Face Representation and Recognition with Local Curvelet Patterns

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

SUMMARY In this paper, we propose Local Curvelet Binary Patterns (LCBP) and Learned Local Curvelet Patterns (LLCP) for presenting the local features of facial images. The proposed methods are based on Curvelet transform which can overcome the weakness of traditional Gabor wavelets in higher dimensions, and better capture the curve singularities and hyperplane singularities of facial images. LCBP can be regarded as a combination of Curvelet features and LBP operator while LLCP designs several learned codebooks from patch sets, which are constructed by sampling patches from Curvelet filtered facial images. Each facial image can be encoded into multiple pattern maps and block-based histograms of these patterns are concatenated into an histogram sequence to be used as a face descriptor. During the face representation phase, one input patch is encoded by one pattern in LCBP while multi-patterns in LLCP. Finally, an effective classifier called Weighted Histogram Spatially constrained Earth Mover’s Distance (WHSEMD) which utilizes the discriminative powers of different facial parts, the different patterns and the spatial information of face is proposed. Performance assessment in face recognition and gender estimation under different challenges shows that the proposed approaches are superior than traditional ones.

key words: face recognition, gender estimation, local curvelet binary patterns, learned local curvelet patterns, WHSEMD

1. Introduction

During the last several decades, 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 advantages due to its 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 so on. However, due to the fact that the facial appearances are easily affected by the variations of expression, illumination, pose and other factors, it is still an active and challenging research topic.

According to how the elements of the face representation are calculated, face methods can be coarsely categorized into global feature based and local feature based methods. Specifically, each element of a global feature is related to the whole input facial image, while each element of local feature is extracted from some local region of the facial image. It is worth pointing out that, there is no clear boundary between them.

Most of the global methods are subspace methods that reduce the dimension of the image. The Principle Component Analysis (PCA) [2], which uses an eigenvalue subspace to project the whole image into several weights and uses the distances between these weights to recognize faces, is the famous method in this class. Independent Component Analysis (ICA) [3] takes higher-order statistics into account, and is suitable for learning complex structure in the dataset. Linear Discriminant Analysis (LDA) [4], which considered the difference both between-class and within-class matrix, and Discrete Cosine Transform (DCT) [5], which can remain more linear property. They are another three well-known methods in this category. Recently, in order to reduce computational time cost and preserve two dimension information of images, 2DPCA [6], 2DLDA [7], 2DPCA-L1 [8] are researched. However, all of these methods are global representation of images which are sensitive to global changes of images, such as, illumination and expression.

In order to solve these difficulties, local matching approaches are presented in face recognition [9], [10] and other visual recognition tasks [11] with 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 is called Local Binary Patterns (LBP) [9]. Recently, since Gabor wavelet has good characteristics in space frequency, space position and direction selectivity, local patterns based on Gabor feature have also been proposed for face representation, such as Local Gabor Binary Patterns (LGBP) [12], Histogram of Gabor Phase Patterns (HGPP) [13] and Learned Local Gabor Patterns (LLGP) [14]. Unlike LBP or LGBP, in LLGP, the patterns are learned by applying the clustering approach to the set of patches, which are sampled from the Gabor filtered facial images. Then, the facial image is encoded into multiple pattern maps based on the learned patterns. However, the common issue of these methods is that the feature dimension is very large due to Gabor decomposition, and there is no clear discuss about why the Gabor transform is the best frequency space for representing facial images? Actually, Gabor transform cannot well represent curve singularity of human facial images since Gabor wavelets are very effective in representing objects with isolated point singularities, but failed to represent line or curve singularities.
To overcome the weakness of Gabor wavelets in higher dimensions, and to better capture the curve singularities and hyperplane singularities of high dimensional signals, Candes and Donoho [15] proposed Curvelet transform. Curvelet transform directly takes edges as the basic representation elements and is strongly anisotropic. It is optimal in representing curved singularities in images or higher dimensional signals. The detail and fine coefficients of Curvelet are strongly orientation-sensitive, which is a useful property for detecting curves in images. In [16], comparison of wavelet, Gabor wavelet and Curvelet transform for face recognition under illumination and expression changes is discussed and concluded that Curvelet is a better choice compared with wavelet and Gabor wavelet, since the Curvelet transform has a more sparse representation of the image than wavelet, thus offering a description with higher time frequency resolution and high degree of directionality and anisotropy, which is particularly appropriate for many images rich with edges and curves.

