FACE LIVENESS DETECTION AND RECOGNITION USING SHEARLET BASED FEATURE DESCRIPTORS

Yuming Li, Lai-Man Po, Xuyuan Xu, Litong Feng, Fang Yuan

Department of Electronic Engineering, City University of Hong Kong, Kowloon, 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 non-real faces such as a photograph or video of valid user. 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 develop a multifunctional feature descriptor and an efficient framework which can be used to deal with both face liveness detection and recognition. In this framework, new feature descriptors are defined using a multiscale directional transform (shearlet transform). Then, stacked autoencoders and softmax classifier are concatenated to detect face liveness and identify person. We evaluated this approach using CASIA Face Anti-Spoofing Database and the results show that our approach performs better than state-of-the-art techniques following the provided evaluation protocols of this database, and is possible to significantly enhance the security of face recognition biometric system.

Index Terms—Anti-spoofing, Face recognition, Liveness detection, Stacked autoencoders, Softmax classification, Shearlet transform.

1. INTRODUCTION

In the last decade, face detection and recognition technology achieved substantial progress. However, recent works have revealed that face biometrics is vulnerable to spoofing attacks using cheap low-tech equipment, such as the photograph or video of valid user. Therefore, anti-spoof problem for face biometric system has gained great attention to the research community.

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 focus on creating certain features that only live faces can possess. These methods usually analyze certain movements of certain parts of the face, such as eye blinking and lip moving, and will consider those movements as a sign of life and therefore a real face [1] [2]. 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 spoof attempt. These methods mainly differentiate the motion pattern between 3D and 2D faces. The generally idea about this type method is that planar objects move significantly different from real human faces which are 3-D objects [3]. For the third type, image quality assessment based analysis focus on the presence of artefacts intrinsically present at the attack media [4]-[6].

Conventional face liveness detection algorithms usually need to calculate or extract some explicit features using complicated modules. These features focus on representing a specific characteristic which can very well distinguish the real face images and non-real face images. However, because of the specificity, these methods are hard to generalize to other spoofing types. Thus, in this paper, we aim to explore a new general purpose face liveness detection algorithm which is based on shearlet transform. The general idea about this method is that the process of creating fake faces disturbs the statistical property of real face image and discriminate it from real face images to non-real face images. Besides, the proposed feature descriptors are multifunctional descriptors, which means the same descriptors can be applied for both face liveness detection and face recognition. In this paper, extracted descriptors are feed into stacked autoencoders which are concatenated with softmax classifier. In this way, all these goals are achieved based on a unified framework. The framework of the proposed approach is summarized in Fig. 1. An image or a video entering the framework is first subjected to a face detector. Then, shearlet based feature descriptors are extracted from these face images. The extracted descriptors are applied to detect the liveness of the face. If it is a real face, these descriptors can be directly used for face recognition.

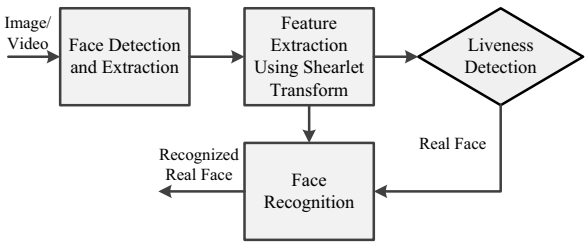

Fig. 1: High-level overview of the proposed framework.

2. SHEARLET BASED FEATURE DESCRIPTORS

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 and wavelets cannot effectively detect their directions and in the sense sparsely approximate them. 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]-[9] is one of them. The proposed feature descriptors are based on shearlet transform. When the dimension $n=2$, the affine systems with composite dilations are the collections of the form:

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

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

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

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.

