2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

Learning Discriminative MspLBP Features Based on Ada-LDA for Multi-Class Pattern Classification

Kwang Ho An and Myung Jin Chung, *Senior Member, IEEE*

Abstract—**This paper presents a novel approach for multi-class pattern classification – face detection and facial expression recognition, which is based on discriminative multi-scale and multi-position Local Binary Pattern (MspLBP) features selected by a boosting technique called the AdaBoost+LDA (Ada-LDA) method. From a large pool of MspLBP features within a face image, the most discriminative MspLBP features trained by two alternative LDA methods depending on the singularity of the within-class scatter matrix, are selected under the framework of AdaBoost. To verify the feasibility of our approach, we performed two extensive experiments on the famous face databases in terms of face detection and facial expression recognition. First, face detection, a typical example of two-class pattern classification, was carried out on the MIT-CBCL and MIT+CMU face test sets. Second, facial expression recognition, a typical problem of multi-class pattern classification, was performed on the JAFFE face database. Given the same number of features, the proposed face detector shows over 25% higher detection rate than the well-known Viola's detector at a given false positive rate of 10%. It can also provide real-time operation with over 10 frames per second rate. For facial expression recognition, our approach also shows a better performance over at least 21% recognition rates than other linear subspace-based methods such as PCA, DCV, and PCA+LDA. Our proposed approach provides a considerable performance improvement with only a small number of discriminative MspLBP features in the multi-class pattern classification problem.**

I. Introduction

HUMAN faces provide a variety of different communicative functions such as identification, the perception of emotional expression, and lip-reading. Face analysis is a fundamental task for many applications in Human-Robot Interaction (HRI) such as face detection, face recognition, and face expression recognition. For this reason, face analysis has been actively studied by many researchers over the past two decades [1]-[6].

Most of the above algorithms directly use pixel values as features. However, they are very sensitive to illumination changes and image noises.

Papageorgiou et al. proposed new features called Haar-like features [7]. These features encode differences in average intensities between two rectangular regions, and they are able to extract texture without depending on absolute intensities.

Manuscript received September 15, 2009.

K. H. An and M. J. Chung are with the Electrical Engineering and Computer Science Department, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea (e-mail: akh@cheonji.kaist.ac.kr, mjchung@ee.kaist.ac.kr).

Viola and Jones proposed an efficient system for evaluating these features through an integral image [8]. And, they also introduced an efficient scheme for constructing a strong classifier by cascading a small number of distinctive features using the AdaBoost learning algorithm. Its result is more robust and computationally efficient.

However, these Haar-like rectangle features seem to be too simple, and the detector often contains extremely large set of rectangle features for good performance guarantee.

Ojala et al. introduced a powerful method of texture description [9]. As an efficient non-parametric method representing the local structure of an image, Local Binary Pattern (LBP) has been introduced for facial analysis. The most important properties of LBP features are their tolerance against monotonic illumination changes and their computational simplicity.

They later made two extensions of the original LBP [10]. First, the basic LBP was extended to use neighborhood of different sizes to capture dominant features at different scales. Second, they proposed to use a small subset of the patterns to describe the texture of images. These patterns are called uniform patterns. In this way, the number of patterns is greatly reduced without losing too much information.

Ahonen et al. recently proposed a novel approach for face recognition, which takes advantage of the LBP histogram that is proved to be an effective texture description [11]. In their method, the face image is equally divided into small sub-windows from which LBP histograms are extracted and concatenated to represent the local texture and global shape of face images. Weighted Chi square distance of these LBP histograms is used as a dissimilarity measure of different face images.

However, there is obviously an aspects could be improved in their method. In their approach, a face image is equally divided into sub-regions from which LBP histograms are extracted, which means the variety of the size and position of the obtained features are limited. By scaling and shifting a sub-window much more features could be obtained, which yields a larger and better description of face images.

To address this problem, Shan and Gritti presented a more promising method [12]. By shifting and scaling a sub-window over face images, many more sub-regions are obtained, and then AdaBoost is adopted to select the most discriminative sub-regions in terms of LBP histogram.

