Local Directional Pattern Variance (LDPv): A Robust Feature Descriptor for Facial Expression Recognition

Hasanul Kabir, Taskeed Jabid, and Oksam Chae
Department of Computer Engineering, Kyung Hee University, South Korea

Abstract: *Automatic facial expression recognition is a challenging problem in computer vision, and has gained significant importance in the applications of human-computer interactions. The vital component of any successful expression recognition system is an effective facial representation from face images. In this paper, we have derived an appearance-based feature descriptor, the Local Directional Pattern Variance (LDPv), which characterizes both the texture and contrast information of facial components. The LDPv descriptor is a collection of Local Directional Pattern (LDP) codes weighted by their corresponding variances. The feature dimension is then reduced by extracting the most discriminative elements of the representation with Principal Component Analysis (PCA). The recognition performance based on our LDPv descriptor has been evaluated using Cohn-Kanade expression database with a Support Vector Machine (SVM) classifier. The discriminative strength of LDPv representation is also assessed over a useful range of low resolution images. Experimental results with prototypic expressions show that the LDPv descriptor has achieved a higher recognition rate, as compared to other existing appearance-based feature descriptors.*

Keywords: *Facial expression recognition, feature descriptor, LDP, LDPv, PCA, SVM classifier.*

Received March 7, 2010; accepted May 20, 2010

1. Introduction

Facial expression is one of the most powerful, natural, and immediate means for human beings to communicate their emotions and intentions [24, 26]. Automatic facial expression recognition has attracted much attention from behavioral scientists since the work of Darwin in 1872 and has gained significant importance in applications of human-computer interactions [25]. Although much work has been done with automatic facial expression analysis, recognition with high accuracy remains difficult due to the complexity and variety of facial expressions [34]. A survey of existing research on facial expression analysis can be found in [11, 22].

Extracting an effective facial representation from human face images is a vital component of any successful facial expression recognition system. The derived representation should retain essential information possessing high discrimination power and stability which minimizes within-class variations of expressions whilst maximizes between-class variations [24]. Classification performance is heavily influenced by the information contained in the expression representations. Two types of facial feature extraction approaches are commonly found: the geometric feature-based system and the appearance-based system [27]. Geometric feature vectors represent the shapes and locations of facial components by encoding the face geometry from the position, distance, angle, and other geometric relationships between these components. Zhang *et al*. [33] represented facial images using the geometric positions of 34 fiducial points as facial features. A widely used facial description is the facial action coding system, where facial expressions are decomposed into one or more Action Units (AUs) [10]. Valstar *et al*. [31, 32] detected AUs by tracking several fiducial points on face and urged that geometric approaches have similar or better performance than appearance-based approaches in facial expression analysis. However, geometric feature-based methods require accurate and reliable facial component detection which is difficult to accommodate in many situations [25].

Recent psychological research suggests that the whole spatial relationship of the facial features can be an additional source of information in the perception of facial emotions [20, 30]. Therefore, in appearance-based methods a single image filter or filter bank is applied to the whole face or some specific region of the face to extract appearance changes. Among the holistic methods, Principal Component Analysis (PCA) has been widely applied to facial images to extract features for recognition purposes [29]. PCA is also used for dimensionality reduction in feature space. Lately, Independent Component Analysis (ICA) [5, 7], Enhanced ICA (EICA) [30], and Zernike Moments (ZM) [18, 23] have been utilized to extract local features and facial changes. Donato *et al.* [9] performed a comprehensive analysis of different techniques, including PCA, ICA, Local Feature Analysis (LFA), Gabor-wavelet and local Principal Components (PCs), to represent face images for facial action recognition. The best performance was achieved by ICA and Gabor-wavelet. Since then Gabor-wavelet representations have been widely adopted in face image analysis by other methods. However, convoluting a facial image with multiple Gabor filters of many scales and orientations makes the Gabor representation time and memory intensive. Lajevardi and Hussain [17], have utilized log-Gabor filters to overcome some limitations of Gabor-wavelet representations but the dimensionality of resulting feature vector is still high.

Recently, Local Binary Pattern (LBP) [21] and its variants [34] have been introduced as a feature descriptor for facial expression representation [24, 25]. Originally, LBP was introduced for texture analysis. A comprehensive study of LBP in facial expression recognition can be found in [25]. Although LBP is computationally efficient and shows robustness to monotonic illumination change, it is sensitive to non-monotonic illumination variation and also shows poor performance in the presence of random noise [14, 35]. A more robust facial descriptor, named as Local Directional Pattern (LDP), was devised by Jabid *et al.* [14], where the LDP representation of face demonstrated better recognition performance than LBP. The LDP feature overcomes the limitations of LBP features since LDP is derived from the edge responses which are less sensitive to illumination changes and noises.

In this work, we propose the LDP variance (LDPv), which characterizes both spatial structure LDP and contrast variance of local texture information for more accurate facial expression recognition performance. Figure 1 shows an overall flow of the expression recognition system based on our LDPv descriptor coupled with PCA and SVM. We empirically study the facial representation based on LDPv for human expression recognition. The performance of LDv representation is evaluated with two machine learning methods: Template matching and Support Vector Machines (SVM) with different kernels. Extensive results from the standard expression database Cohn-Kanade facial expression database [15], demonstrate that LDPv feature is more robust in extracting facial features, and have a superior recognition rate, as compared to LBP, Gabor-wavelet features, and other appearance-based methods. LDPv descriptor also performs stably and robustly over a useful range of low resolution face images.

