A compact local binary pattern using maximization of mutual information for face analysis

Bongjin Jun, Taewan Kim, Daijin Kim *

Department of Computer Science and Engineering, Pohang University of Science and Technology, San 31, Hyoja-Dong, Nam-Gu, Pohang 790-784, Republic of Korea

ARTICLE INFO

ABSTRACT

Although many variants of local binary patterns (LBP) are widely used for face analysis due to their satisfactory classification performance, they have not yet been proven compact. We propose an effective code selection method that obtain a compact LBP (CLBP) using the maximization of mutual information (MMI) between features and class labels. The derived CLBP is effective because it provides better classification performance with smaller number of codes. We demonstrate the effectiveness of the proposed CLBP by several experiments of face recognition and facial expression recognition. Our experimental results show that the CLBP outperforms other LBP variants such as LBP, ULBP, and MCT in terms of smaller number of codes and better recognition performance.

1. Introduction

Face analysis, including face recognition and facial expression recognition, is a very active research area in computer vision, human computer interaction (HCI), and biometrics. There have been many well-known statistical approaches for face analysis, which include such techniques as principal component analysis (PCA) [1], linear discriminant analysis (LDA) [2], independent component analysis (ICA) [3], and support vector machine (SVM) [4]. However, these methods suffer from the generalization problem due to the unpredictable distribution of the face images in real environment, which might be far different from that of the training face images. To avoid this problem, non-statistical face analysis method using local binary pattern (LBP) has been proposed. It has been proven that the non-statistical face analysis methods outperform the statistical face analysis methods in terms of recognition performance and the robustness to illumination change [8,9].

Initially, LBP was first introduced by Ojala et al. [5], which showed a high discriminative power for texture classification due to its invariance to monotonic gray level changes. Recently, Ojala et al. [6] introduced the uniform local binary pattern (ULBP), which extended their original LBP operator to a circular neighborhood of different radius size and selected a small subset of LBP patterns.

After that, many variants of LBPs have been introduced by many other researchers and applied to many areas such as face detection [7,8], face recognition [9–12], face authentication [13,14], facial expression recognition [15], gate recognition [17], image retrieval [18], and object detection [19].

However, LBP contains many less informative codes. Fig. 1 shows the LBP codes sorted in the order of occurrence rates, which is obtained from more than 64,000 face images. From Fig. 1, we know that most LBP codes are rarely occurred, so that they do not have the discriminative characteristics at all. Moreover, the original LBP is not efficient because it has a fixed size of feature dimension. In this circumstance, it is observed that certain LBP codes exhibiting transitions from 1 to 0 or 0 to 1 in a circularly defined code are at most two, have been occurred frequently (more than 90%) in the natural images. Based on this observation, Ojala et al. [6] proposed the uniform LBP (ULBP) and applied it to face recognition. Lahdenoja et al. [10] proposed the symmetry ULBP which reduces the number of codes in the ULBP using the symmetry level of the code. However, it has not been proven that these patterns are effective in both minimizing the number of codes and reducing the classification error.

There have been some studies in the feature (pixel) selection for face recognition [16,20]. Frank et al. [16] proposed an automatic pixel selection for optimal facial expression recognition based on PCA-based eigenfaces. Choi et al. [20] proposed a pixel selection for optimal face recognition based on LDA-based discriminative positions (pixels), where they used PCA and LDA method for feature extraction, and then found discriminative positions (pixels) in face images using eigenspace. These methods can be

---

* Corresponding author.
E-mail addresses: simple21@postech.ac.kr (B. Jun),
taey16@postech.ac.kr (T. Kim), dkim@postech.ac.kr (D. Kim).

very sensitive to illumination variations because they used the pixel intensity value directly. Moreover, both PCA and LDA methods inherently assume the second order statistics of Gaussian distribution. This prior assumption may not be guaranteed well in the case of real face recognition task. To overcome these problems, we proposed to use the information theoretic feature selection method given below.

Information theoretic feature selection, which utilizes the maximization of mutual information (MMI) between feature and class label, has been widely used for determining the compact features because it guarantees minimal classification error (details are described in Section 3). Battiti [21] proposed mutual information based feature selection (MIFS) and Kwak and Choi [22] proposed an improved version of MIFS (MIFS-U). Both MIFS and MIFS-U determine the compact features one by one using the MMI. Principe [23] and Korkkola [24] used the steepest descent method to find the optimal projection basis vectors using the information theoretic error measures. Qiu and Fang [25] used the MMI to determine the optimal projection basis vectors, and applied these to face and car detection systems.

To find a compact LBP (CLBP) without redundancy, we propose to use the information theoretic feature selection method based on the maximization of mutual information between the codes and the class labels. This approach to code selection iteratively selects the LBP codes which maximize the mutual information with respect to the class label, conditioned to codes previously selected.

This paper is organized as follows. Section 2 describes several non-parametric local kernel-based image representation such as LBP, ULBP, and MCT. Section 3 describes a feature selection method using MMI. Section 4 describes a method of determining the compact LBP (CLBP) codes using the MMI-based feature selection method. Section 5 explains the experimental results to demonstrate the effectiveness of the proposed CLBP in terms of

minimizing the number of codes and minimizing the classification error. Section 6 describes a successful application of the proposed CLBP method in gender recognition. Finally, we draw our conclusion in Section 6.

2. Non-parametric local kernel-based image representation

The original LBP [5] uses a 3 by 3 kernel that summarizes the local structure of an image. At a given pixel position (x_c, y_c), it takes the 3 by 3 neighborhood pixels surrounding of the given pixel and generates a binary 1 if the neighbor of the given pixel has a value greater than or equal to the given pixel or a binary 0 if the neighbor of the given pixel has a value smaller than the given pixel. The decimal form of the resulting 8-bit word (LBP code) can be expressed as

$$LBP_{P,R}(x_c, y_c) = \sum_{n=0}^{7} \delta(i_c - i_n)2^n, \quad (1)$$

where i_c is a pixel value positioned at (x_c, y_c), i_n is one of the eight surrounding pixel values, and a sign function $\delta(\cdot)$ is defined such that

$$\delta(x) = \begin{cases} 1 & \text{if } x < 0, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where the subscripts P and R represent the number of neighboring pixels and the radius in multi-scale LBP, respectively [6]. For an example, $LBP_{8,2}$ denotes the LBP with eight equally spaced pixels on a circle of radius 2.

Fig. 2 illustrates how to obtain the LBP code and the LBP feature vector. First, we transform an original image into the LBP transformed image by using Eq. (1). The LBP code has a certain value from 0 to 255, which is a binary coded decimal value. When we read the binary value in the 3×3 window in the circular manner, it gives a 8-bit binary number (e.g., 11001011) and the LBP code has a value of 203. Second, an original face image is divided into M concatenated blocks, where each block has a size of 8×8 pixels commonly. We compute the histogram of LBP codes within each block, i.e., count the number of pixels with a certain LBP code. So, each block provides a 256 dimensional LBP histogram. Then, we concatenate the LBP histograms of M blocks, which is the LBP feature vector that represents the input image.

Liao et al. [12] proposed the multi-scale block LBP (MB-LBP). It captures an n by n block-based local structure rather than pixel-based local structure, so that it is less sensitive to noisy information.

