Journal of Information Hiding and Multimedia Signal Processing ©2015 ISSN 2073-4212
Ubiquitous International Volume **6**, Number **1**, January **2015**

Improving Face Recognition Performance Using Similarity Feature-based Selection and Classification Algorithm

Chi-Kien Tran[1], Tsair-Fwu Lee, Ph.D[1,*],

[1] Department of Electronics Engineering,
National Kaohsiung University of Applied Sciences,
Kaohsiung, Taiwan, ROC

Pei-Ju Chao, Ph.D[2,*]

[2] Department of Radiation Oncology,
Kaohsiung Chang Gung Memorial Hospital,
Kaohsiung, Taiwan, ROC
[*] **Corresponding authors**: tflee@kuas.edu.tw; pjchao99@gmail.com;

Received November, 2013; revised July, 2014

Abstract. *In this paper, we propose the effective similarity feature-based selection and classification algorithm to select similarity features on the training images and to classify face images in face recognition system. The experiments were conducted on the ORL Database of Faces, which consists of 400 images of 40 individuals, and the Yale Face Database, which is made up of 11 images per 15 classes. Three face recognition systems, one based on the histogram-based feature, the second based on the feature which is the mean of pixel values in window with size of 4×4, and the last based on the local directional pattern feature, were developed. Euclidean distance, Manhattan distance and Chi-square distance were taken as distance metrics for the classification method. The results indicated that the proposed algorithms not only reduced the dimensions of feature space but also achieved a mean recognition accuracy that was 1.55%÷11.31% better compared to conventional algorithms.*
Keywords: Face recognition, similarity feature, histogram, pixel values, local directional pattern.

1. **Introduction.** Face recognition has a wide variety of applications such as in identity authentication, access control and surveillance [1]. Engineering started to show interest in face recognition in the 1960s. One of the first researchers of this subject was Woodrow W. Bledsoe [2]. Since Bledsoe, there has been a lot of research to deal with different aspects of this field. Despite achievements, face recognition challenges remain in computer vision research [1, 3, 4]. One of these is how to extract features from face images. These features are important in the later step of identifying the subject with an acceptable error rate. Feature extraction involves in several steps - dimensionality reduction, feature extraction and feature selection. In these steps, the selection a subset of the extracted features is an important step that can cause the smallest of classification errors.

Feature selection transforms or combines the data in order to select a proper subspace in the original feature space. In other words, a feature selection algorithm selects the best subset of the input feature set. Feature selection is an important stage of training and is one of two ways of avoiding the curse of dimensionality (the other is feature extraction). There are two approaches in feature selection known as "forward selection" and "backward selection". Forward selection will start with no features and add them one by one, at each step adding the one that decreases the error most, until any further addition does not significantly decrease the error. Backward selection will start with all the features and remove them one by one, at each step removing the one that decreases the error most (or increases it only slightly), until any further removal increases the error significantly. From the perspective of selection strategy, feature selection algorithms broadly fall into three models: filter, wrapper or embedded [5]. The filter model evaluates features without involving in any learning algorithm. The wrapper model requires a learning algorithm and uses its performance to evaluate the goodness of features. The embedded model incorporates feature selection as a part of the learning process, and use the objective function of the learning model to guide searching for relevant features such as decision trees or artificial neural networks.

In recent studies, many researchers have done much work on feature selection and have presented multiple class separability criterion and algorithms which are essentially based on the concept of 'Similarity Preserving Feature Selection'. These feature selection criterion and algorithms include Relief [6, 7] and ReliefF [7], Laplacian Score [8], Fisher Score [9], SPEC [10], HSIC [11] and Trace Ratio [12], in which, Fisher Score and ReliefF were designed to select features that assign similar values to the samples from the same class and different values to samples from different classes, Laplacian Score was designed to retain sample locality, and HSIC was designed to maximize feature-class dependency. However, these algorithms have the common drawback of being unable to handle feature redundancy, therefore it wastes lots of time computing, the accuracy is not high in face recognition applications [13].

Face images of the same person in a class have small changes in translation, rotation and illumination. Based on these characteristics, our idea is to keep similarity features in the set of training images of the same person, so we propose a feature selection algorithm to select a subset of the extracted features that cause the smallest classification error. Three face recognition systems were developed, the first based on the histogram-based feature, the second based on the feature which is the mean of pixel values in window with size of 4×4 ($M4 \times 4$), and the last based on the local directional pattern feature [14]. Euclidean distance, Manhattan distance and Chi-square distance were taken as distance metrics for the classification method [15, 16]. We also compared the proposed algorithms that used the similarity features and the conventional algorithms that did not use the similarity features. The proposed algorithms showed improvement on the recognition accuracy over the conventional algorithms.

2. Materials and Methods.

2.1. Input data.

The algorithms were implemented in Visual C# and then tested on two face databases, the ORL Database of Faces [17] and the Yale Face Database (cropped images of MIT Media Lab) which is publicly available for this research aims at the URL http://vismod.media.mit.edu/vismod/classes/mas622-00/datasets/. The ORL Database of Faces consists of 400 images of 40 individuals. The images contain a high degree of variability in expression, pose, and facial details are stored as a 112×92 pixel array with 256 gray levels (see Figure 1). The Yale Face Database is made up of 11 images per 15 classes (165 total images). The images are gray scale and are cropped with a resolution of 231×195 pixels (see Figure 2).

2.2. **Features used.** There are many approaches to extract the facial feature for face recognition, such as local binary patterns (LBP) [18], local Gabor binary pattern histogram sequence (LGBPHS) [19], local phase quantization (LPQ) [20], and local directional pattern (LDP) [14]. In this paper, we chose bin-based histogram feature,the mean of pixel values in window with size of 4×4 feature, and LDP-based feature in order to illustrate the potential of the proposed algorithms.

FIGURE 1. Example images of the ORL Database of Faces.

