Shape Primitive Histogram: A Novel Low-Level Face Representation for Face Recognition

Sheng Huang[a], Dan Yang[a,b,*], Haopeng Zhang[c], Luwen Huangfu[d], Xiaohong Zhang[b,e]

[a]*College of Computer Science at Chongqing University,400044,Chongqing,P.R.C*
[b]*School of Software Engineering at Chongqing Univeristy,400044,Chongqing,P.R.C*
[c]*School of Astronautics at Beihang University, 100191, Beijing, P.R.C*
[d]*State Key Laboratory of Management and Control for Complex Systems*
Institute of Automation, Chinese Academy of Sciences, 100190,Beijing, P.R.C
[e]*Ministry of Education Key Laboratory of Dependable Service Computing in Cyber Physical Society, 400044, Chongqing, P.R.C*

Abstract

Human face contains abundant shape features. This fact motivates a lot of impressive shape feature-based face detection and 3D face recognition approaches. However, as far as we know, there is no prior low-level face representation which is purely based on shape feature proposed for conventional 2D (image-based) face recognition. In this paper, we present a novel low-level shape-based face representation named *Shape Primitives Histogram* (SPH) for face recognition. In this approach, the face images are separated into a number of tiny shape fragments and we reduce these shape fragments to several uniform atomic shape patterns called *Shape Primitives*. Then the face representation is obtained by implementing a histogram statistic of shape primitives in a local image region. In order to take scale information into consideration, we also produce Multi-scale Shape Primitive Histograms (MSPH) by concatenating the SPHs extracted from different scales. Moreover, we experimentally study the influences of each stage of SPH computation on performance, concluding that a small cell with 1/2 overlap and a fine size block with 1/2 overlap are important for good results. Four popular face databases, namely ORL, AR, YaleB and LFW-a databases, are employed to evaluate SPH and MSPH. Surprisingly, such seemingly naive shape-based face representations outperform the state-of-the-art low-level face representations.

Keywords: Face Recognition, Haar Wavelet, Image Representation, Image Descriptor, Image classification

1. Introduction

Face recognition is a fundamental task in biometrics, which is widely applied in our life. As the core of face recognition, the quality of the face representation is the key for improving face recognition performance, since it is generally considered that the representation always determines the upper limit of the classification accuracy.

Many researchers have made efforts to find the effective face representations in recent decades. Generally speaking, the face representation approaches can be roughly categorized into two classes. The first one is called appearance-based approach which uses multivariate statistic analysis technique to learn a specific subspace from original high-dimensional sample space. This approach may start with the influential Eigenface [1] and has produced many classical methods such as Fisherface [2] and Laplacianface [3, 4, 5, 6]. Moreover, this approach is also known as a dimensionality reduction technique for data analysis perspective. The second one is the low-level image feature-based face representation. This approach utilizes the pattern among the local pixels to represent and distinguish the faces. Garbor feature [7, 8, 9] and Gradient feature [10, 11] are deemed as two most common adopted low-level face representations. Both Gabor and gradient features are good at capturing the edge information of faces. Although extensive studies have proved their effectiveness, these methods are sensitive to the noise and local geometric transformation. In order to address the previous issues, some researchers develop a popular branch of low-level image representation named local histogram descriptor [12, 13, 14, 15, 16, 17, 18]. This approach will do the histogram statistics of the low-level image feature in a local image regions after the

---

(a) (b) (c)

Figure 1: The faces represented by Pixels,Gradient,Shape Primitive respectively

low-level image feature extraction. The main merit of local histogram descriptor over the conventional low-level image feature is that it is more insensitive to the local geometrical transformations and noises. Currently, Local Binary Patterns (LBP) [15, 19, 20, 21] and Histogram of Oriented Gradient (HOG) [11, 12, 13, 14, 22] are the two most influential local histogram descriptors for face recognition. LBP exploits the local binary pattern among the pixels in a local circular region of the image. It is originally designed for texture description [15]. HOG is used to exploit the gradient orientation patterns in the image. Before being used as a face representation, it is known as a successful human detection feature [12]. Although so many impressive low-level face representation methods have been proposed, most of those methods are based on gradient or edge information. However, human face contains abundant shape feature and the shape features are widely used in face detection [23, 24, 25, 26] and 3D face recognition [27, 28, 29, 30]. So, in this paper, we intend to present a low-level face representation purely based on shape features for conventional 2D face recognition.

A popular way for shape feature extraction is Haar wavelet. Such approach has been demonstrated to be considerably successful in object detection field, especially in the face detection field [23, 25, 26]. This is mainly due to two facts. The first one is that human face contains abundant static shape characteristics [23, 28]. The second one is Haar feature, which provides an effective way to extract shape characteristics. Consequently, it can provide an attractive trade-off between accuracy and detection speed. The Haar feature based object detection has been in vogue for the past decade or so. Many impressive object detection systems and new Haar wavelet features have been proposed [23, 24, 25, 26]. However, as far as we know, there is no prior work that studies the representational power of Haar feature in the face recognition area. In this paper, we further exploit the representational power of Haar wavelet features to extract shape features for solving 2D face recognition task .