However, in their method, they used standard AdaBoost to learn the weak classifiers. The weak classifier is designed to select the single LBP histogram bin which best separates the training samples into their own classes. Their scheme pays attention to not the overall distribution of LBP features but the distribution of only a single LBP within a sub-region with a certain size and position. Hence, the scheme of constructing weak classifiers is too simple and may be not optimal.

With consideration of all of these issues, in this paper, we present a novel approach for multi-class pattern recognition, which uses a strong classifier based on the discriminative multi-scale and multi-position LBP (MspLBP) features trained by the AdaBoost+LDA method (Ada-LDA).

By scanning the face image with a scalable sub-window, many sub-regions are obtained from which the MspLBP histograms are extracted to describe the local structure of a face image. These MspLBP histograms can be considered as feature vectors in a vector space. So, various linear subspace-based methods suitable for pattern classification can be easily adopted to train more robust weak classifiers.

In our approach, we use Linear Discriminant Analysis (LDA) to meet this goal. LDA is a popular and powerful method for pattern classification, and so a good candidate for training weak classifiers [13].

From a large pool of MspLBP features, the most discriminative MspLBP features, trained by two alternative LDA methods depending on the singularity of the within-class scatter matrix, are selected under the framework of AdaBoost.

The rest of this paper is organized as follows. In Section II, the MspLBP feature is introduced and is represented as a feature vector in a vector space. In Section III, two alternative LDA methods depending on the singularity of the within-class scatter matrix are introduced to train weak classifiers in terms of the MspLBP features. Section IV presents the Ada-LDA learning algorithm to select the most discriminative MspLBP features and construct a final strong classifier. In Section V, we provide two challenging experimental results on the widely used face databases to verify the feasibility of the proposed approach in terms of face detection and facial expression recognition. Section VI presents conclusions and discussions.

II. Multi-scale and Multi-position LBP Features

The basic LBP operator is a powerful texture descriptor [9]. The basic LBP labels the pixels of an image by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number as shown in Fig. 1.

In order to capture large scale structure that may be the dominant features of some textures, later the basic LBP was extended to multi-resolution LBP which is denoted by $LBP_{P,R}$ as shown in Fig. 2 [10]. The notation (P,R) denotes a neighborhood of P equally spaced sampling points on a circle of radius of R. Using circular neighborhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighborhood.

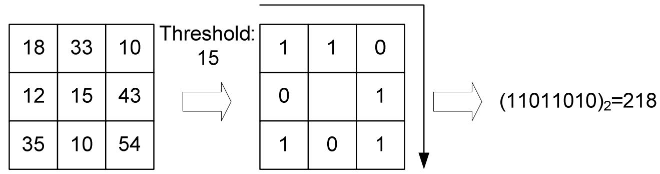

Fig. 1. The basic LBP operator.

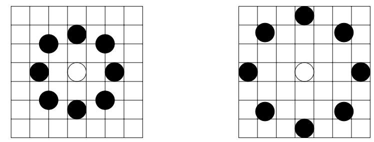
Fig. 2. Multi-resolution LBP operators, $LBP_{8,2}$ and $LBP_{8,3}$.

There are only 256 patterns for the basic LBP operator with $P=8$. According to the statistics, the probability of each pattern is not the same in images. From this observation, a uniform pattern operator was also introduced in [10].

A LBP is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. For example, 00000000, 001110000, and 11100001 are uniform patterns of $LBP_{8,R}^{u2}$. The superscript $u2$ stands for using only uniform patterns and labeling all remaining patterns with a single label.

The uniform patterns represent local primitives such as edges and corners. It was observed that most of the texture information was contained in the uniform patterns. All of non-uniform pattern is replaced by a single fixed pattern. In this way, the number of patterns is reduced greatly.