We start the derivation of our Shearlet Based Feature Descriptors (SBFD) in a gray scale image. The calculation process of SBFD is summarized in Fig. 2. Each element in the red box is defined as

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

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

$$SBFD = \log_2(x_1, \cdots, x_N) \quad (4)$$

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

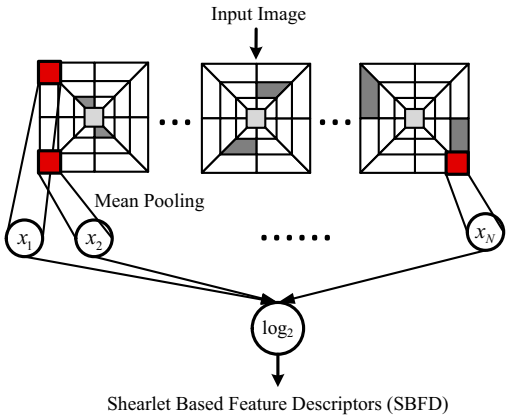

Fig. 2: The calculation process of SBFD.

3. STACKED AUTOENCODERS AND SOFTMAX CLASSIFIER

As previously mentioned, the extracted SBFD can be feed into stacked autoencoders (SAE) and the final face liveness and face type are predicted by a softmax classifier. Before being sent into the stacked autoencoders, the input SBFD is normalized by subtracting the mean and dividing by the standard deviation of its elements, and Zero Components Analysis (ZCA) whitening is performed to the normalized SBFD. Stacked autoencoders is a kind of deep neural networks which contain multiple hidden layers and allow us to compute much more complex features of the input signal [10] [11]. Different from training the traditional Back Propagation (BP) neural network, two steps are implemented to obtain good parameters for a stacked autoencoder. The first step is called pre-training, which is a kind of unsupervised training. In this step, each layer is 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 BP algorithm. Fine-tuning is a strategy that is commonly used in deep learning. Through this step, the performance of a stacked autoencoder can be significantly improved. From a high level perspective, fine-tuning treats all layers of a stacked autoencoder as a single model, so that in one iteration, we are improving upon all the weights in the stacked autoencoder. The final output layer of this deep neural network is softmax classifier. When performing the fine-tuning process, the parameters of softmax are also updated. The output is defined as

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

where K is the class number and θ is the softmax parameter vector. For liveness detection and face recognition, K is 2 and 50 respectively.

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of proposed algorithm and other liveness detection algorithms, the CASIA Face Anti-Spoofing Database [12] are considered. This database contains 50 genuine subjects, and fake faces are made from the high quality records of the genuine faces. There include three imaging qualities (low, normal and high) and three fake face attacks which include warped photo, cut photo (eyeblink) and video attacks. Besides, a baseline algorithm is also provided which is based on multiple Difference of Gaussian (DoG) filters and SVM. In addition, a suggested test protocol is also provided which consists of 7 scenarios and can be summarized as

Quality Test. This test is 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 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 which include liveness detection and face recognition. In liveness detection experiment, we only identify real face images and non-real face images. Therefore, when doing quality test and overall test, in order to make sure that the number of these two type face images is identical, we randomly select 30 face frames from each real face video and 10 face frames from each non-real face video. However, when conducting fake face test, we randomly select 10 face frames for each video. The final label for the video is determined by averaging the selected face image scores. In face recognition experiment, we identify the face image for 50 subjects. Quality test and overall test are considered for this experiment. In this experiment, we randomly select 10 face frames for each video. 5 face images are used for training and another 5 face images are used for testing. There are no overlap between training and testing face images. As discussed previously, most state-of-the-art works apply Local Binary Patterns (LBP) [13] as feature extraction method and use SVM to identify real face images and non-real face images. Therefore, when conducting liveness detection experiments, we first apply DoG, LBP and SBFD as feature extraction methods and send the extracted features into SVM. Besides, in order to provide a rational and fair comparison, we also send the corresponding features into stacked autoencoders.

To give a statistical evaluation, we have not adopted the suggested training and testing set. Instead, we mix the 50 subjects together, for each training and testing process, we randomly select 20 subjects as training set and the other 30 subjects as testing set. Totally 100 train and test iterations are performed and the median classification accuracy is reported for each algorithm.

Table 1: Median classification accuracy for 100 iterations of liveness detection test on the CASIA database.

