Robust Local Features for Remote Face Recognition

Jie Chen[1], Vishal M. Patel[2], Li Liu[1,4], Vili Kellokumpu[1], Guoying Zhao[1], Matti Pietikäinen[1], and Rama Chellappa[3]

[1] Center for Machine Vision and Signal Analysis, University of Oulu, Finland
[2] Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ, USA
[3] Department of Electrical and Computer Engineering, University of Maryland, College Park, USA
[4] College of Information System and Management, National University of Defense Technology, China
{jiechen, lili, kello, gyzhao, mkp}@ee.oulu.fi; vishal.m.patel@rutgers.edu; rama@umiacs.umd.edu

Abstract

In this paper, we propose a robust local descriptor for face recognition. It consists of two components, one based on a shearlet-decomposition and the other on local binary pattern (LBP). Shearlets can completely analyze the singular structures of piecewise smooth images, which is useful since singularities and irregular structures carry useful information in an underlying image. Furthermore, LBP is effective for describing the edges extracted by shearlets even when the images contain high level of noise. Experimental results using the Face Recognition Grand Challenge (FRGC) dataset show that the proposed local descriptor significantly outperforms many widely used features (e.g., Gabor and deep learning based features) when the images are corrupted by random noise, demonstrating the strong noise robustness of our approach. In addition, experimental results show promising results for two challenging datasets which have poor image quality, i.e., a remote face dataset and the Point and Shoot Face Recognition Challenge (PaSC) dataset.

Index Terms—**robust local feature, remote face recognition**

1. Introduction

Feature descriptor is a key factor in the performance of many computer vision and pattern recognition applications. A plethora of feature descriptors has been developed to improve the performance for these applications. There are several studies that evaluate the performance of these methods, such as [41, 42]. These methods can be divided into two classes: one is learning-based deep features in supervised, weakly supervised or unsupervised way, attempting to model high-level abstractions in data by using architectures composed of multiple non-linear transformations [24]. The other one is traditional local features, nowadays often called hand-crafted features inspired by neuroscience studies, e.g., Weber local descriptor (WLD) [8], Gabor [12], scale-invariant feature transform (SIFT) [38], and local binary pattern (LBP) [43].

For learning-based deep features, a typical approach is deep learning (DL), introduced by Hinton [24]. It performs very well for hand-written digit recognition [25], face recognition [47, 48], human pose estimation [50] and object recognition [30]. The main criticism of deep learning comes from the observation that it requires tons of annotated training data. Despite the power of deep learning methods, they still lack much of the functionality needed for realizing this goal entirely [40].

For the class of hand-crafted local features, typical examples are LBP [43], Gabor [12], and SIFT [38]. These local features achieved very good performance for texture classification, face recognition and object recognition. Specifically, Ojala et al. proposed a simple but very powerful local descriptor, i.e., local binary pattern (LBP). It is one of the best performing texture descriptors and has been widely used in various applications, such as textures classification and face recognition [43]. Chen et al. developed a robust local binary pattern [9]. Zhang et al. proposed to use Gabor and LBP for face recognition [56]. Lowe introduced SIFT, which performs well for matching and recognition tasks [38].

However, one issue of feature descriptors (learning-based deep features and hand-crafted local features) is that they are not robust to the noise present in images. Vincent proposed the denoising Autoencoders (dA) to improve deep learning-based methods [51], which showed good results when the input data to the autoencoders was contaminated with noise. In this paper, we propose a local descriptor robust to noise. It is a hand-crafted local feature, which achieves good performance when the scale of the training set is not sufficient for deep learning (see Section 3 for details.). The proposed local descriptor consists of two components. One is based on the shearlet decomposition and the other on local binary patterns (LBP). Shearlets can detect the edges in images even when they have high level of noise. Furthermore, LBP is effective in describing the edges extracted by shearlets. The proposed descriptor is called LSF since it combines LBP and shearlet and the shearlet transformation is performed in the Fourier domain. Experimental results on the Face Recognition Grand Challenge (FRGC) dataset show that the proposed local descriptor significantly outperforms many widely used features (e.g., Gabor and deep learning based features) when noise is present in the images. In addition, experimental results show promising results for two challenging datasets

which have poor image quality, i.e., a remote face dataset [5] and the Point and Shoot Face Recognition Challenge (PaSC) dataset [2].

1.1 Related works

In this section, we introduce some related works and discuss the difference between these works and the proposed method.

Lim developed the discrete shearlet transform (DST) which provides efficient multiscale directional representation and showed how to implement the shearlet transform by multiresolution analysis (MRA). He assessed the performance of DST in image denoising and approximation applications [35]. Dong et al., [14] and He et al., [23] used shearlets for texture classification. Different from them, we use shearlets and LBP for face recognition. In addition, we tested different thresholds and different filters for the shearlet transformation (see Table 2). We found that only the real part of the shearlet transformation in the Fourier domain is good for face recognition. We also tested the robustness of shearlets and LBP over high level of noise and compared to state-of-the-art methods, such as the deep learning-based method [24]. Furthermore, we tested the proposed method for two challenging datasets, i.e., a remote face dataset [5] and PaSC [2], and obtained promising results.

Face recognition, as one of the most typical applications of image analysis and understanding, has attracted significant attention in many areas such as entertainment, information security, law enforcement, and surveillance [26, 32, 58]. There are quite a lot of methods presented recently, e.g., [1, 3, 4, 7, 11, 13, 27, 45, 54, 56]. Specifically, Danti and Zeng used shearlets for face recognition and achieved promising results for some face datasets, i.e., ORL and FERET [11, 54]. In contrast, we use shearlet transform for face recognition in a different setting, that is, a remote face dataset [5] and PaSC [2]. Zhang et al. [56] used Gabor plus LBP for face recognition. Zhang et al. [33] used the phase patterns for face recognition. Vu and Caplier [52] enhanced patterns of oriented edges magnitude for face recognition and image matching. Tan and Triggs [49] proposed local texture feature sets for face recognition under difficult lighting conditions.

In this paper, we employ shearlets and LBP for dealing with heavy noise in face recognition because the shearlets form robust features and can detect edges elegantly even in the presence of high level of noise. Technically, we divide each face into blocks and use an individual classifier for each block and then combine the similarity scores from all the blocks for better performance.

2. Method

In this section, we first introduce the shearlet transformation and LBP. We then discuss methods for combining them to represent face images.

2.1 Shearlet transform

The continuous wavelet transform provides a decomposition of a signal over dilated and translated versions of a fixed waveform ψ. Specifically, for a fixed $\psi \in L^2(R^2)$, this is defined as the mapping W_ψ with domain $L^2(R^2)$ such that for $g \in L^2(R^2)$

$$W_\psi g(a,t) = \int_{R^2} g(u)\overline{\psi_{a,t}(u)}du, \qquad (1)$$

where $\psi_{a,t}(u) = a^{-1}\psi(a^{-1}(u-t))$, $a > 0$, $t \in R^2$ and $\bar{\psi}$ is the complex conjugate. If the function ψ satisfies the admissibility or Calderòn condition $\int_0^\infty |\Psi(a\omega)|^2 \frac{da}{a} = 1$ for a.e. $\omega \in R^2$ (where Ψ denotes the Fourier transform of ψ), then ψ is referred to as a *wavelet*, and any $g \in L^2(R^2)$ can be recovered via the reproducing formula:

$$g = \int_0^\infty \int_{R^2} \langle g, \psi_{a,t} \rangle \psi_{a,t} dt \frac{da}{a^4}. \qquad (2)$$

Despite the success of wavelets in signal and image processing applications, it is known mathematically that traditional wavelets are not very effective in dealing with multidimensional signals containing discontinuities such as edges. This is due to the fact that this transform is isotropic (the analyzing elements $\psi_{a,t}$ are obtained by applying the same dilation factor for all coordinate directions) and, as a result, it has a very limited ability to resolve edges and other distributed discontinuities which usually occur in multidimensional data.

In this section, we briefly describe a multi-scale and multi-directional representation called the *shearlet* transform [16]. The shearlet transform combines the power of multi-scale methods with the ability to capture the geometry of multidimensional signals and is essentially optimal in representing images containing edges.

The shearlet construction can be considered as a natural extension of wavelets into two-dimensions [16]. Its representative elements are defined by the two-dimensional affine system

$$\left\{ \tilde{\psi}_{ast}(x) = |\det M_{as}|^{-\frac{1}{2}} \tilde{\psi}(M_{as}^{-1}x - t) : t \in R^2 \right\}, \qquad (3)$$

where

$$M_{as} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$$

is a product of a shearing and anisotropic dilation matrix for $(a, s) \in R^+ \times R$. The generating function $\tilde{\psi}$ is such that

$$\tilde{\Psi}(\zeta) = \tilde{\Psi}(\zeta_1, \zeta_2) = \tilde{\Psi}_1(\zeta_1)\tilde{\Psi}_2(\frac{\zeta_2}{\zeta_1}), \qquad (4)$$

where $\tilde{\Psi}_1$ is a continuous wavelet for which $\tilde{\Psi}_1 \in C^\infty(R)$ with $supp\, \tilde{\Psi}_1 \subset [-2, 1/2] \cup [1/2, 2]$, and $\tilde{\Psi}_2$ is chosen so that $\tilde{\Psi}_2 \in C^\infty(R)$, $supp\, \tilde{\Psi}_2 \subset [-1, 1]$, with $\tilde{\Psi}_2 > 0$ on (-1, 1), and $\|\tilde{\Psi}_2\|_2 = 1$. Under these assumptions, a function $f \in L^2(R^2)$ can be represented as

$$f(x) = \int_{R^2} \int_{-\infty}^{\infty} \int_0^{\infty} \langle f, \tilde{\psi}_{ast} \rangle \tilde{\psi}_{ast}(x) \frac{da}{a^3} ds\, dt , \qquad (5)$$

for $a \in R^+$, $s \in R$, and $t \in R^2$. The operator SH defined by

$$SH(f; a, s, t) = \langle f, \tilde{\psi}_{ast} \rangle \qquad (6)$$

is referred to as the *continuous shearlet transform* of $f \in L^2(R)$. It is dependent on the scale variable a, the shear s, and the location t. Frequency support of the shearlets for different values of a and s is shown in Fig. 1. In addition, the shearlets are defined on the Cartesian domain and the various directions are obtained from the action of shearing transformations.
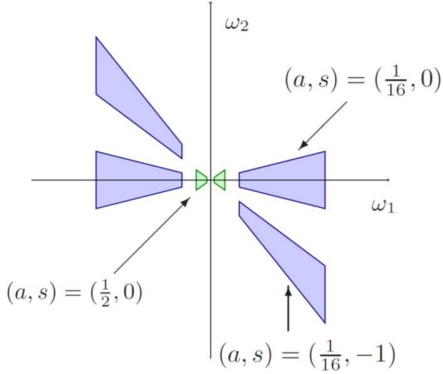