Ojala et al. [6] observed that the natural images generally contain a small number of LBP codes, which are called the uniform LBP (denoted by superscript LBP^{u2}). ULBP contains a maximum of
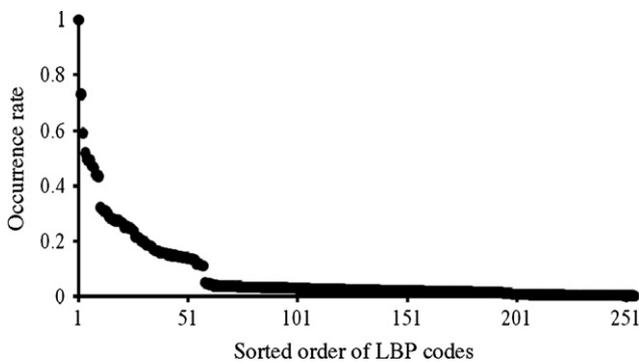
Fig. 1. The LBP codes sorted in the order of occurrence rates.
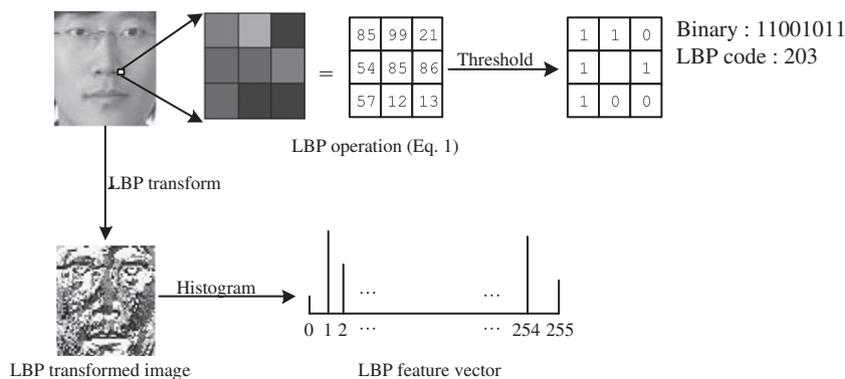
Fig. 2. An illustration of LBP operation, LBP code, LBP transform and LBP feature vector.

two bitwise transitions: 0–1 or 1–0. These uniform patterns represent the majority of microstructures such as lines, edges, and corners.

Lahdenoja et al. [10] proposed a method for reducing the number of the codes in the ULBP by using the level of symmetry L_{sym} of ULBP, which is expressed as

$$L_{sym} = \min\left[\sum_{i=1}^{P} B(i), \sum_{i=1}^{P} \overline{B}(i)\right], \tag{3}$$

where the first and second terms are the number of neighboring pixels with a binary value 1 and 0 in ULBP, respectively. They observed that the codes with high level of L_{sym} are more discriminative than those with low level of L_{sym} by qualitative visual inspection.

Zabih et al. [26] proposed the census transform (CT) that summarizes the local image structure as a bit string, where it is 0 if the intensity value at a position in one image is less than the intensity value at the corresponding position in another image. This census transform has been extended to the modified census transform (MCT) [27] as

$$\Gamma_{MCT}(x_c, y_c) = \sum_{n=0}^{8} \delta(i_n - \bar{i}_c)2^n, \tag{4}$$

where $\bar{i}_c$ denotes the mean of pixel values in a 3 by 3 local kernel positioned at (x_c, y_c), and i_n is one of the nine pixel values in the local kernel. The function $\delta(\cdot)$ is the same as Eq. (2). MCT can be referred to as an enlarged version of the original LBP, which means one pixel in the image is represented by 9 bit length. Hence, there are 512 codes in MCT and 256 codes in the LBP.

Fig. 3 illustrates that the LBP, ULBP, and MCT provide the transformed output images that are invariant to monotonic gray level changes.

3. Feature selection using MMI

Maximization of mutual information (MMI) is a powerful way of selecting the optimal features that simultaneously minimizes both the lower and upper bound of the Bayes error. Let $\mathbf{X} \in \mathbb{N}^{D \times N}$ be a data matrix, where D is the number of images and N is the feature size of each image. Let $F = \{f_1, \ldots, f_N\}$ and C be a discrete valued random variable for representing features and class labels, respectively. Fig. 4 illustrates a typical example of the data matrix $\mathbf{X}$, where $\mathbf{T}$ is a set of the ordered pairs of feature vector and class label vector as $\mathbf{T} = \{(f_1, C), \ldots, (f_N, C)\}$.

Fig. 3. Robustness to monotonic gray level changes: (a) original image, (b) LBP, (c) ULBP, and (d) MCT.
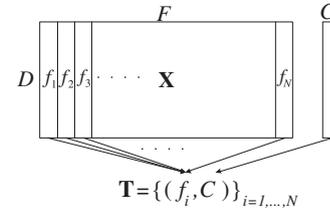
Fig. 4. A typical example of representing data matrix $\mathbf{X}$.

Let a function be $G(F) = \hat{C}$, where $\hat{C}$ is an estimate of C and the C has the class labels as $v = \{1, \ldots, N_c\}$, where N_c is a total number of classes. Then, the lower and upper bounds of Bayes error probability $P_e \triangleq P(\hat{C} \neq C)$ are proven by Fano [28] and Hellman and Raviv [29] as

$$\frac{H(C) - I(C;F) - 1}{\log|v|} \leq P_e \leq \frac{1}{2}(H(C) - I(C;F)), \tag{5}$$

where $H(\cdot)$ denotes entropy, $I(\cdot;\cdot)$ denotes mutual information, and $|v|$ is the number of classes. The Bayes error probability P_e can be directly reduced by maximizing mutual information. From Eq. (5), we know that the maximization of mutual information (MMI) is equivalent to the minimization of the Bayes error probability. Also, we know that the maximization of mutual information (MMI) is equivalent to the minimization of the conditional entropy $H(C|F)$, since $I(C;F) = H(C) - H(C|F)$. Therefore, the optimal feature selection problem can be formulated to select the best feature whose mutual information is largest or Bayes error is smallest as

$$\operatorname*{argmax}_{f_i \in F} I(C; f_i) \quad \text{or} \quad \operatorname*{argmin}_{f_i \in F} H(C|f_i). \tag{6}$$

In practical applications, we usually take the best $k \ll N$ features whose joint mutual information is largest or joint Bayes error is smallest as

$$\operatorname*{argmax}_{F_k \subseteq F} I(C; F_k) \quad \text{or} \quad \operatorname*{argmin}_{F_k \subseteq F} H(C|F_k), \tag{7}$$

where F_k is a feature subset of F whose number of elements is k. Thus, the feature subset F_k can be represented as $F_k = \{f'_1, f'_2, \ldots, f'_k\}$, where f'_i is the ith selected feature.

However, it is impossible to compute the joint mutual information equation (7) because all possible combinations of feature sets is huge (the exact number of possible subsets of selected features is $N!/(N-k)!k!$). To overcome this problem, many researchers have tried to find approximated solutions. Battiti [21] proposed an iterative greedy feature selection strategy called 'mutual information feature selection (MIFS)'. He used the greedy feature selection criterion as

$$\operatorname*{argmax}_{f_i \in F} \left[I(C; f_i) - \beta \sum_{f_s \in S} I(f_s; f_i) \right], \tag{8}$$

where f_i is a candidate feature in the feature set F, f_s is a previously selected feature, S is a set of the previously selected features, and β is a regularization parameter that adjusts the amount of redundance between the candidate features f_i and previously selected features f_s. Table 1 shows a typical algorithm of the MIFS method.

Kwak et al. [22] proposed the MIFS-U method that modifies the selection criterion as

$$\operatorname*{argmax}_{f_i \in F} \left[I(C; f_i) - \beta \sum_{f_s \in S} \frac{I(C; f_s)}{H(f_s)} I(f_s; f_i) \right]. \tag{9}$$
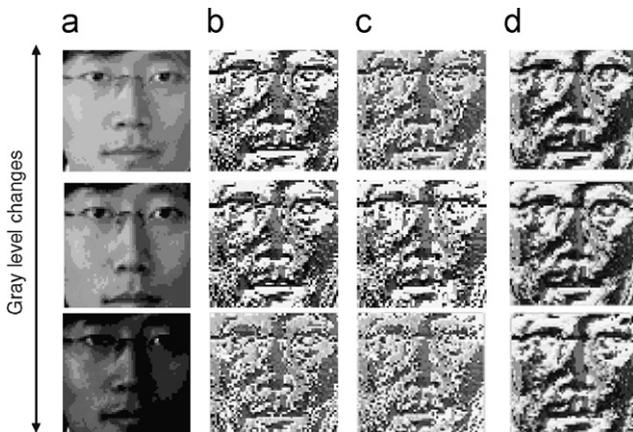
Table 1
Battiti's MIFS method.

