Image Description with Local Patterns: An Application to Face Recognition

Wei ZHOU†(a), Nonmember, Alireza AHRARY††(b), and Sei-ichiro KAMATA††(c), Members

SUMMARY In this paper, we propose a novel approach for presenting the local features of digital image using 1D Local Patterns by Multi-Scans (1DLPMS). We also consider the extenstions and simplifications of the proposed approach into facial images analysis. The proposed approach consists of three steps. At the first step, the gray values of pixels in image are represented as a vector giving the local neighborhood intensity distributions of the pixels. Then, multi-scans are applied to capture different spatial information on the image with advantage of less computation than other traditional ways, such as Local Binary Patterns (LBP). The second step is encoding the local features based on different encoding rules using 1D local patterns. This transformation is expected to be less sensitive to illumination variations besides preserving the appearance of images embedded in the original gray scale. At the final step, Grouped 1D Local Patterns by Multi-Scans (G1DLPMS) is applied to make the proposed approach computationalmally simpler and easy to extend. Next, we further formulate boosted algorithm to extract the most discriminant local features. The evaluated results demonstrate that the proposed approach outperforms the conventional approaches in terms of accuracy in applications of face recognition, gender estimation and facial expression.

key words: face recognition, gender estimation, local presentation, multi-scans, 1DLPMS, G1DLPMS

1. Introduction

Over the last fifteen years or so, face recognition has become a popular area of research in computer vision. Compared with other biometrics [1], such as fingerprint and iris, face recognition has great advantage in high-universality, high-collectability, high-acceptability, and low-circumvention. Hence, face recognition is widely used in a variety of fields such as image analysis, classification, forensic identification and access control. Although face recognition systems show striking improvement in successive competitions, face recognition problem is still considered to be unsolved due to the variations with illuminance, facial expression and so on [2], [3]. According to how the elements of the face representation are calculated, current researches can be coarsely categorized into global feature based and local feature based methods. Specifically, in global feature, its each element is related to the whole input face image, while in local feature, its every dimension is extracted from some local region of the face image. It is worth pointing out that, there is no clear boundary between them: for local feature, with the increase of the size of the region for feature extraction, its “globality” increases correspondingly. Most of the global methods are subspace methods that reduce the dimension of the image. The Principle Component Analysis (PCA) [4] is the famous method in this class, which uses an eigenvalue subspace to project the whole image into several weights, and employs the distances between these weights to recognize faces. Independent Component Analysis (ICA) takes higher-order statistics into account, and is suitable for learning complex structure in the dataset [5]. Linear Discriminant Analysis (LDA) [6], which considered the difference both between-class and within-class matrix, and Discrete Cosine Transform (DCT) [7], which can remain more linear property, are another two well-known methods in this class. Recently, in order to reduce computational time cost and preserve two dimension information of images, 2DPCA [8], 2DLDA [9], 2DPCA-L1 [10] are researched. However, all of these methods are global representation of images which are sensitive to global changes of images, such as, illuminimtation and expression.

In order to overcome these drawbacks of global feature based methods, local matching approaches are presented in face recognition [11], [12] and other visual recognition tasks [13] that are invariant to illumination and expression issues. The general idea of local matching methods is first to locate several features, and then classify the images by comparing and combining the corresponding local statistics. The most famous method in this field is Local Binary Patterns (LBP) [11]. In our previous work, some extensions of LBP were proposed, such as Local Ternary Patterns (LTP) [14], Local Quaternion Patters (LQP) [15] which can make the local patterns more reasonable and effective. Recently, Zhang [16] proposed Local Derivative Pattern (LDP) which is a general framework to encode directional pattern features based on local derivative variations and can capture more detailed information than the first order local pattern used in LBP. However, all of these patterns use circle to encode the local neighborhood pixels, that means, just one kind of spatial information around is considered. Specially for LDP, in previous steps, local derivative variations is used to capture more detailed information, however it has high time complexity. From the reconstructed images in ref. [16] by LDP, more detailed information also takes more noise into account and key parts of our faces can not sep-
arate clearly. At final step, also circle encoding is applied. Another problem of the existing approached in the field is that if the number of the neighborhood pixels enlarged a lot, which makes the number of patterns be huge, then the encoding and calculation of these patterns are time consuming. For example, in LBP[11], assigns a label to every pixel of an image by thresholding the $3 \times 3$-neighborhood of each pixel with the center pixel value and considering the result as a binary number. Generally, the notation $(P, R)$ will be used for pixel neighborhoods which means $P$ equally sampling points on a rectangle with inscribed circle of radius of $R$[11]. If $P$ is small, then we can not get discriminant and spatial information in this label as much as possible, while $P$ is large, encoding seems difficult and lots of histogram of the patterns are equal or similar to zero which can be noted as noise and is not useful for representation. As an example, if $P = 16$. then $2^{16} = 65536$ patterns will be generated. When the resolution of an image is $100 \times 100$, the average magnitude of the histogram is just $10000/65536 \approx 0.15$. Under this condition, many patterns will be useless.

To address these issues, we propose a novel, theoretically and computationally simple approach which is called 1D local patterns by multi-scans (1DLPMS) [17]. First, multi-scans are used to capture the different spatial information on the image. Compared to the traditional ones [11] which only used circle to encode the neighborhood pixels, multi-scans can keep more spatial information. Then based on varied encoding rules, some 1D local patterns are proposed to encode local features with less time consuming. To simplify the computation of 1DLPMS, grouped 1D local patterns by multi-scans (G1DLPMS) is proposed, which separates 1DLPMS into several groups and only the co-occurrences of these groups are used. Boosted algorithm is further formulated to extract the most discriminant local features with less calculation. In order to evaluate the proposed methods, face recognition, gender estimation and facial expression recognition are studied. Experimental results on face recognition based on two famous and challenging databases-ORL and FERET show that the proposed methods outperform some classical algorithms. The same result can be derived from gender estimation and facial expression problem.