		Low	Normal	High
Quality Test	DoG+SVM	0.6833	0.7167	0.6833
	LBP+SVM	0.7667	0.8500	***0.9000***
	SBFD+SVM	***0.9333***	0.9000	0.8167
	DoG+SAE	0.6333	0.7500	0.7500
	LBP+SAE	0.7500	0.8167	***0.9000***
	SBFD+SAE	***0.9333***	***0.9167***	0.8667
		Warped	**Cut**	**Video**
Fake Face Test	DoG+SVM	0.6278	0.6444	0.7056
	LBP+SVM	0.8278	0.7944	0.8167
	SBFD+SVM	0.8333	***0.9389***	***0.9278***
	DoG+SAE	0.6389	0.6889	0.7111
	LBP+SAE	***0.8500***	0.8389	0.8889
	SBFD+SAE	***0.8500***	0.9333	0.9167
Overall Test	DoG+SVM	0.6611		
	LBP+SVM	0.8333		
	SBFD+SVM	0.8444		
	DoG+SAE	0.7167		
	LBP+SAE	0.8556		
	SBFD+SAE	***0.8889***		

Table 2: Median classification accuracy for 100 iterations of face recognition test on the CASIA database.

	Low	Normal	High	Overall
DoG	0.2320	0.2680	0.1640	0.1480
LBP	0.4040	0.3560	0.3320	0.2320
SBFD	***0.9840***	***0.9960***	***0.9880***	***0.9720***

Table 1 and Table 2 list the median classification accuracy for liveness detection and face recognition respectively. It can be seen that the proposed method achieves competitive performance for both liveness detection and face recognition task. In addition, for liveness detection test, we also plot the Detection-Error Trade-off (DET) curves and the box plot of the classification accuracy of the overall test, which are show in Fig. 3(a) and (b) respectively. For face recognition test, the box plot of the classification accuracy of the overall test is show in Fig. 3(c), and Fig. 4(a) to (c) shows the mean confusion matrixes across 100 trails for overall test. These plots further demonstrate the proposed method is suitable for these two tasks.

5. CONCLUSIONS

In this paper, we have proposed a multifunctional feature descriptor and an efficient framework which can be used to deal with face liveness detection and face recognition. This unified framework is based on shearlet transform, stacked autoencoders and softmax classifier. We evaluated this approach using CASIA Face Anti-Spoofing database. The results show that our approach is suitable for both of the two tasks.

6. ACKNOWLEDGEMENTS

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.

7. REFERENCES

[1] Sun, L., Pan, G., Wu, Z., & Lao, S. (2007). Blinking-based live face detection using conditional random fields. In *Advances in Biometrics* (pp. 252-260). Springer Berlin Heidelberg.

[2] Jee, H. K., Jung, S. U., & Yoo, J. H. (2006). Liveness detection for embedded face recognition system. *International Journal of Biomedical Sciences*, 1(4), 235-238.

[3] Bao, W., Li, H., Li, N., & Jiang, W. (2009, April). A liveness detection method for face recognition based on optical flow field. In *Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on* (pp. 233-236). IEEE.

[4] Tan, X., Li, Y., Liu, J., & Jiang, L. (2010). 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.

[5] Maatta, J., Hadid, A., & Pietikainen, M. (2011, October). Face spoofing detection from single images using micro-texture analysis. In *Biometrics (IJCB), 2011 International Joint Conference on* (pp. 1-7). IEEE.

[6] Li, J., Wang, Y., Tan, T., & Jain, A. K. (2004, August). Live face detection based on the analysis of fourier spectra. In *Defense and Security* (pp. 296-303). International Society for Optics and Photonics.

[7] Easley, Glenn, Demetrio Labate, and Wang-Q. Lim. "Sparse directional image representations using the discrete shearlet transform." *Applied and Computational Harmonic Analysis* 25, no. 1 (2008): 25-46.

[8] Kutyniok, Gitta, Wang-Q. Lim, and Xiaosheng Zhuang. "Digital Shearlet Transforms." In *Shearlets*, pp. 239-282. Birkhäuser Boston, 2012.

[9] Kutyniok, Gitta, Morteza Shahram, and Xiaosheng Zhuang. "Shearlab: A rational design of a digital parabolic scaling algorithm." *arXiv preprint arXiv*:1106.1319 (2011).