---

1. **Initialization**
   (1) $S \leftarrow \phi$.
2. **Computation of the mutual information**
   (1) $\text{argmax}_{f_i \in F} I(C; f_i)$.
3. **Selection of the first feature**
   (1) $F \leftarrow F \backslash \{f_i\}$, $S \leftarrow \{f_i\}$.
4. **Greedy selection**
   repeat until $|S| = k$
      (1) For all pairs (f_i, f_s) with $f_i \in F$ and $f_s \in S$,
         select the feature $f_i \in F$ using Eq. (8).
      (2) Set $F \leftarrow F \backslash \{f_i\}$, and $S \leftarrow \{f_i\}$.
   end

Peng et al. [30] proposed a modified MIFS method that uses the max-relevance and min-redundancy (mRMR) criterion as

$$\underset{f_i \in F}{\text{argmax}} \left[I(C; f_i) - \frac{1}{|S|} \sum_{f_s \in S} I(f_s; f_i) \right], \qquad (10)$$

where $|S|$ is the cardinality of the set S.

Estevez et al. [31] proposed another modified MIFS method that uses the selection criterion as

$$\underset{f_i \in F}{\text{argmax}} \left[I(C; f_i) - \frac{1}{|S|} \sum_{f_s \in S} NI(f_i; f_s) \right], \qquad (11)$$

where NI is the normalized mutual information such that

$$NI(f_i; f_s) = \frac{I(f_i; f_s)}{\min\{H(f_i), H(f_s)\}}. \qquad (12)$$

4. Compact LBP code selection using MMI

We applied the MMI-based feature selection method to find the CLBP code. In this work, we took Peng's modified MIFS method that used the max-relevance and min-redundancy (mRMR) as the selection criterion because it showed outstanding classification performance in comparison to the other iterative feature selection methods [32,33]. According to Peng's modified MIFS method, we can regard the LBP codes and the class label as two discrete random variables F and C, respectively. So, we can easily compute the joint distribution $p(C,F)$, and each distribution of them $p(C)$ and $p(F)$ from the histogram (occurrence) based density estimation. Subsequently, the mutual information can be computed as

$$I(C; f) = \sum_{f \in F} \sum_{c \in C} p(c,f) \log \left(\frac{p(c,f)}{p(c)p(f)} \right). \qquad (13)$$

Fig. 5 shows the overall procedure of the proposed MMI-based CLBP code selection method, which consists of three consecutive stages given below.

4.1. Stage I: MMI-based feature reduction

Suppose that we have a set of D training images with N_c classes. Each of the images has a size of $N = h \times w$. All training images are transformed into LBP features. Then, we have a LBP feature matrix with a size of $D \times N$, $F_{LBP} = \{f_1, f_2, \ldots, f_N\}$, where f_i is a D dimensional LBP feature vector at the ith pixel position.

We compute the mutual information $I(C; f_i)$ between the class label C and the feature vector f_i for $i = 1, 2, \ldots, N$ and obtain the selected feature index set $S_{LBP} = \{p_1, p_2, \ldots, p_M\}$ by applying the maximization of the mutual information given in Eq. (10) sequentially, where M is the number of selected LBP feature vectors and the p_i denotes the index of the selected LBP feature
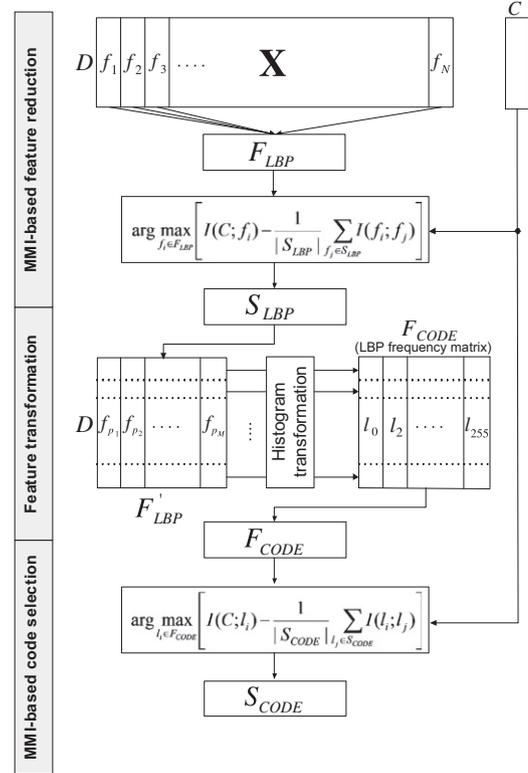
Fig. 5. Overall procedure of the proposed MMI-based CLBP code selection method.

vector at the ith iteration. Thus, we have a reduced LBP feature matrix with a size of $D \times M$, $F'_{LBP} = \{f_{p_1}, f_{p_2}, \ldots, f_{p_M}\}$, where f_{p_i} is a D dimensional LBP feature vector at the p_ith pixel position. Because $M \ll N$, we reduce the dimensionality of the original LBP feature matrix while removing less discriminative features greatly.

Fig. 6 shows two examples of the feature dimensionality reduction using the MMI-based feature reduction method on the POSTECH Face 2007 database (PF07) [34], where the horizontal axis denotes the number of selected features and the white pixels denote the selected features. As you can see, some pixels corresponding to eyes, eyebrows, and mouth are selected due to their discriminative characteristics.

4.2. Stage II: feature transformation

Each row of the reduced LBP feature matrix F'_{LBP} is a LBP transformed training image with the reduced size of M. Each dimensionality reduced training image is transformed into a histogram of LBP (we call it histogram transformation). Then, we have a LBP code frequency matrix with a size of $D \times 256$, $F_{CODE} = \{l_0, l_1, \ldots, l_{255}\}$, where l_i is a D dimensional LBP code frequency vector at a specific LBP code i. Hence, the jth column of l_i has the number of pixels whose LBP code is i in the jth training image.

4.3. Stage III: MMI-based code selection

We again compute the mutual information $I(C; l_i)$ between the class label C and the LBP code frequency vector l_i for $i = 0, 1, \ldots, 255$ and then obtain the selected LBP code frequency set $S_{CODE} = \{c_1, c_2, \ldots, c_K\}$ using the maximization of the mutual information given in Eq. (10), where K is the number of the selected LBP code frequency vectors and the c_i is the index of the selected LBP frequency vector at the ith iteration. Therefore, S_{CODE} is CLBP codes we want to find.

Fig. 7 illustrates how to obtain the CLBP transformed image. First, we transform an original face image into the LBP transform image
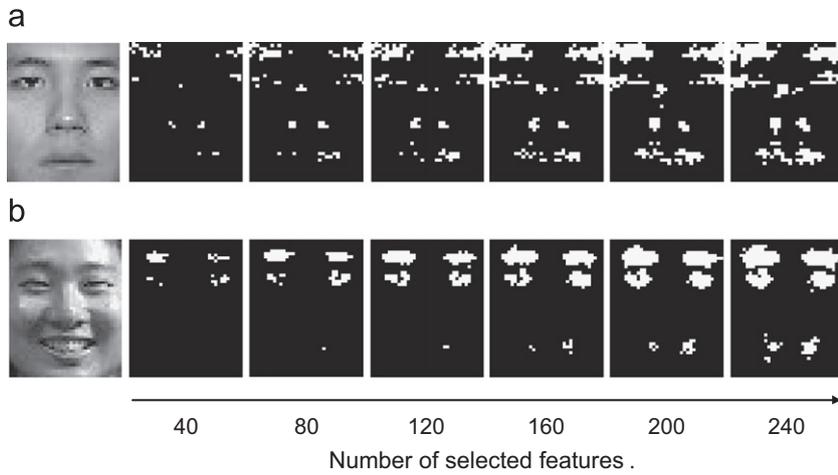
Fig. 6. Examples of feature dimensionality reduction.
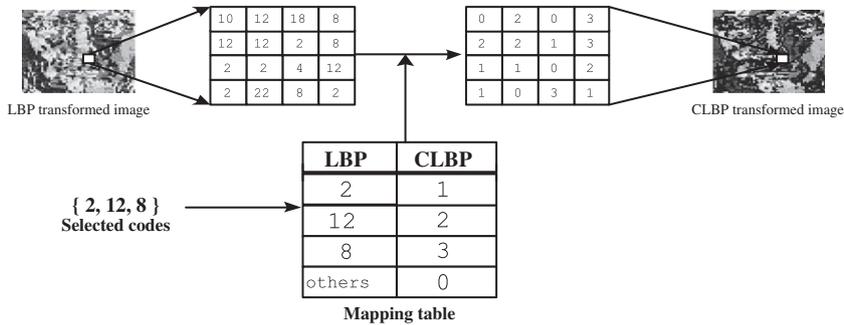
Fig. 7. An illustration of CLBP transformation.
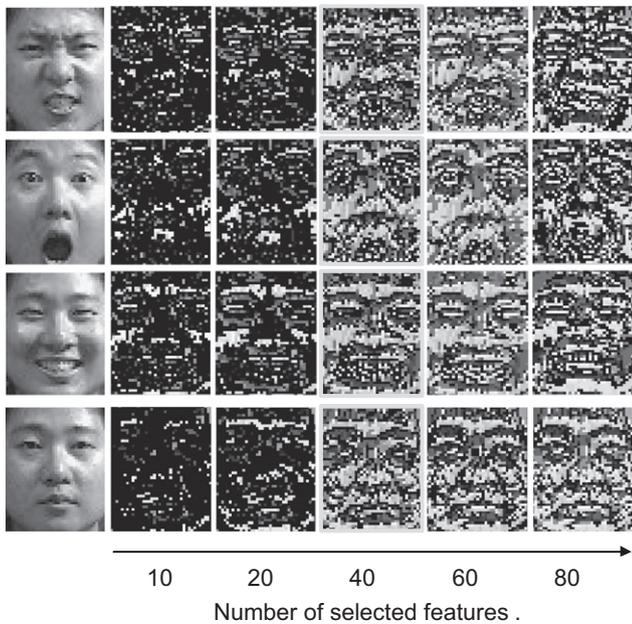
Fig. 8. Some examples of the CLBP transformed images.