The remainder of this paper is organized as follows: In Sect. 2, related works will be introduced briefly and 1D local patterns by multi-scans will be described in Sect. 3. The extension descriptors - grouped 1D local patterns by multis-scans will be introduced in Sect. 4. Experimental results are presented in Sect. 5. Finally, conclusions and future work are discussed in Sect. 6.

2. Related Work

The local feature patterns are designed for texture description. The basic idea is that these operators assign a label to every pixel of an image by applying the $(2R + 1) \times (2R + 1)$ neighborhood with each pixel and considering the relationship between the center pixel (we call it pivot in the following) and neighborhood ones. Figure 1 gives a generalized definition for the label of size $(2R + 1) \times (2R + 1)$. Here, $\alpha = (2R - 1) \times (2R - 1)$. First, let $Z_0$ be the center of the label. Then in each rectangle with inscribed circle of radius of $R$ (the center of the circle is also $Z_0$), we can mark this rectangle from lefttop as $Z_a, Z_{a+1}, \ldots, Z_{a+8R-1}$ by clockwise order.

In this area, the most famous one called LBP (details see [11]) was proposed some years ago. LBP provides an invariant description in presence of monotonic illumination variation on face image, however, suffer much in non-monotonic illumination variation, random noise. In order to overcome this issue, Local Ternary Patterns (LTP) [14] and Local Quaternion Patters (LQP) [15] were studied in our previous work. The thresholding function $f_B(\ldots), f_T(\ldots), f_Q(\ldots)$ for the basic LBP, LTP and LQP can be formally represented as

$$f_B(I(Z_0), I(Z_i)) = \begin{cases} 0, & \text{if } I(Z_i) \leq I(Z_0) \\ 1, & \text{if } I(Z_i) > I(Z_0) \end{cases}$$

(1)

$$f_T(I(Z_0), I(Z_i)) = \begin{cases} 0, & \text{if } I(Z_i) < I(Z_0) - \text{interval} \\ 1, & \text{if } I(Z_0) - \text{interval} \leq I(Z_i) \leq I(Z_0) + \text{interval} \\ 2, & \text{if } I(Z_i) > I(Z_0) + \text{interval} \end{cases}$$

(2)

$$f_Q(I(Z_0), I(Z_i)) = \begin{cases} 0, & \text{if } I(Z_i) < I(Z_0) - \text{interval} \\ 1, & \text{if } I(Z_0) - \text{interval} \leq I(Z_i) \leq I(Z_0) + \text{interval} \\ 2, & \text{if } I(Z_i) > I(Z_0) + \text{interval} \end{cases}$$

(3)

where $Z_i$ is the neighborhood point around $Z_0$ (pivot). $I(Z_i)$ and $I(Z_0)$ are the intensity of the neighborhood points and pivot. $(I(Z_i), I(Z_0)) \in \mathbb{R}$ and $0 \leq I(Z_i), I(Z_0) \leq M$, $M$ is the maximal intensity in the image) The definition of interval is noted as Eq. (4). $k$ is a predefined value [14].

![Figure 1](image-url)
Fig. 2  LBP vs LTP, \(k = 4\) (a), (c) are two LBP operators and (b), (d) are two LTP operators.

Fig. 3  LBP vs LQP, \(k = 4\) (a), (c) are two LBP operators and (b), (d) are two LQP operators.

\[
\text{interval} = \begin{cases} 
I(Z_0)/k, & \text{if } I(Z_0) \leq M/2 \\
(M - I(Z_0))/k, & \text{if } I(Z_0) > M/2
\end{cases} \tag{4}
\]

This kind of patterns can also be considered as the concatenation of the binary/ternary/quaternion gradient directions, and are called micropatterns. The histograms of these micropatterns contain information of the distribution of the edges, spots, and other local features in an image. Figure 2 and Fig. 3 show some examples of obtaining LBP, LTP and LQP micropattern. In both figures, the upper 3 by 3 label is the pixel intensity from image and the lower is the corresponding local patterns. From Fig. 2, we can see that, when some light source came from the bottom left corner, LBP micropatterns are changed while LTP micropatterns keep the same value. Thus, LTP is more insensitive to illumination compared to LBP. In Fig. 3, it is clear that the left top 3 by 3 label stands for flat area while the right top 3 by 3 label stands for spot. However, LBP micropatterns can not separate them while LQP micropatterns can discriminate them clearly. Thus, compare to LBP, LQP can get more discriminative information about image.

Some reconstructed face images by decimal number according to corresponding local patterns are illustrated in Fig. 4 (Note that all the decimal number are normalized into [0, 255] for display). From these reconstructed images, we can clearly see that LTP and LQP are more invariant to illumination changes and can get the key parts in faces more efficiently than LBP, such as eyes, nose and mouth. However, one problem is that the micropatterns obtained by LTP and LQP are larger than traditional LBP with the same bit number which can make time consuming. This is one motivation of this research.

In order to take rotation invariant into account, Advanced Local Binary Patterns (ALBP) is proposed by Liao et al. [18]. The basic idea of this patterns is performing a circular anti-clockwise bitwise shift on the bit number bit by bit and selecting the smallest decimal number. For example in Fig. 2(a), 8 Binary numbers will be obtained and at last 00000111 is selected. However, this method can not solve the basic LBP problems with noise and illumination issues and have more time complexity than basic LBP approach. In some sense, the combination of the proposed methods by multi-scans is invariant to rotation issue.

