J. Vis. Commun. Image R. 38 (2016) 451–460

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

# Integration of image quality and motion cues for face anti-spoofing: A neural network approach ☆

Litong Feng [a,*], Lai-Man Po [a], Yuming Li [a], Xuyuan Xu [a], Fang Yuan [a], Terence Chun-Ho Cheung [b], Kwok-Wai Cheung [c]

[a] *Department of Electronic Engineering, City University of Hong Kong, Hong Kong Special Administrative Region*
[b] *Department of Information Systems, City University of Hong Kong, Hong Kong Special Administrative Region*
[c] *Department of Computer Science, Chu Hai College of Higher Education, Hong Kong Special Administrative Region*

ARTICLE INFO

*Article history:*
Received 29 June 2015
Revised 5 January 2016
Accepted 20 March 2016
Available online 1 April 2016

*Keywords:*
Face anti-spoofing
Neural network
Feature fusion
Shearlet
Dense optical flow

ABSTRACT

Many trait-specific countermeasures to face spoofing attacks have been developed for security of face authentication. However, there is no superior face anti-spoofing technique to deal with every kind of spoofing attack in varying scenarios. In order to improve the generalization ability of face anti-spoofing approaches, an extendable multi-cues integration framework for face anti-spoofing using a hierarchical neural network is proposed, which can fuse image quality cues and motion cues for liveness detection. Shearlet is utilized to develop an image quality-based liveness feature. Dense optical flow is utilized to extract motion-based liveness features. A bottleneck feature fusion strategy can integrate different liveness features effectively. The proposed approach was evaluated on three public face anti-spoofing databases. A half total error rate (HTER) of 0% and an equal error rate (EER) of 0% were achieved on both REPLAY-ATTACK database and 3D-MAD database. An EER of 5.83% was achieved on CASIA-FASD database.

## 1. Introduction

Face recognition has been wildly applied in user authentication systems due to the non-intrusive and natural interaction of face biometrics [1–3]. A high security requirement for face authentication is urgent, because only a photography, video replay, or 3D-mask can easily spoof a face recognition system to access secure information illegally [4]. Photos and videos of a valid user can be easily obtained, especially through social network. Hence the face anti-spoofing module is indispensable to a face authentication system besides the face recognition module. Recently, many attentions have been paid to countermeasures to facial spoofing attacks. A variety of face anti-spoofing algorithms were proposed based on different approaches [5]. Several public face anti-spoofing databases were established [6–8]. And competitions on countermeasures to facial spoofing attacks boosted the development in face liveness detection [9].

Face image quality, facial motions, and scenic motions provide different liveness cues for face anti-spoofing. Various hand-engineered features have been proposed to describe liveness cues from the image quality-based or motion-based aspects, such as local binary pattern (LBP) and histogram of optical flow (HOOF) [10,11]. In this paper, shearlet transform is explored to discriminate between real faces and fake faces as an image quality descriptor. In contrast with the popularly used LBP, which is a local texture pattern feature, shearlet transform is more effective in representing distributed discontinuities, providing multi-scale and multi-directional anisotropic descriptions in images as directional wavelets [12]. On the aspect of motion cues, raw optical flow magnitude (OFM) is directly fed into an autoencoder neural network to learn motion-based liveness features in the cropped facial region and the whole scene, respectively. Compared with previous motion-based features, no motion hypothesis or statistic model is utilized to set motion priors in this paper. Therefore, the proposed motion-based liveness feature can achieve more generalization across face anti-spoofing databases with different scenic settings.

Due to varying attack scenarios and environment conditions, there is no absolutely superior face anti-spoofing technique. The combination of liveness features from the image quality-based and motion-based visual cues provides a promising direction to enhance the generalization and stability of a face anti-spoofing classifier. All the state-of-the-art face spoofing countermeasures

☆ This paper has been recommended for acceptance by Susanto Rahardja.
[*] Corresponding author.
*E-mail address:* lightedfeng@gmail.com (L. Feng).

http://dx.doi.org/10.1016/j.jvcir.2016.03.019

take advantage of the feature fusion or score fusion approach, such as spatial–temporal texture descriptor [13], winning algorithms in the 2nd competition on countermeasures to 2D face spoofing attacks [9], and the complementary countermeasures [14]. Currently, a direct concatenation of feature vectors at feature level or a fusion at score level is wildly utilized. There are few attentions paid on developing a fusion strategy of different liveness features from multiple visual cues. In this paper, a feature fusion framework for integrating features from multiple categories of liveness cues is proposed using a neural network approach. The autoencoder neural network adopted is not only a supervised classifier but can also generate the bottleneck feature, which is a compressed sparse representation of the raw input for the neural network [15]. The bottleneck feature can represent raw inputs more effectively in a reduced dimension with a unit-scaled amplitude [16]. Hence, the liveness bottleneck features learned from different visual cues can be concatenated without scaling. The fused bottleneck features can be fed into a succeeding neural network to perform the final liveness detection. This hierarchical neural network can integrate liveness features from multiple visual cues and learn a complementary countermeasure to face spoofing attacks. Considering user-friendliness and convenience, the challenge-response and multimodality approaches are not discussed in this paper [17], because a face anti-spoofing method transparent to users is pursued here.