There exist extensive Haar wavelet templates. It is

impracticable to apply all Haar templates for face representation. Instead, we consider that all the Haar wavelet templates that can be approximated by 14 square Haar wavelet templates and one flat template (see Figure 3). These 15 templates correspond to 15 atomic local shape characteristics (patterns). We call these atomic local shape patterns *shape primitives* (see Figure 1). However, if we simply follow the same representation way as method of Viola and Jones [26], the face representation will have an incredibly high dimensionality and suffer more sensitivity from the local noises and geometrical transformations. Moreover, the shape characteristics which can benefit face recognition should be more local and detailed than the ones utilized by the face detection. So we adopt the form of local histogram descriptor to manage and describe the face shape features extracted by the shape primitives. Finally, as other histogram descriptors, we will get a vector by concatenating histograms of all the local image blocks. We name this new image descriptor *Shape Primitive Histogram* (SPH). ORL face database is employed to experimentally learn the optimal parameters of SPH and Three larger face databases, namely AR, Yale-B and LFW-a face databases, are employed for evaluating the different face representations. Extensive experimental results demonstrate that our proposed method outperforms the state-of-the-arts.

The rest of paper is organized as follows: Section 2 presents the generations of SPH and its multi-scale version; Section 3 describes the experiments for evaluating the SPH; Finally, the conclusion is summarized in Section 4.

2. Methodology

This section introduces shape primitive histogram in detail. The generation procedure of shape primitive histogram can be divided into three main steps: image blocking, shape primitive extraction (matching) and histogram computation. Figure 2 depicts a face recognition system based on shape primitive histogram. After getting the SPH features, a dimensionality reduction can be employed for obtaining a more compact representation. And finally, the classifier is implemented for recognizing the faces (see Figure 2).

2.1. Image Blocking

In order to incorporate the local spatial information, we divide the entire image into several same size blocks (local image patches, see the illustration of the second step in Figure 2). These blocks are the smallest unit
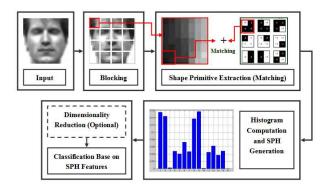
2

Figure 2: the overview of shape primitive histogram based face recognition system

for histogram statistics of shape primitives and determines the locality of SPH. Therefore, Shape primitive histogram is a local descriptor and the size of block can affect its performance. More specifically, if the block size is too small, the representation will be very sparse and thus leads to higher dimensionality. On the contrary, if its size is too big, the representation will be too rough to capture the local shape characteristics. In our case, the block is square. Moreover, the adjacent blocks are overlapping with each other. This strategy can reduce sensitivity to the geometric and photometric transformations [12].

2.2. Shape Primitives Extraction (Matching)

After image blocking, the second step is to extract shape primitives in each obtained image block. No matter how complex a shape is, it can be composed by 15 shape primitives. These shape primitives can be represented by the Haar wavelet templates in Figure 3. The first 14 templates called non-flat shape primitive are used to describe the shape characteristics and the last template is a virtual template named flat shape primitive. The flat shape primitive is applied to handle the case that does't exist shape information.

As same as image blocking procedure, each local image block is divided into dozens of tiny 4^n pixels square fragments. Each fragment is an unit for extracting the shape primitive features. For simplicity, we name such fragment *Cell*. As same as the templates in Figure 3, the cell has four bins. Each bin can be a pixel or a tiny square area which contains 2^n pixels. In order to keep continuity of shape information in a local block, each cell also has 1/2 overlap with the neighbor ones. Generally speaking, the size of cell determines the fineness of extracted local shape patterns. A small size of cell can capture more detailed shape information but also more

sensitive to the noise. In our case, the cell size is fixed to 2×2 pixels and 4×4 pixels, since the size of face image is small.
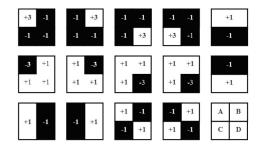

Figure 3: The Shape Primitives corresponding to the Haar wavelet templates.

The convolution operations based on 14 non-flat shape primitive templates are applied to each cell to find its belonging shape pattern. During procedure, 14 matching scores corresponding to 14 non-flat shape primitives are generated as follows:

$$S = \frac{1}{n} \sum_{i=1}^{4} P_i h_i, \quad n = \max(|h_1|, \cdots, |h_4|) \quad (1)$$

where P_i indicates the sum of gray values of bin i and h_i indicates the weighting value corresponding to the bin i. The matching score represents the similarity between the cell and the shape primitive. Therefore, the cell must belong to the shape primitive which owns the maximum matching score, i.e.

$$\hat{i} = \arg\max_i(\mathcal{S}), \quad \mathcal{S} = (S_1, \cdots, S_i, \cdots, S_{14}) \quad (2)$$