More recently, Dominant Local Binary Pattern (DLBP)[19] and Local Derivative Pattern (LDP)[16] are proposed. DLBP considers more complicated shapes, which can contain high curvature edges, crossing boundaries, corners. However, in our face application, most of shapes are fundamental, such as line, edge, spot, flat area. And the conventional LBP can catch about 90 percent of the shapes in case of preprocessed FERET facial images [11]. Thus, in our study, most fundamental information is considered since it is enough for face recognition as well as it can obtain faster speed. LDP encodes directional pattern features based on local derivative variations. It can capture more detailed information than the first order LBP. LBP is always considered first-order local pattern operator, because LBP encodes all-direction first-order derivative binary result whereas LDP encodes the higher-order derivative information. So it contains more discriminative features than LBP.
However, LDP generated more patterns than LBP which make it difficult to be applied in practice. In some cases, higher-order derivative information also catches more noise information which is useless. Compared to these two methods, our proposed method has amazing speed while keeping discriminative and fundamental information, which is less sensitive to noise in face application.

3. 1D local patterns by Multi-Scans (1DLPMS)

Some traditional local patterns which are based on 2D, such as LBP [11], assign a circle to every pixel of an image for encoding to present the micropatterns. However, when some noise is imported in the label, the pattern of LBP will be changed as shown in Fig. 5 where the dash pixel contains some noise. Based on this point, multi-scans are applied for encoding in our study as illustrated in Fig. 6 where several patterns can be obtained. compared Fig. 6 (a) with Fig. 6 (b), just one pattern is changed while other three patterns are same. In this scheme, encoding by scan can not only keep the texture information but also reduce the affectation of noise.

On the other hand, just the circle spatial information around the pixel is considered in traditional methods. As shown in Fig. 1, eight positions from $Z_1$ to $Z_8$ around $Z_0$ are taken into account and the number of obtained patterns is $2^8 = 256$ for $P = 8$ and $R = 1$. Generally speaking, the spatial information and the number of patterns are related to accuracy and time consuming respectively. So in order to get more spatial information around $Z_0$ while reduce the number of patterns at the same time, multi-scans are applied to encode the neighborhood pixels in our study.

Let $S$ be a set of multi-scans, in our study, five scans including raster (S1), raster-type2 (S2), Zig-Zag (S3), Zig-Zag-type2 (S4) and Hilbert (S5) are used. Thus, $S = \{S1, S2, S3, S4, S5\}$ in our application. Based on Fig. 1, we can generalize the order of each scan as follows: (Fig. 7 for $R = 2$

$S1: Z_0, Z_{10}, Z_{11}, Z_{12}, Z_{13}, Z_{14}, Z_5, Z_6, Z_7, Z_{21}, Z_{22}, Z_{20}, Z_{12}, Z_{13}, Z_{14}, Z_{15}, Z_{16}, Z_{17}.$

$S2: Z_0, Z_{24}, Z_{23}, Z_{22}, Z_{21}, Z_{20}, Z_7, Z_8, Z_1, Z_{10}, Z_{11}, Z_{12}, Z_{13}, Z_5, Z_6, Z_7, Z_{12}, Z_{13}, Z_{14}, Z_{15}, Z_{16}, Z_{17}.$

$S3: Z_0, Z_{24}, Z_{10}, Z_{11}, Z_{12}, Z_{23}, Z_{22}, Z_8, Z_2, Z_{12}, Z_{13}, Z_3, Z_4, Z_5, Z_6, Z_7, Z_{12}, Z_{13}, Z_{14}, Z_{15}, Z_{16}, Z_{17}.$

$S4: Z_{24}, Z_{21}, Z_{20}, Z_{22}, Z_{23}, Z_7, Z_{19}, Z_{18}, Z_6, Z_5, Z_{24}, Z_9, Z_1, Z_{10}, Z_{11}, Z_{12}, Z_{13}, Z_5, Z_6, Z_7, Z_{12}, Z_{13}, Z_{14}, Z_{15}, Z_{16}, Z_{17}.$

$S5: Z_{24}, Z_2, Z_1, Z_{10}, Z_{11}, Z_{12}, Z_{13}, Z_{14}, Z_{23}, Z_2, Z_6, Z_5, Z_{16}, Z_{17}, Z_{18}, Z_{19}, Z_{20}, Z_7, Z_2, Z_{21}.$

Here, simple Hilbert Scan order is used for symmetry. The origin of S1, S2, S3 and S5 is from left top while S4 is from left bottom.

Since scan order itself contains the spatial information of the image while different scan orders can get different spatial information, only little neighborhood pixels encoding in multi-scans can also keep enough discriminant and spatial information in that label. For general, the notation $(P, h)$ will be used for pixel neighborhoods which means $P$ sampling points on a interval of $h$ between sampling points.

As mentioned above, there are eight locations marked as $Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8$ around $Z_0$ in LBP, while multi-scans are applied in this label ($P = 4$ and $h = 1$). S1 can cover position $Z_{23}$, $Z_8$ and $Z_{15}$, S2 can cover position $Z_{11}, Z_2, Z_6$ and $Z_{19}$, S3 can cover position $Z_{13}, Z_1, Z_7$ and $Z_{21}$, S4 can cover position $Z_0, Z_1, Z_5$ and $Z_{17}$ while S5 can cover
Each element and the one within circle is the from a transformed label. The number is the intensity of constructed images (Fig. 14). Figure 9 gives an subsequence smoother than 1DLBP [14]. Also we can see this point in re-

turns (1DLBP), 1D Local Ternary Patterns (1DLTP), 1D Lo-

cal Quaternion Patterns (1DLQP) are obtained based on var-

terns (1DLP) is assigned to every pixel by thresholding the 

image.

