Expert Systems With Applications 142 (2020) 113002

Contents lists available at ScienceDirect

Expert Systems With Applications

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

SLNet: Stereo face liveness detection via dynamic disparity-maps and convolutional neural network

Yasar Abbas Ur Rehman*, Lai-Man Po, Mengyang Liu

Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, Kowloon, China

ARTICLE INFO

Article history:
Received 19 June 2019
Revised 6 September 2019
Accepted 2 October 2019
Available online 5 October 2019

Keywords:
Convolutional neural networks
Face anti-spoofing
Face liveness detection
Face-biometrics
Stereo camera
APCER
ACER

ABSTRACT

Current state-of-the-art dual camera-based face liveness detection methods utilize either hand-crafted features, such as disparity, or deep texture features to classify a live face and face Presentation Attack (PA). However, these approaches limit the effectiveness of classifiers, particularly deep Convolutional Neural Networks (CNN) to unknown face PA in adverse scenarios. In contrast to these approaches, in this paper, we show that supervising a deep CNN classifier by learning disparity features using the existing CNN layers improves the performance and robustness of CNN to unknown types of face PA. For this purpose, we propose to supervise a CNN classifier by introducing a disparity layer within CNN to learn the dynamic disparity-maps. Subsequently, the rest of the convolutional layers, following the disparity layer, in the CNN are supervised using the learned dynamic disparity-maps for face liveness detection. We further propose a new video-based stereo face anti-spoofing database with various face PA and different imaging qualities. Experiments on the proposed stereo face anti-spoofing database are performed using various test case scenarios. The experimental results indicate that our proposed system shows promising performance and has good generalization ability.

1. Introduction

Face recognition techniques have gained widespread attention from biometric communities in the recent decade. Along with other ubiquitous biometric recognition, such as fingerprint, iris, and palm, face recognition has been a dominant mode of biometric identification for authentication and security purposes. The advantages of using face biometric-based systems for access control in various electronic applications are its convenient use, user-friendliness, fast response, cleanliness, and involve minimal human interaction with the device (Galbally, Marcel & Fierrez, 2014a).

Without face anti-spoofing (also known as face liveness detection) support, face recognition systems are vulnerable to varieties of face spoofing attacks, also known as face Presentation Attacks (PA) (Li, Correia & Hadid, 2018; Rehman et al., 2019). These face PA can be easily generated to gain illegal access to the user device such as cellphone, computer, or intangible assets such as bank accounts. Face PA can be classified into three main types: printed photo PA, replay-video PA, and face-mask PA. While face-mask PA is not readily available because of higher production cost, the former two face PA can be easily generated using high-definition photography or videography. Due to the availability of high-end cameras and printers, and the easy access to social media platforms like Twitter, Facebook, Instagram, WeChat, and YouTube, it has become easier to obtain a photograph and a video of a person's face for face PA production. As a result, face anti-spoofing has become indispensable for face recognition based authentication and verification systems.

In recent years, a multitude of techniques has been developed to detect face PA in face recognition systems. These techniques can be grouped into sensor-based face liveness detection systems and software-based face liveness detection systems. The sensor-based face liveness detection systems utilize RGB-D (Sun, Huang & Liu, 2018) cameras, Near Infrared Imaging (NIR) (Song & Liu, 2018), thermal cameras (Seo & Chung, 2019), and Kinetics (Erdogmus & Marcel, 2013) to detect liveness cues in the input face image or video, that can help to identify a live face and face PA. Although the performance of these approaches is quite remarkable compared to software-based approaches, their implementation and maintenance costs are high, which limit its use in portable and handheld electronic devices. On the other hand, software-based face liveness detection methods utilize off the shelf cameras for face image capturing and performing face liveness detection. These techniques either exploit hand-crafted features, deep Convolutional Neural Net-

* Corresponding author.
E-mail addresses: yaurehman2-c@my.cityu.edu.hk (Y.A.U. Rehman), eelmpo@my.cityu.edu.hk (L.-M. Po), Mengyaliu7-c@my.cityu.edu.hk (M. Liu).

https://doi.org/10.1016/j.eswa.2019.113002
0957-4174/© 2019 Elsevier Ltd. All rights reserved.

