Journal of Electronic Imaging

JElectronicImaging.org

Face liveness detection using shearlet-based feature descriptors

Litong Feng
Lai-Man Po
Yuming Li
Fang Yuan

SPIE. IS&T imaging.org

Litong Feng, Lai-Man Po, Yuming Li, Fang Yuan, "Face liveness detection using shearlet-based feature descriptors," *J. Electron. Imaging* **25**(4), 043014 (2016), doi: 10.1117/1.JEI.25.4.043014.

Face liveness detection using shearlet-based feature descriptors

Litong Feng,* Lai-Man Po, Yuming Li, and Fang Yuan
City University of Hong Kong, Department of Electronic Engineering, 83 Tat Chee Avenue, Hong Kong, China

Abstract. Face recognition is a widely used biometric technology due to its convenience but it is vulnerable to spoofing attacks made by nonreal faces such as photographs or videos of valid users. The antispoof problem must be well resolved before widely applying face recognition in our daily life. Face liveness detection is a core technology to make sure that the input face is a live person. However, this is still very challenging using conventional liveness detection approaches of texture analysis and motion detection. The aim of this paper is to propose a feature descriptor and an efficient framework that can be used to effectively deal with the face liveness detection problem. In this framework, new feature descriptors are defined using a multiscale directional transform (shearlet transform). Then, stacked autoencoders and a softmax classifier are concatenated to detect face liveness. We evaluated this approach using the CASIA Face antispoofing database and replay-attack database. The experimental results show that our approach performs better than the state-of-the-art techniques following the provided protocols of these databases, and it is possible to significantly enhance the security of the face recognition biometric system. In addition, the experimental results also demonstrate that this framework can be easily extended to classify different spoofing attacks. *© 2016 SPIE and IS&T* [DOI: 10.1117/1.JEI.25.4.043014]

Keywords: antispoofing; liveness detection; stacked autoencoders; softmax classification; shearlet transform.

Paper 15583 received Jul. 20, 2015; accepted for publication Jul. 5, 2016; published online Jul. 21, 2016.

1 Introduction

Face verification and recognition are a widely used biometric technology due to its convenience and nonintrusive interaction. In the last decade, face detection and recognition technology has achieved substantial progress. However, recent works revealed that face biometrics is vulnerable to spoofing attacks using cheap low-tech equipment, such as the photograph or video of a valid user. Therefore, the antispoofing problem for the face biometric system has gained great attention in the research community.

In recent years, liveness detection has been a very active topic and received significant development in the fingerprint recognition and iris recognition communities. However, there is still a lack of effective approaches to deal with problems in face liveness detection. Usually, imposters will present a large number of spoofed biometrics into the system. In face recognition, the usual attack methods may be classified into several categories. The classification is based on what verification proof is provided to the face verification system, such as stolen face photos, recorded videos, and three-dimensional (3-D) face masks with the abilities of blinking and lip moving. The aim of face liveness detection is to differentiate between real faces and nonreal faces. In practice, the security level of a face biometric system will be significantly improved with the help of liveness detection. Face liveness detection is an important and challenging issue, which determines the trustworthiness of the biometric system's security against spoofing.

Most of the conventional face liveness detection algorithms can be classified into three types as (1) presence of vitality, (2) differences in motion patterns, and (3) differences in image quality assessment. For the first type, the presence of vitality detection techniques focuses on creating certain features that only live faces can possess. These methods usually analyze certain movements of certain facial components, such as eye blinking and lip moving, and will consider those movements as a sign of life and therefore a real face. For example, Sun et al.[1] proposed a blinking-based live face detection using conditional random fields. In addition, Jee et al.[2] proposed a method for detecting eyes in sequential input images and then variation of each eye region is calculated to determine the liveness status. For the second type, differences in motion patterns-based analysis mainly rely on the fact that real faces display a different motion behavior compared to a spoofing attempt. These methods mainly differentiate motion patterns between 3-D and two-dimensional (2-D) faces. The general idea of this type method is that planar objects move significantly different from real human faces, which are 3-D objects. Bao et al.[3] proposed a liveness detection method for face recognition based on an optical flow field. It analyzed the differences and properties of optical flow generated by 3-D objects and 2-D planes. The motion of an optical flow field is a combination of four basic movements: translation, rotation, moving, and swing. For the third type, image quality assessment-based analyses focus on the presence of artifacts intrinsically presented at attack medium. Tan et al.[4] developed two strategies to extract the essential information about different surface properties of a live human face or a photograph, in terms of latent samples. In addition, inspired by image quality assessment, characterization of printing artifacts, and differences in light reflection,

*Address all correspondence to: Litong Feng, E-mail: lightedfeng@gmail.com

Maatta et al.[5] presented a face liveness detection method using the microtexture analysis, which analyzed textures of facial images using multiscale local binary patterns (LBP). In addition, Li et al.[6] proposed a live face detection method that is based on the analysis of Fourier spectra of a single face image or face image sequences. It hypothesized that fraudulent photographs have less high-frequency components compared with real ones.

Conventional face liveness detection algorithms usually need to calculate or extract some explicit features using complicated modules, such as modeling the texture, the motion, or the life sign in the face images. These features focus on representing a specific characteristic that can distinguish between the real face images and nonreal face images very well. However, because of the specificity, these methods are hard to be generalized to other spoofing types. Thus, in this paper, we aim to explore a new general purpose face liveness detection algorithm that is based on shearlet transform. The general idea of this method is that the statistical property of real face images is usually constant. Nevertheless, nonreal face images usually contain more or less distortions in all directions. That is, the process of creating fake faces disturbs the statistical property of real face images and discriminates real face images from nonreal face images. Thus, we propose new feature descriptors based on shearlet transform and these descriptors can effectively distinguish between real face images and nonreal face images. Shearlet is a multiscale and multidirectional image descriptor, which is good at capturing anisotropic features. Compared with LBP, shearlet can better describe curvilinear singularities, including edges, textures, and artifacts.