using Eq. (1) (see Fig. 2 for details). Second, we prepare the mapping table that relates the ULBP code with CLBP code, which is the result of the stage III in the proposed MMI-based CLBP code selection method. All other LBP codes that are not selected in the CLBP, are set to a pre-specified value (e.g., set to 0). In this illustration, the LBP code values 2, 12 and 8 correspond to the CLBP code values 1, 2 and 3,

respectively, and all other LBP code values correspond to the CLBP code value 0. Finally, we transform the LBP image into the CLBP image using the mapping table by a simple correspondence.

Fig. 8 shows examples of four CLBP transformed images using the CLBP codes that are obtained from the MMI-based CLBP code selection method, where the horizontal axis denotes the number of selected features. From Fig. 8, we know that the transformed images represent many local structures in greater detail up to 40 features but they do not show a significant improvement of representing local structures after 40 features. From this, we guess that there exists an appropriate number of LBP codes that balance the number of codes and the classification error.

5. Experimental results and discussion

To validate the effectiveness of the proposed MMI-based CLBP code selection method, we performed two kinds of experiments: face recognition and facial expression recognition. In these experiments, all input images were transformed using the CLBP codes obtained by the proposed MMI-based CLBP code selection method. We did not use any preprocessing methods or postprocessing methods to show that effectiveness of only the proposed CLBP code selection method in terms of a number of codes and classification performance.

5.1. Database

For the experiments, we used the PF07 database[1] [34] due to its rich variants and publicly availability, which consists of 100 male

---

Fig. 9. Example images of PF07.

and 100 female subjects, with 320 images per subject. The 320 images all have a size of 40×50 pixels that consist of 16 illuminations, five poses and four facial expressions. Fig. 9 shows some example images from the PF07 database. All of these images were aligned according to the two eye locations, which were obtained manually. This database used for our experiment without any preprocessing or manual labeling.

5.2. Face recognition

For doing the face recognition experiments, we prepared the databases for training, gallery, and probe set as follows. For a strict performance evaluation, we separated 200 subjects into two disjoint groups: GROUP1 and GROUP2. The training set consisted of 32,000 images (= 100 GROUP1 subjects $\times$ 16 different illuminations $\times$ 5 different poses $\times$4 different facial expressions). A gallery set consisted of 100 images of GROUP2 subjects under normal illumination, frontal pose, and neutral expression. The probe set consisted of three different types of images: TYPE1 probe set consisted of 1500 images (= 100 GROUP2 subjects $\times$ 15 non-normal illuminations) under frontal pose and neutral expression. TYPE2 probe set consisted of 400 images (= 100 GROUP2 subjects $\times$ 4 non-frontal poses) under normal illumination and neutral expression. TYPE3 probe set consisted of 300 images (= 100 GROUP2 subjects $\times$3 non-neutral facial expressions) under normal illumination and frontal pose. Hence, we performed three different experiments with three different types of probe sets and the gallery set.

Fig. 10 illustrates the overall process of the face recognition experiment, which consists of three stages.

1. Training stage: We obtained the LBP transformed images by converting the training images with LBP, and determined the K CLBP codes by using the MMI-based CLBP code selection method described in the Section 4. In this experiment, we empirically reduced the dimensionality of the training images from 2000 to 240 in the MMI-based feature reduction stage. The 240 features

included discriminative facial components such as eyes, eyebrows, mouth, and nose (see Fig. 4).

2. Enrollment stage: We transformed the gallery images into CLBP transformed gallery images with the CLBP codes which are obtained in the training stage. Then, we converted the CLBP transformed gallery images into a set of gallery feature vectors using the spatially enhanced histogram method [11], which is explained as follows: (1) each CLBP transformed gallery image was divided into 30 local regions $R_i, i = 1, \ldots, 30$ (We empirically divided the image into 30 regions.), (2) the CLBP histograms of these 30 local regions were computed independently in each CLBP transformed gallery image, and (3) the gallery feature vector was obtained by sequentially concatenating the CLBP histograms of 30 local regions.

3. Recognition stage: By using the CLBP codes, we transformed the probe image into the CLBP transformed probe image. Then, we converted the CLBP transformed probe image into a probe feature vector computed by the same method used in the enrollment stage. Then, we measured the weighted χ^2 distance between the probe feature vector ($F_{P_{k,l}}$) and the gallery feature vector ($F_{G_{k,l}}^i, i = 1, \ldots, 100$) to find the best matched gallery image as

$$\forall i \; \underset{i}{\arg\min}[\chi_w^2(F_{P_{k,l}}, F_{G_{k,l}}^i)], \qquad (14)$$

where the weighted χ^2 distance is computed as

$$\chi_w^2(F_{P_{k,l}}, F_{G_{k,l}}^i) = \sum_{k,l} w_k \frac{(F_{P_{k,l}} - F_{G_{k,l}}^i)^2}{(F_{P_{k,l}} + F_{G_{k,l}}^i)}, \qquad (15)$$

where the indices k and l refer to the lth bin in the histogram within the kth local region, and w_k is the weight that represent the importance of the kth block for face recognition. The value of w_k is set empirically. Initially, we set all weights w_k to the same value ($\forall k, w_k > 0$). Then, we obtain the value of w_1 by finding the value that provides the best recognition accuracy while changing w_1 except the other weights and set the optimal value of w_1 to the found value. Then, we obtain the value of w_2 by finding the value that provides the best recognition accuracy while changing w_2 except the other weights (w_1 is fixed) and set the optimal value of w_2 to the second found value. We continue this process until all weight values are found.

The classification error for the face recognition could be computed such that

$$\text{Classification error (\%)} = \left(1 - \frac{N_{\text{CRI}}}{N_{\text{T}}}\right) \times 100, \qquad (16)$$

where N_{CRI} and N_{T} are the number of correctly recognized images and the number of test images, respectively.

To prove the robustness of the proposed CLBP code to the illumination changes, we performed the face recognition experiment using the TYPE1 probe set and the gallery set. Fig. 11 compares the classification errors of face recognition among four different representation methods such as CLBP, ULBP, LBP, and MCT using the TYPE1 probe set, when the number of CLBP codes changes. The horizontal axis denotes the number of CLBP codes. From Fig. 11, we noticed that the classification error decreases drastically as the number of CLBP codes increases up to 23, and does not change much as the number of CLBP codes increases from 23 to 59. From the viewpoint of the number of codes and the classification error, the compact number of CLBP code was 31. The classification error using 31 CLBP codes was 16.00%, which is the smallest than those of other representation methods such as ULBP (18.73%), LBP (18.07%), and MCT (18.27%).