Since most shape information in our facial images are fundamental [11], such as line, edge, multi-scans can capture them effectively and efficiently. For example, S1 and S2 can capture vertical and horizontal line and edge information, S3 and S4 can capture diagonal line while S5 can capture some more complex shapes, respectively. These fundamental information in face can be coded by multi-scans accurately and fast while it is less sensitive to noise. Also, this point can be seen in Fig. 14, especially, S2 can get more clear information of eyebrows, eyes and mouth which are the key parts in the faical image.

Some representative samples by 1DLTP using S2 scan (a) Original image (b) $R = 1$, $P' = 2$, $h = 1$ (c) $R = 1$, $P' = 4$, $h = 1$ (d) $R = 2$, $P' = 4$, $h = 2$ (e) $R = 2$, $P' = 4$, $h = 3$.

![Fig. 7] Multi-Scans for one label ($R = 2$) (a) raster (S1) (b) raster-type2 (S2) (c) Zig-Zag (S3) (d) Zig-Zag-type2 (S4) (e) Hilbert (S5).

position $Z_0$, $Z_1$, $Z_2$ and $Z_3$ ($Z_0$ is set at the center of each scan). In this case, the number of patterns is just $5 \times 2^4 = 80$, which is much less than LBP which needs $2^8 = 256$ patterns. And multi-scans can cover 16 positions while circle can just cover 8 positions. Therefore, we can use less number of patterns to represent enough discriminant and more spatial information in that label. In other words, multi-scans can not only keep more spatial information than LBP, but also use less number of patterns. Note if $P' = 6$, then S1 can cover position $Z_{34}$, $Z_{23}$, $Z_8$, $Z_4$, $Z_{15}$, $Z_{16}$: S2 can cover $Z_{10}$, $Z_{11}$, $Z_2$, $Z_6$, $Z_{19}$, $Z_{18}$ while S3 can cover position $Z_{12}$, $Z_{13}$, $Z_5$, $Z_7$, $Z_{21}$ and $Z_{20}$.

After getting the transformed sequence, 1D local patterns (1DLP) is assigned to every pixel by thresholding the $P'$-neighborhood of each pixel with pivot and considering the result as a binary, ternary or quaternion number respectively. The local neighborhood is defined as a set of sampling points which evenly space on the scan order and center at the pixel to be labeled. It allows any interval and number of sampling points. Bilinear interpolation is used when a sampling point does not fall in the center of a pixel. Figure 8 gives an example.

Some 1D local patterns called 1D Local Binary Patterns (1DLBP), 1D Local Ternary Patterns (1DLTP), 1D Local Quaternion Patterns (1DLQP) are obtained based on varied encoding rules shown in Eqs. (1), (2), (3) respectively.

Images reconstructed by 1DLTP and 1DLQP are smoother than 1DLBP [14]. Also we can see this point in reconstructed images (Fig. 14). Figure 9 gives an subsequence from a transformed label (the number is the intensity of each element).

![Fig. 8] Bilinear interpolation (a) subsequence from transformed label (b) (4, 1), (c) (4, 1.5), and (d) (6, 1) neighborhoods. black circle without filling is the location of pivot and all black circles are the locations of sampling points.

| Table 1 | 1D local patterns for Fig. 9 ($P' = 4$, $h = 1$, $k = 2$). |
|---------|--------------------------------------------------------|
| 1DLBP   | 1010                                                   |
| 1DLTP   | 2110                                                   |
| 1DLQP   | 3120                                                   |

![Fig. 9] Subsequence from a transformed label (the number is the intensity of each element).

![Fig. 10] Some representative samples for different $R$, $P'$ and $h$ by 1DLTP and S2 scan (a) Original image (b) $R = 1$, $P' = 2$, $h = 1$ (c) $R = 1$, $P' = 4$, $h = 1$ (d) $R = 2$, $P' = 4$, $h = 2$ (e) $R = 2$, $P' = 4$, $h = 3$.
In the proposed 1D local patterns, especially 1DLQP, the parameter $P$ determines the number of patterns. A large $P$ produces a long histogram which means lots of useless information will be generated. When the number of sampling points increases, the number of patterns for basic 1D local patterns will become very large, for 1DLQP it will be $4P$, as shown in Fig. 11. Due to this rapid increase, it is difficult to extend them when needing a large number of sampling points. This limits its applicability.

To address this problem, we simplify the descriptor into Grouped 1D Local Patterns (G1DLP) by concatenating 1D local patterns on some groups. Let $D_i$ be the distance between $i$-th sampling point and pivot in the scan sequence and $D'_i$ be the sorted increasing distance with the rise of $i$ for $i = 1, 2, \ldots, P'$. Then, a basic group (called group 1) can be set, where the sampling points with number $P'_G$ are close to the pivot. For symmetry, the distribution on both side of pivot is same. Thus one basic distance set $D(1) = \{D'_1, D'_2, \ldots, D'_{P'_G}\}$ which contains top $P'_G - th$ smallest distance can be obtained. Some other distance set $D(t)$ is defined as Eq. (5)

$$D(t) = \{D'_1 + (t-1) \times D_M, \ldots, D'_{P'_G} + (t-1) \times D_M\}$$

where $D_M = \max\{D'_1, D'_2, \ldots, D'_{P'_G}\}$ \hspace{1cm} (5)

and $t = 1, 2, \ldots$

The $i$-th sampling point can be set as group $t$ if and only if $D_i \in D(t)$. Figure 12 shows a simple example.