The shearlet-based feature descriptors we proposed are multifunctional descriptors. We can apply the same descriptors for face liveness detection, spoofing attack classification, and face recognition. The extracted descriptors are then fed into stacked autoencoders (SAEs) that are concatenated with a softmax classifier. In this way, all these goals are achieved using a unified framework. In this paper, we focus on face liveness detection and spoofing attack classification problems.

The remainder of the paper is organized as follows. Section 2 introduces the detailed structure and related techniques about the proposed framework. In Sec. 3, experimental results and a thorough analysis of this framework are presented. Finally, a conclusion and future works are given in Sec. 4.

2 Methodology

A high-level overview of the proposed framework is shown in Fig. 1. An image or a video entering the framework is first subjected to a face detector, and each extracted face frame or frame sequence is gray-scaled. Then, shearlet-based feature descriptors are extracted from these face images. The extracted descriptors are applied to detect face liveness. For a real face, these descriptors can be directly used for face recognition. Therefore, the final output of this framework is a recognized real face. However, if it is a nonreal face, these descriptors can be also directly utilized for spoofing attack classification and the output indicates which type of spoofing attack is used. As previously described, we use SAEs and a softmax classifier to serve as a liveness detection system, a face recognition system, and a spoofing attack classification system. In this way, we can use the same features and a unified framework to deal with all three tasks. More details about this framework will be described in the following sections.

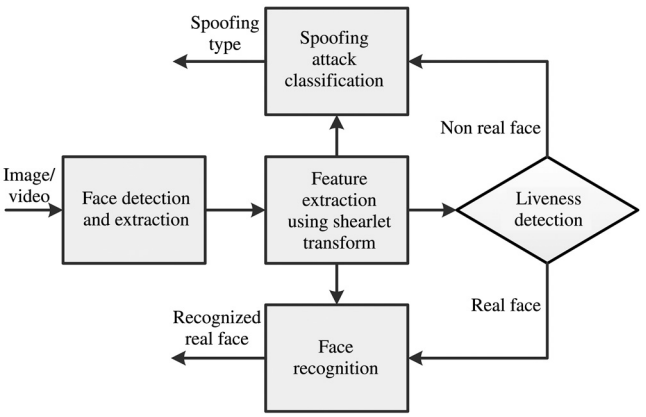

Fig. 1 High-level overview of the proposed framework.

2.1 *Shearlet Transform*

It is known that traditional wavelets and their associated transforms are highly efficient when approximating and analyzing one-dimensional signals. However, these frameworks have some limitations when extended to process multidimensional data such as images or videos. Typically, multidimensional data exhibit curvilinear singularities, which cannot be sparsely approximated using wavelet, because wavelet is ineffective in describing directions. To overcome the drawbacks of wavelets, a new class of multiscale analysis methods has been proposed in recent years, which is defined as the third generation wavelet. A noteworthy characteristic of these new methods is their ability to efficiently capture anisotropic features in multidimensional data and the shearlet representation[7–12] is one of them. The proposed feature descriptors are based on shearlet transform. When the dimension is $n = 2$, the affine systems with composite dilations are the collections of the form

$$SH_\phi f(a,s,t) = \langle f, \phi_{a,s,t} \rangle, \quad a > 0, \quad s \in R, \quad t \in R^2, \tag{1}$$

where the analyzing factor $\phi_{a,s,t}$ is called shearlet basis, which is defined as

$$\phi_{a,s,t}(x) = |\det M_{a,s}|^{-\frac{1}{2}} \phi(M_{a,s}^{-1} x - t), \tag{2}$$

where $M_{a,s} = B_s A_a = \begin{pmatrix} a & \sqrt{a}s \\ 0 & \sqrt{a} \end{pmatrix}$, and $A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$, $B_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$. A_a is the anisotropic dilation matrix and B_s is the shear matrix. The analyzing functions associated to the shearlet transform are anisotropic and are defined at different scales, locations, and orientations. Thus, shearlets have the ability to detect directional information and account for the geometry of multidimensional functions, which overcome the limitation of the wavelet transform.

2.2 *Shearlet-Based Feature Descriptors*

We start the derivation of our shearlet-based feature descriptors (SBFD) in a gray-scale image. The calculation process of SBFD is shown in Fig. 2. Each element in the red box is defined as

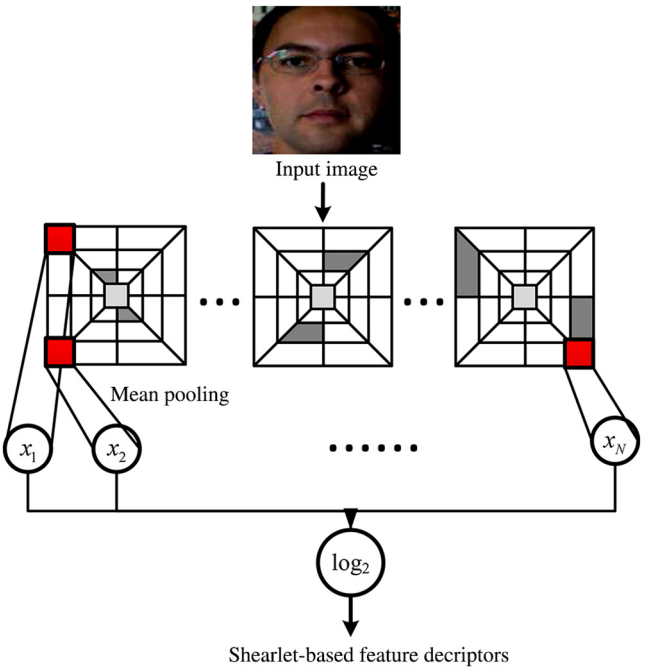

Fig. 2 The calculation process of SBFD.

$$x(a,s,b)=\frac{\sum|SH_{\phi}f(a,s,b)|}{m^2}, \tag{3}$$

where $a=1,\ldots,A$ is the scale index (exclude coarsest scale), $s=1,\ldots,S$ is the direction index and $b=1,\ldots,(M/m)^2$ is the block index of each subband. M represents the size of square image and m indicates the size of the red block. $SH_{\phi}f(a,s,b)$ are the shearlet coefficients of each red block.