FIGURE 2. Example cropped images of the Yale Face Database.

2.2.1. *Histogram.* A histogram is a type of graph that has wide applications in statistics. The horizontal axis depicts the range and scale of observations involved, and the vertical axis shows the number of data points in various intervals i.e. the frequency of observations in the intervals. The histogram allows visualizing numerical data by indicating the number of data points that lie within a range of values, called a class or a bin. The frequency of the data that falls in each class is depicted by the use of a bar. Histograms are invariant to image manipulations such as rotations, translations but they also change slightly with a change in scale, angle of view or with occlusion. Despite these advantages, histograms perform poorly under different imaging or lighting conditions. They are also ineffective in distinguishing different images that have similar color distributions and suffer from inefficient computation due to their dimensionality. Some histogram-based face recognition systems have been introduced in [21-25].

Given a 256 gray image and H is the histogram feature vector of it. H can be defined by,

$$H\left[I\left(x,y\right) \times N_{bin}\ div\ 256\right] = H\left[I(x,y) \times N_{bin}\ div\ 256\right] + 1, \tag{1}$$

where $I(x,y)$ is the value of pixel at coordinate x, y and N_{bin} is the number of the bin.

2.2.2. *The mean of pixel values in window with size of* 4×4 *(*$M4 \times 4$*).* In this paper, we use a simple feature to test the proposed algorithms so-called the mean of pixel values in window with size of 4×4. It is defined by dividing a face image into non-overlapping windows (regions) with size of 4×4 and computing the mean value for these pixels.

$$m = \frac{\sum the\ pixel\ values\ in\ window\ with\ size\ of\ 4 \times 4}{16}. \tag{2}$$

2.2.3. *Local directional pattern (LDP).* LDP [14] 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. Given a central pixel in the image, applying the Kirsch compass edge detector, we obtain eight edge response values $m_0, m_1, , m_7$, each representing the edge significance in its respective direction (see Figure 3). We find the top k values $|\, m_i \,|$ and set them to 1. The remaining (8-k) bits of 8-bit LDP pattern are set to 0. Finally the LDP code is derived which is calculated as follows.

$$LDP_k = \sum_{i=0}^{7} s(m_i - m_k) 2^i, \tag{3}$$

where m_k is the k−th most significant directional response and the step function $s(x)$ is defined as Equation(4). Figure 4 shows an example of LDP code with k=3.

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}. \tag{4}$$

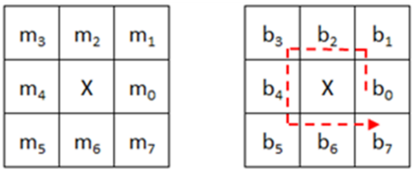
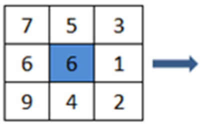


FIGURE 3. Edge response and LDP binary bit positions.



Mask index	m_7	m_6	m_5	m_4	m_3	m_2	m_1	m_0
Mask value	55	9	41	65	33	9	39	63
Rank	3	8	4	1	6	7	5	2
Code bit	1	0	0	1	0	0	0	1
LDP code	145							

FIGURE 4. Generating LDP code with k=3.

2.3. **Conventional algorithms.**

2.3.1. *Algorithm 1.* For training, training images are extracted features and stored in vectors for further processing. After that, mean of features from the stored vectors is calculated and stored in another vector for later use in phase classification. This mean vector is used for calculating the absolute differences among the mean of trained images and the test image.

Similarly, the first step of classification is the same as training. The second step, the minimum distance between the feature vector of test image and the mean feature vectors is calculated to find the matched class with test image.

Training algorithm Let the training set of face images be $\{I_1, I_2, \ldots, I_m\}$ and f_{ij} denote the jth feature of the ith image I_i, i=1,..., m; j=1,...,n.

The mean of the jth feature is defined by,

$$\Psi_j = \frac{1}{m} \sum_i f_{ij}. \tag{5}$$

Classification algorithm Let Y be a feature vector of a test image. Calculate the distance between Y and the mean feature vectors of p classes $\{\Psi^1, \Psi^2, \dots, \Psi^p\}$.

$$d_i\left(\Psi^i, Y\right) = L\left(\Psi^i, Y\right), \tag{6}$$

where proposed L metrics are dissimilarity measures such as Manhattan distance, Euclidean distance, Histogram intersection, Chi-square statistics and other distance measures.

Find the minimum distance between Y and $\{\Psi^1, \Psi^2, \dots, \Psi^p\}$

$$s = argmin_i(d_i), \tag{7}$$

and we say that the face with Y vector belongs to a class s.

2.3.2. *Algorithm 2.* Each image is divided into blocks and extracts the histogram from each block. These histograms are concatenated to get a spatially combined histogram which plays the role of a global face feature for the given face image. The recognition is performed using a nearest neighbor classifier with Chi-square statistics as dissimilarity measures. This algorithm is designed as described in detail in [14], but it does not use weight for regions. Figure 5 describes block diagram of the recognition system based on LDP descriptor.