There are 58 uniform patterns and a non-uniform pattern, 59 patterns in all when using $LBP_{8,R}^{u2}$. As shown in Fig. 3, codes from 1 to 7 have 8 rotated patterns respectively, it is totally 56 codes. Additionally, code 0, 8, and the other non-uniform patterns are assigned to each single code. There are 59 codes in total.

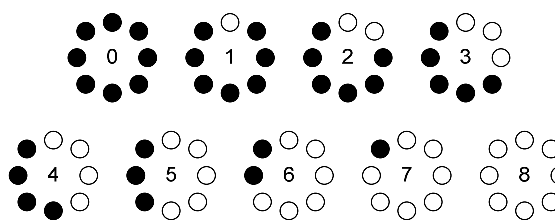

Fig. 3. Uniform patterns. The numbers inside them correspond to their unique $LBP_{8,R}^{u2}$ codes.

We try to represent the local texture and shape of face images with the MspLBP features obtained through $LBP_{8,R}^{u2}$ operator. By scanning the face image with a scalable sub-window, many sub-regions can be obtained. For each sub-region of a face image, a single MspLBP histogram can be constructed by scanning the sub-region with a 3×3 sub-window and extracting the corresponding basic LBP codes within the sub-region. Then, by using $LBP_{8,R}^{u2}$, a single MspLBP histogram, which represents the local structure of a face image, can be represented as a single feature vector $\mathbf{z}$ in the 59-dimensional space as shown in Fig. 4.

And each MspLBP feature can be a candidate of the weak classifier. If then, various linear subspace-based methods suitable for pattern classification can be easily adopted to train more robust weak classifiers which best separate the training samples into their own classes.

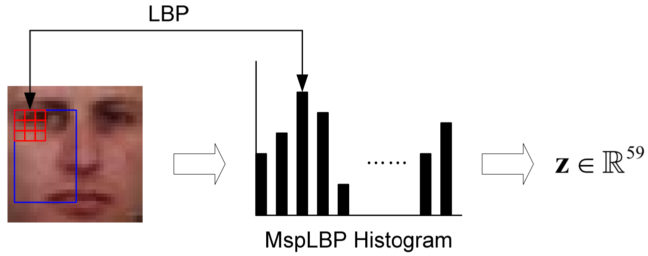

Fig. 4. MspLBP feature vector generation when using $LBP_{8,R}^{u2}$.

III. Weak Learning Algorithm

The weak learning algorithm is designed to select a single classifier which best separates the training samples into their corresponding classes. As we represented a single MspLBP histogram as a feature vector in a 59-dimensional space in Section II, we can easily use well-known and powerful linear subspace-based pattern classification techniques to train more robust weak classifiers.

As mentioned in Section I, to design a more discriminative weak classifier, we use LDA which has shown good performance in the field of face and expression recognition. LDA is used to seek a linear projection from the original sample space to a lower dimensional space, which maximizes the between-class scatter matrix while minimizing the within-class scatter matrix of the projected samples.

We have two alternative approaches to obtain the optimal projection of LDA depending on whether the within-class scatter matrix is nonsingular or singular.

A. Nonsingular $\mathbf{S}_W$

As mentioned above, the optimal solution of LDA is a linear projection which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples, i.e.,

$$\mathbf{W}^* = \arg\max_{\mathbf{W}} \frac{\left|\mathbf{W}^T \mathbf{S}_B \mathbf{W}\right|}{\left|\mathbf{W}^T \mathbf{S}_W \mathbf{W}\right|}, \tag{1}$$

$$\mathbf{S}_W = \sum_{c=1}^{C} \sum_{i=1}^{n_c} (\mathbf{z}_i - \boldsymbol{\mu}_c)(\mathbf{z}_i - \boldsymbol{\mu}_c)^T, \tag{2}$$

$$\mathbf{S}_B = \sum_{c=1}^{C} n_c (\boldsymbol{\mu}_c - \boldsymbol{\mu})(\boldsymbol{\mu}_c - \boldsymbol{\mu})^T, \tag{3}$$

where $\mathbf{z}$ denotes a MspLBP feature vector expressed in a 59-dimensional space. $\mathbf{S}_B$ is the between-class scatter matrix, and $\mathbf{S}_W$ is the within-class scatter matrix. $\boldsymbol{\mu}_c$ is the mean vector of the MspLBP feature vectors from class c, and $\boldsymbol{\mu}$ is the total mean vector of the MspLBP feature vectors from all training samples. n_c is the number of samples in class $c = 1, \ldots, C$.