Generally, Curvelet feature based face recognition is just using some subspace to project the Curvelet coefficients. For example, in [17], [18] and [19], PCA is used while in [20], LDA is considered. None of these methods are face-specific, and do not consider the spatial information about our face. In addition, these methods are sensitive to illumination and facial expression change.

In this paper, unlike the common methods which used Gabor wavelet to transform facial images into frequency space and overcome the problems of traditional Wavelet and Curvelet feature based face recognition, Local Curvelet Patterns are studied. Based on Curvelet filtered facial images, two local Curvelet patterns called Local Curvelet Binary Patterns (LCBP) and Learned Local Curvelet Patterns (LLCP) are proposed. Then, the facial image is encoded into multiple pattern maps based on LBP operator and some learned patterns, respectively. Finally, the block-based histograms of the patterns are concatenated together to describe the input facial image. During face representation part, multi-patterns are used to encode the input patch which is sampled from Curvelet filtered facial image in LLCP, since these encoded multi-patterns have closed similarities with the input patch. In addition, we propose an effective classifier called Weighted Histogram Spatially constrained Earth Mover’s Distance (WHSEMD) which utilizes the discriminative powers of different facial parts, the different patterns and the spatial information of face. In order to evaluate the proposed methods, face recognition and gender estimation are studied. Experiments on three famous and challenging databases-FERET [21] LFW [22] and FRGC [23] show the effectiveness of the proposed methods.

The remainder of this paper is organized as follows: Some relationship between this work and our previous work will be described in Sect. 2. In Sect. 3, Curvelet Transform will be introduced briefly and face representation based on LCBP and LLCP will be described in Sect. 4 and 5, respectively. The classifier called WHSEMD will be introduced in Sect. 6. Experimental results are presented in Sect. 7. Finally, conclusions and future work are discussed in Sect. 8.

2. Previous Work

The proposed methods in this paper belong to local matching group. General speaking, there are many directions to be researched in this area. In our previous work [24], [25], first, multi-scans are used to encode the neighborhood of the patterns and concluded that it can reserve more spatial information with lower cost than the traditional neighborhood encoding, such as circle neighborhood based encoding. Second, in order to improve the robustness of local patterns, not just binary, ternary and quaternion number are used to encode the neighborhood structure, which is more robust to noise and illumination change. However, the multi-scans neighborhood based encoding is also based on gray-level intensity which is sensitive to the outside factors change, and the type of patterns are predefined which are not face-specific. Thus, in this study, first, we utilize Curvelet feature instead of gray-level intensity to learn the local patterns. Since this kind of transformation is expect to capture salient visual properties such as spatial frequency, spatial localization and orientation selectivity, so the patterns defined on Curvelet feature are more effective than those defined on gray feature. Second, clustering method is used to learn a face-specific codebook from the training set, which is more suitable for face perception tasks. Additionally, in order to judge whether multi-scans neighborhood based encoding is effective or not in the frequency space, some experiments are conducted, which have illustrated that multi-scans neighborhood based encoding can also achieve comparable or better performances. One point should be noted is that in this study, for simplicity, just binary number is used for encoding the local patterns. But it is easy to extend to use ternary and quaternion number for encoding.

3. Curvelet Transform

Curvelet aims to deal with interesting phenomena occurring along curved edges in a 2D image. As illustrated in [26], curvelet needs fewer coefficients for representation and the edge produced from Curvelet is smoother than wavelet edge.