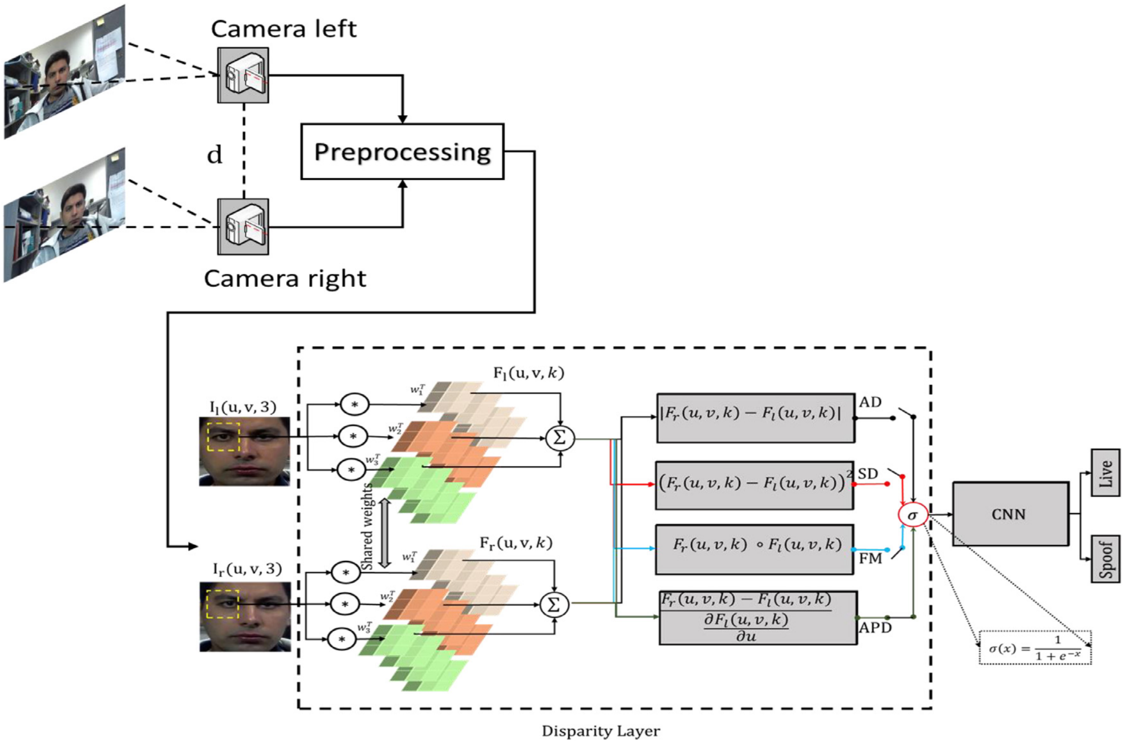

Camera left
d
Preprocessing
Camera right
$I_l(u, v, 3)$
$F_l(u, v, k)$
w_1^T
w_2^T
w_3^T
Shared weights
$I_r(u, v, 3)$
$F_r(u, v, k)$
$|F_r(u,v,k) - F_l(u,v,k)|$
AD
$(F_r(u,v,k) - F_l(u,v,k))^2$
SD
$F_r(u,v,k) \circ F_l(u,v,k)$
FM
$\frac{F_r(u,v,k) - F_l(u,v,k)}{\frac{\partial F_l(u,v,k)}{\partial u}}$
APD
σ
CNN
Live
Spoof
$\sigma(x) = \frac{1}{1+e^{-x}}$
Disparity Layer

Fig. 1. Generalized pipeline for the proposed stereo-camera based face liveness detection using CNN.

works (CNN) (Rehman, Po & Liu, 2018), or a combination of both (Abbas, Rehman, Po, Liu & Zou, 2020; Nguyen, Pham, Baek & Park, 2018) to perform face liveness detection.