It is easy to show that the orthonormal columns of the optimal projection matrix $\mathbf{W}^*$ that maximizes (1) must satisfy (4). This is a generalized eigenvalue problem.

$$\mathbf{S}_B \mathbf{w}_m = \lambda_m \mathbf{S}_W \mathbf{w}_m, \quad m = 1, \ldots, C-1. \tag{4}$$

Hence, if $\mathbf{S}_W$ is nonsingular, the columns of the optimal projection matrix $\mathbf{W}^*$ are the eigenvectors of $\mathbf{S}_W^{-1}\mathbf{S}_B$ corresponding to the largest $C-1$ eigenvalues.

B. Singular $\mathbf{S}_W$

On the other hand, for solving the problem of singular $\mathbf{S}_W$, we use a two-stage PCA+LDA method, also known as the Fisherface method, which was proposed by Belhumeur et al. [5].

In this method, PCA is first used for dimension reduction so as to make the resulting $\mathbf{S}_W$ nonsingular before the application of standard LDA.

$$\mathbf{U}^* = \arg\max_{\mathbf{U}} \left|\mathbf{U}^T \mathbf{S}_T \mathbf{U}\right|, \tag{5}$$

$$\mathbf{S}_T = \sum_{c=1}^{C} \sum_{i=1}^{n_c} (\mathbf{z}_i - \boldsymbol{\mu})(\mathbf{z}_i - \boldsymbol{\mu})^T = \mathbf{S}_W + \mathbf{S}_B, \tag{6}$$

where $\mathbf{S}_T$ is the total scatter matrix of the training samples, and $\mathbf{U}^*$ is the $59 \times r$ matrix whose columns are the orthonormal projection vectors corresponding to nonzero eigenvalues of $\mathbf{S}_T$. $r < 59$ is the rank of $\mathbf{S}_W$. And then, the standard LDA is applied to obtain an optimal projection matrix $\mathbf{V}^*$.

$$\begin{aligned}\mathbf{V}^* &= \arg\max_{\mathbf{V}} \frac{\left|\mathbf{V}^T \left((\mathbf{U}^*)^T \mathbf{S}_B \mathbf{U}^* \right) \mathbf{V}\right|}{\left|\mathbf{V}^T \left((\mathbf{U}^*)^T \mathbf{S}_W \mathbf{U}^* \right) \mathbf{V}\right|} \\ &= \arg\max_{\mathbf{V}} \frac{\left|\mathbf{V}^T \mathbf{S}_B^{'} \mathbf{V}\right|}{\left|\mathbf{V}^T \mathbf{S}_W^{'} \mathbf{V}\right|}.\end{aligned} \tag{7}$$

Also, the orthonormal columns of the optimal projection matrix $\mathbf{V}^*$ that maximizes (7) must satisfy (8).

$$\mathbf{S}_B^{'} \mathbf{v}_m = \chi_m \mathbf{S}_W^{'} \mathbf{v}_m, \quad m = 1, \ldots, C-1. \tag{8}$$

In the same way, the columns of the optimal projection matrix $\mathbf{V}^*$ are the eigenvectors of $\left(\mathbf{S}_W^{'}\right)^{-1} \mathbf{S}_B^{'}$ corresponding to the largest $C-1$ eigenvalues. Then, the final optimal projection matrix $\mathbf{W}^*$ of the PCA+LDA method is given by the following equation.

$$\mathbf{W}^* = \mathbf{U}^* \mathbf{V}^*. \tag{9}$$

C. Weak Classifiers

As mentioned above, each MspLBP feature, which represent the local image structure of the sub-region within a face image, can be a candidate of the potential weak classifiers. Finally, we design a weak classifier as shown below.

$$h = \arg\min_{c} \left\| \mathbf{W}^T \left(\mathbf{z} - \boldsymbol{\mu}_c \right) \right\|^2, \tag{10}$$

where h denotes a weak classifier, and $\mathbf{W}$ is the optimal projection matrix obtained through the two aforementioned alternative LDA methods depending on the singularity of the within-class scatter matrix, which best separates the training samples into their own classes.