The rest of the paper is organized as follows: the proposed LDPv feature is described in section 2. The dimensionality reduction of LDPv and the machine learning techniques used for expression classification are explained in sections 3 and 4, respectively. Section 5 presents the experimental setup used for evaluating the effectiveness of our proposed feature representation, and section 6 lists the expression recognition performances of LDPv compared with existing representations. Finally, section 7 concludes the paper.

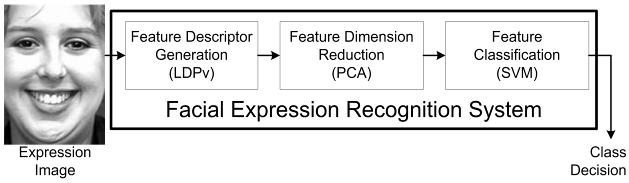


Figure 1. Overview of the facial expression recognition system based on LDPv representation.

2. LDP Variance (LDPv) Descriptor

In this section, we first review the LDP code, and then, the descriptor based on LDPv is explained.

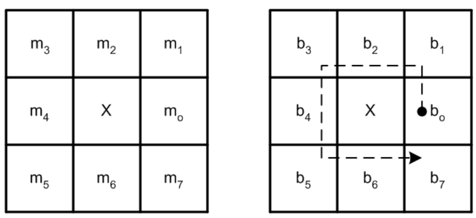
2.1. LDP

LDP is a gray-scale texture pattern which characterizes the spatial structure of a local image texture. A LDP operator computes the edge response values in all eight directions at each pixel position and generates a code from the relative strength magnitude. Since the edge responses are more illumination and noise insensitive than intensity values, the resultant LDP feature describes the local primitives including different types of curves, corners, and junctions, more stably and retains more information. Given a central pixel in the image, the eight directional edge response values $\{m_i\}$, $i = 0,1,...,7$ are computed by Kirsch masks M_i in eight different orientations centered on its position. The masks are shown in Figure 2.

$$\underset{\text{East } M_0}{\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}} \quad \underset{\text{North East } M_1}{\begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}} \quad \underset{\text{North } M_2}{\begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}} \quad \underset{\text{North West } M_3}{\begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}}$$

$$\underset{\text{West } M_4}{\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}} \quad \underset{\text{South West } M_5}{\begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}} \quad \underset{\text{South } M_6}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix}} \quad \underset{\text{South East } M_7}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}}$$

Figure 2. Kirsch edge masks in all eight directions.



a) Eight directional edge response positions. b) LDP binary bit positions.

Figure 3. Mask response and LDP bit positions.

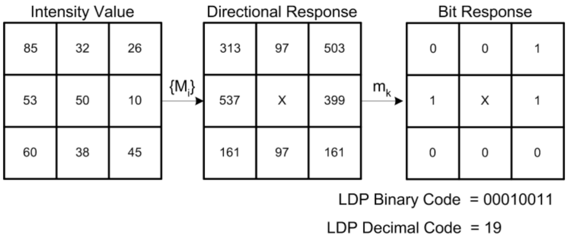


Figure 4. LDP code with $k = 3$.

The response values are not equally important in all directions. The presence of a corner or edge causes high response values in some directions. Therefore, we are interested in the k most prominent 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 using equation 1. Figure 3 shows the mask response and LDP bit positions, and Figure 4 shows an exemplary LDP code with $k=3$.

$$LDP_k = \sum_{i=0}^{7} b_i(m_i - m_k) \times 2^i \tag{1}$$

$$b_i(a) = \begin{cases} 1 & a \geq 0 \\ 0 & a < 0 \end{cases} \tag{2}$$

Where, m_k is the k-th most significant directional response. Since edge responses are more stable than intensity values, LDP pattern provides the same pattern value even presence of noise and non-monotonic illumination changes [14]. After computing the LDP code for each pixel (r, c), the input image I of size $M \times N$ is represented by a LDP histogram H using equation 3. The resultant histogram H is the LDP descriptor of that image.

$$H(\tau) = \sum_{r=1}^{M} \sum_{c=1}^{N} f\left(LDP_k(r,c), \tau\right) \tag{3}$$

$$f(a,\tau) = \begin{cases} 1 & a = \tau \\ 0 & otherwise \end{cases} \tag{4}$$

Where, τ is the LDP code value. For a particular value of k, the histogram H has $C_k^8 = 8!/\left(k! \times (8-k)!\right)$ number of bins. In other words, the LDP descriptor is a C_k^8-element feature vector.

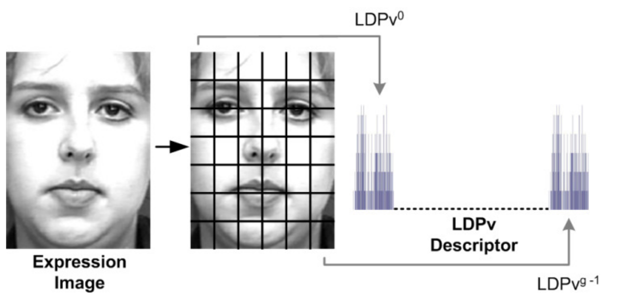


Figure 5. An expression image is divided into small regions from which LDPv histograms are extracted and concatenated into a single LDPv descriptor.