Fig. 1. Frequency support of the shearlets for different values of a and s.
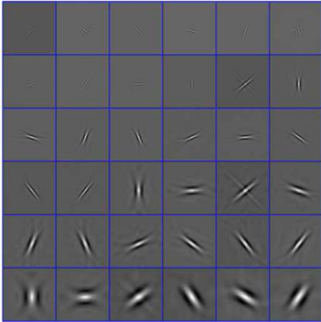

Fig. 2. A few atoms from a shearlet dictionary. Each block represents the result of the shearlet transform for a particular scale and orientation after applying it to a centered impulse response.

The collection of *discrete shearlets* is given by [16, 35]

$$\left\{ \tilde{\psi}_{j,l,k} = |\det A|^{\frac{j}{2}} \tilde{\psi}(B^l A^j x - k) : j, l \in Z, \ k \in Z^2 \right\}, \qquad (7)$$

where

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \ A = \begin{pmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{pmatrix}.$$

Note that the discrete shearlet transform is complex. Please refer to [16, 35] for details.

The matrices A^j (i.e., anisotropic scaling matrices) and B^l (i.e., shear matrices) lead to windows that can be elongated along arbitrary directions and the geometric structures of singularities in images (Fig. 1). In fact, one can approximate 2-D piecewise smooth functions with singularities with nearly optimal approximation rate using shearlets. In addition, shearlets can completely analyze the singular structures of piecewise smooth images. These properties of shearlets are useful in image processing especially since singularities and irregular structures carry essential information in an underlying image. For example, discontinuities in the intensity of an image indicate the presence of edges [35]. For more details about the discrete shearlets and how to implement the shearlet transformation in frequency domain, please refer to [16].

Shearlets form a Parseval frame (tight frame with bounds equal to 1) for $L^2(R^2)$ given the appropriate choice of the generating function $\tilde{\psi}$ [16]. An M-channel filter bank implementation can be done by using the techniques given in [17]. As a consequence, its implementation has a complexity of $O(N^2 \log_2(N))$ for an $N \times N$ image. Fig. 2 shows some basis elements from a shearlet dictionary.

2.2 Local binary pattern

The basic form of LBP is illustrated in Fig. 3 (a) and (b) [43]. The operator takes as input a local neighborhood around each pixel and thresholds the neighborhood pixels at the value of the central pixel. The resulting binary-valued string is then weighted as follows:

$$LBP(I_c) = \sum_{i=0}^{P-1} 2^i s(I_i - I_c), \qquad (8)$$

where the parameter P means the number of the neighbors (e.g., $P = 8$ in Fig. 3), and I_c is the central pixel. I_c and I_i are the gray-level values at c and i, and $s(A)$ is 1 if $A \geq 0$ and 0 otherwise.
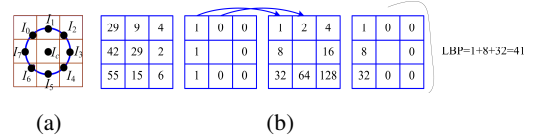

Fig. 3. LBP. (a) A pixel and its eight neighbors; (b) the basic LBP

2.3 Combining shearlets and LBP

We use the shearlet coefficients and LBP to represent faces. For a given image, we perform the shearlet transform and then compute the LBP features of images resulting from the shearlet transform. Specifically, as shown in Fig. 4, given an image or a patch, we have two steps. For the first step, we perform the shearlet transform. In this step, we perform the fast Fourier transform (FFT) first and then carry out the shearlet transformation as discussed in Section 2.1. We keep the real part of the resulting images after taking the shearlet transformation. We then employ a filter derived from 1-D using maximally flat mapping function with two vanishing moments for denoising, better frequency selectivity and regularity [10]. The second step is to compute the LBP feature as discussed in Section 2.2.

For the filter in our framework, we use the nonsubsampled contourlet transform (NSCT) developed by [10]. NSCT is based on a nonsubsampled pyramid structure and nonsubsampled directional filter banks. The resulting decomposition by NSCT is a flexible multiscale,
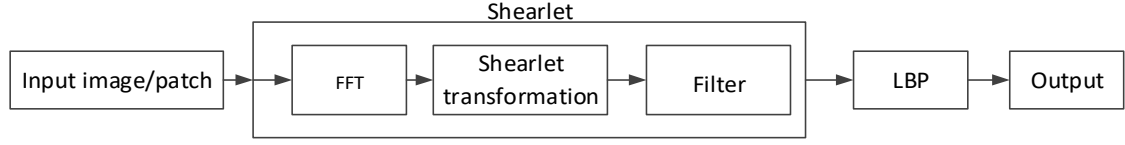
3

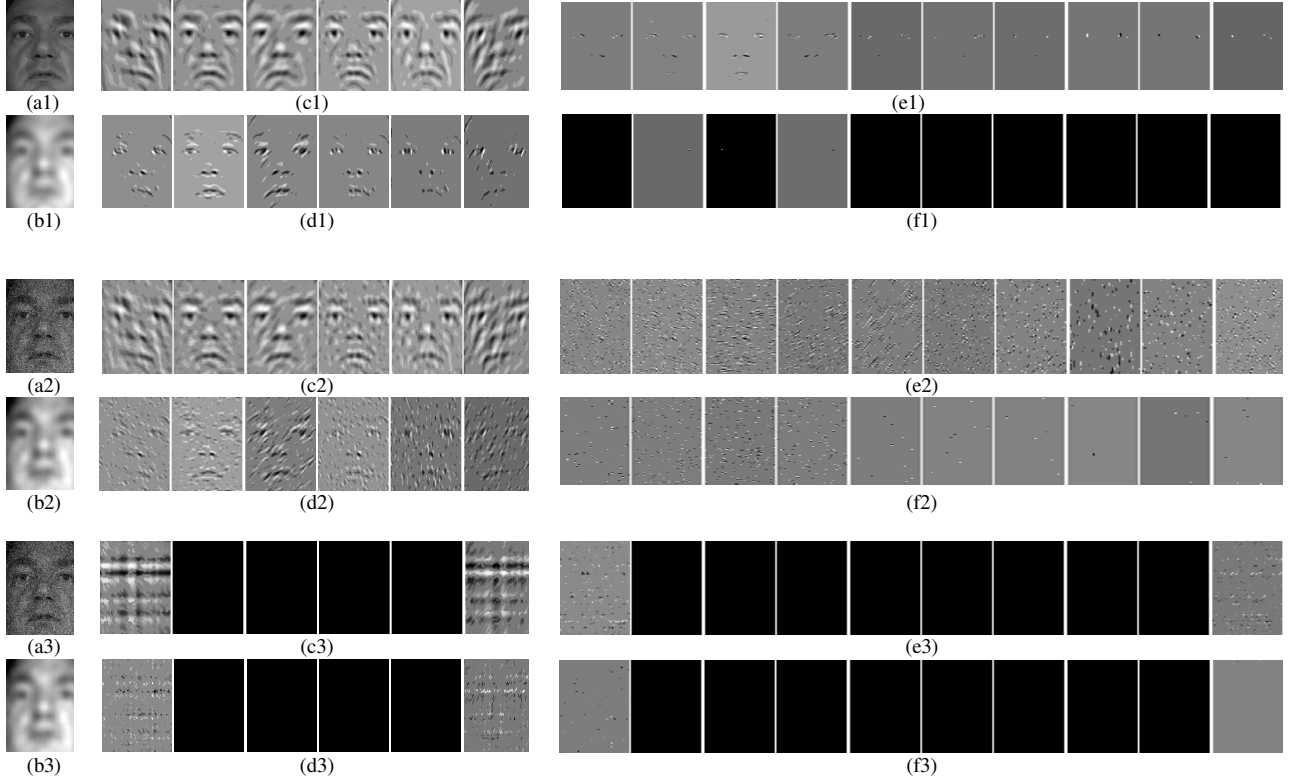Fig. 4. Flowchart of combining shearlet and LBP for faces

Fig. 5 Multi-resolution transformation for one image from FRGC. (a1) is noise free; (a2) and (a3) are with high level of noise; (b1), (c1) (d1), (e1) and (f1) are the real parts after the shearlet transformation for (a1); (b2), (c2) (d2), (e2) and (f2) are the real parts after the shearlet transformation for (a2); (b3), (c3) (d3), (e3) and (f3) are the imaginary parts after the shearlet transformation for (a3). Note that (a2) and (a3) are the same image.

multidirection, and shift invariant image decomposition [10]. The core of NSCT is the nonseparable two-channel nonsubsampled filter bank. NSCT is designed to be with better frequency selectivity and regularity. There are two kind of filters used in our method, i.e., '*pyr*' and '*pyrexc*'. Meanwhile, '*pyr*' means that we use a nonsubsampled pyramid structure to compute the shearlet transform. *'pyrexc'* means that we use nonsubsampled pyramid structure but exchanging the two highpass filters. The highpass at higher scales is filtered by the portion of the directional highpass filter that has "bad" response. Please see [10] for more details about these two kind of filters.