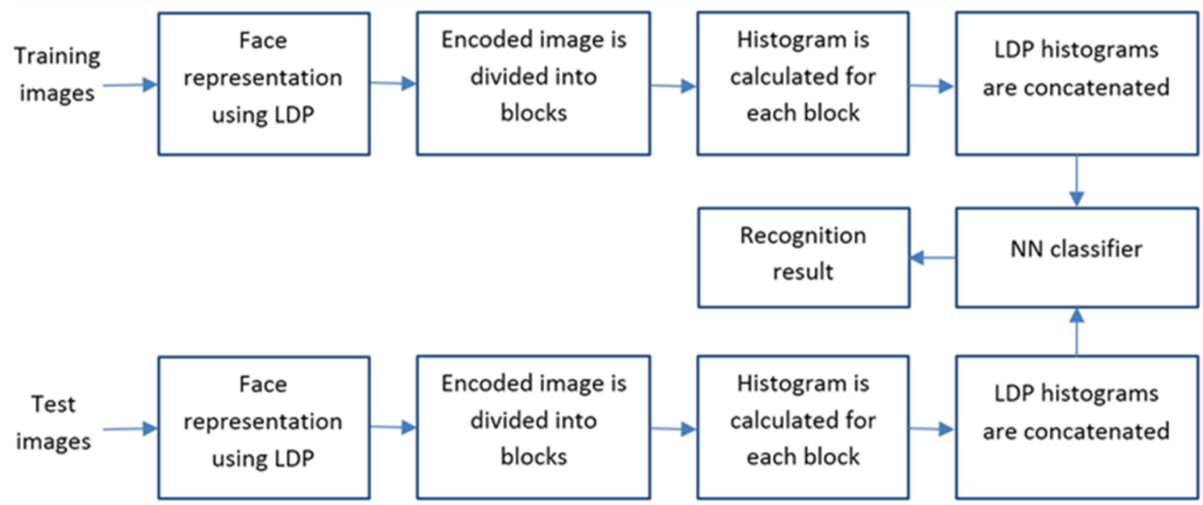


Figure 5. Block diagram of the recognition system based on LDP descriptor.

2.4. **Proposed similarity feature selection and classification algorithms.** Face images of the same person in a subject have small changes in translation, rotation and illumination. From these characteristics, this paper proposes algorithm to retain similarity features having discrimination power and stability which minimizes within-class differences whilst maximizes between-class differences.

In phase training, firstly, the training images of the same person are extracted features and stored in vectors for further processing. See Figure 6 for an illustration of the histogram-based feature extraction. Secondly, the mean of features from the stored vectors of previous step is calculated and is stored in a vector for later use in next step (see Figure 7(a)). Thirdly, the variance of features is calculated and stored in a vector. Fourthly, the mean vector and the variance vector are used to keep the features that have a little variance (the so-called similarity features). It means that the features which tend to be very close to the mean feature are kept and stored in other vectors. Finally (optional step) the mean of similarity features are calculated and stored in other vectors for later use in phase classification. Three similarity feature vectors of three face images are illustrated in Figure 7(b)-(d).

In phase classification, firstly, the test image is extracted features which are the same as phase training. Secondly, the mean distance between the feature vector of a test image and the similarity feature vectors (or mean similarity feature vectors) of classes are calculated. The calculation is based on the distance of the feature pairs which have the same coordinates and the value of similarity feature is greater than -1. Finally the minimum distance found identifies the matched class with test image.

In fact, in phase training, we only retain the similarity features and the corresponding coordinates of it so as to perform in next steps.

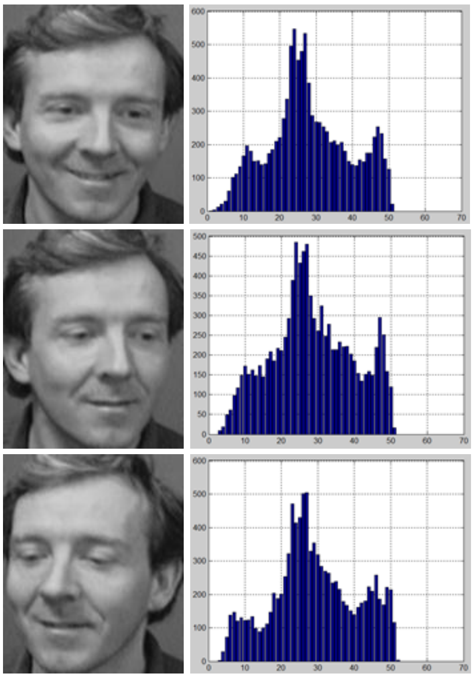


FIGURE 6. Three face images of third class and corresponding histograms with bin size is 4.

2.4.1. *Similarity feature selection algorithm.* Let a face image *I(x,y)* be a two-dimensional array $N_1 \times N_2$. An image may also be considered as a vector of dimension $N_1 \times N_2$, so that a face image with size 112×92 becomes a vector of dimension 10,304, or equivalents to a point in a 10,304-dimensional space.

Let the training set of face images be $\{I_1, I_2, \ldots, I_m\}$ and f_{ij} denote the jth feature of the ith image $I_i, i = 1, \ldots, m; j = 1, \ldots, n$. In which, m is the number of training images and n is the number of the features. The mean of the jth feature is defined by,

$$\Psi_j = \frac{1}{m} \sum_i f_{ij}. \tag{8}$$

Find the variance of the jth feature from the mean value of the jth feature Ψ_j

$$V_j = \frac{1}{m-1} \sum_i (f_{ij} - \Psi_j)^2. \tag{9}$$

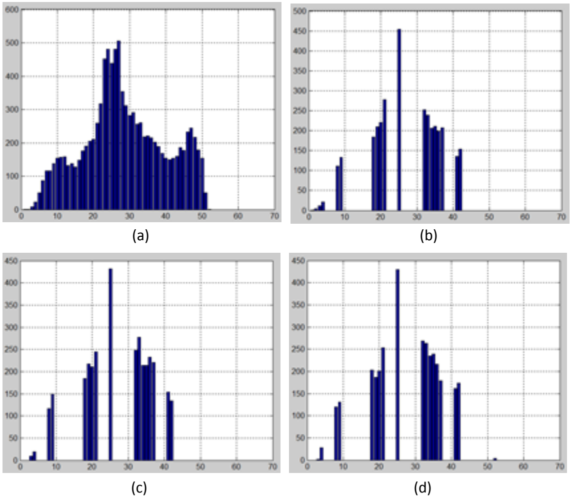


FIGURE 7. The mean histogram of three images and three similarity feature vectors. (a) Mean histogram, (b) similarity feature vector of first image, (c) similarity feature vector of second image, and (d) similarity feature vector of third image.