Curvelet transform is a special member of the multi-scale geometric transforms. It is a transform with multi-scale pyramid with many directions at each length scale. Curvelets will be superior over wavelets in following cases:

1. Optimally sparse representation of objects with edges

2. Optimal image reconstruction in severely ill-posed problems

3. Optimal sparse representation of wave propagators

The newly constructed and improved version of Curvelet transform is known as Fast Discrete Curvelet Transform (FDCT). This new technique is simpler, faster and less redundant than the original Curvelet transform which based on ridgelets. According to Candès et al in [15], two implementations of FDCT are proposed: 1) Unequally-
spaced Fast Fourier transforms (USFFT). 2) Wrapping function. Both implementations of FDCT differ mainly by the choice of spatial grid that used to translate curvelets at each scale and angle. Both digital transformations return a table of digital curvelet coefficients indexed by a scale parameter, an orientation parameter and a spatial location parameter. Wrapping-based transform is based on wrapping a specially selected Fourier samples, and it is easier to implement and understand.

The new implementation of Curvelet transform based on Wrapping of Fourier samples takes a 2D image as an input in the form of a Cartesian array $f[m,n]$, where $0 \leq m < M, 0 \leq n < N$ where $M$ and $N$ are the dimensions of the array. As illustrated in Eq. (1), the outputs will be a collection of Curvelet coefficients $c^D(u, v, k_1, k_2)$ indexed by a scale $u$, an orientation $v$ and spatial location parameters $k_1$ and $k_2$.

$$c^D(u, v, k_1, k_2) = \sum_{m,n} f[m,n] \varphi^D_{u,v,k_1,k_2} [m,n]$$  \hspace{1cm} (1)

Each $\varphi^D_{u,v,k_1,k_2}$ is a digital Curvelet waveform, superscript $D$ stands for “digital” [15]. These approach implementations are the effective parabolic scaling law on the subbands in the frequency domain to capture curved edges within an image in more effective way. Wrapping based Curvelet transform is a multiscale pyramid which consists of several subbands at different scales consisting of different orientations and positions in the frequency domain. At a high frequency level, Curvelets are so fine and looks like a needle shaped element and they are non-directional coarse elements at low frequency level.

In our study, the facial image is decomposed into coarse, detail and fine coefficients and some reconstructed images including coarse layer, two detail layers and one fine layer are illustrated in Fig. 1. Further, reconstructed images by four orientations of detail 2 layer are shown in Fig. 2. Here, CurveLab 2.1.2 which is available at [27] is used.

4. Face Representation with LCBP

The overall framework of the proposed representation approach based on Local Curvelet Binary Pattern is illustrated in Fig. 3. In this approach, a facial image is modeled as a “histogram sequence” by the following procedure: First, the original facial image is transformed to obtain multiple Curvelet Pictures (CPs) in frequency domain by applying Curvelet filters; Second, each CP is converted to Local Curvelet Binary Pattern (LCBP) map by LBP operator; Third, each LCBP Map is further divided into several blocks, and histogram is computed for each region; Fourth, the LCBP histograms of all the LCBP Maps are concatenated to form the final histogram sequence as the model of
the face.

4.1 LCBP

LCBP can be regarded as a combination of Curvelet feature and LBP operator (Note: In our study, in order to reduce feature dimension, only uniform binary patterns are used). The LBP operator [9] labels the pixels of an image by thresholding the $3 \times 3$ neighborhood of each pixel $Z_i (i = 0, 1, \ldots, 7)$ with the center pixel $Z_c$ and considering the result as a binary number. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is considered circular [9].