2.4 Representation for faces

In Fig. 5, we show the resulting images of applying the multi-resolution shearlet transform for an input image in its original form (row a1) and under high level of additive Gaussian noise (a2, a3), respectively. The resulting images in Fig. 5 (b1-f1 and b2-f2) are the real part of the transform, and the transformation is constructed in the Fourier domain. In addition, Fig. 5, (b1-f1) correspond to the different scales, i.e., a in Eq (6). The six images in Fig. 5 (c1) correspond to the different directions, i.e., s in Eq (6). The same is for the other images in Figs. 5 and 6. For more details about how to set up the number of shear orientations, please refer to [36, 22].

Shearlet transform and LBP are complementary, and combining them will improve the performance of the classifier. Firstly, from Figs. 5 and 6, we can find that the multi-resolution shearlet transformation provides good frequency localization and directional selectivity
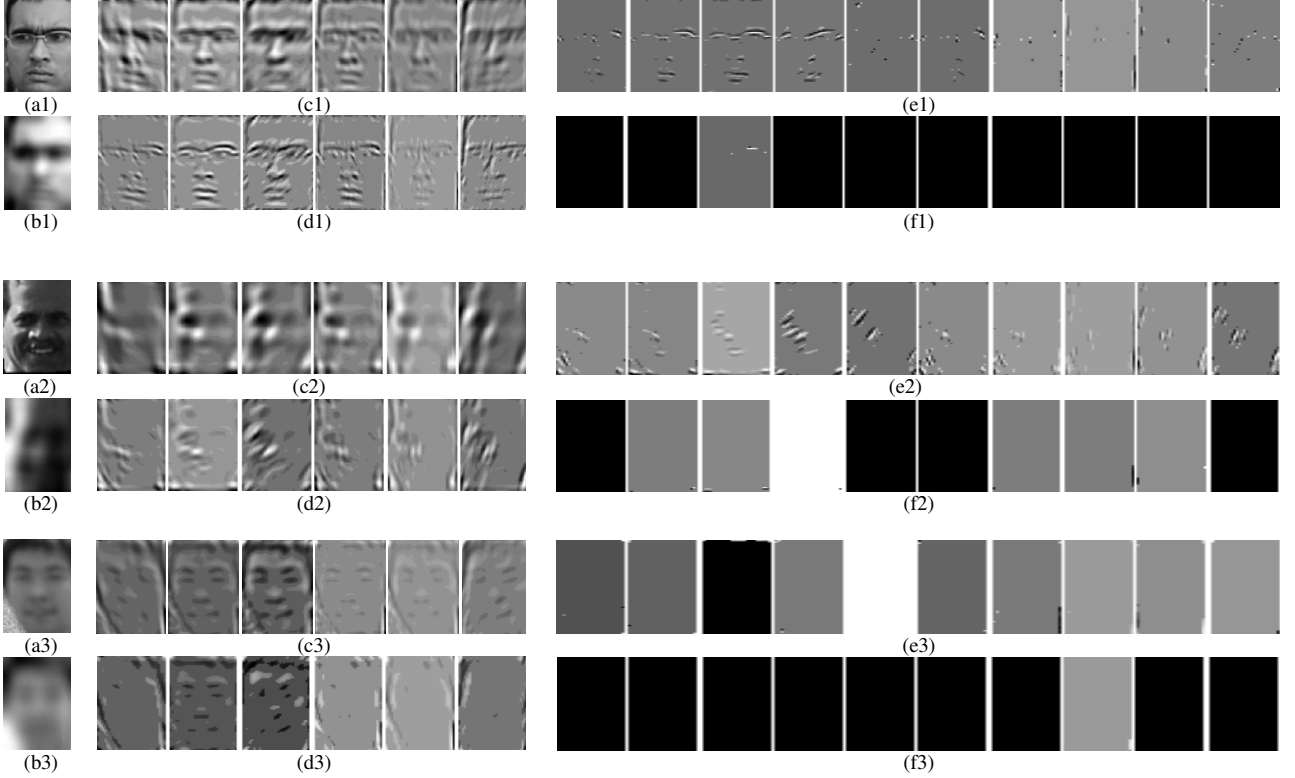
4

Fig. 6 Multi-resolution transformation for face images from remote face dataset by the shearlet transform to provide good frequency localization and directional selectivity. All these multi-resolution transformation of face images are the real parts after the shearlet transformation.

(i.e., edges). Accordingly, LBP is good at describing edges extracted by the shearlet transform. Secondly, LBP is not so robust to noise [43]. However, the shearlet transformation detects the edges very well although the image is contaminated by high level of Gaussian noise. For example, for the image with Gaussian noise as shown in Fig. 5 (a2, b2, c2 and d2), we can find that shearlet is able to completely analyze the singular structures of piecewise smooth images (i.e., edges). For more discussions on the robustness of shearlet to noise, please refer to [19, 35]. Note that images in Fig.5 (b1-f1, b2-f2) are the real parts of the shearlet transform since we find that the imaginary part is not useful for classification. Specifically, Fig. 5 (b3)-(f3) shows the imaginary parts of shearlet transformation in different levels for an image with high level of noise (i.e., Fig. 5, a3). From these images, we find that these imaginary parts contain few discriminative features for face recognition. Later in Section 3, experimental results also confirm it.

We use the shearlet plus LBP to represent faces in low quality images/videos. Given a face image, I, we compute its real part of the shearlet transform $\tilde{\psi}_{j,l,k}$ as shown in Eq. (7), and some resulting example images are shown in Fig 6. After that we use LBP to compute the texture features of the resulting faces from the shearlet transformation $\tilde{\psi}_{j,l,k}$, and then concatenate them into one long histogram, i.e., the multi-resolution transformation as shown in Fig. 5 (b1-f1 and b2-f2) and Fig 6.

In addition, as shown in Fig. 6, we find that this transformation also completely captures the singular structures of face images with poor image quality. From Fig. 5 (f1), we can find that the residual has very little information. In our case we do not use this part of shearlet transformation for face recognition.

When we use the shearlet plus LBP for face recognition, we divide each face into patches as done in [1, 56]. One example is shown in Fig. 7. Unlike [1, 56], we build one classifier for each patch. Specifically, for each patch p_i, where i=1, 2….20, we compute its shearlet transformation $\tilde{\psi}_{j,l,k}$ as in Eq. (7) and then LBP feature as in Eq. (8), i.e., $f_{LBP}(\tilde{\psi}_{j,l,k})$. The window to compute LBP is 3×3 neighbors as in Fig. 3 and we then build the histogram for each patch. Thus we have 256 bins for each $\tilde{\psi}_{j,l,k}$. We then use the PCA to reduce the dimensionality of the feature $f_{LBP}(\tilde{\psi}_{j,l,k})$. We keep 98% energy following [4] and the typical PCA reduced dimensions are around 150. After that, we get $\Phi=f_{PCA}(f_{LBP}(\tilde{\psi}_{j,l,k}))$. For the resulting features Φ, we use Linear Discriminant Analysis (LDA) [18] to perform face recognition. Note that we did not concatenate the LBP features of each patch into one long histogram and both PCA and LDA are applied to per patch histogram here.

During recognition, given two face images S_1 and S_2, we first divide them into patches, $p_{j,i}$ where j=1, 2 and i=1, 2….20. We compute the similarities between two corresponding patches in these two faces, φ_i and i=1, 2….20. The final similarity between these two face images is

$\Delta = \sum_i \varphi_i$.

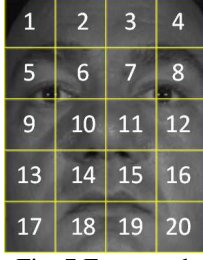| 1 | 2 | 3 | 4 |
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20

Fig. 7 Face patches

3. Experiments

In this section, we present the experimental results for face recognition using the proposed features discussed in Section 2.

(a)

log(1/*SNR*)

Original   0.020   0.030   0.050   0.067   0.100   0.200

white Gaussian noise

DWT
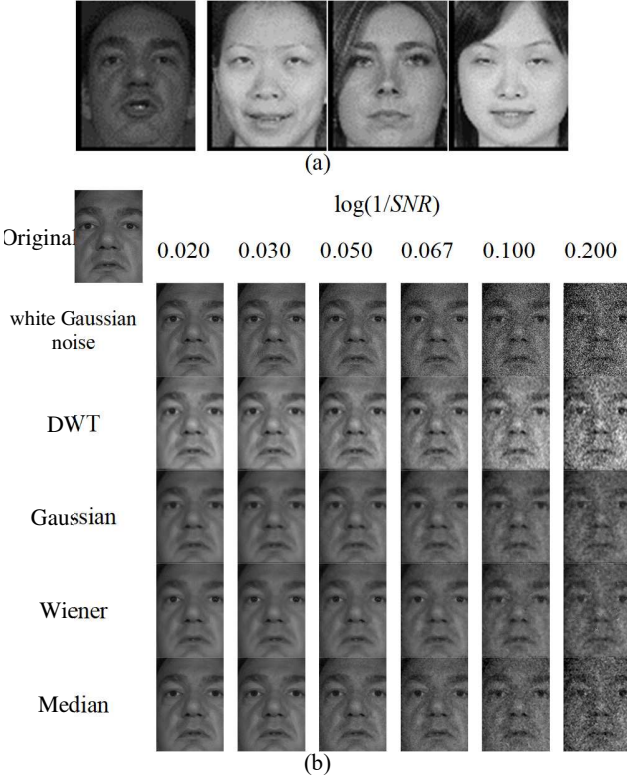
Gaussian

Wiener

Median

(b)

Fig. 8 (a) Example images from the FRGC dataset. (b) The first row is faces with white Gaussian noise; the following rows are filtered by DWT, Gaussian, Wiener and Median filters, respectively.