Find the maximum variance value from corresponding variance vector

$$\lambda = max\{V_j\}. \tag{10}$$

Keep the features of training images (the so-called similarity features) by using scalars λ then store into a set of vectors X_{ij} which denote the features of the ith image. X_{ij} is defined by,

$$X_{ij} = \begin{cases} f_{ij}, & V_j/\lambda \leq \varepsilon \\ -1, & otherwise \end{cases}, \tag{11}$$

where ϵ is the value of proposed threshold.

2.4.2. *Mean feature vector construction algorithm based on similarity feature.* Let $\bar{X}_{\mathrm{kj}}$ be mean similarity feature vector of kth corresponding class, $i = 1, \ldots, m; j = 1, \ldots, n; k = 1, \ldots, p$.

$\bar{X}_{\mathrm{kj}}$ is defined by,

$$\bar{X}_{kj} = \begin{cases} \frac{1}{m}\sum_i X_{ij}, & X_{ij} > -1 \\ -1, & X_{ij} = -1 \end{cases}, \tag{12}$$

where m is the number of training images of a class in phase training, n is the number of the features and p is the number of classes.

2.4.3. *Classification algorithm based on similarity feature.* In phase classification, let V_{kj} be input similarity feature vector of kth object (image or class), $j = 1, \ldots, n; k = 1, \ldots, p$ and Y be a feature vector of a test image.

Calculate the distance between Y and V_{kj}

$$d_k(V_{kj}, Y) = \frac{1}{q^k} L(V_{kj}, Y), \tag{13}$$

where q^k is the number of similarity features of the kth object (the features have value that is greater than -1) and proposed L metrics are dissimilarity measures such as Manhattan distance, Euclidean distance, Chi-square statistics. The calculation is based on the distance of the feature pairs which have the same coordinates and the value of similarity feature is different from -1.

Find the minimum distance between Y and V_{kj}

$$s = argmin_k(d_k), \tag{14}$$

and we say that the face with Y vector belongs to an object s.

3. **Experimental results and discussion.** In our experiments, we illustrated the potential of proposed algorithms on the ORL Database of Faces and the Yale Face Database. With first database, the training was performed by the number of combination of 10, taken m poses from each subject, $m = 2, \ldots, 9$; the others were used for the performance test. With second database, the data set was subdivided into a training set, made up of 5 images per class (75 images), and a test set, made up of 6 images per class (90 images).

3.1. **Experimental Results on the ORL Database of Faces.** Based on the ORL Database of Faces, two face recognition systems were developed. The first system used the histogram-based feature and the second one used M4×4 feature. In each system, Euclidean distance and Manhattan distance were taken as distance metrics for the classification method. In phase training, features were extracted from the training images then a similarity features subset of these ones was selected by the proposed algorithm. Mean feature vectors were calculated for each similarity feature subset for later use in phase classification. In phase classification, when a new image from the test set was considered for recognition, it was extracted the features and classified based on the proposed algorithm. Figure 8 presents the general diagram of the proposed system.

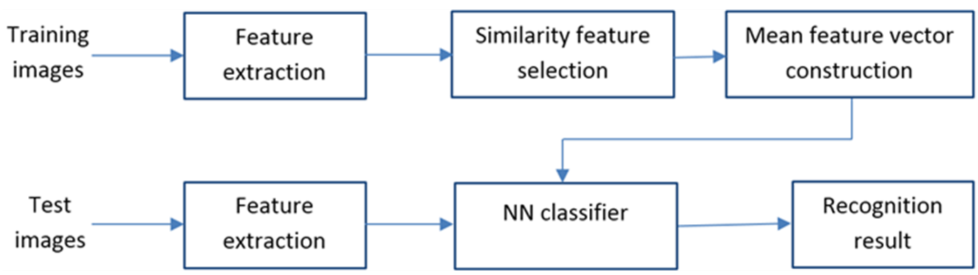


FIGURE 8. Block diagram of the proposed recognition system.

The results of experiments are summarized in Tables 1÷8. In these tables, the first row (text in bold) is result of conventional algorithms and the other rows are results of proposed algorithms. The fitted threshold value ϵ and the corresponding highest results of the proposed algorithms are also shown at the second bold row in these tables.

Tables 1÷3 list the results of the system using the histogram-based feature and the Euclidean distance based classifier. The best mean accuracy of the proposed algorithms achieved 88.74% (bin size = 1, $\epsilon = 0.09$, Table 1), 88.51% (bin size = 2, $\epsilon = 0.1$, Table 2) and 88.09% (bin size = 4, $\epsilon = 0.1$, Table 3) while the corresponding mean accuracy of the conventional algorithm only achieved 77.43%, 77.30% and 77.06%, respectively. We can see from these tables, the number of bins declined from 256 to 64, the mean accuracy also dropped from 77.43% to 77.06% (conventional algorithm) and 88.74% to 88.09% (proposed algorithm). This means a face can be represented by a smaller vector but it also loses much significant information.

Tables 4÷6 show the results of the system using the histogram-based feature and the Manhattan distance based classifier. The best mean accuracy of the proposed algorithms

Table 1. Performance of the histogram-based system with bin size is 1 based on the Euclidean distance based classifier.