4.2 LCBP Histogram Sequence

Since histogram statistical presentation is successfully used in local feature areas, LCBP Histogram Sequence is considered which can be generalized from LCBP maps in our study. First, each LCBP Map is spatially divided into $B$ blocks. And then histogram is extracted from each block. Finally, all the histograms estimated from the blocks of all the LCBP Maps are concatenated into a single histogram sequence for LCBP map $L_a(.)$ ($a = 0, 1, 2, 3$ corresponding to coarse layer, detail 1 layer, detail 2 layer and fine layer in our implementation, respectively), which is formally defined as follows:

$$H_a(.) = [H_{a,0}, H_{a,1}, \ldots, H_{a,B-1}]$$  \hspace{1cm} (2)

Here $H_{a,i}(i = 0, 1, \ldots, B - 1)$ denotes the histogram of the $i$-th block of LCBP map $L_a(.)$ and is formulated as follows:

$$H_{a,i}(j) = \sum_Z B[L_a(Z) = j], j = 0, 1, \ldots, K - 1$$  \hspace{1cm} (3)

where $j$ denotes the index of the patterns, $K$ denotes the number of the patterns, $Z$ is the pixel location in LCBP map and $B(.)$ is defined as follows:

$$B[A] = \begin{cases} 1, & \text{if } A \text{ is true} \\ 0, & \text{otherwise} \end{cases}$$  \hspace{1cm} (4)

Finally, one facial image $I$ is represented as one histogram sequence, which concatenates all the block based histograms of all LCBP pattern maps with various $C$ layer:

$$H(I) = [H_0(I), H_1(I), H_2(I), \ldots, H_C(I)]$$  \hspace{1cm} (5)

5. Face Representation with LLCP

Different from the patterns used in LCBP which are predefined in the same way for face and non-facial images, several particular codebook patterns are learned in LLCP (shown in Fig. 4). First, Images are sampled into many patches and then all the patches are cluttered into $K$ patterns by random-projection tree [28]. Generally speaking, face encoding by LLCP can be divided into learning phase and representing phase. In the learning phase, several particular codebooks are constructed while facial images are encoded in the representing phase.

In the learning phase (shown in Fig. 5), each image in the training set is reconstructed to CPs with different Curvelet coefficients. Then based on all CPs in the same layer, one patch set can be constructed by sampling patches. At last, by using random-projection tree clustering approach to each patch set, LLCP learned codebooks can be constructed. Thus, $C$ LLCP codebooks can be obtained with $C$ layers.

One point should be noted during our implementation is that the sampling methods used are illustrated as Fig. 6, which called sample 1 and sample 2, respectively.

In the representing phase (shown in Fig. 7), first, each facial image is reconstructed to several CPs with different
Curvelet coefficients. Then each CP is encoded into \( t \) pattern maps by mapping its patches to the corresponding LLCP codebook pixel by pixel (Note: here the intensity of mapped image is corresponding to the type of learned patterns, and also the same intensity in different LLCP maps stands for different type of patterns since the codebook is different. More important one is that one CP can be encoded by \( T \) LLCP maps and reason will be explained later). Thus, totally \( C \times T \) LLCP maps can be obtained. Finally, these pattern maps are spatially divided into many blocks and the histograms of all the blocks are concatenated together to form one enhanced histogram sequence which is considered to represent the input facial image.

Here, Euclidean distance \( D_s(L_{a,t}(Z), j) = ||P_{L_{a,t}(Z)} - C_j||_2 \) is used to calculate the similarity score between the input patch and the codewords \( L_{a,t}(\cdot) \) stands for LCBP map in layer \( a \) where the encoded pattern is \( t \)-th smallest distance codeword \( (t = 1, 2, \ldots, T) \). \( C_j \) is the codeword for pattern \( j \) and \( P_{L_{a,t}(Z)} \) is the input patch at position \( Z \) in \( L_{a,t}(\cdot) \). One point should be noted during representing phase is that the input patch can match top \( T \)-th smallest distance codewords in the codebook. The reason is that the input patch is very closed to the top \( T \)-th smallest distance codewords. And at most time, we can not judge which one is the best matching codeword. One example is shown in Fig. 8 where codeword 0 and codeword 2 have the same similarity score, so in such case we can not just select one and ignore the other. That is, the traditional one to one mapping is not useful in this situation and multi-mapping is applied to overcome the problem.