To prove the robustness of the proposed CLBP code to the pose changes, we performed the face recognition experiment using the
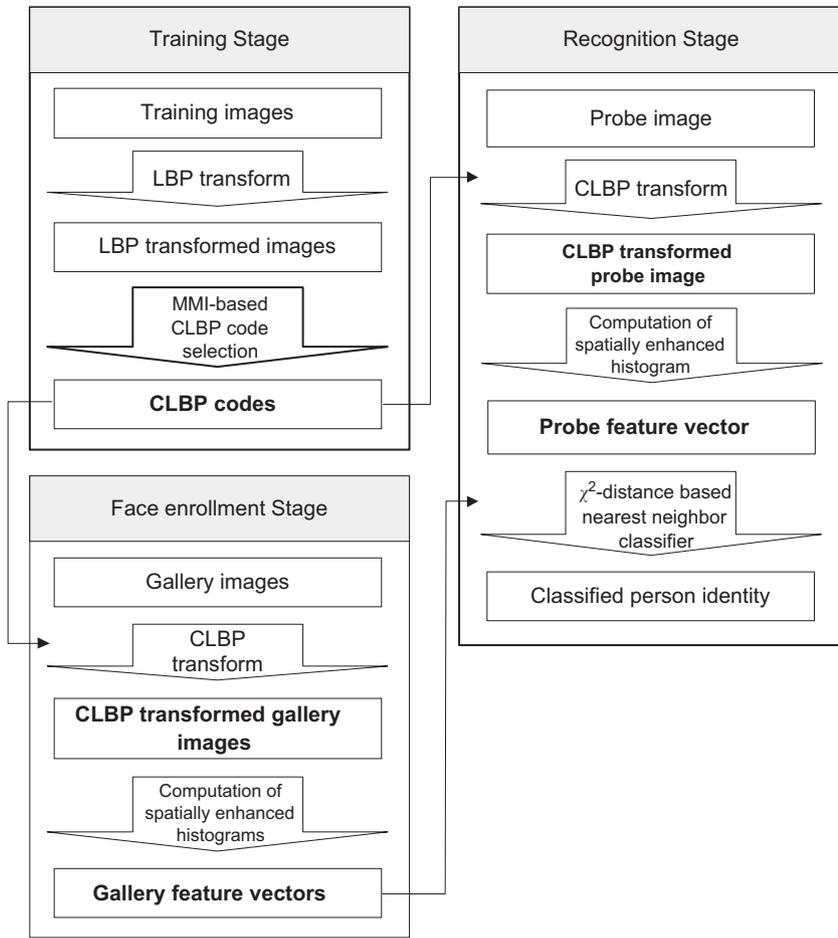
Fig. 10. Overall process of the face recognition.
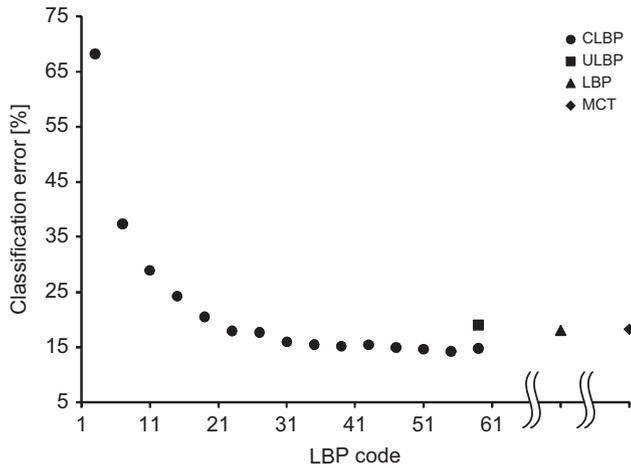
Fig. 11. Comparison of classification errors among four different representation methods using the TYPE1 probe set.
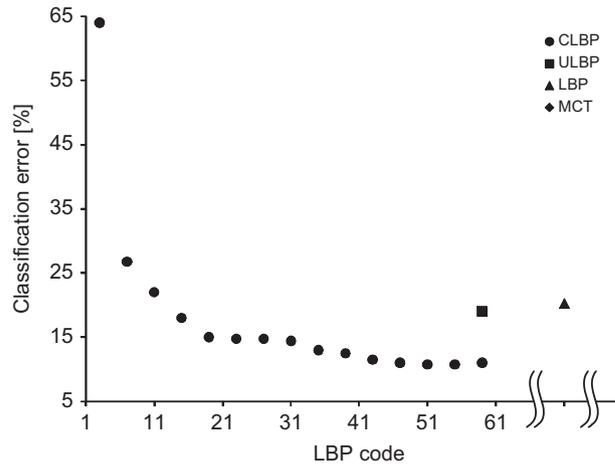
Fig. 12. Comparison of classification errors among four different representation methods using the TYPE2 probe set.

TYPE2 probe set and the gallery set. Fig. 12 compares the classification errors of face recognition among four different representation methods such as CLBP, ULBP, LBP, and MCT using the TYPE2 probe set, when the number of CLBP codes changes. The horizontal axis denotes the number of CLBP codes. From Fig. 12, we noticed that the classification error decreases drastically as the number of CLBP codes increases up to 19, and does not change much as the number of CLBP codes increases from 19 to 59. From

the viewpoint of the number of codes and the classification error, the appropriate number of CLBP code was 27. The classification error using 27 CLBP codes was 14.75%, which is the smallest than those of other representation methods such as ULBP (19.00%), LBP (20.25%), and MCT (20.00%).

To prove the robustness of the proposed CLBP code to the facial expression changes, we performed the face recognition experiments using the TYPE3 probe set and the gallery set. Fig. 13
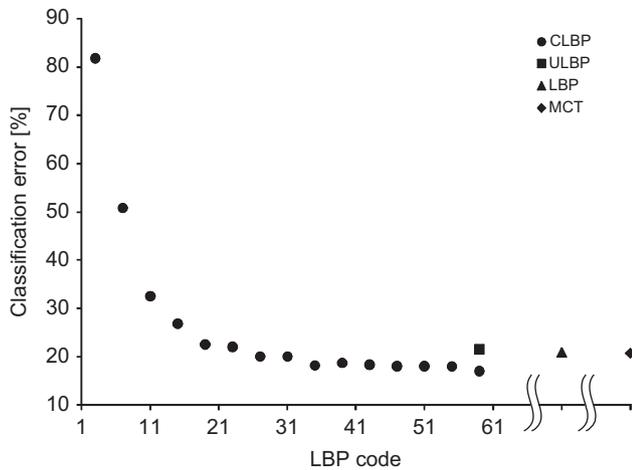
Fig. 13. Comparison of classification errors among four different representation methods using the TYPE3 probe set.

Table 2
Comparison of the classification error and the size of code vectors using PF07 database.

DB	TYPE1		TYPE2		TYPE3	
Method	CLBP	SLBP	CLBP	SLBP	CLBP	SLBP
Classification error (%)	16.00	17.90	14.75	15.18	18.16	19.88
# of codes	31	25	27	25	35	25

compares the classification errors of face recognition among four different representation methods such as CLBP, ULBP, LBP, and MCT using the TYPE3 probe set, when the number of CLBP codes changes. The horizontal axis denotes the number of CLBP codes. From Fig. 13, we noticed that the classification error decreases drastically as the number of CLBP codes increases up to 19, and does not change much as the number of CLBP codes increases from 19 to 59. From the viewpoint of the number of codes and the classification error, the appropriate number of CLBP code was 35. The classification error using 35 CLBP codes was 18.16%, which is the smallest than those of other representation methods such as ULBP (21.50%), LBP (20.83%), and MCT (20.72%).

Furthermore, we compared the classification error between the proposed CLBP and the symmetry ULBP (SLBP) method [10] using TYPE1, TYPE2, and TYPE3 probe sets. In this experiment, the SLBP has 25 codes (24 ULBP codes and 1 dummy code) since we take the symmetry levels 3 and 4 in Eq. (3). Table 2 shows that the error rate of the proposed CLBP code is slight smaller than that of the SLBP ($L_{sym3,4}$) in the cost of slightly larger code size. This improvement is obtained from the compactness of the proposed CLBP that is derived from Eq. (5), which is a direct minimization criterion of Bayes error rate. However, the SLBP ($L_{sym3,4}$) is not justified by any theoretical principle but is only originated from the observation of human's visual inspection.

Finally, we compared the classification error of many different face recognition methods using the well-known FERET [35] face database. In this experiment, we take two different data sets: Fa image set (=1195 randomly selected face image) for a gallery set and Fb image set (=1195 different images corresponding to the selected gallery images but alternative facial expression) for a probe set. Fa image set contains the frontal images and Fb image set contains alternative facial expression than in Fa image set. All of these images are normalized by two eye points. Table 3 compares the classification error and the size of code vectors among many

Table 3
Comparison of the classification error and the size of code vectors using FERET database.