After the mean pooling of shearlet coefficients in each red block, the pooled values are concatenated as a vector and subjected to a logarithmic nonlinearity, which is represented as

$$\text{SBFD}=\log_2(x_1,\cdots,x_N), \tag{4}$$

where $N=A\times S\times(M/m)^2$ is the total number of red block.

Figure 3 shows the SBFD of a 256×256 face image. In this example, the total scale number A is 4, total direction number S in each scale is 6, and the red block size m is 64. Therefore, the length of SBFD is 384.

To illustrate the effect of these descriptors, we randomly select 10 frames for all the high quality videos of 50 subjects in the CASIA face antispoofing database. Each frame of the original frame sequence is gray-scaled and passes through a face detector. The detected faces are geometric normalized to 256×256 pixels. Figure 4 shows the example of extracted and reshaped face images in the CASIA database. There are three types of fake face attacks in this database, which include warped photo, cut photo (eye-blink), and video attacks. The SBFD of each face image is extracted and

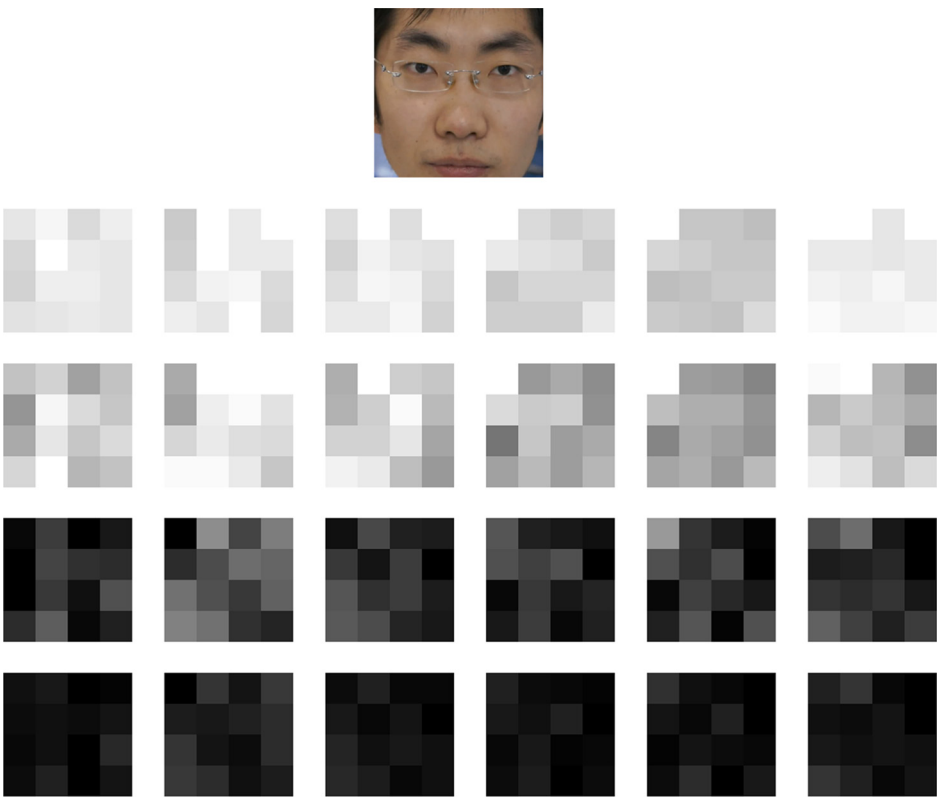

Fig. 3 Visualization of the SBFD for a 256×256 face image.

Fig. 4 Example of extracted and reshaped face images in the CASIA face antispoofing database. (a) Real face image. (b) Warped photo attack image. (c) Cut photo (eyeblink) attack image. (d) Video attacks image.

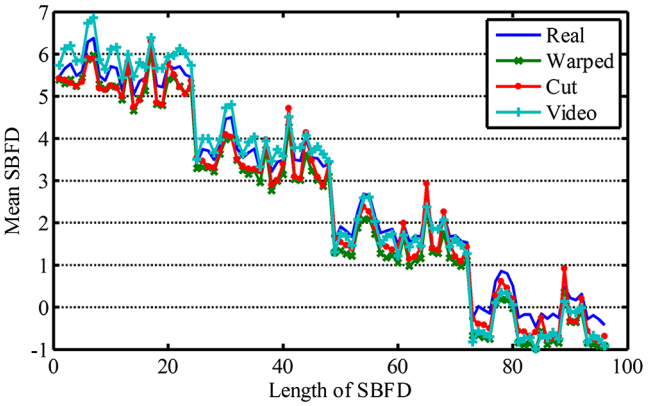

Fig. 5 Plot of the mean SBFD versus the length of SBFD.

the mean SBFD for different spoofing attacks is obtained. Figure 5 shows the mean SBFD versus the length of SBFD where the total length of SBFD is $N = 4 \times 6 \times (256/128)^2 = 96$. It may be observed that the mean SBFD varies with each type of spoofing attack. In addition, we can also observe that the video attack generally increases the low frequency but decreases the high frequency of the image, compared with real face images. However, the warped photo and cut photo attack decrease both low and high frequency of the image. Since the warped photo attack and cut photo are very similar except the eye area, the mean SBFD of these two types of attacks are also similar.

2.3 *Stacked Autoencoders and Softmax Classifier*

As previously mentioned, the extracted SBFD can be fed into SAEs and the final face liveness status, spoofing type, and user identification are predicted by a softmax classifier. The architecture of this framework is shown in Fig. 6.