3.1 Dataset

We use three datasets for face recognition, i.e., FRGC version 2.0 and add noise to the face images, a recently published remote face dataset [5] and PaSC dataset [2]. FRGC 2.0 is designed to promote face recognition in general with emphasis on 3D and high resolution still imaginary. Meanwhile, there are six experimental protocols in FRGC and Experiments 1, 2 and 4 are designed for still images. There are 222 subjects showing 12,776 still images in the training set. In our case, we adopt experiment 1 for

evaluations. Experiment 1 measures performance on 16,028 frontal facial images. These images are taken under controlled illumination and both target and query set have the same number of samples (i.e., 16,028). The performance is reported as Verification Rates (VR) at 0.1% False Acceptance Rate (FAR). Here, we only consider the still images; some example faces in FRGC data set are presented in Fig.8 (a). As shown in Fig. 8 (b), we add white Gaussian noise to face images in FRGC data set. Here, the level of the noise is log(1/*SNR*), i.e., the logarithm of the inverse of the signal-to-noise ratio (SNR). The SNR is computed as: *SNR* =f(I)/f(N), where f(I) and f(N) are the power of the input image I and the noise image N: $f(I) = \frac{1}{n}\sum_{i=0}^{n-1}\left|I_i^2\right|$,

$f(N) = \frac{1}{n}\sum_{i=0}^{n-1}\left|N_i^2\right|$, where n is the dimensionality of the input image and noise image.

Blur   Illum

illum_blur   low_reso

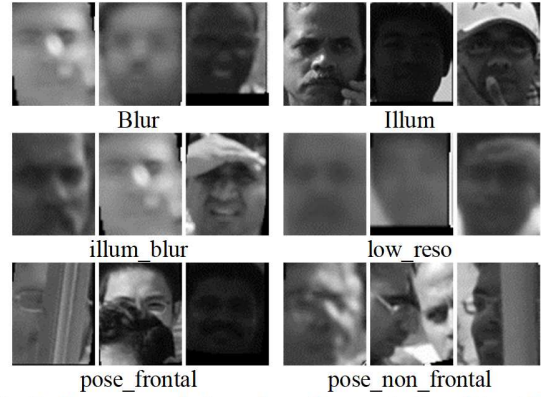pose_frontal   pose_non_frontal

Fig. 9. Some example faces from the remote face dataset [5].

For the remote face recognition problem, we use the dataset from [5]. Some face examples are shown in Fig. 9. All the images were taken in an unconstrained outdoor environment, at a distance varying from 10m to 250m. The face images are manually labeled with five points for each face, including the left pupil, right pupil, nose tip, left mouth corner and right mouth corner. The face images are cropped based on the fiducial points and kept in their original resolution. The resulting database for still color face images contains 17 subjects and 2,102 cropped face images in total. The number of faces per subject falls between 29 to 306. The captured images can be of very low resolution, with a typical resolution of 20 by 30 pixels. Moreover, low resolution images are often coupled with blurring effects. Also, large out-of-plane pose variations are observed. Since the distance between the camera and subjects is large, high magnification blur can be seen. Furthermore, due to the motion between camera and subjects, some of the images also suffer from motion blur. Finally, in some of the images, we see the presence of both blur and poor illumination condition.

This dataset consists of 6 subsets. The first one is *blur*, which shows the variations of blur and has 75 face images. The second one is *Illum*, which shows the variations of illuminations and has 561 face images. The third one is

6

Illum_blur, which shows both variations of illuminations and blur and has 128 face images. The fourth one is *low_reso*, which shows the variations of low resolutions and has 90 face images. The fifth one is *pose_frontal*, which shows the faces close to frontal pose and has 1,166 face images. The last one is *pose_non_frontal*, which shows the variations of poses and has 846 face images.

For the PaSC dataset, it includes 9,376 still images of 293 people balanced with respect to distance to the camera, alternative sensors, frontal versus not-frontal views, and varying location. There are also 2,802 videos for 265 people: a subset of the 293 people. Some examples for the videos and still images are shown in Fig. 10.

Fig. 10. Some example faces from the PaSC. (a) Examples taken during four sessions. Locations were changed between sessions, and sensor, distance to camera and pose were changed within sessions. (b) Cropped face images extracted from still images, showing lighting, motion blur and poor focus variations.
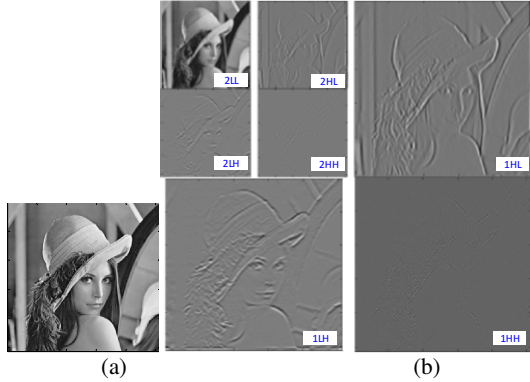

Fig. 11. (a) Original Image, (b) Multi-resolution transformation by DWT for one image

All these face images are cropped using the landmarks of two eyes, and are normalized to 64×80. Each face is divided into patches as in Fig. 7. Each patch is of 16×16. We perform the shearlet transform for each patch $\tilde{\psi}_{j,l,k}$ as shown in Section 2.1. We have three levels, i.e., j=0, 1, 2. The first 0-level is the low frequency part. The 1- and 2-levels of shearlet transformation capture the different texture structures in different frequency parts.

In addition, SCface is a very challenging database with daytime and nighttime protocols [21]. Collected faces in infrared spectrum is a good way for illumination invariant face recognition. We also exploited how to use near infrared for illumination invariant face recognition [7].

For face images in the FRGC dataset, they are manually landmarked by us and then cropped. All the methods in Section 3.2 which are tested over this dataset are re-implemented by us, and we use the same setups for these methods for fair comparison. For the remote face dataset and PaSC, these two datasets provide face landmarks. All the methods in sub-Sections 3.3 and 3.4 which are tested over these two datasets use the exactly same setups.

3.2 Experimental Results for FRGC

As shown in Table 1, we test several combinations between the shearlet and LBP. Here, LBP is the method we described in Section 2.2. DWT is the Discrete Wavelet Transform, and the wavelet used in experiments is the shearlet as mentioned in Section 2.1. In other words, shearlet transformation is equal to performing DWT for an input image plus image reconstruction plus NSCT filtering [35]. One example of performing DWT for one image is shown in Fig. 11. In our case, we perform two levels of transformation and create seven subimages, i.e., HH, LH, HL, and LL subbands for levels 1 and 2.

LBP+DWT means LBP feature computed over the multi-resolution decomposition resulting from DWT (e.g., the seven subbands in Fig. 11(b)). *LBP+DWT_coef_T* means that we perform the multi-resolution decomposition for each image using DWT (see Fig. 11) and then denoise the multi-resolution decomposition by the threshold *T*. After that we compute the LBP feature for the wavelet-denoised images. Here we have two kind of thresholds (*T*), i.e., soft thresholding and hard thresholding [15]. In brief, to suppress the noise, the hard thresholding applies the following nonlinear transform to the empirical wavelet coefficients: $F(x) = x \cdot I(|x| > t)$, where *t* is a certain threshold; $I(x)$ is the input signal and $F(x)$ is output. The choice of the threshold is a very delicate and important statistical problem. For soft thresholding the following nonlinear transform is used: $S(x) = \text{sign}(x) \cdot (|x| - t) I(|x| > t)$, where *t* is a threshold.

We compute the shearlet plus LBP feature for faces as shown in Section 2.4. Experimental results are shown in Tables 2, 3, 4 and 5.