Method	CLBP	SLBP	LBP	PCA	EBGM	LDA
Classification error (%)	3.00	6.00	3.00	15.00	10.00	27.00
# of codes	27	25	256	NA	NA	NA
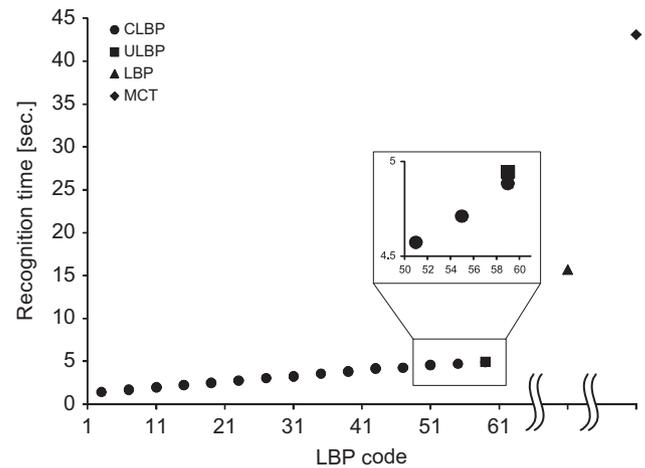
Fig. 14. Comparison of computation times for face recognition among four different representation methods using the TYPE1 probe set.

other methods. In this table, the classification error rates using PCA, elastic bunch graph matching (EGBM) [36] and LDA method were referred from Lahdenoja's paper [10], which were used as the baseline algorithms. From Table 3, we know that the proposed CLBP is optimal because the classification error is smallest and the size of code vectors is the second smallest. The original LBP shows the smallest classification error but relatively a large size of code vectors and the SLBP shows the smallest size of code vectors but a larger classification error.

The proposed CLBP code also reduced the computation time for face recognition. In Ahonen's face recognition method [11], the χ^2 distance based matching equation (15) is the most time-consuming part in the computation time for face recognition. The time complexity of Eq. (15) is represented by a linear function $\mathcal{O}(v)$, where the parameter v is the multiplication of the number of local regions (k) and the number of bins (l) per region.

Fig. 14 compares the computation times for face recognition among four different representation methods such as CLBP, ULBP, LBP, and MCT using the TYPE1 probe set, when the number of CLBP codes changes. The horizontal axis denotes the number of CLBP codes. From Fig. 14, we noticed that the computation time increases almost linearly as the number of CLBP codes increases. The recognition time using 31 CLBP codes was 3.26 s, which is significantly less than those of other representation methods such as ULBP (4.95 s), LBP (15.65 s), and MCT (43.07 s). Therefore, the use of 31 CLBP codes enables us to reduce computation times approximately 1.5 times, 5 times, and 13 times, which were achieved with ULBP, LBP, and MCT, respectively.

5.3. Facial expression recognition

For doing the facial expression recognition experiments, we prepared the training and test image set as follows. First, we prepared 800 randomly selected images (=200 persons × a normal illumination × 4 different facial expressions × a frontal pose). From 800 images, we selected 700 images (=175 persons × a normal

illumination × 4 facial expressions × a frontal pose) for the training image set and selected the remaining 100 images (= the remaining 25 persons × a normal illumination × 4 facial expressions × a frontal pose) for the test image set. We performed the eightfold cross validation method to avoid the data tweak problem.

Fig. 15 illustrates the overall process of the facial expression recognition experiment, which consists of two stages.

1. Training stage: We obtained the LBP transformed images by converting the training images with LBP and found the K CLBP codes using the MMI-based CLBP code selection method described in the Section 4. In this experiment, we empirically reduced the dimensionality of the training images from 2000 to 240 in the feature reduction stage. These 240 features included discriminative facial components such as eyes, eyebrows, and mouth (see Fig. 4). We converted the LBP transformed training images into the CLBP transformed training images by using the obtained K CLBP codes. Then, we converted the CLBP transformed training images into a set of training feature vectors using the spatially enhanced histogram method [11], which is the same as the method described in Section 5.2. Finally, we obtained four person-independent feature vectors, one per facial expression, by averaging all training feature vectors with a specific facial expression [15].

2. Recognition stage: We transformed the input image into the CLBP transformed input images by the K CLBP codes. We converted the CLBP transformed training images into input

feature vectors using the spatially enhanced histogram method [11], which is the same as the method described in the Section 5.2. Finally, we measured the χ^2 distance between the input feature vector (F_I) and the person-independent feature vectors of four facial expressions ($F_e, e \in \{$neutral, happy, surprised, angry$\}$) to find the best matched facial expression such that

$$\forall e \; \underset{e}{\mathrm{argmin}}[\chi^2(F_I, F_e)], \tag{17}$$

where the χ^2 distance is computed as

$$\chi^2(F_I, F_e) = \sum_l \frac{(F_I(l) - F_e(l))^2}{(F_I(l) + F_e(l))}, \tag{18}$$

where l is the component index of the feature vectors F_I and F_e.

To achieve reliable facial expression recognition performance, we performed the eightfold cross-validation method. In the ith trial ($i = 1, \dots, 8$), we had 100 test input images, 25 images for each facial expression and we counted the number of correctly recognized images in facial expression. Then, the classification error for the facial expression recognition could be computed such that

$$\text{Classification error } (\%) = \left(1 - \frac{1}{8}\sum_{i=1}^{8} \frac{N_{\mathrm{CRI}_i}}{N_{\mathrm{T}_i}}\right) \times 100, \tag{19}$$

where N_{CRI_i} and N_{T_i} is the number of correctly recognized images and the number of test images, respectively, at the ith trial. Fig. 16 compares the classification errors of facial expression recognition
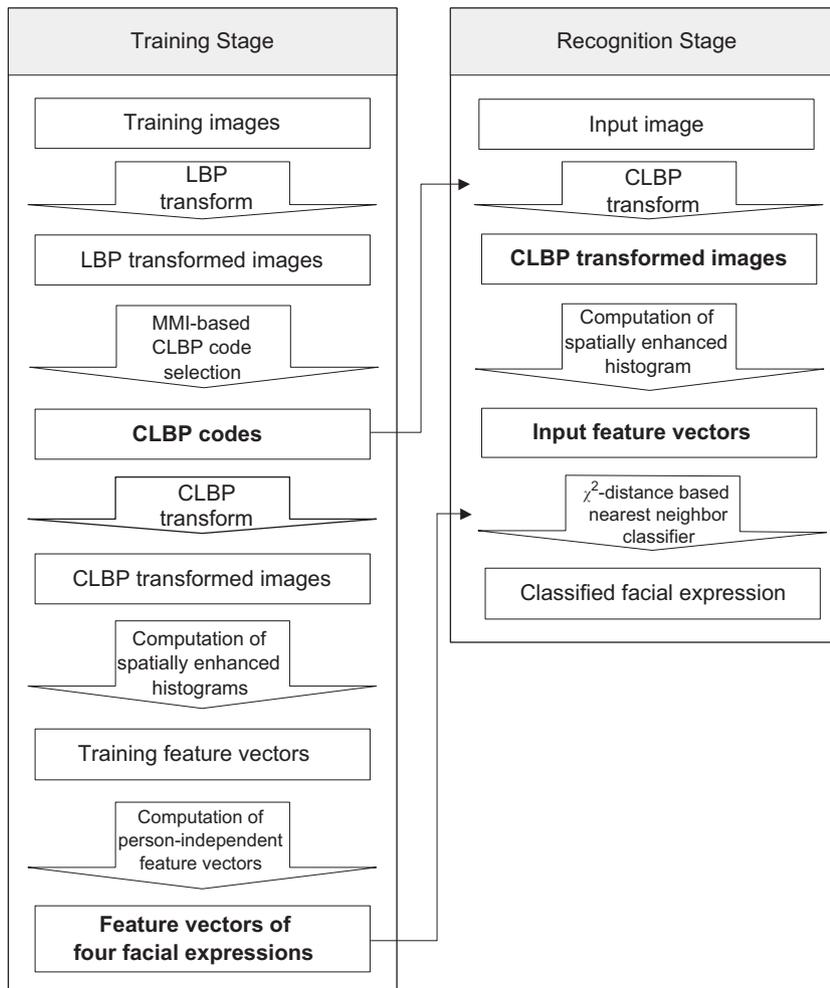
Fig. 15. Overall process of the facial expression recognition.

among four different representation methods such as CLBP, ULBP, LBP, and MCT when the number of CLBP codes changes. The horizontal axis denotes the number of CLBP codes. From Fig. 16, we noticed that the classification error decreases drastically as the number of CLBP codes increases up to 23, and does not change much as the number of CLBP codes increases from 23 to 80. From the viewpoint of the number of codes and the classification error, the appropriate number of CLBP code was 23. The classification error using 23 CLBP codes was 8.00%, which is the smallest than those of other representation methods such as ULBP (12.73%), LBP (14.20%), and MCT (15.07%).