Before being sent into the SAEs, the input SBFD is normalized by subtracting the mean and dividing by the standard deviation of its elements, and zero components analysis whitening is performed to the normalized SBFD. SAEs are a kind of neural network that contains multiple hidden layers and allows us to compute much more complex features of the input signal.[13–21] Since each hidden layer computes a nonlinear transformation of the previous layer, a deep network can achieve significantly greater representational power than a shallow one. Different from training the traditional back propagation (BP) neural network, two steps are implemented to obtain good parameters for an SAE. The first step is called pretraining, which is a kind of unsupervised training. In this step, each layer is

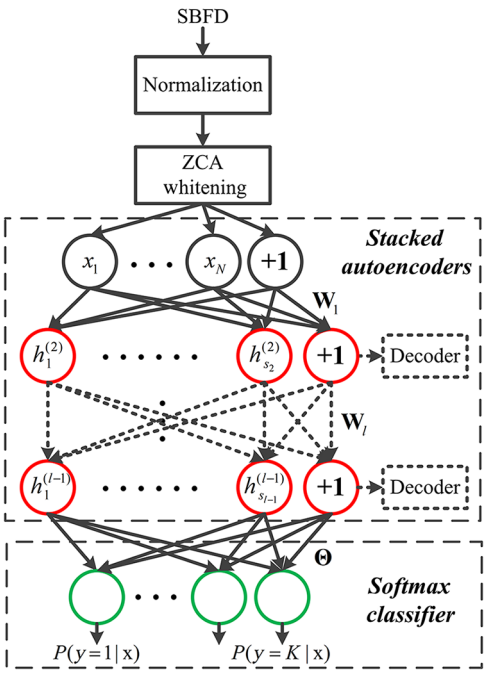

Fig. 6 The architecture of SAEs and softmax classifier.

treated as an individual autoencoder and the optimized encoding weights are obtained as the initial weights instead of random initialization. The second step is called fine-tuning, which is a kind of supervised training using a BP algorithm. Fine-tuning is a strategy that is commonly used in deep learning. Through this step, the performance of an SAE can be significantly improved. From a high level perspective, fine-tuning treats all layers of an SAE as a single model. For each iteration, all the weights in the SAE can be optimized.

The final output layer of this deep neural network is a softmax classifier. When performing the fine-tuning process, the parameters of softmax are also updated. The output is defined as

$$p(y^{(i)} = j|x^{(i)};\theta) = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{K} e^{\theta_l^T x^{(i)}}}, \tag{5}$$

where K is the class number and θ is the softmax parameter vector. For liveness detection and spoofing classification, K is 2 and 4, respectively.

3 Experiments and Related Analysis

3.1 *Experimental Protocol*

In order to effectively evaluate the performance of the proposed algorithm and other liveness detection algorithms, the following two publicly available databases are used, which contain multiple types of spoofing attack.

(1) The CASIA face antispoofing database.[22] This database contains 50 genuine subjects, and fake faces are made from the high quality records of the genuine faces. The database includes three imaging qualities (low, normal, and high) and three fake face attacks that consist of warped photo, cut photo (eye-blink), and video attacks.

Figure 7 shows one complete video set for a subject. There are a total of 600 video clips, and the subjects are divided into subsets for training and testing (240 and 360, respectively). A suggested test protocol is also provided that consists of seven scenarios and can be summarized as:

Quality test. This test is designed to evaluate the performance when image quality is fixed. The samples are:

1. Low (L) quality test: {L1, L2, L3, L4}.
2. Normal (N) quality test: {N1, N2, N3, N4}.
3. High (H) quality test: {H1, H2, H3, H4}.

Fake face test. This test is designed to evaluate the performance when fake face types are fixed. The samples are:

1. Warped photo attack test: {L1, N1, H1, L2, N2, H2}.
2. Cut photo attack test: {L1, N1, H1, L3, N3, H3}.
3. Video attack test: {L1, N1, H1, L4, N4, H4}.

Overall test. In this test, all data are combined together to give a general and overall evaluation.

Based on this suggested protocol, we design our experiments into two main parts that include liveness detection and spoofing attack classification.

In the liveness detection experiment, we identify only real face images and nonreal face images. In the spoofing attack classification experiment, we classify four different spoofing types. Therefore, we conduct only a quality test and overall test. We randomly select 10 face frames for each video and average the selected face image scores as the final label.

(2) The replay-attack database.[23] This database consists of 1300 video clips of photo and video attack attempts of 50 clients, under different lighting conditions. The spoofing attacks are generated in three different scenarios with two different lighting conditions and support conditions. The three types of spoofing attacks include: print (2),

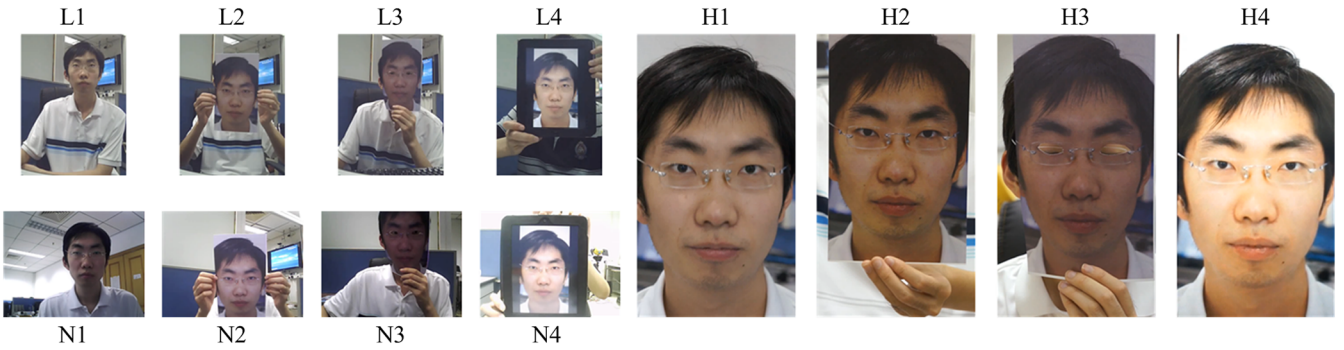

Fig. 7 Example images of real face and the corresponding spoofing attacks in the CASIA database.