2.2. LDPv Descriptor

Generally, texture can be well represented when characterized by a spatial structure along with its contrast [21]. The LDP feature only contains the distribution of local structures. A low contrast structure contributes equally with a high contrast one in the LDP histogram. However, texture with significant contrast should impact more since human eyes are more sensitive to high contrast regions. Hence, we account for the contrast information within the feature descriptor. The variance of a structure is related to the texture. Generally, high frequency texture regions have higher variances and contribute more to the discrimination of texture images [19]. Therefore, the variance σ is introduced as an adaptive weight to adjust the contribution of the LDP code in the histogram generation. The proposed $LDPv$ descriptor is computed as:

$$LDPv(\tau) = \sum_{r=1}^{M} \sum_{c=1}^{N} w\left(LDP_k(r,c), \tau\right) \tag{5}$$

$$w(LDP_k(r,c),\tau) = \begin{cases} \sigma(LDP_k(r,c)) & LDP_k(r,c) = \tau \\ 0 & otherwise \end{cases} \tag{6}$$

$$\sigma\left(LDP_k(r,c)\right) = \frac{1}{8}\sum_{i=0}^{7}(m_i - \overline{m})^2 \tag{7}$$

Where, $\overline{m}$ is the average of all directional responses $\{m_i\}$ calculated for a position (r,c). When LDP and variance σ are treated as the two orthogonal axes in a coordinate system, the LDPv can be regarded as the integral projection [6] along the σ axis. LDPv generated from the whole image loses some location information, but for face images, some degree of location and spatial relationship well represent the image content [1, 2, 12]. Hence, the basic histogram is modified to an extended histogram, where the image is divided into g number of regions $R_0, R_1, \ldots, R_{g-1}$ shown in Figure 5, and the $LDPv^i$ histogram is built for each region R_i using equation 8. Finally, concatenating all of the basic $LDPv^i$ distributions with equation 9 yields the descriptor vector of size $p(=g \times n)$, where n is the size of each basic LDPv histogram.

$$LDPv^i(\tau) = \sum_{r=1}\sum_{c=1} w(LDP_k(r,c),\tau) \;\; where, (r,c) \in R_i \tag{8}$$

$$LDPv = [LDPv^0, LDPv^1, \ldots, LDPv^{g-1}] \tag{9}$$

This extended feature vector represents both texture and contrast information with some extent of spatial relationship. Two parameters can be adjusted for better feature extraction:

1. The prominent directions to encode in the LDP pattern.
2. The number of regions. The optimal parameters are selected from a good trade-off between recognition performance with feature representation and feature-

vector length. A detailed discussion of these two parameter settings can be found in section 6.2.

3. Feature Dimensionality Reduction Using PCA

A feature vector should contain essential information to make the classification task easier. With an inadequate number of features, a good classifier may even fail. On the other hand, too many features increase time and space complexities with no apparent advantage in the classification process. Therefore, Dimensionality Reduction (DR) techniques are proposed as a preprocessing step to address the curse of dimensionality [16]. DR techniques try to find a suitable low-dimensional representation of original data. Mathematically, the DR problem can be viewed as: given a p-dimensional random vector $Y=(y_1, y_2, \ldots, y_p)$, the objective is to find a representation in the lower dimension $Z=(z_1, z_2, \ldots, z_q)$ where, $q<p$ which preserves the content of the original data as much as possible. DR functions can be broadly clustered into two groups:

1. Functions which transform the existing features to a new reduced set of features.
2. Functions those select a subset of existing features.

In this paper, we utilize a DR function, PCA, which transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called Principal Components (PCs). PCA falls into the first category of DR techniques which successfully uncovers the latent structures in the datasets and shows optimality in the case of dimension reduction of the input feature space.

The first step is to compute the eigen-vectors or PCs from the covariance data matrix Q. Then, each input feature is approximated by a linear combination of the top-most few eigen-vectors. These weight-coefficients form a new representation of the feature-vector. The covariance matrix Q and its PCs are computed as follows:

$$Q = \frac{1}{L}\sum_{i=1}^{L}\left(\hat{Y}_i \hat{Y}_i^T\right) \tag{10}$$

$$E^T Q E = \Lambda \tag{11}$$

Where, $\hat{Y}$ is the shifted input feature with the empirical mean subtracted from the original feature vector Y, L is the total number of feature vectors, E contains the orthonormal eigen-vectors and Λ the diagonal matrix of eigen-values. The matrix E represents the eigen-space defined by all the eigen-vectors, and each eigen-value defines its corresponding axis of variance. Usually, some eigen-values are close to zero and can be discarded as they do not contain much information. The selected q eigen-vectors associated with the topmost q eigen-values defines the newly reduced subspace. The LDPv feature vectors Y are projected onto the new subspace defined by the q eigen-vectors found using PCA. Figure 6 illustrates the top 400 eigen-values from all expression images. It can be seen that few dimensions defined by eigen-values contain a significant amount of discriminative information. Thus, the principal component representation of facial expression image can be computed from:

$$Z_i = E_q^T \hat{Y}_i \tag{12}$$

Where, Z_i is the PCA projection of the original feature fector Y_i. The matrix E_q contains the leading q eigen-vectors of Q. Therefore, we obtain a q-element feature vector from the original p-element LDPv representation, where $q<p$.