However, the cell may not contain any shape information. In this case, the pixels in the cell have the same gray value. So, the matching scores of 14 shape primitives are all zero. In such case, the cell is actually a flat and we assume it is accord with the virtual template known as flat shape primitive. However, in the natural images, there seldom exists the absolute flat. So, for handling this case, we assume that the cell belongs to the flat shape primitive when all the matching scores are smaller than the nonnegative loose factor ϵ ($\epsilon \geq 0$). Finally, the shape primitive extraction (matching) scheme can be further expressed as:

$$\hat{i} = \begin{cases} argmin_i(\mathcal{S}), & M > \epsilon \\ 15, & M \leq \epsilon \end{cases} \quad s.t. \quad \epsilon \geq 0 \quad (3)$$

where $M = \max(\mathcal{S})$ is the maximum among the matching scores and $\hat{i}$ indicates the index of the matched shape

primitive. Since each non-flat shape primitive can find a complementary shape primitive among these 14 shape primitives, M is always a positive or equal to zero, $M \geq 0$.

2.3. Histogram Computation

The histogram of each block has 15 bins corresponding to the 15 shape primitives. We can calculate the matching scores of the first 14 non-flat shape primitives. The flat one cannot achieve a matching score via convolution directly. However, these scores are the weighted votes of the histogram and will be accumulated into related bins during the histogram computation. In order to address this issue, we provide an empirical way to calculate matching score corresponding to the flat. The matching score (the weighting vote) of flat, S_{15}, can be assigned as follows:

$$S_{15} = \begin{cases} \epsilon - M + 1, & M \leq \epsilon \\ 0, & M > \epsilon \end{cases} \quad (4)$$

by this way, each cell can get the weighted vote and the weighted histogram statistics becomes feasible. So, the whole histogram computation procedure is denoted as follows

$$\mathcal{H}_i = \begin{cases} \mathcal{H}_i + \epsilon - M + 1, & M \leq \epsilon, \ i = 15 \\ \mathcal{H}_i + M, & M > \epsilon, \ i = \arg\min_i(\mathcal{S}) \end{cases} \quad (5)$$

where $\mathcal{H}_i$ is the value of the ith bin of the histogram and its initial value is zero. After finishing the histogram computation in each local block, a linear normalization, as $\mathcal{H} = \mathcal{H}/\sqrt{\mathcal{H}^T\mathcal{H}}$, is presented to every block for improving robustness to variation in illumination. Next, all the histograms are concatenated into a 1-D feature vector and this is the SPH feature.

2.4. Multi-Scale Shape Primitive Histogram

The cell size and block size are very important parameters, because they determine the fineness and locality of extracted features. More specifically, the SPH feature extraction via a smaller block with smaller cells can better capture more detailed shape characteristics, but it is hard to capture the more global shape characteristics and is more sensitive to the local noise. However, the quality and scale of images in a practical application are always extremely varied. Thus, a fixed size SPH may not be suitable to the image in an uncontrolled environment. For this case, we provide a very simple way to incorporate the scale information with SPH to improve its robustness. We hypothesize that a scale-robust SPH can be obtained by concatenating all SPH vectors extracted

at different scales together. We name this new feature Multi-Scale Shape Primitives Histogram (MSPH). Note that this combined way is not optimal since the SPH features from different scales have correlations.

3. Experiments

In this section, we conduct several experiments to learn the optimal parameters of SPH on ORL [31] database, which is a smaller database in comparison with the other three databases. AR [32], YaleB [33], LFW-a [34] datasets are employed for evaluating the performances of SPH and MSPH in comparison with three state-of-the-art face representations, namely Gabor Feature [7], Local Binary Patterns (LBP) [19] and Histogram of Oriented Gradient (HOG) [13]. Principal Component Analysis (PCA) [1] and Linear Discriminant Analysis (LDA) [2] are used as dimensionality reduction algorithms. The conventional classifier, Nearest Neighbor Classifier (NNC) and two more advanced classifiers, namely Collaborative Representation Classifier (CRC) [35] and Support Vector Machine (SVM) [36], are adopted for classification.

3.1. Datasets

1. The ORL database contains 400 images from 40 subjects [31]. Each subject has ten images acquired at different times. In this database, the subjects' facial expressions and facial details are varying. And these images are also taken with a tolerance for some tilting and rotation of the face of up to 20°. The size of the face image is 32×32 pixels. Compared to the other three databases, this database is much smaller and we use it to learn the optimal parameters.

2. The AR database consists of more than 4,000 images of 126 subjects [32]. The database characterizes divergence from ideal conditions by incorporating various facial expressions, luminance alterations, and occlusion modes. Following paper [37], a subset contains 1680 images with 120 subjects are constructed in our experiment. All these images are 50×40 pixels.

3. The Extended YaleB database [33] consists of 2,414 frontal face images of 38 subjects under various lighting conditions. In our experiment, all of these images are 32×32 pixels.

