- [4] M.C. Hwang et al., "Person Identification System for Future Digital TV with Intelligence," *IEEE Trans. Consum. Electron.*, vol. 53, no. 1, 2007, pp. 218-226.
- [5] P. Corcoran et al., "Biometric Access Control for Digital Media Streams in Home Networks," *IEEE Trans. Consum. Electron.*, vol. 53, no. 3, 2007, pp. 917-925.
- [6] M. Pantic and L.J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, 2000, pp. 1424-1445.
- [7] B. Fasel and J. Luetin, "Automatic Facial Expression Analysis: A Survey," *Pattern Recog.*, vol. 36, no. 1, 2003, pp. 259-275.
- [8] G. Zhao and M. Pietikainen, "Boosted Multi-resolution Spatio-Temporal Descriptors for Facial Expression Recognition," *Pattern Recognit. Lett.*, vol. 30, no. 12, 2009, pp. 1117-1127.
- [9] C. Shan, S. Gong, and P.W. McOwan, "Facial Expression Recognition based on Local Binary Patterns: A Comprehensive Study," *Image Vision Comput.*, vol. 27, no. 6, May 2009, pp. 803-816.
- [10] K.H. Choi et al., "A Probabilistic Network for Facial Feature Verification," *ETRI J.*, vol. 25, no. 2, Apr. 2003, pp. 140-143.
- [11] K.H. Kim et al., "Facial Feature Extraction Based on Private Energy Map in DCT Domain," *ETRI J.*, vol. 29, no. 2, Apr. 2007, pp. 243-245.
- [12] Y. Tian, T. Kanade, and J.F. Cohn, "Facial Expression Analysis," *Handbook of Face Recognition*, Springer, Oct. 2003.
- [13] Z. Zhang et al., "Comparison between Geometry-Based and Gabor-wavelets-based Facial Expression Recognition Using Multi-layer Perceptron," *Proc. IEEE Int. Conf. Auto. Face Gesture Recog.*, Apr. 1998, pp. 454-459.
- [14] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for Measurement of Facial Movement*, Consulting Psychologists Press, 1978.
- [15] M. Valstar, I. Patras, and M. Pantic, "Facial Action Unit Detection using Probabilistic Actively Learned Support Vector Machines on Tracked Facial Point Data," *IEEE CVPR Workshop*, vol. 3, 2005, pp. 76-84.
- [16] M. Valstar and M. Pantic, "Fully Automatic Facial Action Unit Detection and Temporal Analysis," *IEEE CVPR Workshop*, June 2006, p. 149.
- [17] M.A. Turk and A.P. Pentland, "Face Recognition Using Eigenfaces," *Proc. Comput. Vision Pattern Recog.*, 1991, pp. 586-591.
- [18] C. Padgett and G. Cottrell, "Representation Face Images for Emotion Classification," *Advances Neural Inf. Process. Syst.*, vol. 9, Cambridge, MA, MIT Press, 1997.
- [19] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "Face Recognition by Independent Component Analysis," *IEEE Trans. Neural Networks*, vol. 13, no. 6, 2002, pp. 1450-1464.
- [20] C.C. Fa and F.Y. Shin, "Recognizing Facial Action Units using Independent Component Analysis and Support Vector Machine," *Pattern Recog.*, vol. 39, no. 9, 2006, pp. 1795-1798.
- [21] M.J. Lyons, J. Budynek, and S. Akamatsu, "Automatic Classification of Single Facial Images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 12, 1999, pp. 1357-1362.
- [22] G. Donato et al., "Classifying Facial Actions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 10, 1999, pp. 974-989.
- [23] T. Ojala and M. Pietikainen, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, 2002, pp. 971-987.
- [24] X. Feng, M. Pietikainen, and A. Hadid, "Facial Expression Recognition with Local Binary Patterns and Linear Programming," *Pattern Recog. Image Anal.*, vol. 15, no. 2, 2005, pp. 546-548.
- [25] G. Zhao and M. Pietikainen, "Dynamic Texture Recognition using Local Binary Patterns with An Application to Facial Expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, 2007, pp. 915-928.
- [26] H. Zhou, R. Wang, and C. Wang, "A Novel Extended Local Binary Pattern Operator for Texture Analysis," *Inf. Science*, vol. 178, no. 22, 2008, pp. 4314-4325.
- [27] T. Jabit, M.H. Kabir, and O.S. Chae, "Local Directional Pattern (LDP) for Face Recognition," *IEEE Int. Conf. Consum. Electron.*, 2010, pp. 329-330.
- [28] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis," *IEEE Int. Conf. Autom. Face Gesture Recog.*, Mar. 2000, pp. 46-53.
- [29] S. Zhao, Y. Gao, and B. Zhang, "Sobel-LBP," *Int. Conf. Image Process.*, 2008, pp. 2144-2147.
- [30] R. Mattivi and L. Shao, "Human Action Recognition Using LBP-TOP as Sparse Spatio-Temporal Feature Descriptor," *Int. Conf. Comput. Anal. Image Pattern*, 2009, pp. 740-747.
- [31] W.K. Pratt, *Digital Image Processing*, Wiley, New York, 1978.
- [32] S.W. Lee, "Off-line Recognition of Totally Unconstrained Handwritten Numerals Using Multilayer Cluster Neural

Network,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 6, 1996, pp. 648-652.