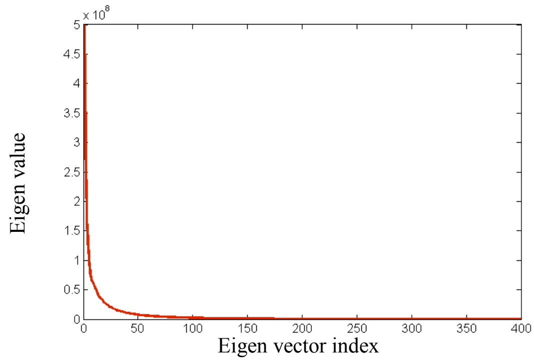


Figure 6. Top 400 eigen-values associated with their eigen-vectors.

4. Facial Expression Recognition Using LDPv

Different techniques have been proposed to classify facial expressions. A comparative analysis of four machine learning technique, namely Template matching, Linear Discriminant Analysis, Linear programming, and SVM, are examined in [25], and SVM performed the best. However, template matching is commonly adopted for its simplicity. In this section, both template matching and SVM are explained for expression classification based on LDPv features.

4.1. Template Matching (TM)

A template for each class of expression images is formed to model that particular expression. During the training phase, the 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. Chi square statistics (x^2) is usually used as the dissimilarity measure as given below:

$$\chi^2 = \sum_{\tau} \frac{\left(S(\tau) - M(\tau)\right)^2}{S(\tau) + M(\tau)} \tag{13}$$

Where, S is the test sample and M is the template LDPv histogram feature.

4.2. Support Vector Machine (SVM)

SVM is a well founded statistical learning theory that has been successfully applied in various classification tasks in computer vision. 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 $T = \{(s_i, l_i), i = 1, 2, \ldots, L\}$ where $s_i \in \mathbb{R}^q$, and $l_i \in \{-1, 1\}$, a new test data x is classified by:

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{L} \alpha_i l_i K(x_i, x) + b\right) \quad (14)$$

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 sample x_i with $\alpha_i > 0$ is called the support vectors, and the separating hyper-plane maximizes the margin with respect to these support vectors. Given a non-linear mapping function Φ that transforms the input data to the higher dimensional feature space, kernels have the form $K(x_i, x_j) = \langle \Phi(x_i, \Phi(x_j) \rangle$. Of the various kernels found in the literature, linear, polynomial and Radial Basis Function (RBF) kernels are the most frequently used.

SVM makes binary decisions and multi-class classification can be achieved by adopting the one-against-rest or several two-class (anger-happiness, anger-fear, fear-sadness, etc.,) 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 grid-search on the hyper-parameters in a cross-validation approach for selecting the parameters, as suggested in [13]. The parameter setting producing the best cross-validation accuracy was picked.

5. Experimental Setup

Facial expressions can be described at different levels [27]. Most facial expression recognition systems attempt to recognize a set of prototypic emotional expressions including anger, disgust, fear, joy, sadness, and surprise [25]. In this work, we also try to recognize the basic six prototypic expressions. Including the neutral expression, the 6-class prototypic expression set is extended to 7-class expression problem.

The performance of the proposed concept is evaluated with the well-known image dataset from the Kanade *et al.* [15] facial expression database. This database consists of 100 university students from 18 to 30 years in age, of which 65% were female, 15% were African-American, and 3% were Asian or Latino. Subjects were instructed to perform a series of 23 facial displays, six of which were based on descriptions of prototypic emotions (i.e., anger, disgust, fear, joy, sadness, and surprise). Image sequences from neutral to target display were digitized into 640×490 pixel arrays of gray-scale frames. In our setup, we selected 408 image sequences, each of which are labeled as one of the six basic emotions. These sequences come from 96 subjects, with 1-6 emotions per subject. For 6-class prototypic expression recognition, three peak frames were used from each sequence that resulted into 1224 number of expression images. In order to build the neutral expression set, the first frame from all 408 sequences is taken to make the 7-class expression dataset 1632 images. Facial images of size 150×110 pixels were cropped from the original image using the positions of two eyes. Figure 7 shows an example of a cropped facial image. No further alignment of facial features such as alignment of mouth [33] is performed in our algorithm. Since LDP is robust in illumination change, no attempt is made to remove illumination changes. Following Shan *et al.* [25], we adopted a cross validation test to evaluate the recognition results. In our experiment, we carried out a 7-fold cross-validation scheme where the dataset is randomly partitioned into seven groups. Six groups were used as training dataset to train the classifiers or model their template, while the remaining group was used as testing dataset. The above process is repeated seven times for each group, and the average recognition rate is calculated.

Figure 7. The original face and cropped region as an expression image.

6. Results and Discussion

In this section, we first show the generalization performance of LDPv descriptor with the optimal parameter settings. Here, each 150×110 pixel facial image is divided into $g=42(=7\times6)$ regions and the value of k is set to 3. To support these parameter values, we also provide an empirical analysis in determining the optimal parameter values. Next, the effects and benefits of DR are also demonstrated with PCA. Finally, we present the achieved expression recognition rate at low resolution images.