In Table 2, we test the performance of the methods using shearlet to remove the noise in the images. We perform this group of experiments since shearlet can be used to denoise the images very well [35, 17]. Here, *LSF* means *LBP+shearlet+FFT* and we compute the feature as shown in Section 2.3 and using Fourier based shearlet transform to improve the efficiency, where FFT is the fast Fourier transform. 'pyr' means that we use a

Table 1 LBP+DWT feature computation

	LBP Feature computation	Threshold	band of DWT
LBP+DWT	over DWT reconstructed images	soft	
LBP+DWT+coef_hard	over DWT coefficients	hard	level_2*(LL+ HH+ LH)+level_1*(HL+HH+HL+LH)
LBP+DWT+coef_soft	over DWT coefficients	soft	level_2*(LL+ HH+ LH)+level_1*(HL+HH+HL+LH)

Table 2 Performance comparison (%) of the components of shearlet plus LBP over FRGC by adding white Gaussian noise
(LSF = LBP+shearlet+FFT and real = real part of shearlet)

Methods	log(1/*SNR*)										
	0	0.020	0.022	0.025	0.028	0.033	0.040	0.050	0.067	0.100	0.200
LBP	95.53	93.64	91.54	91.86	91.46	90.37	89.67	86.42	78.64	56.48	25.68
LBP+DWT	96.34	96.15	96.23	96.37	96.78	95.43	94.82	91.76	86.42	76.18	58.94
LBP+DWT+coef_hard	97.86	97.98	97.43	97.57	97.66	**98.21**	97.32	97.16	94.21	85.49	71.48
LBP+DWT+coef_soft	97.86	97.84	97.65	97.42	97.81	98.24	97.93	97.87	94.69	85.76	72.45
LSF +'pyr'-real	**98.13**	**98.16**	**98.27**	**98.31**	**98.34**	**97.56**	**97.38**	**97.24**	**97.16**	**92.18**	**86.18**

Table 3 Performance comparison (%) using different filters and real/imaginary part for shearlet
over FRGC by adding white Gaussian noise

methods	log(1/*SNR*)										
	0	0.020	0.022	0.025	0.028	0.033	0.040	0.050	0.067	0.100	0.200
LBP	95.53	93.64	91.54	91.86	91.46	90.37	89.67	86.42	78.64	56.48	25.68
LSF +'pyr'-real	98.13	98.16	98.27	98.31	98.34	97.56	97.38	**97.24**	**97.16**	**92.18**	**86.18**
LSF + 'pyrexc '-real	**98.32**	**98.25**	**98.31**	**98.45**	**98.47**	**97.93**	**97.65**	97.04	97.13	92.08	85.43
LSF+'pyr'-real+imag	98.27	98.05	98.13	98.24	98.17	97.21	97.06	96.87	96.76	91.56	83.74
LSF+'pyr '-imag	78.64	78.42	77.96	76.58	76.42	76.85	75.86	74.86	72.68	68.45	55.37

non-subsampled pyramid structure [10] to compute the shearlet transform. 'real' means that we only use the real part of shearlet transform.

From Table 2, we can find that *LBP+DWT_coef_T* works much better than *LBP+DWT* and *LBP*, especially for high level of noise. However, the different thresholding methods (i.e., soft and hard) change the performance marginally. One explanation is that shearlets detect the edges well even with high level of noise, and LBP describes these edges in faces well. Further, *LSF+'pyr'-real* outperforms significantly compared to *LBP+DWT_coef_T, i.e.,* 86.18% vs 71.48% and 72.45% for the high level of noise. It shows that *LSF +'pyr'-real* works well for high level of noise and the 'pyr' (i.e., nonsubsampled pyramid structure) is good for the face recognition when the faces are with noise.

In Table 3, we test *LSF* with different filter, i.e., *'pyr'* and + *'pyrexc'*. Here 'pyr' means that we use a non-subsampled pyramid structure as mentioned before. *'pyrexc'* means that we use nonsubsampled pyramid structure but exchanging the two highpass filters [10]. 'real' means we only use the real part and 'imag' means we only use the imaginary part. *LSF+'pyr'-real+imag* means we use both the real and imaginary parts.

From Table 3, we can find that *LSF+'pyr'-real* works the best. The performances between *LSF +'pyr'-real* and *LSF + 'pyrexc '-real* are close. It means that by exchanging the two

highpass filters for the pyramid structure the performance can be improved marginally. However, if we use the imaginary part of shearlet transform, the performance changes significantly. Even when we combine both real and imaginary parts, the performance is not as good as that of only using the real part. One explanation is that the imaginary part cannot detect the edges well in images and is also easily affected by the noise in the images. In the following, we will use LSF to represent *LSF +'pyr'-real* for short.

In Table 4, we compare our method with others for FRGC, e.g., LBP plus several filters for noisy faces, e.g., Wiener, Gaussian, and Median. Specifically, we use Matlab function for these filters. For the median filter, the neighborhood size takes the default value (3×3). For the Wiener filter, it automatically estimates the additive noise power before doing the filtering and the neighborhood size takes the default value (3×3). For Gaussian filter, it filters an image with a 2-D Gaussian smoothing kernel with standard deviation of 0.5. For the DWT denoising, we use the method NSCT proposed in [10].

Table 4 Performance comparison (%) using different filters to filter out the noise over FRGC by adding white Gaussian noise

methods	log(1/*SNR*)										
	0	0.020	0.022	0.025	0.028	0.033	0.040	0.050	0.067	0.100	0.200
LBP	95.53	93.64	91.54	91.86	91.46	90.37	89.67	86.42	78.64	56.48	25.68
LBP+Gabor	97.46	97.35	96.89	96.87	94.56	93.16	91.67	86.54	83.16	78.35	64.25
LBP+wiener	92.46	92.17	91.97	91.93	92.06	92.04	91.82	89.16	85.07	73.82	53.57
LBP+Gaussian	87.65	87.34	88.76	87.95	88.63	87.26	81.39	80.34	81.02	75.62	60.42
LBP+Median	93.54	92.84	91.27	91.89	92.68	92.34	91.38	89.46	83.63	72.96	50.31
LSF	**98.13**	**98.16**	**98.27**	**98.31**	**98.34**	**97.56**	**97.38**	**97.24**	**97.16**	**92.18**	**86.18**

Table 5 Performance comparison with exiting methods over FRGC by adding white Gaussian noise

methods	log(1/*SNR*)										
	0	0.020	0.022	0.025	0.028	0.033	0.040	0.050	0.067	0.100	0.200
LBP	95.53	93.64	91.54	91.86	91.46	90.37	89.67	86.42	78.64	56.48	25.68
LBP+Gabor	97.46	97.35	96.89	96.87	94.56	93.16	91.67	86.54	83.16	78.35	64.25
CLBP	96.17	95.89	95.69	94.58	93.82	92.17	91.58	88.76	78.93	58.94	30.58
DLBP	**98.46**	50.67	52.48	49.87	48.64	40.29	39.87	38.16	35.24	30.68	17.59
LTP	98.13	98.11	97.89	97.78	95.64	93.15	91.59	84.28	82.49	65.26	48.28
LQP	92.75	92.46	91.97	91.83	90.46	88.67	87.95	87.38	85.48	76.49	54.89
DFD	98.03	98.23	98.15	97.15	97.14	96.89	95.67	92.48	72.39	58.64	26.53
shearlet	90.26	90.35	90.63	90.36	89.56	89.37	88.56	88.47	88.67	84.61	75.39
DL	97.32	96.95	94.68	93.29	93.57	92.64	92.46	91.56	72.36	47.26	26.18
dA	97.43	97.09	97.12	96.34	95.67	94.38	92.65	87.17	82.63	76.52	68.35
LSF	**98.13**	**98.16**	**98.27**	**98.31**	**98.34**	**97.56**	**97.38**	**97.24**	**97.16**	**92.18**	**86.18**

edges with the high level of noise as show in Fig. 5.

In addition, we also compare with other existing descriptors in Table 5, such as *LBP+Gabor* [56], Completed Local Binary Pattern (*CLBP*) [20], Dominant Local Binary Patterns (*DLBP*) [34], Local Ternary Patterns (*LTP*) [49], Local Phase Quantization (*LQP*) [44], Discriminant Face Descriptor (*DFD*) [31], *DL* [24], and denoising autoencoders (*dA*) [51]. For *DL*, we use Caffe. The other methods are implemented by us.

From Table 5, one can find that LSF works comparable to state-of-the-art methods for low level of noise (e.g., log(1/*SNR*) = 0.020) and outperforms them significantly for the high level of noise (e.g., log(1/*SNR*) = 0.200). *LBP+Gabor*, *CLBP*, *DLBP*, *LTP* and *DFD* work quite well for low level of noise but their performances drop with the increase in the level of noise, especially for high level of noise since the faces are blurred seriously (see Fig. 8). DL also works well for low level of noise but the performance
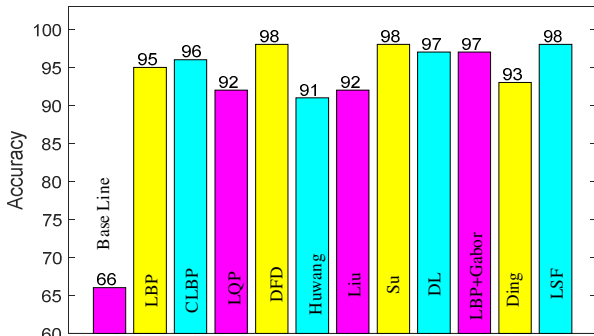

Fig. 12. Performance comparison with existing methods over FRGC face dataset.