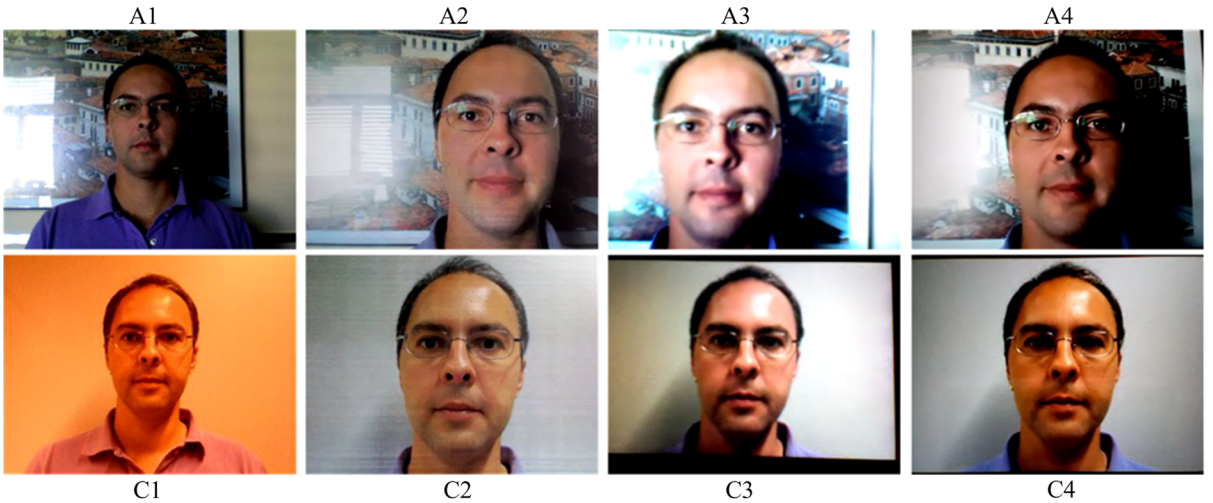

Fig. 8 Example images of real face and the corresponding spoofing attacks in the replay-attack database.

mobile (3), and highdef (4). The two lighting conditions include: adverse (A) and controlled (C). The two different support conditions include hand-based and fixed. In addition, this database also defines three nonoverlapping partitions for training, development, and testing. Figure 8 shows some example frames of real face and corresponding attacks. The testing protocols adopted in this paper are as following:

Fake face test. This test is designed to evaluate the performance when fake face types are fixed. The samples are:

1. Print photo attack test: {A1, C1, A2, C2}.
2. Mobile video attack test: {A1, C1, A3, C3}.
3. Highdef video attack test: {A1, C1, A4, C4}.

Overall test. In this test, all data are combined to give a general and overall evaluation. The samples are: {A1, C1, A2, C2, A3, C3, A4, C4}.

We use this database to test the performance of liveness detection and spoofing attack classification. For liveness detection, we apply the fake face test and overall test protocol. For spoofing attack classification, only the overall test protocol is adopted, similar to testing on the CASIA database.

As discussed previously, most state-of-the-art works apply LBP[24] as a feature extraction method and use SVM to identify real face images and nonreal face images. In this paper, when conducting liveness detection experiments, we first apply DoG, LBP, and SBFD as feature extraction methods and send the extracted features into SVM. Radial basis function kernel is selected for SVM. The parameters of SVM kernels were set using grid-search. In addition, in order to provide a rational and fair comparison, we also send the corresponding features into SAEs.

For both the CASIA database and the replay-attack database, we strictly train each framework using the suggested training set and report the classification accuracy for the testing set.

Parameters of algorithms: Shearlet transform is applied on each 256×256 gray-scale face image. The face image is decomposed into four scales (exclude approximation component) and the direction number for each scale is six. The pooling block size is 64 and the final SBFD length is 384. For LBP algorithm, block-based multiscale LBP is used.[23] For DoG algorithm, four DoG filters are considered: $\sigma_1 = 0.5$, $\sigma_2 = 1$; $\sigma_1 = 1$, $\sigma_2 = 1.5$; $\sigma_1 = 1.5$, $\sigma_2 = 2$; and $\sigma_1 = 1$, $\sigma_2 = 2$. The downsampling size is 16. The SAE consists of four layers, one input layer, two hidden layers, and one output layer. The neuron number for the hidden layers is 16 and 8, respectively. In addition, the weight decay parameter λ for both SAEs and softmax classifier is 5e-5. Sparsity parameter ρ is 0.1 and weight of sparsity penalty term β is 5.

In order to provide an intuitive demonstration about the proposed method, we made a demo video which shows some experimental results on the CASIA database. In this video, liveness detection and spoofing attack classification results for four randomly selected subjects are demonstrated. As previously described, SBFD also can be directly applied for face recognition. Therefore, in this video, we also demonstrate a preliminary experiment about the face recognition test. In the face recognition experiment, we identify the face image for 50 subjects in the CASIA database. The quality

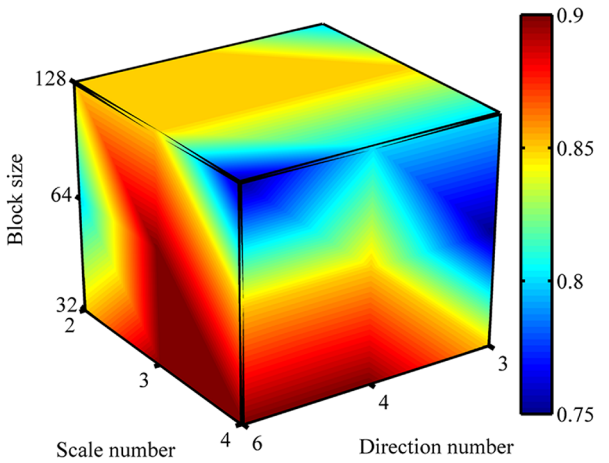

Fig. 9 Plot of classification accuracy for different combinations of scale number, direction number, and block size in high quality test.

Table 1 Classification accuracy of liveness detection test on the CASIA database.