- [33] T. Jabid, M.H. Kabir, and O.S. Chae, “Local Directional Pattern (LDP): A Robust Image Descriptor for Object Recognition,” *IEEE Int. Conf. Adv. Video and Signal-Based Surveillance*, 2010, pp. 482-487.
- [34] D. Lowe, “Distinctive Image Features from Scale Invariant Key Points,” *Int. J. Comput. Vision*, vol. 60, no. 2, 2004, pp. 91-110.
- [35] T. Ahonen, A. Hadid, and M. Pietikainen, “Face Description with Local Binary Patterns: Application to Face Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, 2006, pp. 2037-2041.
- [36] S. Gundimada and V.K. Asari, “Facial Recognition Using Multisensor Images Based on Localized Kernel Eigen Spaces,” *IEEE Trans. Image Process.*, vol. 18, no. 6, 2009, pp. 1314-1325.
- [37] C.A. Kumar, “Analysis of Unsupervised Dimensionality Reduction Techniques,” *Comput. Sci. Inf. Syst.*, vol. 6, no. 2, Dec. 2009, pp. 217-227.
- [38] Y. Freund and R.E. Schapire, “A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting,” *Computational Learning Theory*, 1995, pp. 23-37.
- [39] R.E. Schapire and Y. Singer, “Improved Boosting Algorithms using Confidence-Rated Predictions,” *Maching Learning*, vol. 37, no.3, 1999, pp. 297-336.
- [40] C. Cortes and V. Vapnik, “Support Vector Networks,” *Machine Learning*, vol. 20, no. 3, 1995, pp. 273-297.
- [41] C.W. Hsu and C.J. Lin, “A Comparison on Methods for Multiclass Support Vector Machines,” *IEEE Trans. Neural Networks*, vol. 13, no. 2, 2002, pp. 415-425.
- [42] Z. Niu et al., “2D Cascaded AdaBoost for Eye Localization,” *Proc. IEEE Int. Conf. Pattern Recog.*, 2006, pp. 1216-1219.
- [43] M.S. Bartlett et al., “Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior,” *IEEE Conf. Computer Vision and Pattern Recog.*, 2005, pp. 568-573.
- [44] Y. Tian, “Evaluation of Face Resolution for Expression Analysis,” *CVPR Workshop Face Process. Video*, 2004, p. 82.
- [45] Y. Ono, FT. Okabe, and Y. Sato, “Gaze Estimation from Low Resolution Images,” *IEEE Pacific-Rim Symp. Image Video Technol.*, 2006, pp. 178-188.
- [46] J.N. Bassili, “Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Areas of the Face,” *J. Personality Social Psychology*, vol. 37, no. 11, 1979, pp. 2049-2058.



Taskeed Jabid received his BS in computer science from East West University, Bangladesh, in 2001. After his graduation, he worked as a lecturer in the Computer Science and Engineering Department of East West University, Bangladesh. Currently he is pursuing his PhD in the Department of Computer Engineering, Kyung Hee University, Rep. of Korea. His research interests include texture analysis, image processing, computer vision, and pattern recognition. He is a member of IEEE.



Md. Hasanul Kabir received his BS degree in computer science and information technology from the Islamic University of Technology, Bangladesh, in 2003. After his graduation, he worked as a lecturer in the Computer Science and Information Technology Department, Islamic University of Technology, Bangladesh.

Currently he is pursuing his PhD in the Department of Computer Engineering, Kyung Hee University, Rep. of Korea. His research interests include motion estimation for mobile devices, image enhancement, computer vision, and pattern recognition. He is a member of IEB and IEEE.



Oksam Chae received his BS in electronics engineering from Inha University, Rep. of Korea, in 1977. He completed his MS and PhD in electrical and computer engineering from Oklahoma State University, USA, in 1982 and 1986, respectively. From 1986 to 1988, he worked as research engineer for Texas

Instruments, USA. Since 1988, he has been working as a professor in the Department of Computer Engineering, Kyung Hee University, Rep. of Korea. His research interests include multimedia data processing environments, intelligent filters, motion estimation, intrusion detection systems, and medical image processing in dentistry. He is a member of IEEE, SPIE, KES, and IEICE.