IV. Ada-LDA Learning Algorithm

Although the complete feature set of MspLBP features is much smaller than Haar-like features, it also contains much redundant information, and not all extracted features are effective for pattern classification. Within a face image the total number of MspLBP features is very large, much larger than the number of pixels. So we propose the Ada-LDA learning algorithm to learn the most discriminative MspLBP features from a large pool of MspLBP features generated by shifting and scaling a sub-window over face images.

At each iteration, a single MspLBP feature trained through the two alternative LDA methods discussed in Section III, which minimizes the weighted error rate, is selected as a weak classifier, and the distribution is updated to increase the weights of the misclassified samples and reduce the importance of the others.

The main procedure of the proposed Ada-LDA learning algorithm is as follows:

- Given example images $(x_1, y_1), ..., (x_n, y_n)$, where $y_i = 1, ..., C$ for each class-belonging. y_i is the class label.
- Initialize the weights $w_{1,i} = 1/Cn_c$ for $y_i = 1, ..., C$ respectively, where n_c is the number of samples belonging to class c.
- For $t = 1, ..., T$:
  - ■ Normalize the weights $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{i=1}^{n} w_{t,i}}$ so that w_t is a probability distribution.
  - ■ For each MspLBP feature j, train a classifier h_j, which is restricted to using a single feature. The error is evaluated with respect to w_t. δ is the Kronecker delta function.

    $\varepsilon_j = \sum_{i=1}^{n} w_i \left(1 - \delta\left[h_j(x_i) - y_i\right]\right),$

    $\delta[k] = \begin{cases} 1, & if \quad k = 0 \\ 0, & if \quad k \neq 0 \end{cases},$

    $h_j(x_i) = \arg\min_{c} \left\| \mathbf{W}_j^T \left(\mathbf{z}_j(x_i) - \boldsymbol{\mu}_c^j \right) \right\|^2.$
  - ■ Choose the classifier h_t with the lowest error ε_t as the weak classifier.
  - ■ Update the weights, $w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$, where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$.
- The final strong classifier is defined as shown below.

  $H(x) = \arg\max_{c} \left(\sum_{t=1}^{T} \alpha_t \delta\left[h_t(x) - c\right] \right), \quad \alpha_t = \log\frac{1}{\beta_t}.$

V. Experimental Results

To verify the feasibility of our approach, we performed two challenging experiments on the famous face databases in terms of face detection and facial expression recognition. First, face detection, a typical example of two-class pattern classification, was carried out on the MIT-CBCL and MIT+CMU face test sets [14], [2]. Second, facial expression recognition, a typical problem of multi-class pattern classification, was performed on the JAFFE face database.

A. Two-class Problem : Face Detection

Face detection is a typical problem of two-class pattern classification. We used the MIT-CBCL face database for the training and test of our face detector. The MIT-CBCL training set contains 2,429 face images and 4,548 non-face images in 19 × 19 grayscale images. The training faces are roughly aligned. They were cropped manually around each face just above the eyebrows and about half-way between the mouth and the chin.

Fig. 5 shows the most discriminative 10 MspLBP features selected by the proposed Ada-LDA. As shown in this figure, most of significant features for face detection distribute in eyes and mouth, and nose regions.

We compared the performance of our detector with that of the well-known Viola's detector [8] on the MIT-CBCL face test set, which consists of 472 faces and 23,573 non-faces. The test images are of the same size as the training images, and are also cropped similarly. Considerable pose and lighting variations are represented by the test set. The test face images are clearly more challenging to identify as compared to the training ones.