Threshold	Training/Test images								Mean
	2/8(%)	3/7(%)	4/6(%)	5/5(%)	6/4(%)	7/3(%)	8/2(%)	9/1(%)	(%)
CA	**73.82**	**73.87**	**76.10**	**77.12**	**78.70**	**79.02**	**80.11**	**80.75**	**77.43**
0.50	75.25	78.18	80.89	82.52	84.20	85.13	85.94	87.25	82.42
0.40	75.43	79.32	82.19	84.03	85.83	86.84	87.83	89.25	83.84
0.30	75.62	80.51	83.70	85.79	87.65	88.73	90.05	92.00	85.51
0.20	75.68	81.98	85.63	87.94	89.47	90.41	90.77	91.25	86.64
0.10	75.72	83.43	87.44	89.73	91.48	92.56	93.30	94.00	88.46
0.09	**75.72**	**83.50**	**87.59**	**89.89**	**91.58**	**92.89**	**93.75**	**95.00**	**88.74**
0.08	75.59	83.50	87.74	89.97	91.72	92.87	93.94	94.50	88.73
0.07	75.69	83.61	87.60	89.94	91.52	92.63	93.69	93.50	88.52
0.06	75.37	83.52	87.63	89.69	91.13	92.12	92.61	93.50	88.20
0.05	75.24	83.36	87.33	89.08	89.98	90.40	90.33	89.75	86.93

m/n: means 'Training/Test images'; the training is performed by m poses from each subject and the performance testing is performed by n (n=10-m) poses of the same subjects.

Abbreviation- CA: Conventional Algorithm; the experimental results are performed based on conventional algorithm.

Table 2. Performance of the histogram-based system with bin size is 2 based on the Euclidean distance based classifier.

Threshold	Training/Test images								Mean
	2/8(%)	3/7(%)	4/6(%)	5/5(%)	6/4(%)	7/3(%)	8/2(%)	9/1(%)	(%)
CA	**73.74**	**73.74**	**75.99**	**76.97**	**78.51**	**78.83**	**79.86**	**80.75**	**77.30**
0.50	75.06	78.62	81.48	83.03	84.72	85.86	87.52	88.75	83.13
0.40	75.17	79.66	82.79	84.56	86.31	87.24	88.44	90.00	84.27
0.30	75.38	80.82	84.22	86.26	87.94	88.84	89.77	90.00	85.40
0.20	75.32	82.28	85.79	88.03	89.32	90.04	93.75	91.50	87.00
0.10	**75.51**	**83.11**	**87.38**	**89.97**	**91.68**	**92.90**	**93.55**	**94.00**	**88.51**
0.09	75.38	83.10	87.43	90.04	91.72	93.04	93.61	93.75	88.51
0.08	75.33	83.11	87.37	89.82	91.63	92.65	93.69	93.75	88.42
0.07	75.25	83.14	87.12	89.44	91.10	92.00	92.44	92.75	87.90
0.06	74.85	82.86	86.84	88.91	90.14	90.64	90.41	90.50	86.89
0.05	74.34	82.40	86.38	88.02	88.39	88.63	88.61	89.25	85.75

Abbreviation- CA: Conventional Algorithm

Table 3. Performance of the histogram-based system with bin size is 4 based on the Euclidean distance based classifier.

Threshold	Training/Test images								Mean
	2/8(%)	3/7(%)	4/6(%)	5/5(%)	6/4(%)	7/3(%)	8/2(%)	9/1(%)	(%)
CA	**73.65**	**73.58**	**75.80**	**76.76**	**78.12**	**78.56**	**79.52**	**80.50**	**77.06**
0.50	74.62	78.68	81.07	83.06	84.30	85.00	85.91	86.25	82.36
0.40	74.84	79.59	82.52	84.44	86.03	86.88	88.44	89.25	84.00
0.30	74.97	80.97	84.22	86.16	87.75	88.47	89.05	89.25	85.10
0.20	75.34	82.06	85.78	88.05	89.66	90.53	90.61	89.75	86.47
0.10	**74.81**	**82.61**	**86.87**	**89.51**	**91.04**	**92.15**	**93.25**	**94.50**	**88.09**
0.09	74.63	82.44	86.70	89.20	90.77	91.72	92.94	93.25	87.71
0.08	74.42	82.39	86.28	88.58	89.87	91.13	91.58	92.25	87.06
0.07	74.19	82.00	85.90	87.85	89.11	89.81	90.61	91.25	86.34
0.06	73.86	81.64	85.10	86.74	87.50	87.82	88.61	90.75	85.25
0.05	73.34	80.95	83.55	84.84	85.39	85.60	85.63	84.50	82.98

Abbreviation- CA: Conventional Algorithm achieved 84.34% (bin size = 1, $\epsilon = 0.2$, Table 4), 83.65% (bin size = 2, $\epsilon = 0.2$, Table 5) and 83.54% (bin size = 4, $\epsilon = 0.2$, Table 6) while the corresponding mean accuracy of the conventional algorithm achieved 78.94%, 78.77% and 78.13%, respectively. Similarly, when the number of bins dropped from 256 to 64, the mean accuracy also dropped from 78.94% to 78.13% (conventional algorithm) and 84.34% to 83.54% (proposed algorithm).

TABLE 4. Performance of the histogram-based system with bin size is 1 based on the Manhattan distance based classifier.

Threshold	Training/Test images								Mean
	2/8(%)	3/7(%)	4/6(%)	5/5(%)	6/4(%)	7/3(%)	8/2(%)	9/1(%)	(%)
CA	**74.90**	**75.70**	**77.59**	**79.30**	**79.93**	**80.83**	**81.25**	**82.00**	**78.94**
0.50	75.70	78.52	80.76	82.20	83.38	84.34	84.77	85.25	81.86
0.40	75.91	79.31	81.70	83.25	84.58	85.41	85.50	86.00	82.71
0.30	75.79	80.22	82.88	84.25	85.58	86.22	87.27	87.50	83.71
0.20	**75.72**	**80.84**	**83.84**	**85.29**	**86.05**	**86.72**	**87.25**	**89.00**	**84.34**
0.10	75.25	80.93	83.38	84.38	85.20	85.90	87.00	87.50	83.69
0.09	75.27	80.70	83.06	83.84	84.58	85.22	86.00	87.00	83.21
0.08	75.02	80.50	82.39	83.11	83.45	83.76	84.33	85.50	82.26
0.07	74.93	80.08	81.68	81.87	81.80	81.61	81.52	80.50	80.50
0.06	74.75	79.63	80.50	80.04	79.33	78.15	76.52	73.75	77.83
0.05	74.41	78.75	78.56	77.01	75.10	72.74	69.91	67.50	74.25

Abbreviation- CA: Conventional Algorithm

TABLE 5. Performance of the histogram-based system with bin size is 2 based on the Manhattan distance based classifier.