Table 1. Recognition performance (%) with template matching.

Feature Descriptor	6-Class Recognition	7-Class Recognition
LBP [21, 25]	84.5 ± 5.2	79.1 ± 4.6
LDP [14]	85.7 ± 2.5	81.9 ± 2.8
LDPv	87.2 ± 4.1	83.6 ± 5.8

6.1. Generalization Performance

Basic template matching with LDPv achieved a recognition rate of 87.2% and 83.6% for the 6-class and 7-class expression recognition problems. In Table 1, the comparative results are provided and contrasted with LBP and LDP features. Although, both LBP and LDP based descriptors previously performed better than the Geometric feature 73.2% with Tree-Augmented Naïve Bayes (TAN) classifier [8], the proposed LDPv shows the best recognition rate. The confusion matrix for the 6-class and 7-class expression recognition with TM is given in Tables 2 and 3, respectively. It is observed that with the inclusion of Neutral expression in the 7-class recognition problem, the accuracy of other six expressions decreases as more facial expression samples are confused as Neutral expression.

We preferred to use SVM with different kernels to classify the facial expressions. The comparative generalization performances achieved with SVM based on different features are shown in Tables 4 and 5. With the SVM (RBF kernel) classifier, our proposed LDPv representation achieved a recognition rate of 96.7% and 93.1% for the 6-class and 7-class recognition problems. The improvement in recognition rate with LDPv was due its extended capabilities in encoding both the spatial structures and contrast of facial components. It is observed that, LDPv representation has the same feature dimensionality as LDP or LBP representation but performs more stably and robustly. Furthermore, to show the discriminative strength of different representations, the generalized performances of the proposed LDPv feature and four existing approaches are listed in Table 6.

Table 2. Confusion matrix of 6-class expression recognition using LDPv and TM.

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sad (%)	Surprise (%)
Anger	**81.2**	8.7	0.0	0.5	3.4	6.3
Disgust	10.6	**84.1**	1.5	1.5	2.3	0.0
Fear	12.3	3.6	**67.2**	6.7	5.6	4.6
Joy	4.2	4.2	1.9	**87.5**	0.0	2.3
Sad	25.3	0.5	1.1	0.0	**68.3**	4.8
Surprise	8.3	0.0	3.8	0.0	1.3	**86.7**

Table 3. Confusion matrix of 7-class expression recognition using LDPv and TM.

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sad (%)	Surprise (%)	Neutral (%)
Anger	70.0	8.7	0.0	0.8	0.8	5.5	14.3
Disgust	3.8	82.2	1.3	0.6	2.2	0.0	9.8
Fear	9.7	3.4	66.9	6.3	2.9	4.0	6.9
Joy	2.4	1.0	1.7	85.1	0.0	2.7	7.1
Sad	10.0	0.4	1.7	0.0	65.2	3.3	19.5
Surprise	2.0	0.0	1.4	4.0	0.7	84.8	6.1
Neutral	12.2	0.0	2.0	0.5	0.5	4.0	80.7

Table 4. 6-Class expression recognition: SVM with different kernels.

	Kernels		
Feature Descriptor	**Liner (%)**	**Polynomial (%)**	**RBF (%)**
Gabor feature [4]	89.4 ± 3.0	89.4 ± 3.0	89.8 ± 3.1
LBP [21, 25]	91.5 ± 3.1	91.5 ± 3.1	92.6 ± 2.9
LDP [14]	92.8 ± 2.3	92.8 ± 2.3	94.5 ± 1.8
LDPv	95.2 ± 1.2	95.2 ± 1.2	96.7 ± 0.9

Table 5. 7-Class expression recognition: SVM with different kernels.

	Kernels		
Feature Descriptor	**Liner (%)**	**Polynomial (%)**	**RBF (%)**
Gabor feature [4]	86.6 ± 4.1	86.6 ± 4.1	86.8 ± 3.6
LBP [21, 25]	88.1 ± 3.8	88.1 ± 3.8	88.9 ± 3.5
LDP [14]	89.8 ± 1.9	89.8 ± 1.9	91.3 ± 1.7
LDPv	92.5 ± 1.8	92.5 ± 1.8	93.1 ± 1.6

Table 6. Recognition rate of different methods for the 6-class expression recognition problem.

Feature Descriptor	Recognition Rate (%)
ICA [30]	60.4
EICA [30]	65.8
Zernike Moment (10^{th} order) [18]	73.2
log-Gabor feature [17]	91.8
LDPv	96.7

6.2. Optimal LDPv Parameter Value Determination

In order to determine the optimal values of the two LDPv parameters, we first fix the number of regions *g*, and find the value of *k* which gives the best recognition performance. Next, with the determined value of *k*, we search for the optimal value of *g*. For a particular value of *k*, the LDP produces C_k^8 number of possible codes. The number of possible codes decides the number of bins in LDPv. Due to commutative property, *k*=1 gives the same number of bins as *k*=7, i.e., $C_1^8 = C_7^8$. Similarly, $C_2^8 = C_6^8$ and $C_3^8 = C_5^8$. Therefore, the parameter *k* can take any value from $\{1, 2, 3, 4\}$. Table 7 shows the generalized performance for different *k* values with the facial images divided into 42 (7×6) regions. The best recognition rate, with respect to the dimension of the LDPv feature, is achieved when *k*=3. With *k*=4, the LDPv descriptor's dimension 2940 is higher than that 2352 with *k*=3, but the recognition rate does not improve much. This supports the fact that larger descriptor does not always contain more discriminative information, sometimes even degrades the classification task.