So when histogram encoded by LLCP is computed, Eq. (3) should be changed as following

\[
H_{a,j}(f) = \sum_i \sum_{j \in N'(i)} w_{a,i,f} g(H_{a,j}^{i_p}, H_{a,j}^{i_c})
\]

(7)

where \( w_{a,i} \) is the weighting value for different facial parts corresponding to layer \( a \) and block \( i \), \( f_j \) represents the flow (weighting value) from the histogram \( H_{a,j}^{i_p} \) and histogram \( H_{a,j}^{i_c} \) as the index set of \( r \) nearest blocks in the spatial domain from block \( i \) in image \( I_P \). Function \( g(\cdot) \) is the distance between two histogram defined as

\[
g(H_{a,j}^{i_p}, H_{a,j}^{i_c}) = \sum_{k=0}^{K-1} \bar{\omega}_k \min(H_{a,j}^{i_p}(k), H_{a,j}^{i_c}(k))
\]

(8)

here, \( \bar{\omega}_k \) is the weighting value corresponding to pattern \( k \). The calculation of \( w_{a,i} \) and \( \bar{\omega}_k \) is presented in [12] and [14], respectively. Note: in face representation with LCBP, \( \bar{\omega}_k \) is always set to 1.

Formally, for each block in image \( I_P \), we got nearest blocks in the spatial domain from image \( I_G \), bounded by a distance of \( \sqrt{2} \) in this work. Low pass filter (Gaussian Kernel is used in our study) is applied around to reduce noise and get a smoother score for classification. Thus, WHSEMD is slightly expect to be invariant to pose change.

### 6. WHSEMD

In this section, a new variant called Weighted Histogram Spatially constrained Earth Mover’s Distance (WHSEMD) is proposed. It is an improvement version of Spatially constrained Earth Mover’s Distance (SEMD) [29]. To reduce the computation cost, we represent probe Image \( I_P \) and gallery Image \( I_G \) with non overlapping and overlapping blocks, respectively, as well as constrain each block in image \( I_P \) to be matched only to blocks within a local spatial neighborhood in image \( I_G \). With the spatial neighborhood constraint and non overlapping representation for image \( I_P \), the total number of parameters is reduced. Thus, the computational cost is significantly reduced.

The (directional) distance from probe images \( I_P \) to gallery \( I_G \) is defined as

\[
d(I_P \rightarrow I_G) = \sum_a \sum_{i \in N'(i)} \sum_{j \in N'(i)} w_{a,i,f \in j} g(H_{a,j}^{i_p}, H_{a,j}^{i_c})
\]

(7)

where \( w_{a,i} \) is the weighting value for different facial parts corresponding to layer \( a \) and block \( i \), \( f_j \) represents the flow (weighting value) from the histogram \( H_{a,j}^{i_p} \) and histogram \( H_{a,j}^{i_c} \) as the index set of \( r \) nearest blocks in the spatial domain from block \( i \) in image \( I_P \). Function \( g(\cdot) \) is the distance between two histogram defined as

\[
g(H_{a,j}^{i_p}, H_{a,j}^{i_c}) = \sum_{k=0}^{K-1} \bar{\omega}_k \min(H_{a,j}^{i_p}(k), H_{a,j}^{i_c}(k))
\]

(8)

here, \( \bar{\omega}_k \) is the weighting value corresponding to pattern \( k \). The calculation of \( w_{a,i} \) and \( \bar{\omega}_k \) is presented in [12] and [14], respectively. Note: in face representation with LCBP, \( \bar{\omega}_k \) is always set to 1.

Formally, for each block in image \( I_P \), we got nearest blocks in the spatial domain from image \( I_G \), bounded by a distance of \( \sqrt{2} \) in this work. Low pass filter (Gaussian Kernel is used in our study) is applied around to reduce noise and get a smoother score for classification. Thus, WHSEMD is slightly expect to be invariant to pose change.