4. The LFW-a database [34], which aims at studying the problem of the unconstrained face recognition, is considered as one of the most challenging databases, since it contains 13233 images with
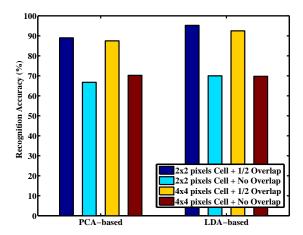
Figure 6: The face recognition performances under different Cell Sizes and Overlaps.

great variations in terms of lighting, pose, age, and even image quality. We cropped these images to 120×120 pixels around their center and resize these images to 32×32 pixels for computational efficiency.

Note, in our experiments, all the images on these four databases are grayscale images. With regard to the color image, the SPH feature can be extracted from each channel and yield them together as the new SPH feature.

3.2. Parameters Learning

SPH has several important parameters such as block size, overlap region, and loose factor which can influence the face recognition performance of SPH. Following [22], the ORL database which is relatively smaller among four databases is utilized for experimentally learning the optimal values of these parameters. For a practical application, the training samples can be used for learning optimal parameters. Note, NNC is chosen as the classifier in these experiments. We adopt a learning strategy that fixes other parameters when one parameter is being experimentally learned.

3.2.1. Sizes and Overlaps of Blocks

In this experiment, four groups of blocks whose sizes are respectively 4×4, 6×6, 8×8, 10×10 pixels and four overlaps include no overlap, 1/4 overlap, 1/2 overlap, 3/4 overlap are adopted to produce a total of $4 \times 4 = 16$ combinations of blocks and overlaps for finding the best parameters of block and overlap. In this experiment, the cell size is fixed to 2×2 pixels and LDA and PCA are applied for dimensionality reduction. Two-fold and five-fold cross-validations are employed for evaluating the

recognition performance. The *n*-fold cross-validation in our paper is defined as: one part for training and the rest *n*-1 parts for testing.

Figure 4 describes the influences of block size and overlap to the recognition performance on the ORL database. Figure 5(a) depicts the comprehensive recognition performance, which is the mean of recognition rates under different cross-validation schemes with different block sizes and overlaps. Figure 5(b) depicts the dimensions of different SPHs under different combinations of block size and overlap. From Figures 4 and 5, it can be obviously concluded that SPH with a smaller block size and larger overlapping area can achieve higher recognition accuracy. However, the dimension of SPH also rapidly increases along with the block size reduction and the overlapping region expansion. In order to balance the recognition accuracy and recognition speed, a block size of 8×8 pixels with 1/2 overlap are deemed as the optimal parameter group.

3.2.2. Sizes and Overlaps of Cells

Since the size of cell should be smaller than the size of block and the optimal block size is 8×8 pixels, we choose two groups of cells whose sizes are respectively 2×2 and 4×4 pixels. Two overlaps include no overlap and 1/2 overlap are combined with these two groups of cells to produce four combinations. We apply two-fold cross-validation to evaluate the produced combinations. These experiments are all conducted under 8×8 pixels block with 1/2 overlap. The results of experiments are shown in Figure 6.

From the observations, we can know that a smaller cell slightly outperforms a larger cell while the overlap in cells is a more important factor to improve the performance. The gain of the cell with 1/2 overlap over the cell without any overlap is around 20%. According to the results, we recommend to adopt the 2×2 pixel size cell with 1/2 overlap for SPH extraction.

3.2.3. The values of Loose Factor epsilon

The loose factor ϵ controls the boundary between flat shape primitive and non-flat shape primitives and influence the flat shape primitive histogram weighting vote. So, it is very important to learn the optimal ϵ. This parameter is related to the cell size f (pixels), since the summation of the gray values of each bin of a bigger cell is greater. It can be denoted as follows:

$$\epsilon = k \times (f/2)^2, k \geq 0 \qquad (6)$$

We can learn the value of k to further achieve the optimal ϵ. In this experiment, two groups of SPH: 8×8 pixels block with 1/2 overlap, 2×2 pixels cell and 16×16

(a) LDA-based two-fold cross-validation

(b) PCA-based two-fold cross-validation

(c) LDA-based five-fold cross-validation
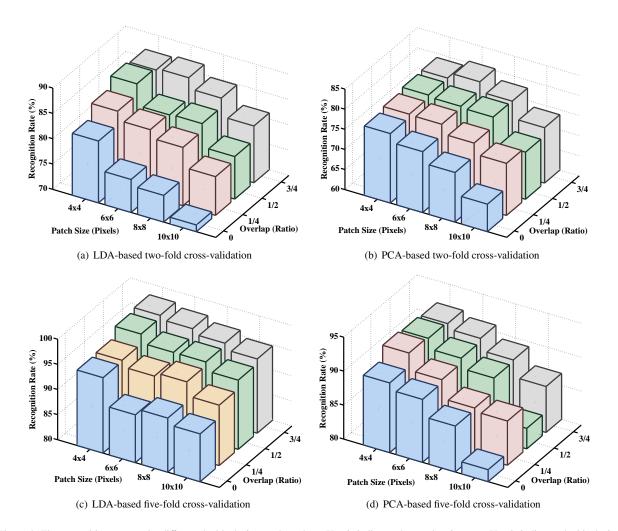
(d) PCA-based five-fold cross-validation

Figure 4: The recognition rates under different the block sizes and overlaps. X axis indicates the overlapping area, Y axis indicates the block size and Z axis indicates the recognition rate.

pixels block with 1/2 overlap, 4×4 pixels cell are used to study the effect of ϵ. The two-fold and five-fold cross-validation schemes are adopted in these experiments.