Table 7. Recognition performance for different values of *k*.

	6-Class Expression (%)	7-Class Expression (%)	Vector Length of LDPv Feature
***k* = 1**	95.0	91.2	336
***k* = 2**	95.9	91.8	1176
***k* = 3**	96.7	93.1	2352
***k* = 4**	96.7	92.5	2940

Now, we would like to determine the optimal number of regions on the facial images. The commonly used numbers of regions are 3×3, 5×5, 7×7, 7×6, 9×8 etcetera. In our experiment, we evaluated four cases: 3×3, 5×5, 7×6, 9×8. Table 8 lists the effects of different number of regions on the recognition performance. With small number of regions, the expression recognition rate is low (below 83%). As we increase the number of regions, the recognition performance starts to increase as the descriptor feature starts to incorporate more local and spatial relationship information. But at a certain point, too many regions incorporated unnecessary local information that degraded performance. From our observation, 7×6 number of regions gives a good trade-off between recognition performance and feature vector length. Therefore, we concluded that k=3 and g= 7×6 are the optimal parameter values in the proposed LDPv descriptor for representing facial expression images.

Table 8. Recognition performance for different number of regions.

	6-Class Expression (%)	7-Class Expression (%)	Vector Length of LDPv Feature
g = 3×3	83.1	80.3	504
g = 5×5	95.6	90.4	1400
g = 7×6	96.7	93.1	2352
g = 9×8	96.2	93.1	4032

6.3. Effect of Dimensionality Reduction

All LDPv facial representations are projected onto the sub-space for dimension reduction defined by the significant principal components from PCA. The dimension of the sub-space 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 section, the optimal number of PCs is determined and the new feature space is found from those PCs. Figure 8 shows the recognition rate for different number of PCs varying from 60-260. With 240 PCs, the projected features achieved a recognition performance of 96.7% and 93.1% for 6-class and 7-class facial expression recognition, respectively. With more number of transformed features, the recognition rate shows a nearly constant performance. Therefore, for our expression recognition system, we opted to choose the topmost q=240 element representation instead of the original p=2352 dimensional feature vector, and this 240-element feature descriptor still provides the same recognition rate as the original descriptor.

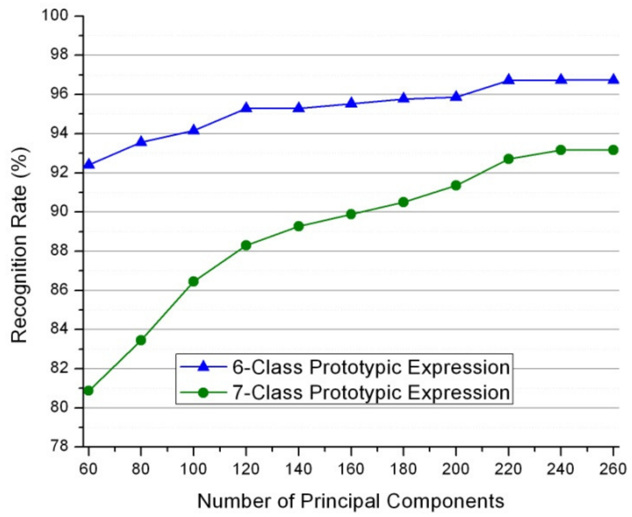


Figure 8. Recognition rate of prototypic facial expressions for various numbers of PCs in eigen-space.

Table 9. Confusion matrix of 6-class expression using LDPv+PCA and SVM (RBF).

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sad (%)	Surprise (%)
Anger	93.8	2.5	0.5	0.0	2.8	0.5
Disgust	0.0	97.2	2.8	0.0	0.0	0.0
Fear	1.0	0.0	96.5	2.0	0.0	0.5
Joy	0.0	0.5	0.5	98.0	1.0	0.0
Sad	1.6	0.0	0.0	0.0	97.5	1.0
Surprise	0.0	0.0	2.0	0.0	0.0	98.0

Table 10. Confusion matrix of 7-class expression using LDPv+PCA and SVM (RBF).

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sad (%)	Surprise (%)	Neutral (%)
Anger	79.4	0.5	0.5	0.0	2.3	0.9	16.5
Disgust	2.2	92.8	0.7	0.0	0.0	0.0	4.4
Fear	1.5	0.0	93.1	0.0	0.0	0.5	4.9
Joy	0.0	0.0	0.4	99.6	0.0	0.0	0.0
Sad	1.6	0.5	0.0	0.0	92.0	0.0	5.9
Surprise	0.8	0.0	0.0	0.0	0.0	98.7	0.4
Neutral	6.0	0.5	0.5	0.0	2.1	0.5	90.3

Tables 9 and 10 list the confusion matrix for 6-class and 7-class expression recognition, respectively, using LDPv with PCA and SVM classifier. Compared to the recognition results for template matching in Tables 2 and 3, the recognition performance has increased for every expression. For the 6-class problem, all expressions have a good recognition rate (above 93%). In case of 7-class problem, it is observed that Disgust, Fear, Joy, Sad, Surprise, and Neutral can be recognized with a high accuracy (90-99%). However, the recognition rate for Anger expression is slightly below 80%.