### 7. Experimental Results

In this section, the proposed methods are evaluated on face recognition, gender estimation problems. Three well-known databases - FERET [21], Labeled Faces in the Wild (LFW) [22] and FRGC [23] are considered.
7.1 FERET Database

The FERET database consists of a total of 14,051 gray-scale images representing 1,199 individuals. The images contain variations in lighting, facial expressions, pose angle, etc. In this work, only frontal faces are considered. These facial images can be divided into five sets as follows: 1) Fa set, used as a gallery set, contains frontal images of 1,196 people. 2) Fb set (1,195 images). The subjects were asked for an alternative facial expression than in the Fa photograph. 3) Fc set (194 images). The photos were taken under different lighting conditions. 4) Dup I set (722 images). The photos were taken later in time. 5) Dup II set (234 images). This is a subset of the dup I set containing those images that were taken at least a year after the corresponding gallery image. And in our study all facial images are normalized to 80 × 88 pixels and the procedure of the FERET evaluation principle [21] is followed.

In order to determine how the parameters affect the final recognition rate, several experiments are evaluated on Fb and Dup I datasets. Figure 9 and Fig. 10 show the recognition rates change with the size of codebook \( K \) and different sampling methods (Here, the block size is fixed to 8 × 8 and just histogram intersection is used for classification). From these two figures, we can find that the performance becomes better with the increase of \( K \). When \( K \) is from 256 to 512, the recognition rate stay stable and just 0.5% improved from 128 to 256. So for the sake of trade-off between precision and computation cost, \( K = 128 \) is used in the following. For the sampling methods, sample 2 is better than sample 1 while the combination of them can achieve the best result.

Next experiment is based on the change of block size and sampling methods while the \( K \) is fixed to 128. The results are shown in Fig. 11 and Fig. 12. From these two figures, we can seen that small or large block size can decrease the performance, especially with larger one. When the block size is about 8 × 8, the best accuracy can be obtained. The possible reason may be that larger block size can not preserve enough spatial information in the facial images while the patterns in smaller block can not discriminate effectively. For the sampling method, same as the previous one, the combination can get the toppes rate.

The third experiment is evaluated on the size of training set and different sampling methods and the results are shown in Fig. 13 and Fig. 14. We can see that only small dataset can be conducted robust recognition rate while larger ones do not increase the final performance.

The fourth experiment is designed for judge whether our multi-mapping is useful or not. The evaluation based on Dup I is list in Fig. 15. We can see that a little larger \( \beta \)
can improve our performance which is same as our thinking. And if $\beta$ is so large, that means the input patch is also encoded by some dissimilar patterns which can confuse the distribution of final histogram and decrease our final result.

The fifth experiment is designed to judge how the recognition rate changes with variation of image size. The evaluation result based on FERET dataset is list in Fig. 16 where the combined sampling method is used. We can see that when the image size is larger than $80 \times 88$, the performance is almost same, while just a little improved in Dup I and Dup II subsets with the image size from $100 \times 100$ to $128 \times 128$. Thus, image size with $80 \times 88$ is sufficient for this kind evaluation in this dataset.

At last, the total performance on FERET dataset is list in Table 1 (Note that: in LCBP and LLCP, just histogram intersection is used for classification and proposed classifier is applied in Weighted LCBP and Weighted LLCP methods). From Table 1, we can get the effectiveness of Curvelet transform and the proposed classifier. LCBP+multi-scans and Weighted LCBP+multi-scans mean that when encoding the neighborhood of the patterns, multi-scans are used instead of the traditional circular based [25]. We can see that multi-scans neighborhood based encoding is effective in the Curvelet based frequency domain.

### 7.2 FRGC Database

The FRGC [23] dataset consists of over 50,000 frontal facial recordings of more than 4,00 subjects including frontal views with different facial expressions, lighting conditions. For the experiments reported in this section, 200 different individuals were randomly selected from this database and each subject has 8 images. Then there are totally 1600 images in our experiments. All the images are manually cropped and resized to $80 \times 88$ pixels and divided into $8 \times 8$
blocks in our study. Some examples are shown in Fig. 17 (b).

In this evaluation, some \( d(d = 1, 2, 3, 4) \) images of each person are randomly chosen for training, while the remaining images for testing. To compare our method with LBP, LGBP, five tests are performed with a varying number of training samples and mean rate is recorded. Table 2 shows the accuracy. From this table, we can see again that Curvelet based local patterns are more effectiveness than the traditional local patterns and Weighted LLCP is the outstanding one.