According to the observations of Figure 7, which indicates the recognition accuracies under different k, we can conclude that two groups of SPH are all insensitive to k. Consequently, we let $k = 1$. In other words, $\epsilon = f^2/4$ in this paper.

Feature	Parameters Configurations
SPH	8×8 pixels block, 1/2 overlap, 2×2 pixels cell, $k = 1$
MSPH	8×8 pixels block, 1/2 overlap, 2×2 pixels cell, $k = 1$ 16×16 pixels block, 1/2 overlap, 4×4 pixels cell, $k = 1$ 32×32 pixels block, 1/2 overlap, 8×8 pixels cell,$k = 1$

Table 1: Parameters settings of SPH and MSPH in experiments

3.3. Face Recognition

Three larger face databases including AR, YaleB and LFW-a databases are employed for evaluating the face recognition performance. Among these three face databases, the LFW-a database is a face database in uncontrolled environment which is a very challenging database aiming at the evaluation of face recognition in the wild. The sample number of each subject in this database is very different. Following paper [38], we divide the LFW-a database into two subsets. The first subset (147 subjects, 1100 samples) is constructed by the subjects whose sample numbers are ranged from 5 to 10 and the second subset (127 subjects, 2891 samples) is constructed by the subjects whose sample numbers are all over 11. We apply leave-one-out cross-evaluation scheme to these two subsets. With
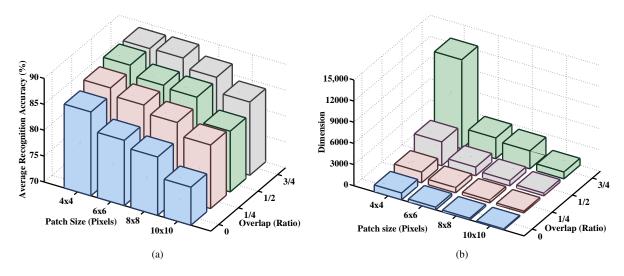
(a)

(b)

Figure 5: (a) the comprehensive recognition rate, which is the mean of recognition rates under different cross-validations (see Figure 4). Y axis indicates block sizes, X axis indicates overlaps and Z indicates comprehensive recognition rates. (b) the dimensions of SPH under different block sizes and overlaps, Z indicates the dimension.
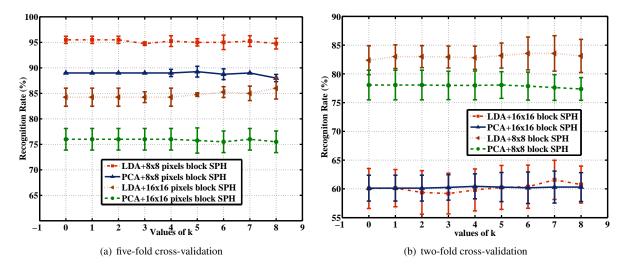

(a) five-fold cross-validation

(b) two-fold cross-validation

Figure 7: the recognition rates under different *k*.

regard to AR database, we employ 7-fold and 4-fold cross-validation schemes. While, the cross-validation schemes of YaleB are 8-fold and 5-fold. The parameters configurations of SPH and MSPH in these experiments are introduced in Table 1. The parameter settings of LBP, Gabor and HOG are mainly following [19], [7] and [13] respectively. But the block sizes of HOG and LBP are slightly tuned for fitting the face image size in our experiments (the sizes of images in their experiments are quite different to us). The block sizes of LBP and HOG are all 16×16 pixels and each block has 1/2 overlap with the neighbour ones.

Tables 2, 3 and 4 show the face recognition accuracies using different face representations on AR, YaleB and LFW-a databases respectively. It is obvious that SPH outperforms the other three state-of-the-art representations on AR and YaleB databases under all three classifiers. For example, SPH obtains averagely 3.5%, 4.5% and 2.25% gain over the second top representation using NNC, CRC and SVM classifiers respectively on AR database. On YaleB database, these numbers are respectively 8.25%, 7.75% and 12%. Additionally, the MSPH can also maintain the second place on these two databases in the most of time. With regards to the ex-