Threshold	Training/Test images								Mean
	2/8(%)	3/7(%)	4/6(%)	5/5(%)	6/4(%)	7/3(%)	8/2(%)	9/1(%)	(%)
CA	**74.68**	**75.53**	**77.39**	**79.28**	**79.73**	**80.40**	**81.16**	**82.00**	**78.77**
0.50	75.20	78.49	81.17	82.44	83.88	84.65	85.50	86.00	82.17
0.40	75.41	79.17	82.01	83.40	85.01	85.55	86.41	85.50	82.81
0.30	75.38	80.00	82.98	84.21	85.26	86.20	86.52	86.25	83.35
0.20	**75.49**	**80.40**	**83.54**	**85.03**	**85.50**	**86.09**	**86.38**	**86.75**	**83.65**
0.10	74.63	80.56	83.00	84.19	84.93	85.44	86.19	85.50	83.05
0.09	74.43	80.28	82.50	83.57	84.02	84.34	84.16	85.50	82.35
0.08	74.27	79.95	81.88	82.66	82.83	82.79	83.19	83.50	81.38
0.07	74.29	79.43	80.92	81.32	80.89	79.98	78.30	76.50	78.95
0.06	73.86	78.57	79.48	79.04	77.90	75.82	72.88	70.50	76.00
0.05	73.25	77.13	77.31	75.51	73.24	70.33	67.69	66.25	72.59

Abbreviation- CA: Conventional Algorithm

Tables 7÷8 show the results of two systems using the M4x4 feature, one uses the Euclidean distance-based classification method (see Table 7); the other uses the Manhattan distance-based classification method (see Table 8). The highest mean accuracy of first system reached 91.77% at $\epsilon = 0.2$ while this mean accuracy was only 90.22% in case of the conventional algorithm (see Table 7). Similarly, the second system reached 94.36% at $\epsilon = 0.1$ while the mean accuracy was 90.68% in case of the conventional algorithm (see Table 8). The highest mean accuracy of the second system achieved 94.36% whereas the first system gave 91.77%. This shows that the Manhattan distance-based classification method is better than the Euclidean distance-based classification method in case of the M4×4 feature.

TABLE 6. Performance of the histogram-based system with bin size is 4 based on the Manhattan distance based classifier.

Threshold	Training/Test images								Mean
	2/8(%)	3/7(%)	4/6(%)	5/5(%)	6/4(%)	7/3(%)	8/2(%)	9/1(%)	(%)
CA	**74.22**	**75.05**	**76.83**	**78.64**	**79.09**	**79.86**	**80.11**	**81.25**	**78.13**
0.50	74.80	78.13	80.58	82.12	83.13	83.66	83.88	83.50	81.23
0.40	74.97	78.77	81.43	82.98	84.21	84.93	85.63	86.25	82.39
0.30	75.03	79.53	82.35	83.58	84.96	85.59	86.75	87.00	83.10
0.20	**74.86**	**80.26**	**83.12**	**84.62**	**86.00**	**86.40**	**86.80**	**86.25**	**83.54**
0.10	73.68	79.25	81.74	83.02	83.74	84.35	84.41	85.50	81.96
0.09	73.54	78.75	80.97	82.16	82.97	83.38	83.80	83.50	81.13
0.08	73.20	78.21	79.77	80.82	81.30	80.95	80.33	79.75	79.29
0.07	72.56	77.33	78.57	79.03	79.17	78.67	77.69	77.50	77.56
0.06	72.37	76.22	76.64	76.39	75.86	75.26	74.41	74.00	75.14
0.05	71.61	74.81	73.86	73.40	72.89	72.21	71.41	68.75	72.37

Abbreviation- CA: Conventional Algorithm

TABLE 7. Performance of the mean pixel value in window of size 4×4 based system based on the Euclidean distance based classifier.

Threshold	Training/Test images								Mean
	2/8(%)	3/7(%)	4/6(%)	5/5(%)	6/4(%)	7/3(%)	8/2(%)	9/1(%)	(%)
CA	**82.06**	**86.63**	**88.93**	**90.66**	**91.96**	**92.95**	**94.02**	**94.50**	**90.22**
0.5	82.47	87.48	89.91	91.86	93.25	94.31	95.16	96.00	91.31
0.4	82.59	87.80	90.36	92.16	93.62	94.74	95.55	95.50	91.54
0.3	82.70	88.18	90.87	92.59	93.90	94.76	95.47	95.25	91.71
0.2	**82.82**	**88.40**	**91.15**	**92.76**	**93.87**	**94.65**	**95.25**	**95.25**	**91.77**
0.1	82.50	87.92	90.67	92.38	93.38	94.01	94.44	94.50	91.22
0.09	82.27	87.81	90.47	92.20	93.20	93.78	94.38	94.75	91.11
0.08	82.13	87.52	90.22	91.90	92.94	93.56	94.11	95.00	90.92
0.07	81.97	87.22	90.02	91.60	92.69	93.34	94.11	94.75	90.71
0.06	81.70	86.89	89.64	91.22	92.33	93.06	93.66	93.75	90.28
0.05	81.31	86.36	89.11	90.79	91.95	92.80	93.66	94.75	90.09

Abbreviation- CA: Conventional Algorithm

Table 9 presents a comparison of the best mean recognition accuracies between two recognition systems using the proposed algorithms, one based on the Euclidean distance-based classification method and the other based on the Manhattan distance-based classification method. As we can see from this table, the accuracy of system using the Euclidean distance-based classification method was better than using the Manhattan distance-based classification method in the same histogram-based feature; however, the result was opposite to the M4×4 feature. The findings also showed that the M4×4 feature is superior to the histogram-based feature.