The ROC curves of our detector on the MIT-CBCL face test set with the discriminative 50 MspLBP features for using $LBP_{8,1}^{u2}$ and $LBP_{8,2}^{u2}$ operators are shown in Fig. 6. As can be observed, our detector using the $LBP_{8,1}^{u2}$ operator produces slightly better performance than that of using $LBP_{8,2}^{u2}$. This may be due to a low resolution of the MIT-CBCL database.

Viola used Haar-like features under the framework of AdaBoost, and we used MspLBP features constructed by using the $LBP_{8,1}^{u2}$ operator under the framework of the proposed Ada-LDA. The ROC curves of the two detectors for the number of selected features are shown in Fig. 7. Detection rates with 50 selected features for various false positive rates are listed in Table I. It is shown that at the given false positive rate of 10%, our detector provides over 25% higher detection rate than the Viola's detector when using 50 selected features, and that our detector with only 10 MspLBP features is superior to Viola's detector with 50 Haar-like features. It is mainly because the MspLBP features can capture more information about the image structure, and so are more distinctive, and because the proposed scheme of constructing the final strong classifier through Ada-LDA yields more discrimination power.

Some of face detection results on the MIT+CMU test set, which is widely used to evaluate the performance of face detection algorithm, are shown in Fig. 8. Since the final strong classifier is insensitive to small changes in scale and shift, multiple detections occur around each face. Simple post-processing is needed to combine overlapping detections into a single detection as in [8]. The processing time of our detector for a 320×240 image is less than 0.1s on a 3.16 GHz Intel Core2 Duo PC.

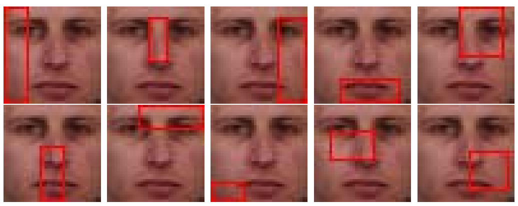

Fig. 5. The most discriminative 10 MspLBP features selected by Ada-LDA (left to right, top to bottom).

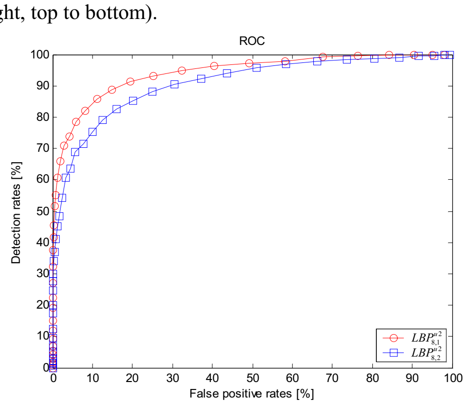

Fig. 6. ROC curves with the 50 selected MspLBP features for the $LBP_{8,1}^{u2}$ and $LBP_{8,2}^{u2}$ operators on the MIT-CBCL test set.

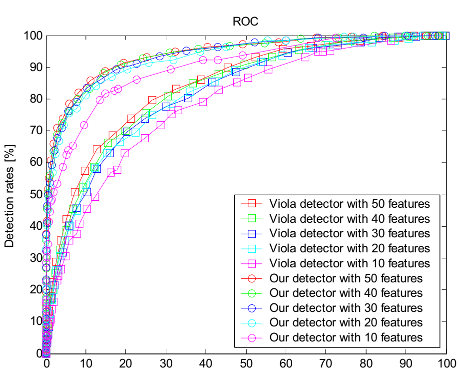

Fig. 7. ROC curves of two detectors on the MIT-CBCL test set for the number of selected features. "□" and "○"denote Viola's detector and ours respectively. Lines "red", "green", "blue", "cyan", and "magenta" represent using 50, 40, 30, 20, and 10 features respectively.