7

Representation		Cross-Validation Schemes-Recognition Rate (%)					
		NNC		CRC		SVM	
		7-fold	4-fold	7-fold	4-fold	7-fold	4-fold
PCA	**SPH**	82.06±4.5%	84.51±5.6%	83.73±5.3%	86.93±6.4%	**83.79±5.3%**	**87.03±6.5%**
	MSPH	80.72±4.8%	83.12±5.3%	82.54±6.0%	85.28±7.0%	82.94±6.3%	85.70±7.7%
	LBP [19]	71.27±6.7%	76.69±6.9%	76.30±8.2%	81.69±9.7%	76.43±8.8%	81.95±9.9%
	Gabor [7]	76.17±5.9%	78.16±7.5%	77.00±7.9%	79.89±10.2%	77.24±8.6%	80.15±11.2%
	HOG [13]	80.30±4.4%	83.69±4.3%	81.34±5.7%	85.59±6.3%	81.21±5.9%	85.02±6.8%
LDA	**SPH**	**83.16±6.0%**	**86.67±7.3%**	**84.56±5.2%**	**88.28±6.1%**	80.71±6.4%	86.25±7.3%
	MSPH	79.29±7.1%	85.01±8.2%	81.84±6.1%	84.92±8.4%	80.39±7.2%	84.32±8.6%
	LBP [19]	74.38±8.4%	82.48±8.0%	76.67±8.4%	82.63±8.5%	76.40±8.4%	83.01±8.5%
	Gabor [7]	79.74±7.2%	83.30±10.2%	79.59±7.7%	84.09±9.5%	77.84±8.2%	83.54±10.0%
	HOG [13]	75.18±6.8%	80.66±8.0%	75.59±7.6%	75.02±10.3%	75.11±8.0%	74.94±10.3%

Table 2: Recognition accuracies on AR Database

Representation		Leave-one-out Cross-Validation-Recognition Rate (%)					
		NNC		CRC		SVM	
		8-fold	5-fold	8-fold	5-fold	8-fold	5-fold
PCA	**SPH**	52.51±11.2%	60.00±3.2%	**57.49±11.6%**	**66.74±6.0%**	60.99±12.4%	**70.02±5.6%**
	MSPH	45.68±11.2%	53.56±3.4%	49.78±11.5%	59.65±4.7%	52.27±11.0%	63.15±4.5%
	LBP [19]	33.27±12.3%	40.59±7.2%	37.58±12.3%	47.40±7.6%	37.74±11.5%	46.26±6.0%
	Gabor [7]	43.30±9.1%	51.63±7.5%	49.99±12.3%	58.70±6.2%	48.22±8.3%	57.53±6.0%
	HOG [13]	39.31±11.2%	46.64±2.2%	43.33±10.6%	51.98±3.9%	43.00±8.3%	51.28±3.4%
LDA	**SPH**	**58.88±9.3%**	**66.65±6.0%**	53.95±9.3%	55.35±4.5%	55.40±9.0%	56.60±5.1%
	MSPH	55.67±9.0%	62.58±6.7%	53.18±9.0%	57.25±5.7%	54.24±9.0%	57.04±5.1%
	LBP [19]	41.03±11.2%	49.11±3.4%	37.47±10.3%	35.42±4.6%	38.05±10.5%	35.01±4.8%
	Gabor [7]	50.45±4.5%	57.29±6.4%	48.04±6.9%	53.98±2.5%	38.05±10.5%	35.01±4.8%
	HOG [13]	39.31±11.2%	48.03±3.6%	32.06±8.3%	31.21±3.6%	32.55±7.9%	31.31±3.8%

Table 3: Recognition accuracies on YaleB Database

Representation		Leave-one-out Cross-Validation-Recognition Rate (%)					
		NNC		CRC		SVM	
		subset1	subset2	subset1	subset2	subset1	subset2
PCA	**SPH**	19.05±2.5%	19.83±3.6%	25.62±5.6%	37.65±4.4%	26.53±2.5%	40.09±4.1%
	MSPH	19.05±2.7%	20.83±3.6%	26.53±5.7%	**38.23±4.2%**	27.66±2.8%	41.26±3.9%
	LBP [19]	22.79±2.5%	23.77±4.5%	31.63±1.1%	38.11±5.1%	29.82±0.7%	42.66±4.6%
	Gabor [7]	15.42±1.6%	19.40±2.4%	17.91±2.2%	30.42±3.1%	19.61±1.6%	32.87±5.4%
	HOG [13]	19.27±3.5%	22.62±2.8%	25.17±2.2%	34.38±4.2%	23.81±3.4%	35.20±4.3%
LDA	**SPH**	24.94±3.5%	35.00±4.3%	29.71±3.4%	37.65±4.2%	30.84±3.4%	44.38±3.6%
	MSPH	**27.66±2.8%**	**35.79±3.6%**	**32.09±2.7%**	36.95±4.3%	**32.31±2.7%**	**44.60±4.4%**
	LBP [19]	23.58±2.9%	30.35±5.3%	28.68±4.2%	32.40±4.7%	28.91±3.5%	41.66±4.4%
	Gabor [7]	22.56±3.0%	33.65±5.6%	26.30±3.0%	33.57±3.3%	26.30±2.6%	43.59±4.9%
	HOG [13]	18.71±2.3%	24.84±3.8%	22.68±2.8%	27.04±4.2%	22.90±3.1%	32.86±3.1%

Table 4: Recognition accuracies on LFW-a Database

perimental results on the LFW-a database, MSPH gets a better performance than SPH and defeats other compared methods under all three different classifiers. The reason why the MSPH gets a better performance than SPH using the LFW-a database may be the fact that the samples in the LFW-a database suffer from more variation in image resolution. Thus, the MSPH can perform better. Besides that, Table 4 demonstrates that SPH also obtains very promising performance.