3.2. **Experimental Results on The Yale Face Database.** In the experiments, each image of Yale face database was divided into 19×19 blocks. The conventional algorithm 2 was used to compare to the proposed algorithm. Both of them used Chi-square distance as dissimilarity measure. Figure 9 presents the general diagram of the proposed system.

Table 10 draws comparisons between two algorithms based on LDP feature with the Chi-square distance is used for the classification method. The proposed algorithm obtained better result than the conventional algorithm 1.9%. The suitable threshold value ϵ could

Table 8. Performance of the mean pixel value in window of size 4×4 based system based on the Manhattan distance based classifier.

Threshold	Training/Test images								Mean
	2/8(%)	3/7(%)	4/6(%)	5/5(%)	6/4(%)	7/3(%)	8/2(%)	9/1(%)	(%)
CA	**84.13**	**87.78**	**89.42**	**91.04**	**92.13**	**92.92**	**93.77**	**94.25**	**90.68**
0.50	84.52	88.56	90.28	92.04	93.09	93.77	94.88	95.00	91.52
0.40	84.61	88.83	90.69	92.38	93.59	94.46	95.33	95.75	91.95
0.30	84.84	89.34	91.49	93.06	94.35	95.14	95.72	96.00	92.49
0.20	85.09	90.11	92.69	94.28	95.56	96.37	97.02	97.25	93.55
0.10	**85.50**	**90.96**	**93.78**	**95.40**	**96.46**	**97.22**	**97.55**	**98.00**	**94.36**
0.09	85.42	90.93	93.83	95.47	96.42	97.07	97.47	98.25	94.36
0.08	85.48	90.96	93.91	95.50	96.49	97.03	97.44	98.00	94.35
0.07	85.54	91.00	93.89	95.47	96.37	96.97	97.30	97.25	94.22
0.06	85.63	91.01	93.86	95.34	96.25	96.80	97.13	97.50	94.19
0.05	85.56	90.89	93.72	95.15	96.06	96.75	97.00	97.50	94.08

Abbreviation- CA: Conventional Algorithm

Table 9. Performance of the mean pixel value in window of size 4×4 based system based on the Manhattan distance based classifier.

Metric	Bin(256) (%)	Bin(128) (%)	Bin(64) (%)	M(4x4) (%)
Euclidean Distance	88.74	88.51	88.09	91.77
Manhattan Distance	84.34	83.65	83.54	94.36

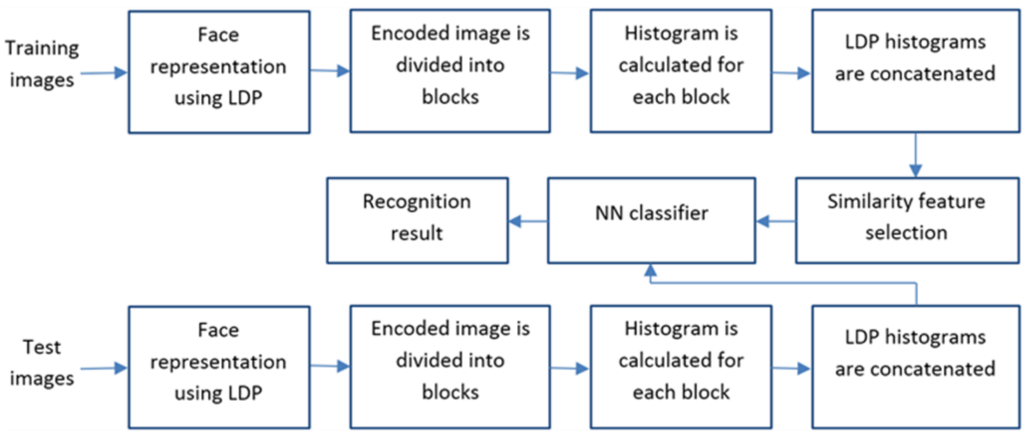


Figure 9. Block diagram of the proposed recognition system based on LDP descriptor.

get from 0.05 to 0.5 because the essential features (useful bins) which helped system to improve recognition rate, had similar values whereas other features had high distribution.

The results also showed that if the threshold value ϵ was too small, face image was presented in a lower dimension but it also lost many useful features, so recognition accuracy was also decreased. On the other hand, if the threshold value ϵ was high then the dispersion of feature in a class increased and the discrimination among different classes collapsed; therefore, the recognition accuracy was also decreased.

In this paper, we used three feature types and the classifier based on Euclidean distance, Manhattan distance and Chi-square distance. The results showed that the suitable

Table 10. Performance of the LDP-based system with the Chi-square distance is used for the classification method.

Set	Training set	Testing set	*Recognition rate (%)*	
			CA	*ε(0.05÷0.5)*
1	1,2,3,4,5	6,7,8,9,10,11	94.44	91.11
2	2,3,4,5,6	1,7,8,9,10,11	91.11	91.11
3	3,4,5,6,7	1,2,8,9,10,11	95.55	100.00
4	4,5,6,7,8	1,2,3,9,10,11	95.55	100.00
5	5,6,7,8,9	1,2,3,4,10,11	91.11	92.22
6	6,7,8,9,10	1,2,3,4,5,11	91.11	92.22
7	7,8,9,10,11	1,2,3,4,5,6	86.66	92.22
Average			**92.22**	**94.12**

threshold values for two face databases are $\epsilon \in [0.09, 0.5]$. However, if other systems use other features or classifiers then we have to find the suitable threshold value, because the results depend on the four factors: feature type, the number of training images, threshold value and distance measure.