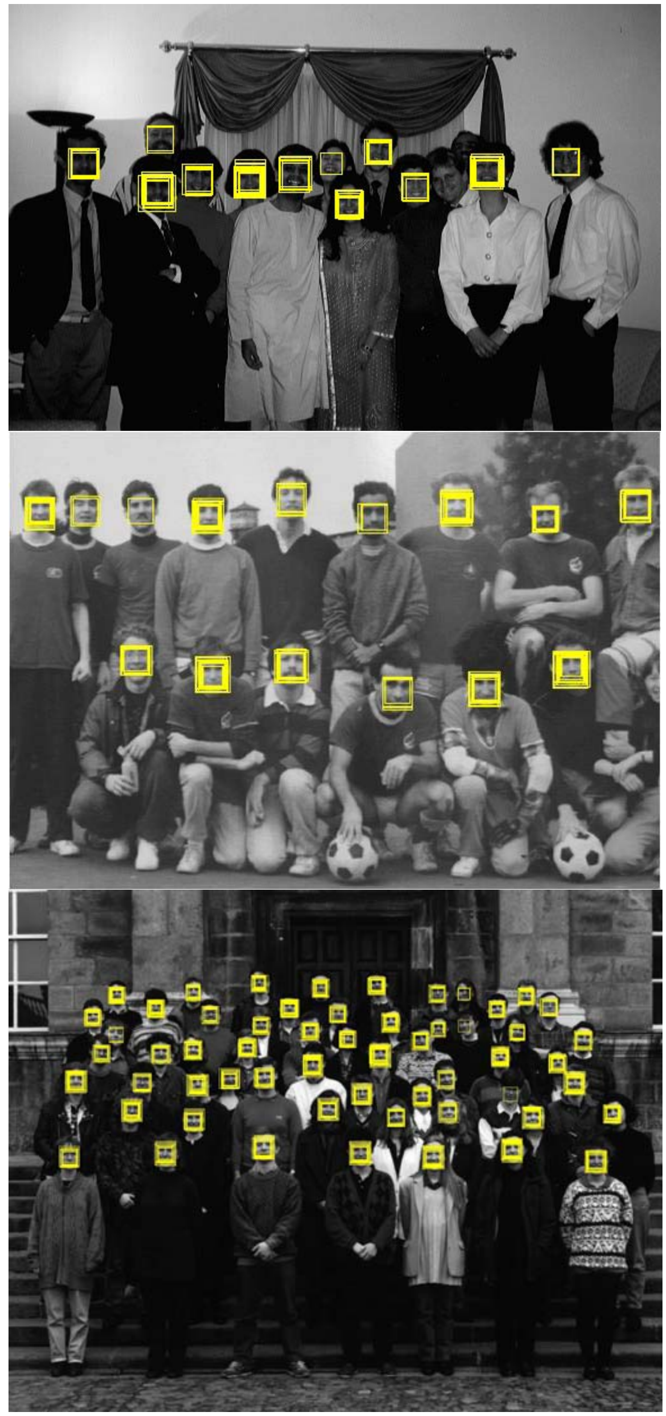

Fig. 8. Some face detection results on the MIT+CMU test set.

TABLE I
DETECTION RATES WITH 50 FEATURES FOR VARIOUS FALSE POSITIVE RATES ON MIT-CBCL TEST SET

	1.4%	1.8%	11.2%	12.8%	19.4%	21.2%
Ours	60.6%	-	85.8%	-	91.3%	-
Viola	-	22.3%	-	64.2%	-	73.7%

B. *Multi-class Problem : Facial Expression Recognition*

Facial expression recognition is a typical example of multi-class pattern classification. We used JAFFE database to verify the feasibility of the proposed algorithm for facial expression recognition [15]. Available JAFFE database contains 213 images of 7 facial expressions (6 basic facial expressions including happiness, sadness, surprise, anger, disgust, and fear and a neutral facial expression) posed by 10 subjects. 7 typical facial expression images from the JAFFE database are shown in Fig. 9.