From Table 4, we can find that *LSF* works the best, especially with the high level of noise, 86.18% compared to *LBP+Gabor* (64.25%), and *LBP* (25.68%). In addition, LBP plus different filters, i.e., Wiener, Gaussian and Median are not as good as *LSF*. One explanation is that these filters blur the edges in the image, while these edges help face recognition. However, shearlets perform well in detecting

Table 6 Rank-1 analysis on the remote face dataset [5] using five face images of each subject as gallery

| Condition	Descritpors	Acc (%)	Condition	Descritpors	Acc (%)	Condition	Descritpors	Acc (%)
Blur	LBP	45.9	Illum	LBP	78.4	Illum_blur	LBP	73.2
	LPQ	58.1		LPQ	79.5		LPQ	75.6
	BSIF	62.2		BSIF	79.3		BSIF	74.8
	Gabor	65.3		Gabor	85.5		Gabor	75.0
	DFD	63.5		DFD	83.4		DFD	75.2
	DL	48.6		DL	80.4		DL	71.8
	shearlet	62.5		shearlet	81.6		shearlet	74.3
	LSF	**67.3**		LSF	**92.5**		LSF	**76.0**
Low_reso	LBP	12.4	Pose_frontal	LBP	67.9	Pose_non_frontal	LBP	49.9
	LPQ	13.5		LPQ	70.6		LPQ	47.5
	BSIF	11.2		BSIF	70.1		BSIF	49.2
	Gabor	15.9		Gabor	76.8		Gabor	50.2
	DFD	14.5		DFD	78.6		DFD	52.5
	DL	11.5		DL	80.3		DL	49.8
	shearlet	13.8		shearlet	77.8		shearlet	51.7
	LSF	**19.9**		LSF	**83.8**		LSF	**57.2**

(a) Control videos (b) Handheld videos
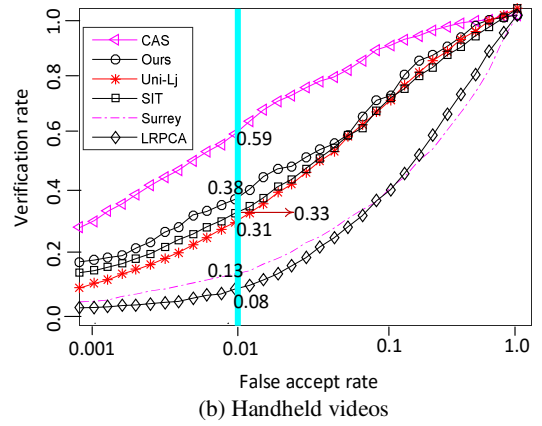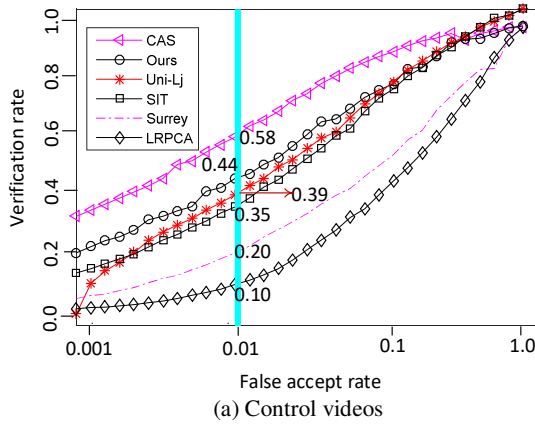
Fig. 13. ROC curves for the control and handheld video evaluations.

also drops with the increase of the level of noise. *dA* improves the performance since we use the noisy images as input and the noise free images as output to fine tune the neural network. However, the performance is not as good as *LSF* for high level of noise. One reason that *DL* and *dA* work not so well as the other descriptors for high level of noise might be that the training set is not large enough to train a well-performed network (i.e., 16,028 images for training and 16,028 images for testing).

DLBP works the best for noise free images but its performance drops quickly when white Gaussian noise is added. A possible reason is that noise may change the pattern type of a dominant pattern present in the noise-free case. Therefore, the noise may possibly reshape the dominant pattern histogram in some cases by assigning some of the dominant pattern's occurrence to other types of patterns.

In addition, the performance comparison of *LSF* and existing methods for FRGC without noise is shown in Fig. 12. Here, *LBP*, *CLBP*, *LQP*, *DFD*, *LBP+Gabor* [56], and *LSF* use the same setups for classification but using different features. The results of Ding [13], Huwang [28], Liu [37], and Su [46] are quoted directly from the original papers. From this figure, one can find that *LSF* gets the best results

over this dataset. In addition, *LBP*, *CLBP, DFD*, Su, *DL* and *LBP+Gabor* also achieve very good results. We also use *CLBP* to replace *LBP* to combine with shearlet and find that the performance difference between *CLBP+shearlet* and *LBP+shearlet* is not significant. Thus, we use *LBP+shearlet* in the following experiments.

3.3 Remote face recognition

The performance comparison of *LSF* and several existing methods is shown in Table 6. Here, *LBP*, *LPQ* [44], *BSIF* [29], *Gabor* [56], and *DFD* [31] and *LSF* use the same setups for classification but using different features. In addition, the remote face dataset [5] does not provide a training set. We thus use the training set of FRGC 2.0 Experiment 1, i.e., 16,028 frontal facial images and they are quite different from the images in the remote face dataset [5]. In other words, we train our classifier over the dataset in FRGC 2.0 Experiment 1 and test over the remote face dataset.

From Table 6, one can find that LSF gets the best results over this dataset, especially for the subset *Illum*. One reason is the illumination does not change the texture/edge in faces. In addition, for subsets, *Blur*, *Illum_blur*, *Pose_frontal* and

10

Pose_non_frontal, *LSF* achieves promising results. *LBP*, *CLBP, DFD*, Su, *DL* and *LBP+Gabor* also achieve very good results. For the subset *Low_reso*, all the features do not work very well since the face images are blurred seriously and the faces are of small size, e.g., with a typical resolution of 20 by 30 pixels. Detecting textures/edges from these faces is very difficult. Here, *DL* does not work well. One explanation is that the test set is different from the face images for training. However, *DL* achieve comparable performance to other methods for the subset *Pose_frontal* since this subset has similar distribution as the training set in FRGC 2.0 Experiment 1.

3.4 PaSC

For PaSC dataset, we firstly crop the facial region from the given video frame based on the eye coordinates provided by the organizers [2]. The cropped facial region is aligned and scaled to a size of 64×80 pixels and transformed to gray-scale. The gray-scale images are then used as input for the feature extraction procedure as discussed on Section 2. During feature extraction, LSF feature sets are generated for each still image and each processed frame of a given video sequence.

To ensure that our framework produces fixed size templates regardless of the number of frames in the video, we use the same procedure as that in [45]. Before extracting features, we partition the frames of a given video into two groups depending on the extent of the head rotation (yaw) of the person shown in the video. Here, the first group contains frames with yaw angles below 15°, and the second group contains frames with yaw angles greater than 15°. Frames with negative yaw angles are mirrored prior to feature extraction to ensure that two frame-groups are sufficient to cover all rotation-dependent variability of the faces.

In Fig. 13, we compare our methods with state-of-the-art methods. Here, *CAS* was proposed by [27], which is based on Hybrid Euclidean-and-Riemannian Metric Learning combined with deeply learned features (abbr. to HERML-DeLF). It is for image set classification with image features learned by a deep neural network. *Uni-Lj* was proposed by [45], which used the MODEST framework to represent facial images (or frames) with various texture descriptors and using the computed descriptors as input to a probabilistic modeling technique capable of deriving low-dimensional representations from the extracted texture representations. *SIT* was developed by [33], which use the Hierarchical - Probabilistic Elastic Part (PEP) model to approach the video face recognition problem. The Hierarchical-PEP model builds pose-invariant face representation by applying the PEP model hierarchically to decompose a face image into face parts at different levels of detail and thus to build pose-invariant part-based face representations. *Surrey* was developed by [4], which tackles the PaSC video-to-video matching by combining a dynamic video frame selection method with a multiscale local phase quantization (MLPQ) based frame-to-frame matching algorithm and a simple voting strategy. *LRPCA* is the baseline, which uses principal component analysis for local region (LRPCA) [3]. Video

person recognition evaluation for all these methods are presented in [3].

From Fig. 13, we can find that our methods achieve promising results and *CAS* performs the best. Here, *CAS* is a deep learning based method, which is heavily trained. Firstly, it is pre-trained using "Celebrities on the Web" (CFW) database [57] and then fine-tuned using other two datasets. The first is the training portion of the PaSC [2]. The second is the Institute of Computing Technology, CAS-OMRON Social Solutions Co. Ltd-Xinjiang University (COX) face database collected by the members of the CAS group [27].

3.5 Discussion

In this section, we will discuss the time complexity of LSF and the reasons that LSF works for remote face recognition.

As shown in Table 7, we perform a quantitative comparison of timing information for LSF with LBP and Gabor for an image 128×160. We test over 1,000 face images and calculate the average time. All of them are implemented in C++ and tested on an i5-2400 CPU@ 8.0G memory. Both Gabor and LSF are speeded up using FFT. From this table, we can find that both LSF and LBP are very efficient compared to Gabor.

Table 7 Comparison of the Average Time Consumption with LBP and Gabor

Methods	Times (s)
LBP	0.0075
Gabor	1.2
LSF	0.0306

For the remote face dataset, the usual variations which degrade the performance of face recognition are low resolution, blur and poor illumination condition. Some examples showing low resolution, serious illumination variations and blur variations and their multi-resolution transformation by the shearlet transform are shown in Fig. 6. We can see that the shearlet transform provides good frequency localization and directional selectivity, especially the low frequency part and the second level of shearlet transformation. In addition, the shadow in the face shown in Fig. 6 (a2) blurs the edges of this face. The local textures/edges and the global features are well extracted. Likewise, the face in Fig. 6 (a3) is blurred seriously and the local and global features are computed successfully. On the other hand, the small face in Fig. 6 (a1) has a clear appearance, so its low frequency part and the second level of shearlet transformation show good discriminative power. From Table 6, we can see the subset *Pose_frontal* gets very good performance (i.e., 83.8%) although several faces are of low resolution. It is the same for the subset *Illum* (i.e., 92.5%). However, for the blurred faces, as shown in Fig. 6 (a3), its third level of shearlet transformation contains small amount of discriminative information for face recognition. Thus, the performance of the subset *blur* reduces to 67.3%. However, the discriminative information in low frequency part and the second level of shearlet transformation still make the performance of LSF much better than existing

11

methods.