The 1D local patterns are extracted from each group to concatenate grouped 1D local patterns (G1DLP), such as grouped 1DLBP (G1DLPB), grouped 1DLTP (G1DLTP), grouped 1DLQP (G1DLQP). Same as 1DLPMS, for each scan, G1DLP can be obtained, and grouped 1D local patterns by multi-scans (G1DLPMS) is the combination of them. For all pixels, the statistics of these groups are obtained and then concatenated into a single histogram (Fig. 12). So the length of feature vector is linear to the number of sampling points (shown in Fig. 11 with $P'_G = 2$).

5. Experimental Results

In this section, the proposed methods will be evaluated on face recognition, gender estimation and facial expression. Two well-known databases - ORL and FERET are evaluated for face recognition and CHI-square distance is applied for classification with $R = 2$.

5.1 Experiments on ORL Database

The ORL database consists of face images of 40 different people, each individual providing 10 different images. For some subjects, the images were taken at different times. The facial expressions open or closed eyes, smiling or non-smiling and facial details (glasses or no glasses) also vary. The images were taken with a tolerance for some tilting and rotation of the face of up to 20 degrees. Moreover, there is also some variation in the scale of up to about 10 percent. All images are gray scale and normalized to a resolution of $112 \times 92$ pixels. Some examples are shown in Fig. 13.

Each image is divided into 16 regions. Some reconstructed images are list in Fig. 14. Compared to LBP, combined image have clearer edge and less noise, especially in 1DLTP and 1DLQP.

In first experiment, the leave-one-out strategy is used. The result is listed in Table 2. -B, -T, -Q means using different encoding rule under Eqs. (1), (2), (3), respectively. The notation “Num.” means the number of patterns and “time” corresponding to matching time. It can be seen that -T and -Q achieve higher accuracy than LTP with less patterns and computational complexity. And the best one is G1DLPMS.
Table 3  Accuracy on ORL- d (1, 2, 3, 4, 5) for gallery and others for probe.

| Methods       | Accuracy | $P/P'$ | $P'_G$ | Num. | time (ms) |
|---------------|----------|--------|--------|------|-----------|
| PCA           | 66.9     | 84.7   | 88.2   | 90.8 | 93.5      |
| LDA           | 75.2     | 85.6   | 87.5   | 91.7 | 92.5      |
| LBP           | 74.7     | 82.5   | 87.8   | 92   | 94        |
| LGBP [21]     | 77.5     | 87.8   | 88.6   | 92.9 | 96        |
| LTP [14]      | 77.5     | 86.9   | 90.7   | 94.6 | 97        |
| ALBP [18]     | 75       | 82.7   | 88.6   | 92.5 | 94        |
| DLBP [19]     | 73.6     | 83.1   | 88.6   | 92.9 | 95        |
| LDP [16]      | 78.9     | 83.1   | 87.9   | 93.5 | 96        |
| 1DLPMS-B      | 79.7     | 83.2   | 89     | 92.9 | 95.5      |
| G1DLPMS-B     | 77.5     | 85.3   | 88.5   | 93.8 | 96        |
| 1DLPMS-T      | 81.9     | 89.1   | 91.1   | 94.6 | 97.5      |
| G1DLPMS-T     | 86.7     | 93     | 93.6   | 97.1 | 98        |
| 1DLPMS-Q      | 85       | 90     | 92     | 95   | 97.5      |
| G1DLPMS-Q     | 87.2     | 91.9   | 93     | 95.1 | 98        |

T with the trade-off between accuracy and computational complexity. Same parameters $P$, $P'$ and $P'_G$ will be used in the following.

In second experiment, some $d$ ($d = 1, 2, 3, 4, 5$) images of each person are randomly chosen for training, while the remaining images for testing. To compare our method with LBP and other famous approaches, five tests are performed with a varying number of training samples and mean rate is recorded. Table 3 shows the accuracy ($k = 2$ in this case). It can be seen that the proposed methods achieve better performance since it can overcome rotation problems more efficiently.

In third study, we try to find which scan order can keep key significant spatial information. In this case, G1DLPMS-T and G1DLPMS-Q are outstanding since they can keep more discriminable information than others.

Influence of Parameter $P'$ by 1DLPMS-B, one image for training.

| $P'$ | Recognition Rate |
|------|------------------|
| 2    | 79.7             |
| 4    | 82.2             |
| 6    | 83.9             |
| 8    | 83.9             |

Table 4 Influence of Parameter $P'$ by 1DLPMS-B, one image for training.

| $h$  | Recognition Rate |
|------|------------------|
| 1    | 80.3             |
| 1.5  | 80.3             |
| 2    | 82.2             |
| 2.5  | 83.9             |
| 3    | 83.9             |
| 3.5  | 83.6             |
| 4    | 83.5             |

Table 5 Influence of Parameter $h$ by 1DLPMS-B, one image for training.
not discriminant the close area and non-close area clearly. Otherwise, if $k$ is very large, that means the close area is very narrow, also, we can not discriminant the close area and non-close area clearly. Thus, in our experiments, the parameter $k$ is around 4.5.