[10] Erhan, Dumitru, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. "Why does unsupervised pre-training help deep learning?" *The Journal of Machine Learning Research* 11 (2010): 625-660.

[11] Masci, Jonathan, et al. "Stacked convolutional auto-encoders for hierarchical feature extraction." *Artificial Neural Networks and Machine Learning–ICANN 2011*. Springer Berlin Heidelberg, 2011. 52-59.

[12] Zhang, Zhiwei, et al. "A face antispoofing database with diverse attacks." *Biometrics (ICB), 2012 5th IAPR International Conference on*. IEEE, 2012.

[13] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.7 (2002): 971-987.

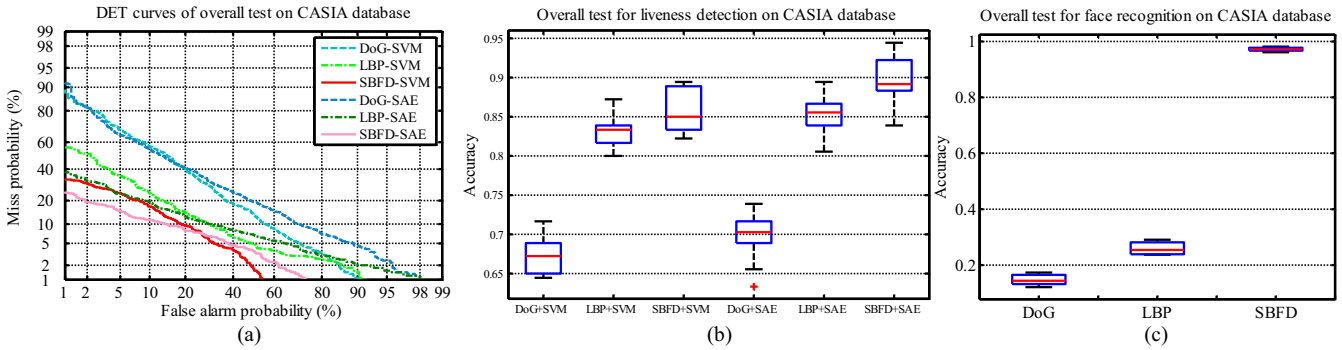

Fig. 3: (a) DET curves of six methods for overall test after 100 iterations. (b) Box plot of liveness detection accuracy of six methods over 100 trials for overall test. (c) Box plot of face recognition accuracy of three methods over 100 trials for overall test.

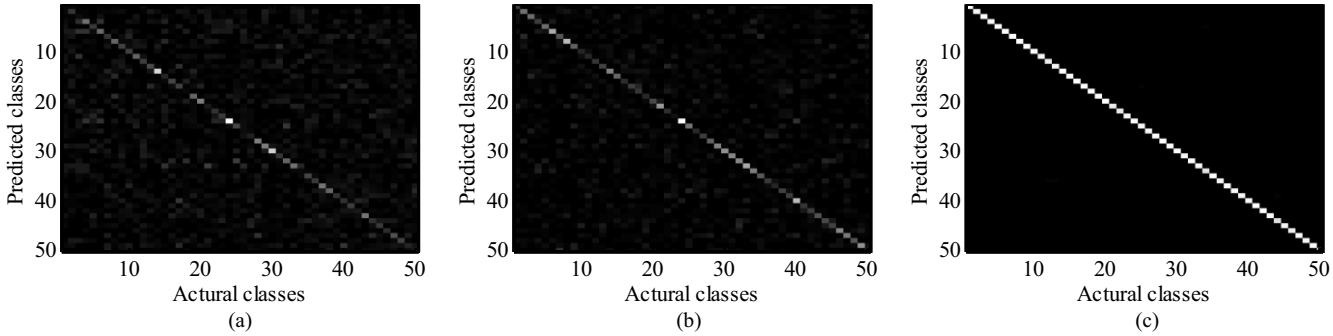

Fig. 4: Mean confusion matrix for face recognition across 100 trails for overall test. (a) Mean confusion matrix of DoG. (b) Mean confusion matrix of LBP. (c) Mean confusion matrix of SBFD.