6.4. Evaluation at Different Resolution

In environments like smart meeting, visual surveillance, old-home monitoring, only low resolution video input is available [28]. Deriving action units from these facial images are critical problems. In this section, we explore the recognition performance on low resolution images with LDPv descriptor. Four different resolutions (150×110, 75×55, 48×36, 37×27) of face images based on the Cohn-Kanade dataset were studied. The low resolution images were formed by down-sampling the original images. All face images were divided into 7×6 number of regions for building the LDPv descriptor. To compare with the methods based on LBP and Gabor features, we conducted similar experiments on the 6-class prototypic expression recognition using a SVM with RBF kernel. Table 11 lists the recognition results with LBP, Gabor and the proposed LDPv feature. Both LBP and LDPv based features performed better than the Gabor feature, but have a lower feature dimension. The proposed LDPv based facial descriptor showed improved recognition performance, as compared to existing appearance-based methods. We also note that a dimension reduction with PCA reduces the dimension of the LDPv descriptor from O(10^3) to O(10^2) with little effect on the recognition performance but requires less computational resources.

When using low resolution images, it is difficult to extract geometric feature, therefore, appearance based methods appear to be a good alternative. Our analysis with LDPv feature demonstrates that the proposed descriptor feature performs robustly and stably over a range of expressions, even for low resolution facial images. The superiority of encoding the directional response of facial components over encoding the intensity values during face detection and recognition is also reported in [14].

Table 11. Recognition performance (%) in low resolution images.

	150×110	75×55	48×36	37×27
Feature				
Gabor [4]	89.8 ± 3.1	89.2 ± 3.0	86.4 ± 3.3	83.0 ± 4.3
LBP[21, 25]	92.6 ± 2.9	89.9 ± 3.1	87.3 ± 3.4	84.3 ± 4.1
LDPv	96.7 ± 0.9	95.6 ± 1.7	93.6 ± 2.0	90.3 ± 2.2
LDPv+PCA	96.7 ± 1.7	95.2 ± 1.6	93.1 ± 2.2	90.6 ± 2.7

7. Conclusions

In this paper, we have presented a facial expression recognition system based on the proposed LDPv representation, which encodes the spatial structure and contrast information of facial expressions. Extensive experiments illustrate that the LDPv features are effective and efficient for expression recognition. The discriminative power of the LDPv descriptor mainly lies in the integration of local edge response pattern and contrast information that makes it robust and insensitive to noise and non-monotonous illumination changes. Furthermore, with DR functions, like PCA, the newly transformed LDPv features also maintain a high recognition rate with lower computational cost. Once trained, our system can be used for human-computer interaction by facial expressions. Psychological experiments by Bassili [3] have suggested that facial expressions can be recognized more accurately from sequence images than from a single image. In the future, we plan to explore the sequence images and incorporate temporal information into the LDPv descriptor.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0015908).

This work was supported by a grant from the Kyung Hee University in 2010 (KHU-20101372).

References

[1] Ahonen T., Hadid A., and Pietikainen M., “Face Description with Local Binary Patterns: Application to Face Recognition,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, 2006.

[2] Ahonen T., Hadid A., and Pietikäinen M., “Face Recognition with Local Binary Patterns,” *in Proceedings of 8th European Conference on Computer Vision*, Berlin, pp. 469-481, 2004.

[3] Bassili N., “Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Area of the Face,” *Journal of Personality and Social Psychology*, vol. 37, no. 11, pp. 2049-2058, 1979.

[4] Bartlett S., Littlewort G., Frank M., Lainscsek C., Fasel I., and Movellan J., “Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior,” *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, California, vol. 2, pp. 568-573, 2005.

[5] Bartlett S., Movellan J., and Sejnowski J., “Face Recognition by Independent Component Analysis,” *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450-1464, 2002.

[6] Brunelli R. and Poggio T., “Face Recognition: Features Versus Templates,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, 1993.

[7] Chao-Fa C. and Shin F., "Recognizing Facial Action Units Using Independent Component Analysis and Support Vector Machine," *Pattern Recognition*, vol. 39, no. 9, pp. 1795-1798, 2006.

[8] Cohen I., Sebe N., Garg A., Chen L., and Huang T., "Facial Expression Recognition from Video Sequences: Temporal and Static Modeling," *Computer Vision and Image Understanding*, vol. 91, no. 1-2, pp. 160-187, 2003.

[9] Donato G., Bartlett S., Hagar C., Ekman P., and Sejnowski J., "Classifying Facial Actions," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989, 1999.

[10] Ekman P. and Friesen W., *Facial Action Coding System: A Technique for Measurement of Facial Movement*, Consulting Psychologists Press, 1978.

[11] Fasel B. and Luettin J., "Automatic Facial Expression Analysis: A Survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259-275, 2003.

[12] Gundimada S. and Asari K., "Facial Recognition Using Multisensor Images Based on Localized Kernel Eigen Spaces," *IEEE Transactions on Image Processing*, vol. 18, no. 6, pp. 1314-1325, 2009.