### 5.2 FERET Database

In our study, regular frontal image called ‘fa’ set is used for making galley set, the size of galley in our study is 581. The ‘fb’ probe set is for analyzing the effect of a different facial expression on recognition performance. The size of ‘fb’ probe set is 580. The size of ‘dup1’, which consists of all other frontal of the subjects taken several days or even years later than ‘fa’, is 475. The ‘dup2’ probe set which taken at least one year later than ‘fa’ is a subset of the ‘dup1’ probe set and the size is 118. The images are copped by eyes and mouth but no normalization methods are applied (Note that these cropped images can also be directly detected by some traditional face detection algorithm). Each face is resized to 100×100 and divided into 25 regions. Figure 15 gives some samples. (the first two images are from ‘fa’ and ‘fb’ set, respectively. The images in second row are located in ‘dup1’ set, while the images in last row come from ‘dup2’ probe set.) In this experiment, we just use one image in set ‘fa’ for training, and ‘fb’, ‘dup1’ and ‘dup2’ for testing. The result is shown in Table 7. G1DLPMS-T and G1DLPMS-Q are two outstanding ones while G1DLPMS-T uses less number of patterns.

The reason for making this experiment is that all the captured face can be directly obtained by traditional face detection method and then face recognition can be performed without any normalization according to the known coordinates of eyes and mouth. And in order to compare to some start-of-art methods, same experimental setup in [16] is followed. Also, gabor feature is considered and the results is show in Table 8. Prefix G means gabor feature is used in that corresponding method.

From Table 8, we can see that gabor feature is effective for the proposed method, especially in dup1 and dup2 set, which can improve performance by about 20 percent, since Gabor wavelet has good characteristics in space frequency, space position and direction selectivity. And G1DLPMS-T and G1DLPMS-Q are two excellent ones.

Figure 16 investigates how the performance will be changed when $R$ and resolution of images are varied.

| $k$ | Recognition Rate |
|-----|------------------|
| 1.5 | 80.6             |
| 2   | 82.5             |
| 2.5 | 83.1             |
| 3   | 84.7             |
| 3.5 | 84.2             |
| 4   | 85.0             |
| 4.5 | 85.3             |
| 5   | 84.4             |
| 5.5 | 83.9             |
| 6   | 83.6             |

Table 7 Precision in FERET [20] database from our cropped images.

|       | fb  | dup1 | dup2 |
|-------|-----|------|------|
| LBP   | 83.5| 44.5 | 28.4 |
| 1DLPMS-B     | 86.2| 47.2 | 30.5 |
| G1DLPMS-B     | 85.2| 45.9 | 28.8 |
| LTP [14]      | 87.6| 48.3 | 31.3 |
| LDP [16]      | 88.3| 47.4 | 30.5 |
| 1DLPMS-T     | 89.3| 49.9 | 33.9 |
| G1DLPMS-T     | 91.4| 53.7 | 39.9 |
| 1DLPMS-Q     | 89.5| 50.3 | 35.6 |
| G1DLPMS-Q     | 93.4| 53.3 | 39.0 |

Table 8 Precision in FERET [20] database by standard principle.

|       | fb  | fc  | dup1 | dup2 |
|-------|-----|-----|------|------|
| LBP   | 91  | 65  | 53   | 38   |
| LGBP [21] | 94  | 97  | 68   | 53   |
| 1DLPMS-B     | 92  | 70  | 59   | 42   |
| G1DLPMS-B     | 97  | 97  | 78   | 75   |
| ALBP [18]     | 93  | 95  | 59   | 41   |
| DLBP [19]     | 89  | 69  | 61   | 43   |
| LDP [16]      | 92  | 88  | 63   | 60   |
| G1DLPMS-T     | 98  | 99  | 80   | 80   |
| 1DLPMS-Q     | 95  | 93  | 67   | 63   |
| G1DLPMS-Q     | 99  | 99  | 82   | 84   |
Fig. 17 Some samples for gender estimation (a) LFW (b) FRGC.

(1DLPMS-B and fb probe set is used). Three kinds of resolution and four kinds of $R$ are selected. From this figure, we can see that higher performance can be obtained by better resolution. For small resolution images, a little larger $R$ can get rapid increasing in accuracy ($R$ from 1 to 3 in $50 \times 50$ resolution) while performance is almost same for different $R$ in large resolution images. ($R$ from 2 to 4 in $100 \times 100$ resolution). The reason may be that in small resolution, larger $R$ can capture more local information, while in large resolution, with the increasing of $R$, little local information will be added.

5.3 Gender Estimation

In this experiment, two large databases-Labeled Faces in the Wild (LFW)[22] and Face Recognition Grand Challenge (FRGC)[23] are used. Some samples are shown in Fig. 17. LFW is a database of face photographs designed for studying the problem of unconstrained face understanding. The database contains more than 13,000 images of faces collected from the web. Each face has been labeled with the name of the person pictured. 1680 of the people pictured have two or more distinct photos in the database. The only constraint on these faces is that they were detected by the Viola-Jones face detector. We randomly select 2000 males and 1000 females for training and using another 2000 males and 1000 females for testing. For FRGC database, 1000 males and 1000 females are selected for training and other 1000 males and 1000 females are used for probe. These images are cropped to $64 \times 64$ and divided into 16 regions. The most different between the two databases is that the images in FRGC are all frontal faces while not in LFW database. A SVM classifier is selected as the classifier in our gender estimation system since it is well founded in statistical learning theory and has been successfully applied to gender estimation. SVM is a binary classification method that finds the optimal linear decision surface based on the concept of structural risk minimization (SRM) principle [24], [25]. In our system, three kernel functions are concerned as Table 9. 