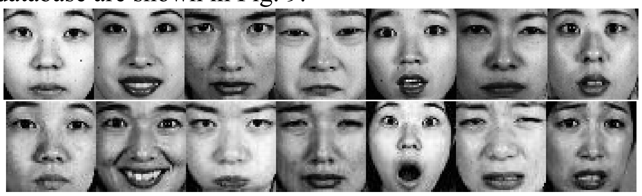

Fig. 9. Example facial expression images from the JAFFE database. Neutral, happy, anger, sad, surprise, disgust, and fear facial expressions are shown from left to right.

First, 210 images were taken from the database for our experiments. Facial images of 50×50 pixels were roughly cropped from original images based on the two eyes location. Then, we created an extended version of the JAFFE database by mirroring, rotating, translating, and scaling the original images by small amounts to obtain a set of 4,200 facial expression images. So, each class contains 600 corresponding facial expression images.

To compare the performance of the proposed algorithm with those of existing methods such as PCA, PCA+LDA, and DCV [4]-[6], we performed two-fold cross validation. We divided the extended data set into two equally-sized sets. One set was used for training and the other for test. This was repeated twice until each of the two sets is used once as the test set.

We only used the most discriminative 5 MspLBP features selected by the proposed Ada-LDA for each test. Also, we used $LBP_{8,2}^{u2}$ operator to describe the MspLBP features.

This experimental results clearly show that our proposed approach outperforms three other methods which use a whole face image as a feature vector over at least 21% recognition rates.

It is also due to the discriminative MspLBP features and the boosting scheme of constructing the final strong classifier through the proposed Ada-LDA.

TABLE II
RECOGNITION RATES ON THE EXTENDED JAFFE DATABASE

	PCA	PCA+LDA	DCV	Ours
Recognition Rates [%]	33.10	50.50	47.95	71.98

VI. CONCLUSION

In this paper, we introduced a novel algorithm for multi-class pattern classification, which is based on the discriminative MspLBP features derived from the proposed Ada-LDA learning algorithm. We proposed the Ada-LDA learning algorithm to select the most discriminative sub-regions, which are represented by MspLBP features, from a large pool of sub-regions generated by shifting and scaling a sub-window over face images. Optimal weak classifiers which compose a final strong classifier are trained by the two alternative LDA methods depending on the singularity of the within-class scatter matrix under the framework of AdaBoost. To verify the feasibility of the proposed approach, two challenging experiments in terms of face detection and facial expression recognition were carried out on the widely used face databases. Face detection on the MIT-CBCL and MIT+CMU test sets and facial expression recognition on the JAFFE face database have shown that MspLBP features can capture more information about image structure than traditional Haar-like features and raw image pixel intensitise, and so are more distinctive, and that the proposed scheme of constructing the final strong classifier with only a small number of discriminative MspLBP features through Ada-LDA yields more discrimination power and better classification performance.

REFERENCES

[1] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: an application to face detection," *Int'l Conf. Computer Vision and Pattern Recognition*, pp. 130-136, 1997.

[2] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23-28, 1998.

[3] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no.1, pp. 39-51, 1998.

[4] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, 1991.

[5] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[6] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative Common Vectors for Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 1, Jan. 2005.

[7] C. P. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," *Int'l Conf. Computer Vision*, pp. 555-562, 1998.

[8] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Int'l Conf. Computer Vision and Pattern Recognition*, pp. 511-518, 2001.

[9] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51-59, 1996.

[10] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.

[11] T. Ahonen , A. Hadid , and M. Pietikainen, "Face recognition with local binary patterns," In *Proc. European Conference on Computer Vision*, pp. 469-481, 2004.

[12] C. Shan and T. Gritti, "Learning discriminative LBP-histogram bins for facial expression recognition," In *Proc. British Machine Vision Conference*, 2008.

[13] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831-836, 1996.

[14] CBCL Face Database #1, MIT Center For Biological and Computation Learning, http://www.ai.mit.edu/projects/cbcl.

[15] M. Lyons and S. Akamatsu, "Coding facial expressions with gabor wavelets," *Int'l Conf. Automatic Face and Gesture Recognition*, pp. 200-205, 1998.