		Low	Normal	High
Quality test	DoG + SVM	0.6767	0.7181	0.6970
	LBP + SVM	0.7806	0.8397	0.9000
	SBFD + SVM	0.9194	0.8996	0.8257
	DoG + SAE	0.6226	0.7477	0.7625
	LBP + SAE	0.7588	0.8305	0.9047
	SBFD + SAE	0.9470	0.9272	0.8797
		Warped	Cut	Video
Fake face test	DoG + SVM	0.6332	0.6521	0.7129
	LBP + SVM	0.8246	0.7991	0.8068
	SBFD + SVM	0.8395	0.9249	0.9211
	DoG + SAE	0.6253	0.6768	0.7208
	LBP + SAE	0.8558	0.8334	0.9024
	SBFD + SAE	0.8360	0.9315	0.9131
Overall test	DoG + SVM		0.6664	
	LBP + SVM		0.8392	
	SBFD + SVM		0.8381	
	DoG + SAE		0.7165	
	LBP + SAE		0.8545	
	SBFD + SAE		0.8918	

test and overall test are considered for this experiment. We randomly select 10 face frames for each video. Five face images are used for training purposes and another five face images are used for testing. There is no overlap between training and testing face images. In addition, we also demonstrate the results of DoG and LBP for comparison purposes. This demo video is available at the link found in Ref. 25.

In addition, in order to show that the proposed method can be directly used in real situations, we completed a real-time implementation of our method and tested it using real data that also include a print photo attack, mobile photo attack, and video attack. In this demo, 21 frames are analyzed for each detection and the final result is the average score of these 21 frames. This demo video is available at the link found in Ref. 26.

Table 2 Classification accuracy of liveness detection test on the testing set of the replay-attack database.

		Testing set		
		Print	Mobile	Highdef
Fake face test	DoG + SVM	0.7375	0.5875	0.6250
	LBP + SVM	0.7625	0.9375	0.8625
	SBFD + SVM	0.9250	1.000	0.8375
	DoG + SAE	0.6500	0.5625	0.5750
	LBP + SAE	0.9625	0.9625	0.9500
	SBFD + SAE	0.9000	1.000	0.9500
Overall test	DoG + SVM		0.5500	
	LBP + SVM		0.8125	
	SBFD + SVM		0.9250	
	DoG + SAE		0.6625	
	LBP + SAE		0.9375	
	SBFD + SAE		0.9500	

3.2 *Effectiveness of Parameters*

Several parameters are involved in the design of SBFD. In this experiment, we will first analyze and discuss how these parameters affect the performance of liveness detection. To examine the performance, we select several different combinations of SBFD parameters and test them using high quality test protocol. From Eq. (3), we can see that there are three adjustable parameters for SBFD. Therefore, in this experiment, we select scale number A as 2, 3, and 4; direction number S as 3, 4, and 6; block size m as 32, 64, and 128. In total, 27 types of combinations are considered. Figure 9 shows the classification accuracy for different combinations. We can observe that the performance shows the rising tendency with the increase of scale number and direction number, and decrease of block size. With the increases of scales and directions, SBFD can describe image quality more discriminatively. With the increase of blocks on the face image, the difference between fake and true faces can be described on different local facial regions using SBFD. Thus this difference can be represented more precisely.

3.3 *Performance Evaluation for Liveness Detection*

In this section, we will first test and compare the performance of each method on the liveness detection task. The experiments are conducted as the testing protocols detailed previously. The classification accuracy for each method under different testing scenarios of the CASIA database is listed in Table 1. It can be seen from the testing results that SBFD outperforms DoG and LBP in most scenarios. In addition, we can also notice that the classification effect of SAE is almost always better than SVM. Since we feed all the features into the same classification framework, the testing results reliably demonstrate that SBFD is more suitable for liveness detection task compared with DoG and LBP. Table 2 shows the classification accuracy for the testing set of the replay-attack database. The good performance of SBFD also can be demonstrated by these datasets.

Similarly,[22,27] we also plot detection-error trade-off (DET) curves[28] of the overall test for both the CASIA database and the replay-attack database, which are shown in Figs. 10(a) and 10(b), respectively. The DET plot also confirms the ability of SBFD.

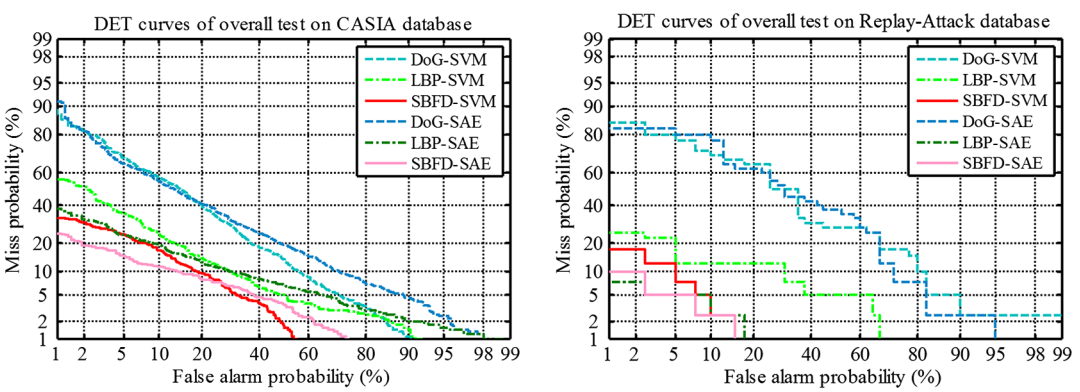

Fig. 10 (a) DET curves of six methods for overall test on the CASIA database. (b) DET curves of six methods for overall test on the replay-attack database.

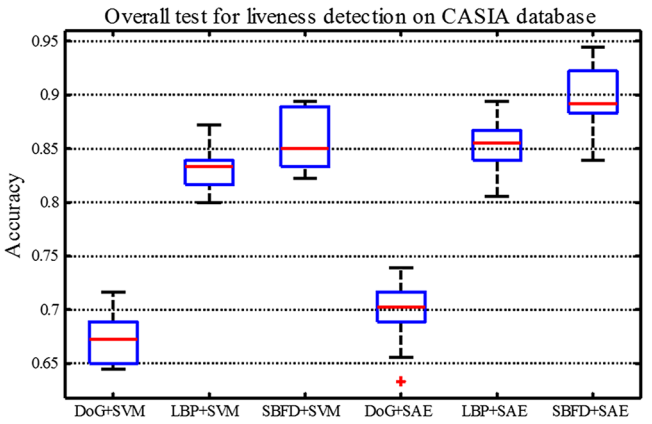

Fig. 11 Box plot of liveness detection accuracy of six methods for overall test on the CASIA database.