In this facial expression recognition, the time complexity is the same as the face recognition experiments because the χ^2 distance is used for finding correct matching. Fig. 17 compares the computation times for face recognition among four different representation methods such as CLBP, ULBP, LBP, and MCT when the number of CLBP codes changes. The horizontal axis denotes the number of CLBP codes. From Fig. 17, we noticed that the computation time increases almost linearly as the number of CLBP codes increases. The recognition time using 23 CLBP codes was 0.5928 s, which is the smallest than those of other representation methods such as ULBP (0.8892 s), LBP (2.3088 s), and MCT (4.3056 s). Hence, the use of 23 CLBP codes enables us to reduce computation times approximately 1.5 times, 4 times, and
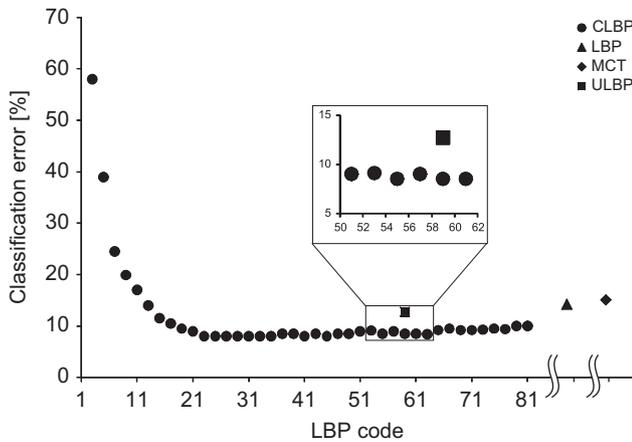
Fig. 16. Comparison of classification errors for facial expression recognition among four different representation methods.
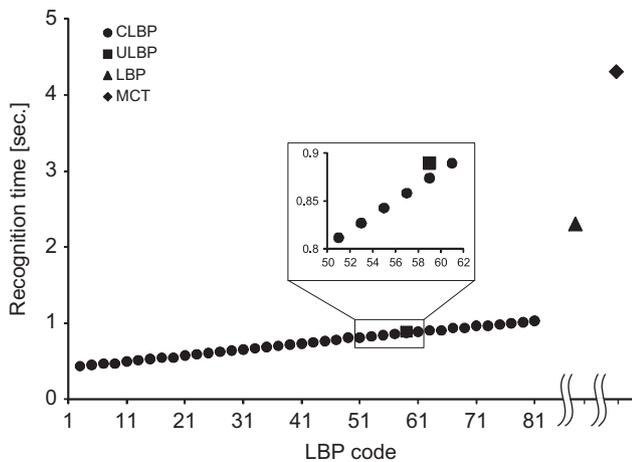
Fig. 17. Comparison of computation times for facial expression recognition among four different representation methods.
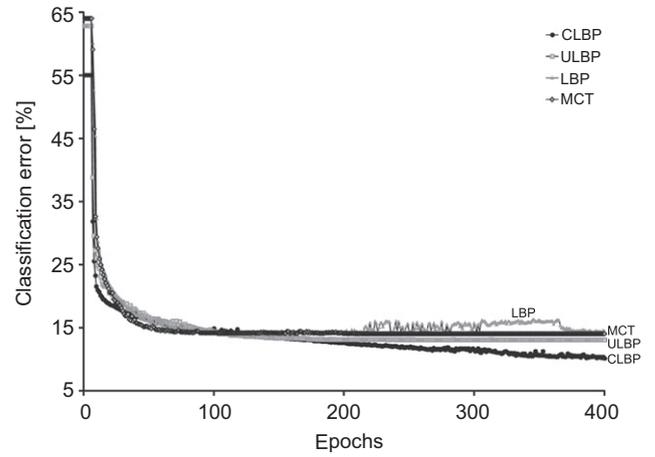
Fig. 18. Comparison of classification errors for gender recognition among four different representation methods.

7.2 times, which were achieved with ULBP, LBP, and MCT, respectively.

6. Real-time gender recognition

To validate the general applicability of the proposed CLBP method, we implemented a real-time gender recognition system. In this experiment, we used 28,800 randomly selected images (=7200 male images on the web+7200 female images on the web+7200 male images on the PF07+7200 female images on the PF07). Among 28,800 images, we selected 14,400 images (=3600 male images on the web+3600 male images on the PF07+3600 female images on the web+3600 female images on the PF07) for the training image set, and selected the remaining 14,400 images for the test image set. We empirically selected 25 CLBP codes and we divided each face image into 20 blocks. This induces 500 dimensional CLBP feature vector per one face image. We took Hinton's deep belief network which is pre-trained by restrict Boltzman machine [37] as a male/female classifier.

Fig. 18 compares the classification errors of gender recognition among four different representation methods such as OLBP, ULBP, LBP, and MCT. The horizontal axis denotes the number of epoches of training Hinton's deep belief network. From Fig. 18, we know that the classification error using 24 CLBP codes was 10.09% which is the smallest than those of other representation methods such as ULBP(13.02%), LBP(14.23%), and MCT(14.00%). Moreover, 25 OLBP codes allowed to reduce dimension of feature vector as 500 which is the smallest size of dimension than other representation methods (ULBP: 1180 dimension, LBP: 5120 dimension, MCT: 10,240 dimension).

Fig. 19 shows several screen-captured images of real-time gender recognition system. All of these images are totally different from the training image set. The resulting images explicitly showed that the real-time gender recognition system is almost invariant to person-identity and scale. It also could recognize genders with slight pose changes ($\pm 20^\circ$). You can see the gender classification result using the proposed CLBP method in the video clip: http://imlab.postech.ac.kr/video/lipreading/demo/gender.wmv.

7. Conclusion

This paper proposes the MMI-based code selection method for the compact LBP, which simultaneously improves recognition

Fig. 19. Classification results of the real-time gender recognition system.

performance and reduces recognition time. Because the maximization of mutual information (MMI) between feature and class label assures the minimal classification error, we selected the codes of the compact LBP iteratively, in the order of mutual information per code.

The proposed CLBP code selection method consisted of three stages: MMI-based feature reduction, feature transformation, and MMI-based code selection. In the first stage, we reduced the dimensionality of training images by using mutual information. The selected LBP feature vectors were the most discriminative than any other remaining features. In the second stage, the dimensionality reduced training images were transformed into histograms of LBP by the histogram transformation. Then, we had a LBP code frequency matrix. Each column vector of the matrix represented a feature vector of corresponding LBP code. In the last stage, we selected several LBP code frequency vectors using MMI-based code selection, allowing to obtain the indices of the selected LBP frequency vectors.

To validate the effectiveness of the MMI-based CLBP code selection method, we applied it to two applications: face recognition and facial expression recognition. First, in face recognition experiments, we used three different probe sets, TYPE1, TYPE2, and TYPE3. In the first experiment using TYPE1 probe set, the appropriate number of CLBP codes was 31. The classification error using 31 CLBP codes was 16.00%, which is the smallest than those of other representation methods such as ULBP (18.73%), LBP (18.07%), and MCT (18.27%). In this experiment, we showed that the use of 31 CLBP codes enables us to reduce computation times approximately 1.5 times, 5 times, and 13 times, which were achieved with ULBP, LBP, and MCT, respectively. In the second experiment using TYPE2 probe set, the appropriate number of CLBP codes was 27.