[13] Hsu C. and Lin C., "A Comparison on Methods for Multi-Class Support Vector Machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415-425, 2002.

[14] Jabid T., Kabir H., and Chae O., "Local Directional Pattern for Face Recognition," *in Proceedings of IEEE International Conference on Consumer Electronics*, NV, pp. 329-330, 2010.

[15] Kanade T., Cohn J., and Tian Y., "Comprehensive Database for Facial Expression Analysis," *in Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, France, pp. 46-53, 2000.

[16] Kumar A., "Analysis of Unsupervised Dimensionality Reduction Techniques," *Computer Science and Information System*s, vol. 6, no. 2, pp. 217-227, 2009.

[17] Lajevardi M. and Hussain M., "Feature Extraction for Facial Expression Recognition Based on Hybrid Face Regions," *Advances in Electrical and Computer Engineering*, vol. 9, no. 3, pp. 63-67, 2009.

[18] Lajevardi M. and Hussain M., "Higher Order Orthogonal Moments for Invariant Facial Expression Recognition," *Digital Signal Processing*, vol. 20, no. 6, pp. 1771-1779, 2010.

[19] Manjunath S. and Ma Y., "Texture Features for Browsing and Retrieval of Image Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837-842, 1999.

[20] Meulders M., Boeck D., Mechelen V., and Gelman A., "Probabilistic Feature Analysis of Facial Perception of Emotions," *Applied Statistic*s, vol. 54, no. 4, pp. 781-793, 2005.

[21] Ojala T. and Pietikainen M., "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.

[22] Pantic M. and Rothkrantz M., "Automatic Analysis of Facial Expressions: the State of the Art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424-1445, 2000.

[23] Qader H., Ramli A., and Haddad S., "Fingerprint Recognition Using Zernike Moments," *International Arab Journal of Information Technology*, vol. 4, no. 4, pp. 372-376, 2007.

[24] Shan C., Gong S., and McOwan P., "Robust Facial Expression Recognition Using Local Binary Patterns," *in Proceedings of IEEE International Conference Image Processing*, pp. 914-917, 2005

[25] Shan C., Gong S., and McOwan P., "Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803-816, 2009.

[26] Tian Y., Brown L., Hampapur A., Pankanti S., Senior A., and Bolle R., "Real World Real-Time Automatic Recognition of Facial Expressions," *in Proceedings of IEEE Workshop on Performance Evaluation of Tracking and Surveillance*, USA, pp. 9-16, 2003.

[27] Tian Y., Kanade T., and Cohn J., *Facial Expression Analysis: Handbook of Face Recognition*, Springer, 2003.

[28] Tian Y., "Evaluation of Face Resolution for Expression Analysis," *in Proceedings of CVPR Workshop on Face Processing in Video*, USA, pp. 82, 2004.

[29] Turk A. and Pentland P., "Face Recognition Using Eigenfaces," *in Proceedings of Computer Vision and Pattern Recognition*, USA, pp. 586-591, 1991.

[30] Uddin Z., Lee J., and Kim T., "An Enhanced Independent Component-Based Human Facial Expression Recognition from Video," *IEEE Transactions Consumer Electronics*, vol. 55, no. 4, pp. 2216-2224, 2009.

[31] Valstar M., Patras I., and Pantic M., "Facial Action Unit Detection Using Probabilistic Actively Learned Support Vector Machines on Tracked Facial Point Data," *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop*, USA, vol. 3, pp. 76-84, 2005.

[32] Valstar M. and Pantic M., "Fully Automatic Facial Action Unit Detection and Temporal Analysis," *in Proceedings of IEEE Conference*

on Computer Vision and Pattern Recognition Workshop, New York, pp. 149, 2006.

[33] Zhang Z., Lyons M., Schuster M., and Akamatsu S., "Comparison between Geometry-Based and Gabor-Wavelets-Based Facial Expression Recognition using Multi-Layer Perceptron," *in Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, USA, pp. 454-459, 1998.

[34] Zhao G. and Pietikainen M., "Boosted Multi-Resolution Spatiotemporal Descriptors for Facial Expression Recognition," *Pattern Recognition Letters*, vol. 30, no. 12. pp. 1117-1127, 2009.

[35] Zhou H., Wang R., and Wang C., "A Novel Extended Local-Binary-Pattern Operator for Texture Analysis," *Information Sciences*, vol. 178, no. 22, pp. 4314-4325, 2008.

Hasanul Kabir received his BSc degree in computer science and information technology from the Islamic University of Technology, Bangladesh in 2003, and PhD degree in computer engineering from Kyung Hee University, South Korea in 2011. Currently he is serving as an assistant professor in the Department of CSE in Islamic University of Technology. His research interests include feature extraction, motion estimation, computer vision, and pattern recognition.

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

Oksam Chae received his BSc degree in electronics engineering from Inha University, South Korea in 1977. He completed his MS and PhD degree in electrical and computer engineering from Oklahoma State University, USA in 1982 and 1986, respectively. In 1986-88 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, South Korea. His research interests include multimedia data processing environments, intelligent filter, motion estimation, intrusion detection system, and medical image processing in dentistry. He is a member of IEEE, SPIE, KES and IEICE.