Table 9 Kernel functions using in the system.

| Kernel  | $k(x', y') = x' \cdot y'$ |
|---------|----------------------------|
| Dot product | $k(x', y') = x' \cdot y'$ |
| Polynomial | $k(x', y') = (x' \cdot y' + 1)^p$ |
| RBF     | $k(x', y') = \exp(-\|x' - y'\|^2 / 2\sigma^2)$ |

$x', y' \in \mathbb{R}^m$.

The experimental results are shown in Table 10 and Table 11, respectively. In order to ignore the results by different coding rules, only binary coding is used here. Top four scans mentioned in Fig. 7 are used. 0, 1, 2 means the kernel used in SVM is dot product, polynomial and RBF, respectively. The last column in the two tables is the feature dimension applied for classification. From these two tables, we can see that proposed methods can get more discriminant features about male and female than LBP under different variations. And frontal faces are more useful for gender estimation than profile.

Next evaluation is tested on male and female set, respectively. We try to find which gender is harder to be recognized compared to the other one. The results evaluated on LFW and FRGC are list in Table 12 and Table 13, respectively. From these two tables, we can see that the recognition rate of male is much higher than female, while the recognition rate of female is almost same as male in FRGC database. So male is expected easier to be recognized in our real life while comparable with female if just frontal face is used to estimation.

5.4 Facial Expression Recognition

The above experiments clearly demonstrate that the 1D local features are effective for face recognition, and performed better than reported existing techniques but with a significant low-computation advantage. In the above investigation, face images are equally divided into small sub-regions from which local histograms are extracted and concatenated into a single feature vector. However, apparently the extracted features depend on the divided sub-regions, so this feature extraction scheme suffers from fixed sub-region size and positions. By shifting and scaling a sub-window over face images, more sub-regions can be obtained, bringing more histograms, which yield a more complete descrip-
Table 12  Precision about male and female in LFW database, respectively.

|       | Male 0 | Female 0 | Male 1 | Female 1 | Male 2 | Female 2 |
|-------|--------|----------|--------|----------|--------|----------|
| LBP   | 94.7   | 64.0     | 92.1   | 76.9     | 95.0   | 65.2     |
| 1DLPMS-B | 94.3   | 71.1     | 91.9   | 81.6     | 95.5   | 68.3     |
| 1DLPMS-T | 94.6   | 70.7     | 92.9   | 84.2     | 95.6   | 69.0     |
| 1DLPMS-Q | 91.0   | 85.0     | 94.2   | 86.6     | 92.6   | 76.7     |

Table 13  Precision about male and female in FRGC database, respectively.

|       | Male 0 | Female 0 | Male 1 | Female 1 | Male 2 | Female 2 |
|-------|--------|----------|--------|----------|--------|----------|
| LBP   | 92.0   | 94.3     | 93.5   | 96.6     | 90.6   | 94.2     |
| 1DLPMS-B | 91.6   | 96.1     | 92.6   | 97.8     | 93.2   | 95.7     |
| 1DLPMS-T | 96.4   | 97.3     | 96.1   | 97.7     | 92.5   | 93.8     |
| 1DLPMS-Q | 97.0   | 97.1     | 96.3   | 96.3     | 92.6   | 95.5     |

Fig. 18  Samples from JAFFE database. The facial expression from left to right is angry (AN), disgust (DI), fear (FE), happy (HA), sad (SA) and surprised (SU).

The JAFFE database (Fig. 18) contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models [9]. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects. There are 2 to 4 images for each expression posed by each individual. 15 images for each expression are selected as gallery randomly while others are for probe. SVM is applied for classification and the resolution of all images is 64 by 64.

As each histogram is calculated from a sub-region, AdaBoost is actually used to find the sub-regions that contain more discriminative information for facial expression classification in term of the histogram. On selecting a weak classifier for AdaBoost, we adopted the histogram-based template matching as same as in [27].

By shifting and scaling a sub-window, 625 sub-regions, i.e., 625 histograms, in total were extracted from each face image. The sub-window was shifted in the whole image with the shifting step of 4 pixels, while its size was scaled between 16 × 16 pixels and 20 × 20 pixels with the scaling step of 4 pixels. AdaBoost was used to learn a small set of effective histograms. We plot in Fig. 19 and Fig. 20 the spatial localization of the 6 sub-regions that corresponded by the top 6 histograms selected by Boosted-LBP and Boosted-1DLPMS-B for two comparable expressions, respectively.

From the two figures, we can see that Boosted-1DLPMS-B can select more powerful regions than Boosted-LBP. Take AN vs DI for an example, in Fig. 19, one region is selected between right eye and nose, and another region is selected in the bottom right cheek. Both of these two regions are not so varied in AN and DI facial expression. But, in Fig. 20, all six regions are located in eyebrows, eyes and mouth where the difference between AN and DI facial ex-
pression is significant.

The comparison of two facial expression accuracy is illustrated in Fig. 21 and average precision is shown in Table 14. In Fig. 21, one against one strategy by SVM classifier is used. For example, in AN vs DI condition, all the probe images in AN and DI facial expression are treated as input, and the average accuracy is recorded. From these results, it can be seen that boosted local patterns can keep more discriminant information but with a significant low time cost advantage. And Boosted-1DLPMS-B can achieve the best performance.

6. Conclusions and Future Work

In this paper, a novel approach for presenting the local patterns is proposed and its simplifications and extensions to facial image are also considered. The proposed method - 1DLPMS can capture more spatial information and key parts clearly in face applications with less time complexity. To make the approach computationally simpler and easy to extend, only the co-occurrences of local patterns in some groups are taken into account. Meanwhile, the proposed measure is more invariant to illuminance changes and noise than traditional method. Experiments on both face recognition and gender estimation show that the proposed methods have better performance than relevant ones and S2 can keep more useful spatial information than other scans. From gender estimation, we can see that male is easier to be recognized compared to female under variant environment in our daily life while it is comparable with female if just frontal faces are considered. Boosted local patterns can keep more discriminant information but with a significant low time cost advantage. The contribution of our approach include local presentation, robustness to gray-scale changes, easy extension and simple computation.