The classification error using 27 CLBP codes was 14.75%, which is the smallest than those of other representation methods such as ULBP (19.00%), LBP (20.25%), and MCT (20.00%). In the third experiment using TYPE3 probe set, the appropriate number of CLBP codes was 35. The classification error using 35 CLBP codes was 18.16%, which is the smallest than those of other representation methods such as ULBP (21.50%), LBP (20.83%), and MCT (20.27%). Moreover, we validated that the Bayes error rate of the proposed CLBP method is better than that of the symmetry ULBP, $L_{sym3,4}$.

In the facial expression recognition experiments, the appropriate number of CLBP codes was 23. The classification error using 23 CLBP codes was 8.00%, which is the smallest than those of other representation methods such as ULBP (12.73%), LBP (14.20%), and MCT (15.07%). Moreover, we showed that the use of 23 CLBP codes enables us to reduce computation times approximately 1.5 times, 4 times, and 7.2 times, which were achieved with ULBP, LBP, and MCT, respectively. From these experimental results, we conclude that the CLBP outperform other features such as LBP, ULBP, and MCT in terms of minimizing the number of codes, minimizing the classification error, and reducing computation time.

References

[1] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1991) 72–86.

[2] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.

[3] M. Bartlett, J. Movellan, T. Sejnowski, Face recognition by independent component analysis, IEEE Transactions on Neural Networks 13 (2002) 1450–1464.

[4] B. Heisele, Y. Ho, T. Poggio, Face recognition with support vector machines: global versus component-based approach, in: Proceedings of International Conference on Computer Vision, 2001, pp. 688–694.

[5] T. Ojala, M. Pietikainen, D. Harwood, A comparative study of texture measures with classification based on feature distributions, Pattern Recognition 29 (1996) 51–59.

[6] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution grayscale and rotation invariant texture classification with local binary patterns, IEEE Transactions of Pattern Analysis and Machine Intelligence 24 (7) (2002) 971–977.

[7] H. Jin, Q. Liu, H. Lu, X. Tong, Face detection using improved LBP under Bayesian framework, in: Proceeding of Third International Conference on Image and Graphics, 2004, pp. 306–309.

[8] L. Zhang, R. Chu, S. Xiang, S. Liao, S.Z. Li, Face detection based on multi-block LBP representation, in: The Second International Conference on Biometrics, 2007, pp. 11–18.

[9] T. Ahonen, A. Hadid, M. Pietikainen, Face recognition with local binary patterns, in: The Eighth European Conference on Computer Vision, 2004, pp. 469–481.

[10] O. Lahdenoja, M. Laiho, A. Paasio, Reducing the feature vector length in local binary pattern based face recognition, in: IEEE International Conference on Image Processing, 2005, pp. 11–14.

[11] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (12) (2006) 2037–2041.

[12] S. Liao, X. Zhu, Z. Lei, L. Zhang, S.Z. Li, Learning multi-scale block local binary patterns for face recognition, in: The Second International Conference on Biometrics, 2007, pp. 828–837.

[13] G. Heusch, Y. Rodriguez, S. Marcel, Local binary patterns as an image preprocessing for face authentication, in: The Seventh International Conference on Automatic Face and Gesture Recognition, 2006, pp. 9–14.

[14] Y. Rodriguez, S. Marcel, Face authentication using adapted local binary pattern histograms, in: The 10th European Conference on Computer Vision, 2006, pp. 321–332.

[15] C. Shan, S. Gong, P.W. McOwan, Facial expression recognition based on local binary patterns: a comprehensive study, Image and Vision Computing 27 (2009) 803–816.

[16] C. Frank, E. Noth, Automatic pixel selection for optimizing facial expression recognition using eigenfaces, in: Lecture Notes in Computer Science, 2003, pp. 378–385.

[17] V. Kellokumpu, G. Zhao, S.Z. Li, M. Pietikainen, Dynamic texture based gait recognition, in: Lecture Notes in Computer Science, 2009, pp. 1000–1009.

[18] V. Takala, T. Ahonen, M. Pietikäinen, Block-based methods for image retrieval using local binary patterns, in: Proceeding of the 14th Scandinavian Conference on Image Analysis, 2005, pp. 882–891.

[19] M. Heikkilä, M. Pietikäinen, J. Heikkilä, A texture-based method for detecting moving objects, in: Proceeding of the 15th British Machine Vision Conference, 2004, pp. 187–196.

[20] S. Choi, C. Choi, G. Jeong, Fixel selection in a face image based on discriminant features for face recognition, in: IEEE International Conference on Automatic Face and Gesture Recognition, 2008.

[21] R. Battiti, Using mutual information for selecting features in supervised neural net learning, IEEE Transactions on Neural Networks 5 (4) (1994) 537–550.

[22] N. Kwak, C. Choi, Improved mutual information feature selector for neural networks in supervised learning, in: The 10th International Joint Conference on Neural Networks, 1999, pp. 1313–1318.

[23] J.C. Principe, Feature extraction using information theoretic learning, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (9) (2002) 1385–1392.

[24] K. Torkkola, Feature extraction by non-parametric mutual information maximization, Journal of Machine Learning Research 3 (2003) 1415–1438.

[25] G. Qiu, J. Fang, Car/non-car classification in an informative sample subspace, in: The 16th International Conference on Pattern Recognition, 2006.

[26] R. Zabih, J. Woodfill, Non-parametric local transforms for computing visual correspondence, in: European Conference on Computer Vision, 2006, pp. 151–158.

[27] B. Froba, A. Ernst, Face detection with the modified census transform, in: IEEE the Sixth International Conference on Face and Gesture Recognition, 2004, pp. 91–96.

[28] R.M. Fano, D. Hawkins, Transmission of information: a statistical theory of communications, American Journal of Physics 29 (1999) 793–794.

[29] M.E. Hellman, J. Raviv, Probability of error, equivocation, and the chernoff bound, IEEE Transactions on Information Theory 16 (4) (1970) 368–372.

[30] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance and min-redundancy, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (8) (2005) 1226–1238.

[31] P.A. Estevez, M. Tesmer, C.A. Perez, J.M. Zuraner, Normalized mutual information feature selection, IEEE Transactions on Neural Networks 20 (2) (2009) 189–201.

[32] C. Yun, J. Yang, Experimental comparison of feature subset selection methods, in: IEEE International Conference on Data Mining, 2007, pp. 367–372.

[33] S. Li, Y. Zhu, J. Feng, P. Ai, X. Chen, Comparative study of three feature selection methods for regional land cover classification using MODIS data, in: IEEE Congress on Image and Signal Processing, 2008, pp. 565–569.

[34] H. Lee, S. Park, B. Kang, J. Shin, J. Lee, H. Je, B. Jun, D. Kim, The POSTECH face database (PF07) and performance evaluation, in: The Eighth IEEE International Conference on Automatic Face and Gesture Recognition, 2008.

[35] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10) (2000) 1090–1104.

[36] L. Wiskott, Jean-Marc Fellous, N. Krüger, Christoph von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 775–779.

[37] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (2006) 504–506.

Bongjin Jun received the B.S. degree in Computer Engineering from Busan National University, Korea, in 2000, and the M.S. degree in Computer Engineering from Pohang University of Science and Technology (POSTECH), in 2002. Now, he is pursuing toward Ph.D. degree at POSTECH.

His research interests include computer vision, face analysis, and human analysis.

Taewan Kim received the B.S. degree in Computer and Information Engineering from Korea Martime University, Korea, in 2006, and the M.S. degree in Computer Engineering from Pohang University of Science and Technology (POSTECH), in 2009. Now, he is working for Daum Communication as a Software Engineer.

His research interests include biometrics, face analysis, and facial expression recognition.

Daijin Kim received the B.S. degree in Electronic Engineering from Yonsei University, Seoul, Korea, in 1981, and the M.S. degree in Electrical Engineering from the Korea Advanced Institute of Science and Technology (KAIST), Taejon, 1984. In 1991, he received the Ph.D. degree in Electrical and Computer Engineering from Syracuse University, Syracuse, NY.

During 1992–1999, he was an Associate Professor in the Department of Computer Engineering at DongA University, Pusan, Korea. He is currently a Professor in the Department of Computer Science and Engineering at POSTECH, Pohang, Korea.

His research interests include biometrics, human computer interaction, and intelligent systems.