LSF descriptor and Gabor get comparable performance for high quality faces (98.13% vs.97.46%, Table 5, without noise for FRGC), but LSF works better than Gabor for faces of poor quality (86.18% vs. 64.25%, Table 5, with high level of noise for FRGC, and also for remote face dataset and PaSC). One explanation is that the Gabor transform can extract local texture features (i.e., edges and their orientations) for faces very well when the faces are in high quality. As the image quality degrades, local textures in the faces become weak. Specifically, local edges reduce and their orientations also change. Thus, the performance of Gabor degrades. However, the shearlet is designed to extract the singular structures of piecewise smooth images although the image quality degrades, as shown in Fig. 5 and 6. Thus, LSF works better than Gabor for the faces in poor quality images.

4.   Conclusion

In this paper, we proposed a robust descriptor LSF, which combines the shearlet coefficients and LBP. We performed the shearlet transform in the Fourier domain and tested its real and imaginary parts. We also tested different filters for the shearlet transform. We then used LBP to encode the textures (i.e., edges) in faces detected by the shearlet transform. The resulting LSF is robust to the noise in images and achieves very good results for face recognition. Specifically, the experimental results over FRGC and the noisy faces show promising results. For example, LSF achieves an accuracy of 98.13% for the dataset without noise and 86.18% with high levels of noise, while LBP gets the accuracy of 95.53% for the dataset without noise and 25.68% with high levels of noise. LSF also outperforms deep learning based method (97.32% without noise and 26.18% with high levels of noise) for this dataset even that we use the denoising autoencoders version of deep learning (97.43% without noise and 68.35% with high levels of noise). In addition, we also test LSF over two challenging datasets, remote face dataset [5] and PaSC. LSF also outperforms the state-of-the-art methods.

References

[1]   T. Ahonen, A. Hadid and M. Pietikäinen, "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, 2006.

[2]   J. R. Beveridge, P. J. Phillips, D. Bolme, B. A. Draper, G. H. Givens, Y-M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer, P. J. Flynn, S. Cheng, The Challenge of Face Recognition from Digital Point-and-Shoot Cameras, *IEEE Sixth* International *Conference on Biometrics: Theory, Applications and Systems (BTAS),* 2013.

[3]   J.R. Beveridge, P.J. Flynn, Z. Feng, P. Huber, J. Kittler, Zhiwu Huang, S. Li, Y. Li, M. Kan, R. Wang, S. Shan, X. Chen, H. Li, G. Hau, V. Struc, J. Krizaj, C. Ding, D. Tao, P.J. Phillips. Report on the FG 2015 Video Person Recognition Evaluation, *The Eleventh IEEE International Conference on Automatic Face and Gesture Recognition*, 2015.

[4]   C. H. Chan, M. Tahir, J. Kittler, and M. Pietikainen. Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 5 pp. 1164-1177, *2013.*

[5]   R. Chellappa, J. Ni, and V. M. Patel, Remote Identification of Faces: Problems, Prospects, and Progress, *Pattern Recognition Letters*, vol. 33, no. 14, 15, pp. 1849–1859, 2012

[6]   D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of Dimensionality: High-dimensional Feature and Its Efficient Compression for Face Verification. *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2013.

[7]   J. Chen, D. Yi, J. Yang, G. Zhao, S. Li, M. Pietikäinen, Learning Mappings for Face Synthesis from Near Infrared to Visual Light Images, *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2009

[8]   J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, W. Gao. WLD: A Robust Local Image Descriptor, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705–1720, 2010.

[9]   J. Chen, V. Kellokumpu, G. Zhao and M. Pietikäinen M, RLBP: Robust local binary pattern. *Proc. the British Machine Vision Conference*, 2013.

[10]   A. L. da Cunha, J. Zhou, M. N. Do, The Nonsubsampled Contourlet Transform: Theory, Design, and Applications, *IEEE Trans. on Image Processing*, vol. 15, no. 10, pp. 3089–3101, 2006.

[11]   A. Danti, K.M. Poornima, Face Recognition using Shearlets*, 7th IEEE International Conference on Industrial and Information Systems*, 2012

[12]   J.G. Daugman, High confidence visual recognition of persons by a test of statistical independence, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148–1161, 1993.

[13]   C. Ding, J. Choi, D. Tao, L.S. Davis, Multi-Directional Multi-Level Dual-Cross Patterns for Robust Face Recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2016

[14]   Y. Dong, D. Tao, X. Li, J. Ma, and J. Pu, Texture Classification and Retrieval Using Shearlets and Linear Regression, *IEEE Trans. on Cybernetics,* vol. 45, no. 3, pp. 358–369, 2013

[15]   D. L. Donoho, De-Noising by Soft-Thresholding, *IEEE Trans. On Information Theory*, vol. 41, no. 3, pp. 613–627, 1995

[16]   G. R. Easley, D. Labate, and W. Q. Lim. Sparse directional image representations using the discrete Shearlet transform. *Appl. Comput. Harmon. Anal.*, vol. 25, no. 1, pp. 25–46, 2008.

[17]   G. R. Easley, D. Labate, V. M. Patel, Directional Multiscale Processing of Images Using Wavelets with Composite Dilations, *J Math Imaging Vis*, vol. 48, no. 1, pp. 13–34, 2012

[18]   K. Fukunaga, *Introduction to statistical pattern classification*. USA: Academic Press. 1990

[19]   K. Guo, D. Labate, and W. Lim, Edge analysis and identification using the continuous shearlet transform, *Appl. Comput. Harmon. Anal.* vol. 27, no. 1, pp. 24–46, 2009

[20] Z. Guo, L. Zhang, D. Zhang, A Completed Modeling of Local Binary Pattern Operator for Texture Classification, *IEEE Trans. on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010

[21] M. Grgic, K. Delac, and S. Grgic, SCface surveillance cameras face database, *Multimedia Tools and Application*s, vol. 51, no. 3, pp. 863–879, 2009

[22] S. Hauser, G. Steidl, Fast Finite Shearlet Transform: a tutorial, *arXiv:1202.1773v2*, July 24, 2014

[23] J. He, H. Ji, and X. Yang, Rotation Invariant Texture Descriptor Using Local Shearlet-Based Energy Histograms, *IEEE Signal Processing Letters*, vol. 20, no. 9, pp. 905–908, 2013

[24] G. E. Hinton, and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks. *Science*, vol. 313, Issue 5786, pp. 504–507, 2006

[25] G. E. Hinton, S. Osindero and Y. Teh, A fast learning algorithm for deep belief nets. *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[26] 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. *Technical Report* 07-49, University of Massachusetts, Amherst, 2007

[27] Z. Huang, R. Wang, S. Shan, and X. Chen. Hybrid Euclidean and-Riemannian Metric Learning for Image Set Classification. *Asian Conference on Computer Vision*, 2014.

[28] W. Hwang, G. Park, J. Lee and S.C.Kee, Multiple Face Model of Hybrid Fourier Feature for Large Face Image Set, *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2006.

[29] J. Kannala and E. Rahtu, BSIF: Binarized Statistical Image Features, *Proc. of International Conference on Pattern Recognition*, 2012

[30] A. Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, *Proc. of Neural Information Processing Systems,* 2012.

[31] Z. Lei, M. Pietikainen, and Stan Z. Li, Learning Discriminant Face Descriptor, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 289–302, 2014

[32] S.Z. Li and A.K. Jain, *Handbook of Face Recognition*, second ed., Springer-Verlag, Aug. 2011.

[33] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic elastic matching for pose variant face verification. *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2013

[34] S. Liao, Max W. K. Law, and Albert C. S. Chung, Dominant Local Binary Patterns for Texture Classification, *IEEE Trans. on Image Processing*, vol. 18, no. 5, pp. 1107–1118, 2009.

[35] W.-Q Lim, The Discrete Shearlet Transform: A New Directional Transform and Compactly Supported Shearlet Frames, *IEEE Trans. on Image Processing*, vol. 19, no. 5, pp. 1166–1180, 2010

[36] W.-Q Lim, Nonseparable Shearlet Transform, *IEEE Trans. on Image Processing*, vol. 22, no. 5, pp. 2056–2065, 2013

[37] C. Liu, Capitalize on Dimensionality Increasing Techniques for Improving Face Recognition Performance, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 725–737, 2006.

[38] D.G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[39] J. Lu; V.E. Liong, X. Zhou; J. Zhou, Learning Compact Binary Face Descriptor for Face Recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 2041–2056, 2015

[40] G. Marcus., Is "Deep Learning" a Revolution in Artificial Intelligence? *The New Yorker*, 25 November 2012.

[41] K. Mikolajczyk and C. Schmid, A Performance Evaluation of Local Descriptors, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615-1630, 2005.

[42] P. Moreels and P. Perona, Evaluation of Features Detectors and Descriptors based on 3D objects, *International Journal of Computer Vision*, vol. 73, no. 3, pp. 263-28, 2007.

[43] T. Ojala, M. Pietikäinen and T. Mäenpää, Multiresolution Gray Scale and Rotation Invariant Texture Analysis with Local Binary Patterns, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971 - 987, 2002.

[44] E. Rahtu, J. Heikkila, V. Ojansivu and T. Ahonen, Local Phase Quantization for Blur-insensitive Image Analysis, Image and Vision Computing, vol. 30, no. 8, pp. 501–512, 2012

[45] J. K. V. Struc and S. Dobrisek. MODEST face recognition. *In International Workshop on Biometrics and Forensics*, 2015.

[46] Y. Su, S. Shan, X. Chen, W. Gao, Hierarchical Ensemble of Global and Local Classifiers for Face Recognition, *Proc. IEEE Int'l Conf. on Computer Vision*, 2007.

[47] Y. Sun, X. Wang, X. Tang, Deep Learning Face Representation by Joint Identification-Verification, *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2014.