### 7.3 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. 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 [31], [32]. In our system, three kernel functions are concerned as Table 3. \( g \) is equal 3 in our system.

The first experiment is to evaluate how the precision changes with variation of the number of regions. Figure 18 and Fig. 19 show the results based on LFW and FRGC datasets, respectively. 0, 1, 2 means the kernel used in SVM is dot product, polynomial and RBF, respectively. From these two figures, we can see than when the number of regions is set about 64, the performance achieves best, and decreases when the number of regions is larger or smaller. The reason may be same as discussed in face recognition.
problem.

The total experimental results are shown in Table 4 and Table 5, respectively. Here, the number of regions is set to 64. From these two tables, we can see that proposed methods can get more discriminant features about male and female than LBP and LGBP under different variations. And frontal faces are more useful for gender estimation than profile.

### 8. Conclusions and Future Work

In this paper, two novel approaches called Local Curvelet Binary Patterns (LCBP) and Learned Local Curvelet Patterns (LLCP) for presenting the local patterns are proposed. The represented facial images can better capture the curve singularities and hyperplane singularities than some traditional methods, such as LGBP, LLGBP. LCBP uses some predefined patterns to encode the facial image while LLCP learned some codebooks from sampled patches which are regraded as face-specific and more desirable for face perception tasks. During face representation phase, multi-mapping is used in LLCP which is more reasonable than traditional one to one mapping. In addition, we propose an effective classifier called Weighted Histogram Spatially constrained Earth Mover’s Distance (WHSEMD) which utilizes the discriminative powers of different facial parts, the different patterns and the spatial information of face. Experiments in both face recognition and gender estimation show that the proposed methods have better performance than relevant ones.

In future, in order to reduce feature dimension further, some feature selection strategies should be considered, such as adaboost. And how to solve the pose problem in our face recognition is the other major direction.

### Acknowledgements

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 Edu-

cation, 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] M.A. Turk and A.P. Pentland, “Face recognition using eigenfaces,” CVPR, 1991.

[3] 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.

[4] 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.

[5] Z.M. Hafed and M.D. Levine, “Face recognition using the discrete cosine transform,” Int. J. Comput. Vis., vol.43, no.3, pp.167–188, 2001.

[6] J. Yang, D. Zhang, A.F. Frangi, 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.

[7] S. Kongsontana and Y. Rangsanseri, “Face recognition using 2dlda algorithm,” Signal Processing and Its Applications of the Eighth International Symposium, vol.2, pp.675–678, Aug. 2005.

[8] X. Li, Y. Pang, and Y. Yuan, “L1-norm-based 2DPCA,” IEEE Trans. Syst. Man Cybern., B, Cybern., vol.40, no.4, pp.1170–1175, Aug. 2009.

[9] 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.

[10] G.Y. Zhao and M. Pietikainen, “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.

[11] S. Ullman, M. Vidal-Naquet, and E. Salì, “Visual features of intermediate complexity and their use in classification,” Nature Neurosci., vol.5, pp.682–687, 2002.

[12] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, “Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition,” IEEE Trans. Neural Netw., vol.18, no.1, pp.1–10, Jan. 2007.

[13] B. Zhang, S. Shan, X. Chen, and W. Gao, “Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition,” IEEE Trans. Image Process., vol.16, no.1, pp.57–68, 2007.

[14] S. Xie, S. Shan, X. Chen, X. Meng, and W. Gao, “Learned local gabor patterns for face representation and recognition,” Signal Process., vol.89, no.12, pp.2333–2344, Dec. 2009.

[15] E. Candès, L. Demanet, D. Donoho, and L. Ying, “Fast discrete curvelet transforms,” Multiscale Modeling and Simulation, vol.5, no.3, pp.861–899, 2006.