Compared to previous work, the contributions of this paper can be summarized as:

(1) Shearlet transform is utilized to perform face image quality assessment, providing a better image quality descriptor than the popularly used LBP.
(2) Motion-based liveness features are automatically learned using the neural network from raw optical flow information in the cropped facial region and the whole scene, respectively. With the pursuit of the generalization of the motion-based feature, no motion assumption or scenic model for face anti-spoofing is adopted.
(3) A feature fusion framework for integrating the image quality-based and the motion-based liveness cues is proposed using a hierarchical neural network. A higher face anti-spoofing classification accuracy is achieved by the proposed approach compared with the state-of-the-art methods.

The remainder of the paper is organized as follows. A brief literature review of the state-of-the-art face anti-spoofing methods is given in Section 2. The proposed feature fusion framework is explained in details in Section 3. Three public face anti-spoofing databases utilized in this paper are introduced in Section 4. Extensive experiments are conducted on the three public databases, and the corresponding results are reported in Section 5. Finally, a conclusion is drawn in Section 6.

## 2. Related work

Existing face anti-spoofing methods can be mainly classified into four categories: extra hardware-aided, image quality-based, motion-based, and multi-cues integration-based.

### *2.1. Extra hardware-aided*

In addition to 2D images captured by regular cameras in the visible spectrum, extra hardware can provide other valuable information to distinguish between genuine faces and fake faces. Zhang et al. [18] selected two groups of light-emitting diodes (LED) with working spectra at 850 nm and 1450 nm as active light sources. Two corresponding photodiodes can detect the discriminant reflectance between real faces and fake materials. Lagorio et al. [19] captured 3D scanned points of face surfaces using an optoelectronic stereo system. 3D curvatures computed from acquired scanned data can distinguish between genuine faces and 2D spoofing mediums. Sun et al. [20] collected images in the visible spectrum and the infrared spectrum from a face simultaneously, with the combination of a visible camera and a thermal camera. Canonical correlation analysis of a pair of visible/thermal images using patched cross-modality was performed to detect face liveness. All the hardware-aided methods mentioned above obtained good performance on their private evaluation databases. However, the commercial applications of hardware-aided face anti-spoofing methods will be limited by requirements of setting up extra hardware.

### *2.2. Image quality-based*

Spatial liveness information is extracted from static face images, with the expectation that fake face images displayed by spoofing mediums (paper, LCD screen, mask, etc.) will have image quality different from that of natural real face images, including sharpness, local artifacts, textureness, and statistical properties in some transform domains [21]. High frequency components of fake face images are much reduced in the Fourier spectrum compared with the cases of real face images [22]. Zhang et al. [7] transformed face images into a series of frequency bands using multiple difference of Gaussian (DoG) filters. Genuine and fake face images possess discriminant distributions in the DoG domain, and an overall equal error rate (EER) of 17% was achieved on the CASIA-FASD database. Micro-texture difference was observed between real and fake face images due to their surface differences [10]. LBP is a successful texture operator to describe micro-texture information. Block-based multi-scale LBP codes can obtain half total error rates (HTER) of 13.87%, 18.21%, and 0.95% on the REPLAY-ATTACK, CASIA-FASD, and 3D-MAD databases, respectively [6,8]. A parameterization combining 25 general image quality measures was proposed to perform the fake biometric detection, including face, iris, and fingerprint biometrics [21], and a comparable result with that of LBP can be achieved on the REPLAY-ATTACK database. Menotti et al. [23] extracted deep representations of raw face images through optimizing a deep convolutional neural network (CNN). This deep learning approach can gain a HTER of 0.75% on the REPLAY-ATTACK database

### *2.3. Motion-based*

Motion patterns on the face or suspicious motion cues in the scenario are valuable temporal liveness information. Based on the region examined, motion-based face anti-spoofing can be grouped into two sub-categories, face motion-based and scene motion-based.

#### *2.3.1. Face motion-based*

Eye-blinking is a typical face motion cue wildly utilized in early work [24]. A human face is a non-rigid 3D object, exhibiting different optical flow trajectories compared with a 2D photo face. Kollreider et al. [25] assumed that the correlation between different facial components' motions is discriminant between real faces and photography faces. Bao et al. [26] proposed a motion model to describe the optical flow field of planar objects, and a divergence from this mode was assumed to exist in a real face's motion. Bharadwaj et al. [11] utilized Eulerian motion magnification to amplify subtle facial motions in a specialized frequency band. Macro- and micro-facial expressions presented by real faces can