For better demonstrating the superiority of our method and studying the influence of dimensionality reduction algorithms to the low-level face representations, we conduct several experiments on AR and YaleB databases to draw the retained dimensions versus recognition accuracies curves in Figure 8. On AR database, the first 10 samples per subject are used for training while the rest samples are for testing. On YaleB database, the first 48 samples per subject are used for training while the rest samples for testing. From the observations in Figure 8, we can know that SPH outper-
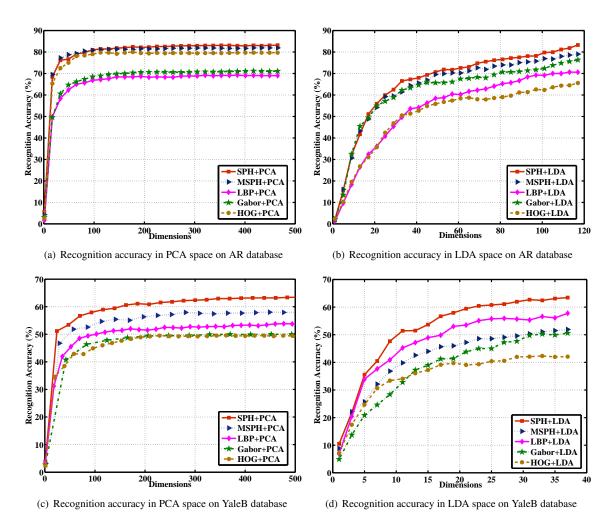
(a) Recognition accuracy in PCA space on AR database

(b) Recognition accuracy in LDA space on AR database

(c) Recognition accuracy in PCA space on YaleB database

(d) Recognition accuracy in LDA space on YaleB database

Figure 8: Recognition Accuracy versus Retained Dimensions.

forms other comparsived methods in all dimensions while MSPH also outperforms the compared methods except in the experiments on YaleB database that uses LDA for dimensionality reduction. The results of experiments clearly demonstrate that SPH is more discriminative than the other three face representations.

3.4. Feature Extraction Efficiency

For more comprehensively evaluating SPH, we also test the feature extraction efficiencies of different low-level face representations on ORL database and report the results in Figure 9. The experimental hardware configuration is CPU: 2.5 GHz, RAM: 8G. The results of experiments in Figure 9 clearly show that feature extraction speeds of SPH and MSPH are competitive.
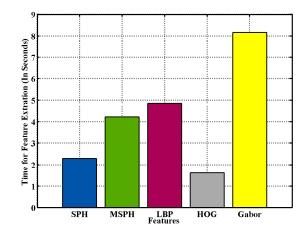

Figure 9: The extraction time of different features on ORL database.

9

4. Conclusion

In this paper, we have proposed a simple but effective low-level face representation for face recognition. This representation focuses on highlighting the shape characteristics of face and supposes human face can be divided into a group of small shape fragments, which share a series of uniform shape patterns named *Shape Primitives*. A histogram of shape primitives is computed in each local block and be concatenated as a 1-D vector to represent the face. Moreover, we have also generated a multi-scale shape primitive histogram via concatenating the different scale SPH vectors together. Three well known face databases were used for validating the proposed methods. The experimental results demonstrate the superiority of SPH in comparison with the state-of-the-art low-level face representation methods.

There are many worthwhile works of our method can be further exploited. For example, this descriptor can be combined with facial landmark localizations or face interesting region selections to increase the recognition accuracy [13, 39, 40]. Currently, face recognition via fusing different features is a very popular trend for face recognition [41]. So, we can also fuse SPH with other state-of-the-art low-level features to present a more powerful representation. Moreover, applying SPH to face expression analysis, face alignment and face spoofing detection [21] are also interesting directions.

Acknowledgement

References

[1] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, January 1991.

[2] Peter N. Belhumeur, P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 711–720, 1997.

[3] Xiaofei He and Partha Niyogi. Locality preserving projections. In *Advances in Neural Information Processing Systems (NIPS)*. MIT Press, 2003.

[4] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong jiang Zhang. Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:328–340, 2005.

[5] Huang Sheng, Ahmed Elgammal, Luwen Huangfu, Dan Yang, and Xiaohong Zhang. Globality-locality preserving projections for biometric data dimensionality reduction. In *IEEE conference on Computer Vision and Pattern Recognition Workshop on Biometrics (CVPRW)*, 2014.

[6] Sheng Huang, Dan Yang, Yongxin Ge, Dengyang Zhao, and Xin Feng. Discriminant hyper-laplacian projections with its applications to face recognition. In *IEEE conference on Multimedia and Expo Workshop on HIM (ICMEW)*, 2014.

[7] Chengjun Liu and Harry Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 11(4):467–476, 2002.

[8] Chengjun Liu and Harry Wechsler. Independent component analysis of gabor features for face recognition. *IEEE Transactions on Neural Networks*, 14:919–928, 2003.