To visualize the statistical significance of the comparison, we show a box plot of the distribution of the classification accuracy values for the overall test on the CASIA database. The plot is shown in Fig. 11. Obviously, the lower the standard deviation with a higher classification accuracy, the better the performance is. The plot intuitively shows that SBFD statistically performs better than DoG and LBP, and SAE performs better than SVM when dealing with binary classification.

3.4 *Performance Evaluation for Spoofing Attack Classification*

To study whether the proposed method has the ability to distinguish different spoofing types, in this section, we conduct the experiments using the spoofing attack classification experiment protocols described previously. The spoofing attack classification task is actually the extension of the liveness detection problem where the class number changes from two to four. We utilize the same features as in the liveness detection, and the classification accuracy for each method under different testing scenarios of the CASIA database is listed in Table 3. In addition, Table 4 also provides the classification accuracy for the replay-attack database. We not only report the overall accuracy, but also list the accuracy of each individual spoofing type. Since the classification difficulty increases, the classification accuracy of all the three methods decreases. However, the performance of SBFD still significantly outperforms DoG and LBP in low quality, normal quality, and overall test of the CASIA database, and is very close to LBP in high quality test of the CASIA database and the replay-attack database.

Figure 12 shows the box plot of spoofing attack classification accuracy of three methods for the overall test on the CASIA database. Compared with the testing results of liveness detection, in the spoofing attack classification test, SBFD is still highly competitive and the testing results demonstrate that SBFD is also suitable for distinguishing different spoofing attacks.

3.5 *Discussion and Future Work*

All the above experimental results demonstrate that SBFD is a kind of multifunctional feature descriptor and we can utilize the same SBFD for many different applications. These good performances may be owed to the excellent mathematical properties of shearlets such as well localization, highly directional sensitivity, and optimal sparseness. Because of these properties, shearlet functions are actually a very good model for the oriented receptive fields of simple cells in the primary visual cortex (V1).[29] Through using shearlet

Table 3 Classification accuracy of spoofing attack classification test on the CASIA database.

	Real	Warped	Cut	Video	All
			Low		
DoG	0.3775	0.2584	0.3655	0.5936	0.3948
LBP	0.5766	0.5324	0.4133	0.6792	0.5551
SBFD	0.8850	0.7278	0.7617	0.9118	0.8251
			Normal		
DoG	0.4793	0.4450	0.5138	0.6147	0.4975
LBP	0.7062	0.5327	0.5202	0.6993	0.6261
SBFD	0.8923	0.7429	0.6655	0.8312	0.7842
			High		
DoG	0.4302	0.2759	0.3922	0.3866	0.3693
LBP	0.8648	0.8092	0.6736	0.8977	0.8159
SBFD	0.7131	0.7953	0.8240	0.8699	0.8009
			Overall		
DoG	0.4973	0.1941	0.4678	0.5743	0.4213
LBP	0.7174	0.6386	0.5552	0.7931	0.6823
SBFD	0.8468	0.7063	0.7685	0.8464	0.8002

Table 4 Classification accuracy of spoofing attack classification test on the replay-attack database.

	Real	Print	Mobile	Highdef	All
			Testing set		
DoG	0.4250	0.3500	0.4500	0.2500	0.3625
LBP	0.7975	0.8425	0.9675	0.8375	0.8831
SBFD	0.8800	0.8300	0.9750	0.6650	0.8375
			Development set		
DoG	0.2500	0.2300	0.4333	0.2923	0.3016
LBP	0.9267	0.8567	0.9533	0.7633	0.8750
SBFD	0.8833	0.8367	0.9967	0.6733	0.8475

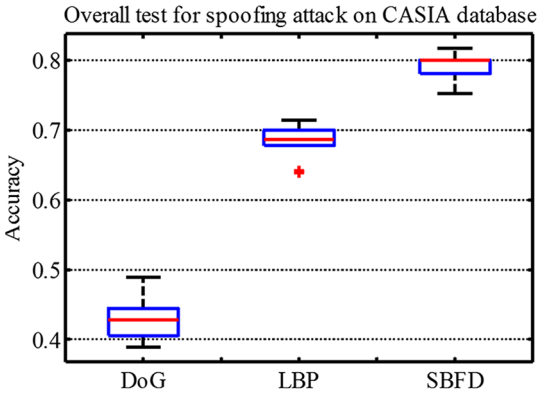

Fig. 12 Box plot of spoofing attack classification accuracy of three methods for overall test on the CASIA database.

functions, the receptive field is reduced to a small number of parameters and these parameters are enough to describe the basic selectivity properties of simple cells. In addition, humans tend to perceive "poor" regions in an image and these "poor" regions heavily affect the subjective impression. Therefore, when designing SBFD, we adopt a pooling method to further reduce the dimensionality of shearlet coefficients and abstract the input image. There are several pooling types usually adopted in deep learning problems,[30,31] such as percentile pooling, max or min pooling, sum pooling, and average (mean) pooling. In this paper, we empirically observed that mean pooling performs better than other pooling methods.

Since a real face is a nonrigid object with contractions of facial muscles that result in temporally deformed facial features, it can be assumed that the specific temporal information should also be detected when a live human face is observed in front of the camera. Therefore, in the future, we will extend the feature descriptors using 3-D shearlet transform,[32] which is the perfect extension of the 2-D shearlet transform. Through 3-D shearlet transform, we can achieve a spatial-temporal feature representation and the proposed face liveness detection method is naturally extended to videos.

4 Conclusion

In this paper, we have proposed a multifunctional feature descriptor and an efficient framework that can be used to deal with face liveness detection and spoofing attack classification. This unified framework is based on shearlet transform, SAEs, and a softmax classifier. We evaluated this approach using the CASIA face antispoofing database and the replay-attack database. The results show that our approach is highly competitive and suitable for both of the two tasks.

Acknowledgments

The work described in this paper was substantially supported by a grant from the City University of Hong Kong, Kowloon, Hong Kong, with Project number of 7004058.