[16] J. Zhang, Y. Wang, Z. Zhang, and C. Xia, “Comparison of wavelet, gabor and curvelet transform for face recognition,” Optica Applicata, vol.XLI, no.1, pp.183–193, 2011.

[17] T. Mandal and Q.I. Wu, “Face recognition using curvelet based pca,” ICPR, pp.1–4, 2008.

[18] H. Huo and E. Song, “Face recognition using curvelet and selective pca,” ICICIP, pp.348–351, 2010.

[19] S. Rahman, S. Motahar, A.A. Farooq, and M.M. Islam, “Curvelet texture based face recognition using principal component analysis,” ICCIT, pp.45–50, 2010.

[20] J. Zhang and Y. Wang, “A comparative study of wavelet and curvelet transform for face recognition,” CISP, pp.1718–1722, 2010.

[21] 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.
[22] G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” Tech. Rep. 07-49, University of Massachusetts, Amherst, Oct. 2007.
[23] 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.
[24] W. Zhou, A. Ahrary, and S. Kamata, “Image description with 1d local patterns by multi-scans: An application to face recognition,” ICIP, Sept. 2010.
[25] W. Zhou, A. Ahrary, and S. Kamata, “Image description with local patterns: An application to face recognition,” IEICE Trans. Inf. & Syst., vol.E95-D, no.5, pp.1494–1505, May 2012.
[26] L. Boubchir and J. Fadili, “Multivariate statistical modelling of images with the curvelet transform,” Image Processing Group, pp.747–750, 2005.
[27] available at http://www.curvelet.org
[28] available at http://cseweb.ucsd.edu/naverma/RPTrees/index.html
[29] D. Xu, S. Yan, and J. Luo, “Face recognition using spatially constrained earth mover’s distance,” IEEE Trans. Image Process., vol.17, no.11, pp.2256–2260, Nov. 2008.
[30] 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.
[31] V. Vapnik, The nature of statistical learning theory, Springer, 1995.
[32] C. Chih-Chung and L. Chih-Jen, “LIBSVM: A library for support vector machines,” ACM Transactions on Intelligent Systems and Technology, vol.2, pp.27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

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 Sep. 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.

Alireza Ahrary received M.E. in Industrial Science and Technology Engineering and PhD in Electronic Engineering from the Niigata University and Kyushu Institute of Technology, Japan, in 2004 and 2007, respectively. In 2004, he joined GMD-Japan as research assistance. He was a research scientist at FAIS- Robotics Development Support Office, which took over GMD-Japan in April 2004. Since 2007, he has been a postdoctoral research fellow of the Japan Society for the Promotion of Science (JSPS) at Waseda University. From 2008 to 2010, he served as researcher of Fukuoka Industry, Science and Technology Foundation (Fukuoka IST) and guest researcher at IPSRC, Waseda University, Japan. From 2010 to 2012, he is an Assistant Professor in the Department of Human and Computer Intelligence, Nagasaki Institute of Applied Science. Currently, he is an Associate Professor in the Department of Computer and Information Science, Sojo University. His research interests are Pattern Recognition, Biometric, Robotics and Computer Vision. He is a member of IEEE, RSJ and IIEEJ.

Sei-ichiro Kamata received the M.S. degree in computer science from Kyushu University, Fukuoka, Japan, in 1985, and the doctor of Engineering degree from the Department of Computer Science, Kyushu Institute of Technology, Kitakyushu, Japan, in 1995. From 1985 to 1988, he was in NEC, Ltd., Kawasaki, Japan. In 1988, he joined the faculty at Kyushu Institute of Technology. From 1996 to 2001, he was an Associate Professor in the Department of Intelligent Systems, Graduate School of Information Science and Electrical Engineering, Kyushu University. Since 2003, he has been a professor in Graduate School of Information, Production and Systems, Waseda University. In 1990 and 1994, he was a Visiting Researcher at the University of Maine, Orono. His research interests include image processing, pattern recognition, image compression, remotely sensed image analysis, Image database and space-filling curve and fractals. Prof. Kamata is a member of the IEEE, and the ITE in Japan.