In future, some other special applications using the proposed patterns will be evaluated, such as age estimation, image searching, object tracking. And Curvelet frequency space will be studied.

Acknowledgment

We would like to thank all the people providing their data for test and all the observers for giving their contributions to this study. This work was supported in part by a grant of Knowledge Cluster Initiative 2nd stage by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

[1] A.K. Jain, “Biometric recognition: How do I know who you are?,” Proc. 12th IEEE Signal Processing and Communications Applications Conference, pp.3–5, April 2004.
[2] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, “The feret evaluation methodology for face-recognition algorithms,” IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.10, pp.1090–1104, Oct. 2000.
[3] S. Zhou, G. Aggarwal, R. Chellappa, and D. Jacobs, “Appearance characterization of linear lambertian objects generalized photometric stereo, and illumination-invariant face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.2, pp.230–245, 2007.
[4] M.A. Turk and A.P. Pentland, “Face recognition using eigenfaces,” IEEE Conference on Computer Vision and Pattern Recognition, 1991.
[5] K.C. Kwak and W. Pedrycz, “Face recognition using an enhanced independent component analysis approach,” IEEE Trans. Neural Netw., vol.18, no.2, pp.530–541, Feb. 2007.
[6] W. Zhao, R. Chellappa, and A. Krishnaswamy, “Discriminant analysis of principal components for face recognition,” 3rd International Conference on Automatic Face and Gesture Recognition, 1998.
[7] Z.M. Hafed and M.D. Levine, “Face recognition using the discrete cosine transform,” Int. J. Comput. Vis., pp.167–188, 2001.
[8] J. Yang, Z.D. Frangi, A. F, and J.Y. Yang, “Two-dimensional pca: a new approach to appearance-based face representation and recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol.26, no.1, pp.131–137, Jan. 2004.
[9] K. S. and R. Y., “Face recognition using 2dlda algorithm,” Signal Processing and Its Applications of the Eighth International Symposium, vol.2, pp.675–678, Aug. 2005.
[10] X. Li, Y. Pang, and Y. Yuan, “L1-norm-based ZPCA,” IEEE Trans. Syst., Man Cybern., B, Cybern., vol.40, no.4, pp.1170–1175, Aug. 2009.
[11] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol.28, no.12, pp.2037–2041, Dec. 2006.
[12] G. Zhao and P. M., “Dynamic texture recognition using local binary patterns with an application to facial expressions,” IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.6, pp.915–928, June 2007.
[13] S. Ullman, M. Vidal-Neto, and E. Sali, “Visual features of intermediate complexity and their use in classification,” Nature Neurosci, vol.5, pp.682–687, 2002.
[14] W. Zhou, A. Ahrary, and S. Kamata, “Face recognition with local feature patterns and histogram spatially constrained earth mover’s distance,” ICSIPA, pp.374–379, Nov. 2009.
[15] W. Zhou, A. Ahrary, and S. Kamata, “Face recognition using local quaternion patters and weighted spatially constrained earth movers distance,” ISCE, pp.285–289, 2009.
[16] B. Zhang, Y. Gao, S. Zhao, and J. Liu, “Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor,” IEEE Trans. Image Process., vol.19, no.2, pp.533–544, Feb. 2010.
[17] W. Zhou, A. Ahrary, and S. Kamata, “Image description with 1d local patterns by multi-scans: An application to face recognition,” ICASSP, pp.1221–1224, 2007.

[18] S. Liao and A. Chung, “Texture classification by using advanced local binary patterns and spatial distribution of dominant patterns,” IEEE Trans. Image Process., vol.18, no.5, pp.1107–1115, May 2009.

[19] S. Liao, M. w.k. Lawand, and A. Chung, “Dominant local binary patterns for texture classification,” IEEE Trans. Image Process., vol.18, no.5, pp.1107–1115, May 2009.

[20] P.J. Phillips, H. Moon, P.J. Rauss, and S. Rizvi, “The feret evaluation methodology for face recognition algorithms,” IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.10, pp.1090–1104, 2000.

[21] P.J. Phillips, P. Flynn, T. Scruggs, and K.W. Bowyer, “Overview of the face recognition grand challenge,” CVPR, vol.1, pp.947–954, 2005.

[22] V. Vapnik, “The nature of statistical learning theory,” Springer, 1995.

[23] C. Chih-Chung and Chih-Jen, “LIBSVM: A library for support vector machines,” ACM Trans. Intelligent Systems and Technology, vol.2, pp:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[24] R. Schapire and Y. Singer, “Improved boosting algorithms using confidence-rated predictions,” Mach. Learn., vol.37, pp.297–336, 1999.

[25] G. Zhang, X. Huang, Y.W.S.Z. Li, and X. Wu, “Boosting local binary pattern (LBP)-based face recognition,” Chinese Conference on Biometric Recognition (SINOBIOMETRICS), pp.179–186, 2004.

Wei Zhou received the B.E. degree in software engineering from Nanjing University, Nanjing, China, in July 2007, and the M.E. degree in information engineering from Waseda University, Japan in Sept. 2009. He is now pursuing the Ph.D. degree at the Graduate School of Information, Production and Systems, Waseda University, Fukuoka, Japan. His current research is focus on face understanding and object classification.