Tables 1÷10 showed that the results of the proposed algorithms were outstanding, because these not only reduced the dimension of feature space, but also achieved a higher mean recognition accuracy than conventional algorithms from 1.55% to 11.31%. The proposed algorithms could perform better than the conventional ones because they kept essential information from training images and so enhanced the power of discrimination among different classes. Thanks to the advantages of the proposed algorithms, storage, performance and communication of face recognition systems will be better.

4. **Conclusion and future works.** In this paper, we propose similarity feature-based selection and classification algorithm. Three face recognition systems, the first system based on the histogram-based feature, the second one based on the feature which is the mean of value pixels in window with size of 4×4 (M4×4), and the third one based on LDP feature, were developed to show that the proposed algorithms outperform the conventional algorithms. The results showed that the our algorithms were a valuable tool for performance improvement of face recognition system.

Although higher recognition rate achieved by the proposed methods, still there are some issues which should be furthered addressed such as finding the optimal threshold for each database automatically, or applying the proposed algorithm with other features, in the purpose of improving the recognition rate.

Competing interests: Part of this study was presented on The Second International Conference on Robot, Vision and Signal Processing (RVSP-2013).

Acknowledgment. This study was supported financially, in part, by grants from NSC-101-2221-E-151-007-MY3, and NSC-102-2221-E-182A-002.

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, Face recognition: A literature survey, *ACM Comput. Surv.*, vol. 35, pp. 399–458, 2003.
[2] M. Ballantyne, R. S. Boyer, and L. Hines, Woody Bledsoe : His life and legacy, *The AI Magazine*, vol. 17, pp. 7–20, 1996.

[3] A. K. Jain, A. Ross, and S. Prabhakar, An introduction to biometric recognition, *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, pp. 4–20, 2004.

[4] Y. Ming-Hsuan, J. K. David, and A. Narendra, Detecting Faces in Images: A Survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 34–58, 2002.

[5] I. Guyon and A. Elisseeff, An introduction to variable and feature selection, *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[6] K. Kira and L. A. Rendell, A practical approach to feature selection, *The proceedings of the ninth international workshop on Machine learning, Aberdeen, Scotland, United Kingdom*, 1992.

[7] M. Robnik-Šikonja and I. Kononenko, Theoretical and Empirical Analysis of ReliefF and RReliefF, *Machine Learning*, vol. 53, pp. 23–69, 2003.

[8] X. He, D. Cai, and P. Niyogi, Laplacian score for feature selection, *Advances in Neural Information Processing Systems*, vol. 17, 2005.

[9] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, *2nd Edition ed. New York: John Wiley & Sons*, 2001.

[10] Z. Zhao and H. Liu, Spectral feature selection for supervised and unsupervised learning, *The proceedings of the 24th international conference on Machine learning, Corvalis, Oregon*, 2007.

[11] L. Song, A. Smola, A. Gretton, J. Bedo, and K. Borgwardt, Feature selection via dependence maximization, *J. Mach. Learn. Res.*, vol. 13, pp. 1393–1434, 2012.

[12] F. Nie, S. Xiang, Y. Jia, C. Zhang, and S. Yan, *Trace ratio criterion for feature selection*, presented at the Proceedings of the 23rd national conference on Artificial intelligence - Volume 2, Chicago, Illinois, 2008.

[13] D. Zhang, S. Chen, and Z.-H. Zhou, Constraint Score: A new filter method for feature selection with pairwise constraints, *Pattern Recognition*, vol. 41, pp. 1440–1451, 2008.

[14] T. Jabid, M. H. Kabir, and O. Chae, Local Directional Pattern (LDP) for face recognition, *in Consumer Electronics (ICCE), 2010 Digest of Technical Papers International Conference on*, 2010, pp. 329–330.

[15] V. Perlibakas, Distance measures for PCA-based face recognition, *Pattern Recognition Letters*, vol. 25, pp. 711–724, 2004.

[16] T. Ahonen, A. Hadid, and M. Pietikäinen, Face Recognition with Local Binary Patterns, *in Computer Vision - ECCV 2004*. vol. 3021, T. Pajdla and J. Matas, Eds., ed: Springer Berlin Heidelberg, 2004, pp. 469–481.

[17] The ORL Database of Faces, *AT&T Laboratories Cambridge*. Available: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

[18] T. Ojala, M. Pietikäinen, and D. Harwood, A comparative study of texture measures with classification based on featured distributions, *Pattern Recognition*, vol. 29, pp. 51–59, 1996.

[19] Z. Wenchao, S. Shiguang, G. Wen, C. Xilin, and Z. Hongming, Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition, *in Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1, pp. 786–791, 2005.

[20] V. Ojansivu and J. Heikkilä, Blur Insensitive Texture Classification Using Local Phase Quantization, *in Image and Signal Processing*. A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass, Eds., ed: Springer Berlin Heidelberg, vol. 5099, pp. 236–243, 2008.

[21] B. Fazl e, M. Y. Javed, and U. Qayyum, Face Recognition using Processed Histogram and Phase-Only Correlation (POC), *in Emerging Technologies, 2007. ICET 2007. International Conference on*, 2007, pp. 238–242.

[22] H. Gulati, D. Aggarwal, A. Verma, and D. P. S. Sandhu, Face Recognition using Hybrid Histogram & Eigen value Approach, *International Journal of Research in Engineering and Technology (IJRET)*, vol. 1, pp. 64–68, 2012.

[23] S. Singh, M. Sharma, and N. S. Rao, Robust & Accurate Face Recognition using Histograms, *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 10, pp. 113–122, 2012.

[24] H. Demirel and G. Anbarjafari, A new face recognition system based on color histogram matching, *in Signal Processing, Communication and Applications Conference, 2008. SIU 2008. IEEE 16th*, 2008, pp. 1–4.

[25] H. Demirel and G. Anbarjafari, High Performance Pose Invariant Face Recognition, *INSTICC - Institute for Systems and Technologies of Information, Control and Communication*, pp. 282–285, 2008.