[48] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, DeepFace: Closing the Gap to Human-Level Performance in Face Verification, *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2014

[49] X. Tan, B. Triggs, Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2007

[50] A. Toshev and C. Szegedy, DeepPose: Human Pose Estimation via Deep Neural Networks, *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2014

[51] P. Vincent, H. Larochelle Y. Bengio and P.A. Manzagol, Extracting and Composing Robust Features with Denoising Autoencoders, *Proc. Int'l Conf. on Machine Learning*, 2008.

[52] N. Vu, A. Caplier, Enhanced Patterns of Oriented Edge Magnitudes for Face Recognition and Image Matching. *IEEE Trans. on Image Processing*, vol. 21, no. 3, pp. 1352–1365, 2012

[53] S. Xie, S. Shan, X. Chen, J. Chen. Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition. *IEEE Trans. on Image Processing*, vol. 19, no. 5, pp. 1349–1361, 2010

[54] Z. Zeng, J. Hu, Face Recognition Based on Shearlets Transform and Principle Component Analysis, *5th International Conference on Intelligent Networking and Collaborative Systems*, 2013

[55] B. Zhang, S. Shan, X. Chen, W. Gao, Histogram of Gabor phase patterns (HGPP): a novel object representation approach for face recognition. *IEEE Trans. on Image Processing*, vol. 16, no. 1, pp. 57–68, 2007.

[56] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang: Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition. *Proc. IEEE Int'l Conf. on Computer Vision*, 2005

[57] X. Zhang, L. Zhang, X.-J. Wang, and H.-Y. Shum. Finding celebrities in billions of web images. *IEEE Transactions on Multimedia*, vol. 14, no. 4, pp. 995–1007, 2012.

[58] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, Face Recognition: A Literature Survey, *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[59] L Liu, PW Fieguth, Texture classification from random features, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 3, 574-586

[60] L Liu, P Fieguth, Y Guo, X Wang, M Pietikäinen, Local Binary Features for Texture Classification: Taxonomy and

Experimental Study, Pattern Recognition vol. 62, no. 2, 135–160

Jie Chen received the M.S. and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 2002 and 2007, respectively. Since September 2007, he has been a senior researcher with the Center for Machine Vision and Signal Analysis at the University of Oulu, Finland. In 2012, he visited the Computer Vision Laboratory at the University of Maryland, USA. His research interests include pattern recognition, computer vision, machine learning, dynamic texture and watermarking. He has authored over 50 papers in journals and conferences and he is a Program Committee Member for many conferences. He is a co-organizer of ACCV 2014 Workshop on RoLoD: Robust local descriptors for computer vision, a co-organizer of CVPR 2016 Workshop on RoF: Robust Features for Computer Vision and a guest editor of Neurocomputing 2014 Special Session on "Robust local descriptors for computer vision". He is a member of the IEEE.

Vishal M. Patel received the B.S. degrees in electrical engineering and applied mathematics (Hons.) and the M.S. degree in applied mathematics from North Carolina State University, Raleigh, NC, USA, in 2004 and 2005, respectively, and the Ph.D. degree in electrical engineering from the University of Maryland College Park, MD, USA, in 2010. He is currently an A. Walter Tyson Assistant Professor in the Department of Electrical and Computer Engineering (ECE) at Rutgers University. Prior to joining Rutgers University, he was a member of the research faculty at the University of Maryland Institute for Advanced Computer Studies (UMIACS). His current research interests include signal processing, computer vision, and pattern recognition with applications in biometrics and imaging. He has received a number of awards including the 2016 ONR Young Investigator Award, the 2016 Jimmy Lin Award for Invention, A. Walter Tyson Assistant Professorship Award, the Best Paper Award at IEEE BTAS 2015, and Best Poster Awards at BTAS 2015 and 2016. He is an Associate Editor of the IEEE Signal Processing Magazine, IEEE Biometrics Compendium, and serves on the Information Forensics and Security Technical Committee of the IEEE Signal Processing Society. He is a member of Eta Kappa Nu, Pi Mu Epsilon, and Phi Beta Kappa.

Li Liu received the B.S. degree in communication engineering, the M.S. degree in photogrammetry and remote sensing and the Ph.D. degree in information and communication engineering from the National University of Defense Technology, China, in 2003, 2005 and 2012, respectively. She joined the faculty at the National University of Defense Technology in 2012, where she is currently an Associate Professor with the College of Information System and Management. During her PhD study, she spent more than two years as a Visiting Student at the University of Waterloo, Canada, from 2008 to 2010. From 2015 to 2016, she visited the Multimedia Laboratory at the Chinese University of Hong Kong. From 2016 to 2018, she is visiting Machine Vision Group at the University of Oulu, Finland. Her research interests include texture analysis, object detection and recognition, and scene classification.

Vili-Petteri Kellokumpu received the Ph.D. degree in Center for Machine Vision and Signal Analysis, University of Oulu, Finland, in 2011. He is currently a senior research scientist with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland. He has authored and co-authored about 20 papers in journals and conferences. He is a Program Committee Member for many conferences. His current research interests include gait analysis, and dynamic-texture recognition.

Guoying Zhao (SM'12) received the Ph.D. degree in computer science from the Chinese Academy of Sciences, Beijing, China, in 2005. She is currently an Associate Professor with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland, where she has been a senior researcher since 2005, and a visiting Professor in Northwest University in China. In 2011, she was selected to the highly competitive Academy Research Fellow position. She has authored or co-authored more than 160 papers in journals and conferences. Her papers have currently over 5800 citations in Google Scholar (h-index 33).She has served as area chairs for FG 2017 and WACV 2017 and is associate editor for Pattern Recognition and Image and Vision Computing Journals. She has lectured tutorials at ICPR 2006, ICCV 2009, and SCIA 2013, authored/edited three books and six special issues in journals. Dr. Zhao was a Co-Chair of 12 International Workshops at ECCV, ICCV, CVPR and ACCV, and two special sessions at FG13 and FG15. Her current research interests include image and video descriptors, facial-expression and micro-expression recognition, gait analysis, dynamic-texture recognition, human motion analysis, and person identification. Her research has been reported by Finnish TV programs, newspapers and MIT Technology Review.

Matti Pietikäinen received his Doctor of Science in Technology degree from the University of Oulu, Finland. He is currently a professor, Scientific Director of Infotech Oulu and Director of Machine Vision Group at the University of Oulu. From 1980 to 1981 and from 1984 to 1985, he visited the Computer Vision Laboratory at the University of Maryland. He has made pioneering contributions, e.g. to local binary pattern (LBP) methodology, texture-based image and video analysis, and facial image analysis. He has authored over 340 refereed papers in international journals, books and conferences. He was Associate Editor of IEEE Transactions on Pattern Analysis and Machine Intelligence and Pattern Recognition journals, and currently serves as Associate Editor of Image and Vision Computing and IEEE Transactions on Forensics and Security journals. He was President of the Pattern Recognition Society of Finland from 1989 to 1992, and was named its Honorary Member in 2014. From 1989 to 2007 he served as Member of the Governing Board of International Association for Pattern Recognition (IAPR), and became one of the founding fellows of the IAPR in 1994. He is IEEE Fellow for contributions to texture and facial image analysis for machine vision. In 2014, his research on LBP-based face description was awarded the Koenderink Prize for Fundamental Contributions in Computer Vision.

Rama Chellappa received the B.E. (Hons.) degree in electronics and communication engineering from the University of Madras, Chennai, India, in 1975, the M.E. (with Distinction) degree from the Indian Institute of Science, Bangalore, India, in 1977, and the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1978 and 1981, respectively. He was a Faculty Member with the Department of Electrical Engineering/Systems, University of Southern California (USC), from 1981 to 1991. Since 1991, he has been a Professor of electrical and computer engineering (ECE) and an affiliate Professor of computer science with the University of Maryland (UMD), College Park, MD, USA. He is also affiliated with the Center for Automation Research and the Institute for Advanced Computer Studies (Permanent Member) and is serving as the Chair of the ECE Department. In 2005, he was named a Minta Martin Professor of Engineering. He holds three patents. His current research interests include face recognition, clustering and video summarization, 3D modeling from video, image and video-based recognition of objects, events and activities, dictionary-based inference, compressive sensing, domain adaptation and hyper spectral processing.

Dr. Chellappa was the recipient of an NSF Presidential Young Investigator Award, four IBM Faculty Development Awards, an Excellence in Teaching Award from the School of Engineering at USC, and two paper awards from the International Association of Pattern Recognition (IAPR). He is a recipient of the K.S. Fu Prize from IAPR. He received the Society, Technical Achievement and Meritorious Service Awards from the IEEE Signal Processing Society. He was also the recipient of the Technical Achievement and Meritorious Service Awards from the IEEE Computer Society. At UMD, he was elected as a Distinguished Faculty Research Fellow and Distinguished Scholar-Teacher, received an Outstanding Innovator Award from the Office of Technology Commercialization, and received an Outstanding GEMSTONE Mentor Award from the Honors College. He received the Outstanding Faculty Research Award and the Poole and Kent Teaching

Award for Senior Faculty from the College of Engineering. In 2010, he was recognized as an Outstanding ECE by Purdue University. He is a Fellow of IAPR, OSA and AAAS.

Prof. Chellappa served as the Editor-in-Chief of the IEEE transactions on Pattern Analysis and Machine Intelligence. He has served as a General and Technical Program Chair for several IEEE international and national conferences and workshops. He is a Golden Core Member of the IEEE Computer Society and served as a Distinguished Lecturer of the IEEE Signal Processing Society. Recently, he completed a two-year term as the President of the IEEE Biometrics Council.