[9] Yong Xu, Zhengming Li, Jeng-Shyang Pan, and Jing-Yu Yang. Face recognition based on fusion of multi-resolution gabor features. *Neural Computing and Applications*, 23(5):1251–1256, 2013.

[10] Taiping Zhang, Yuan Yan Tang, Bin Fang, Zhaowei Shang, and Xiaoyu Liu. Face recognition under varying illumination using gradientfaces. *IEEE Transactions on Image Processing*, 18(11):2599–2606, 2009.

[11] Ngoc-Son Vu. Exploring patterns of gradient orientations and magnitudes for face recognition. *IEEE Transactions on Information Forensics and Security*, 8(2):295–304, 2013.

[12] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 886–893, 2005.

[13] O. Déniz, G. Bueno, J. Salido, and F. De la Torre. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, 32(12):1598–1603, September 2011.

[14] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, September 2010.

[15] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, July 2002.

[16] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[17] Zhimin Cao, Qi Yin, Xiaoou Tang, and Jian Sun. Face recognition with learning-based descriptor. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2707–2714. IEEE, 2010.

[18] Rakesh Mehta, Jirui Yuan, and Karen Egiazarian. Face recognition using scale-adaptive directional and textural features. *Pattern Recognition*, 47(5):1846–1858, 2014.

[19] Abdenour Hadid Timo Ahonen and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:2037–2041, 2006.

[20] Di Huang, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang, and Liming Chen. Local binary patterns and its application to facial image analysis: A survey. *IEEE Transactions on Systems, Man, and Cybernetics Part C*, 41(6):765–781, November 2011.

[21] J Maatta, A Hadid, and M Pietikainen. Face spoofing detection from single images using texture and local shape analysis. *IET*

Biometrics, 1(1):3–10, 2012.

[22] Alberto Albiol, David Monzo, Antoine Martin, Jorge Sastre, and Antonio Albiol. Face recognition using hog-ebgm. *Pattern Recognition Letters*, 29(10):1537–1543, July 2008.

[23] Constantine Papageorgiou and Tomaso Poggio. A trainable system for object detection. *International Journal of Computer Vision*, 38(1):15–33, June 2000.

[24] Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In *IEEE International Conference on Image Processing (ICIP)*, pages 900–903, 2002.

[25] Sri-Kaushik Pavani, David Delgado, and Alejandro F. Frangi. Haar-like features with optimally weighted rectangles for rapid object detection. *Pattern Recognition*, 43(1):160–172, January 2010.

[26] Paul A. Viola and Michael J. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 511–518, 2001.

[27] Yueming Wang, Jianzhuang Liu, and Xiaoou Tang. Robust 3d face recognition by local shape difference boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10):1858–1870, 2010.

[28] Berk Gökberk, M Okan İrfanoğlu, and Lale Akarun. 3d shape-based face representation and feature extraction for face recognition. *Image and Vision Computing*, 24(8):857–869, 2006.

[29] Peijiang Liu, Yunhong Wang, Di Huang, Zhaoxiang Zhang, and Liming Chen. Learning the spherical harmonic features for 3-d face recognition. *IEEE Transactions on Image Processing*, 22(3):914–925, 2013.

[30] Chafik Samir, Anuj Srivastava, and Mohamed Daoudi. Three-dimensional face recognition using shapes of facial curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1858–1863, 2006.

[31] F. S. Samaria, F. S. Samaria, A.C. Harter, and Old Addenbrooke. Parameterisation of a stochastic model for human face identification, 1994.

[32] Aleix Martínez and Robert Benavente. The ar face database, Jun 1998.

[33] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:643–660, 2001.

[34] Lior Wolf, Tal Hassner, and Yaniv Taigman. Similarity scores based on background samples. In *Asian Conference on Computer Vision (ACCV)*, pages 88–97, Berlin, Heidelberg, 2009. Springer-Verlag.

[35] D Zhang, Meng Yang, and Xiangchu Feng. Sparse representation or collaborative representation: Which helps face recognition? In *IEEE International Conference on Computer Vision (ICCV)*, pages 471–478, 2011.

[36] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2, 2011.

[37] Imran Naseem, Roberto Togneri, and Mohammed Bennamoun. Linear regression for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):2106–2112, 2010.

[38] Meng Yang, D Zhang, and Shenlong Wang. Relaxed collaborative representation for pattern classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2224–2231. IEEE, 2012.

[39] Lin Zhong, Qingshan Liu, Peng Yang, Bo Liu, Junzhou Huang, and Dimitris N Metaxas. Learning active facial patches for expression analysis. In *IEEE Conference on Computer Vision and*

Pattern Recognition (CVPR), pages 2562–2569, 2012.

[40] Xiang Yu, Junzhou Huang, Shaoting Zhang, Wang Yan, and Dimitris N Metaxas. Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model. In *IEEE International Conference on Computer Vision (ICCV)*, 2013.

[41] Zhen Cui, Wen Li, Dong Xu, Shiguang Shan, and Xilin Chen. Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3554–3561, 2013.