References

1. L. Sun et al., "Blinking-based live face detection using conditional random fields," in *Advances in Biometrics*, N. K. Ratha and V. Govindaraju, Eds., pp. 252–260, Springer-Verlag, London (2007).
2. H. K. Jee, S. U. Jung, and J. H. Yoo, "Liveness detection for embedded face recognition system," *Int. J. Biomed. Sci.* **1**(4), 235–238 (2006).
3. W. Bao et al., "A liveness detection method for face recognition based on optical flow field," in *Proc. IEEE Int. Conf. on Image Analysis and Signal Processing*, pp. 233–236 (2009).
4. X. Tan et al., "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *Computer Vision–ECCV 2010*, pp. 504–517, Springer, Berlin, Heidelberg (2010).
5. J. Maatta, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *Proc. IEEE Int. Joint Conf. on Biometrics*, pp. 1–7 (2011).
6. J. Li et al., "Live face detection based on the analysis of Fourier spectra," *Proc. SPIE* **5404**, 296–303 (2004).
7. Y. Sheng et al., "A shearlet approach to edge analysis and detection," *IEEE Trans. Image Process.* **18**(5), 929–941 (2009).
8. G. Easley, D. Labate, and W. Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Appl. Comput. Harmon. Anal.* **25**(1), 25–46 (2008).
9. G. Kutyniok and W. Q. Lim, "Image separation using wavelets and shearlets," in *Curves and Surfaces*, J. -D. Boissonnat et al., Eds., pp. 419–430, Springer, Berlin Heidelberg (2012).
10. G. Kutyniok, W. Q. Lim, and X. Zhuang, "Digital shearlet transforms," in *Shearlets*, G. Kutyniok and D. Labate , Eds., pp. 239–282, Birkhäuser, Boston (2012).
11. G. Kutyniok, M. Shaharm, and X. Zhuang, "ShearLab: a rational design of a digital parabolic scaling algorithm," *SIAM J. Imag. Sci.* **5**(4), 1291–1332 (2012).
12. G. Kutyniok, M. Shahram, and D. L. Donoho, "Development of a digital shearlet transform based on pseudo-polar FFT," *Proc. SPIE* **7446**, 74460B (2009).
13. G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science* **313**, 5786, 504–507 (2006).
14. A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks," in *Proc. Annual Conf. on Neural Information Processing Systems*, pp. 1097–1105 (2012).
15. M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional neural networks," in *Computer Vision–ECCV 2014*, pp. 818–833, Springer, International Publishing (2014).
16. V. Jain and S. Seung, "Natural image denoising with convolutional networks," in *Proc. Annual Conf. on Neural Information Processing Systems*, pp. 769–776 (2009).
17. J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Annual Conf. on Neural Information Processing Systems*, pp. 341–349 (2012).
18. H. Larochelle et al., "An empirical evaluation of deep architectures on problems with many factors of variation," in *Proc. ACM Int. Conf. on Machine Learning*, pp. 473–480 (2007).
19. D. Erhan et al., "Why does unsupervised pre-training help deep learning?" *J. Mach. Learn. Res.* **11**, 625–660 (2010).
20. I. Goodfellow et al., "Measuring invariances in deep networks," in *Proc. Annual Conf. on Neural Information Processing Systems*, pp. 646–654 (2009).
21. J. Masci et al., "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Artificial Neural Networks and Machine Learning-ICANN 2011*, pp. 52–59, Springer, Berlin Heidelberg (2011).
22. Z. Zhang et al., "A face antispoofing database with diverse attacks," in *Proc. IEEE IAPR Int. Conf. on Biometrics*, pp. 26–31 (2012).
23. I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. Int. Conf. of the Biometrics Special Interest Group*, pp. 1–7 (2012).
24. T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987 (2002).
25. Y. Li, "Demo of SBFD," 2014, https://www.youtube.com/watch?v=kUCC0hLSJaU (20July2015).
26. Y. Li, "Liveness real-time demo," 2015, https://www.youtube.com/watch?v=q-nhCSt-BWY (20July2015).
27. A. B. Teoh, A. Goh, and D. C. Ngo, "Random multispace quantization as an analytic mechanism for biohashing of biometric and random identity inputs," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(12), 1892–1901 (2006).
28. A. Martin et al., *The DET Curve in Assessment of Detection Task Performance*, National INST of Standards and Technology, Gaithersburg, Maryland (1997).
29. A. Hyvärinen, J. Hurri, and P. O. Hoyer, *Natural Image Statistics: a Probabilistic Approach to Early Computational Vision*, Springer Science & Business Media, London (2009).
30. P. Ye et al., "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1098–1105 (2012).
31. L. Kang et al., "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1733–1740 (2014).
32. P. S. Negi and D. Labate, "3-D discrete shearlet transform and video processing," *IEEE Trans. Image Process.* **21**(6), 2944–2954 (2012).

Litong Feng received his BE degree in electronic science and technology from Harbin Institute of Technology, Harbin, China, in 2008, and his ME degree in optical engineering from Tianjin Jinhang Institute of Technical Physics, Tianjin, China, in 2011. He is currently working toward the PhD in the Department of Electronic Engineering, City University of Hong Kong. His research interests include video processing for vital signs, face liveness detection, and optical design.

Lai-Man Po received his BS and PhD degrees in electronic engineering from the City University of Hong Kong, Kowloon, Hong Kong, in 1988 and 1991, respectively. He is currently an associate professor and lab director of TI Educational Training Centre in the Department of Electronic Engineering, City University of Hong Kong. He has published over 160 technical journal and conference papers. His current research interests are in the areas of video coding and content-based video copy detection.

Yuming Li received his BEng degree from Huazhong University of Science and Technology in 2011 and his MEng degree from Huazhong University of Science and Technology in 2013. He is currently pursuing his PhD at the City University of Hong Kong. His research interests include image and video processing, multiscale analysis, and machine learning.

Fang Yuan received his BSc degree in physics from Central South University, Changsha, China, in 2009, and his ME degree in communication and information system from Sun Yat-sen University, Guangzhou, China, in 2012. He is currently working toward his PhD in the Department of Electronic Engineering, City University of Hong Kong. His research interests include signal processing and machine learning.