Current state-of-the-art software-based face liveness detection techniques that operate on dual cameras exploit the disparity or depth information in the input face image or video to detect the face PA (Sun, Huang & Liu, 2016). Although these methods have shown better performance in intra-database face liveness detection scenarios, their performance degrades in general on unknown face PA, and in adverse conditions. One common reason for the drop in performance is that these methods train a classifier like Support Vector Machine (SVM) and CNN using the disparity or depth information computed directly from the RGB or grayscale images. Since the disparity or depth data in case of RGB and grayscale image have fixed number of channels, training a CNN in particular directly on this data limits its capability to learn discriminative and generalized feature maps that can help in identifying unknown face PA.

In order to circumvent this problem, we propose to supervise a CNN by learning dynamic disparity-maps from the existing convolutional layers of CNN. The dynamic disparity-maps represents multi-channel disparities learned from the convolutional layers in a CNN with stereo RGB data as input. In contrast to the fixed disparity computed from stereo RGB and grayscale data, the learned dynamic disparity-maps represents broad patterns in the stereo RGB face data that can discriminate between a live face and a face PA. Fig. 1 shows the general pipeline of the proposed system. As shown in Fig. 1, the input to the proposed system are preprocessed stereo face images from the left camera and the right camera, followed by the custom-designed disparity layer. The disparity layer consists of two convolutional layers with shared weights that output k feature maps. In contrast to the RGB or gray-scale disparity that has a fixed number feature of channels, the proposed method can incorporate any number of feature channels by varying the parameter k in the disparity layer. In this work, we use $k = 8$. However, in practice, any number of k feature maps can be generated by varying the parameter k. The outputs of the two convolutional layers in the disparity layer are followed by the disparity blocks, with sigmoid activation, to learn the dynamic disparity-maps between the convolutional features learned from the stereo input face images. The learned dynamic disparity-maps supervise the rest of the CNN layers, following the disparity layer, in an end-to-end fashion. Consequently, the disparity between the convolutional features are learned using Square Disparity (SD), Absolute Disparity (AD), Feature Multiplication (FM), and Approximate Disparity (APD) operations. In general, only one disparity block is used each time the network is trained.We performed experiments in each category to analyze its performance for face liveness detection. The main contribution of the proposed work are summarized as follows:

1. Different from previous approaches, we propose a low-cost stereo camera-based face liveness detection method that utilizes dynamic-disparity maps learned from convolutional layers of CNN with stereo RGB data as input. For this purpose, we designed a custom disparity layer, the output of which is used to supervise the rest of CNN layers for face liveness detection.
2. We evaluate various forms of learned dynamic disparity-maps using operation such as SD, AD, FM, and APD for face liveness detection in controlled and adverse scenarios.
3. We also propose a novel stereo camera-based face anti-spoofing database for face liveness detection and provide a detail explanation of generating stereo real face images and stereo face PA.
4. Extensive experiments on three designed protocols are performed to evaluate the effectiveness of the proposed method against unknown face PA. Further, the proposed method is tested by introducing variations in the input data to check its robustness in adverse conditions.

The organization of the rest of the paper is as follows: In Section 2, we review state-of-the-art methods in face liveness detection using both monocular and stereo camera-based techniques. The details of the proposed stereo camera-based face liveness detection method are provided in Section 3. Section 4 provides the details of the proposed stereo camera-based face anti-spoofing database, the experimental protocols, and performance evaluation. Finally, the paper concluded with a conclusion and future work in Section 5.

2. Literature review

Software-based face anti-spoofing techniques can be categorized into two domains: hand-crafted features-based and learnable or dynamic features-based techniques. Hand-crafted features-based face anti-spoofing techniques utilized liveness cues in face images such as motion, texture, and spectral energy contents (Pinto, Pedrini, Schwartz & Rocha, 2015). On the other hand, learnable or dynamic features-based techniques, like CNN, learn features directly from the raw data fed to it. Therefore, the features learned by CNN are dynamic and represent a broad range of patterns in the data as compared to hand-crafted features-based methods (Krizhevsky, Sutskever & Hinton, 2012; Rehman et al., 2018).

Hand-crafted features-based face anti-spoofing techniques can be classified into three domains: motion-based (de Freitas Pereira et al., 2014), texture-based (Määttä, Hadid & Pietikäinen, 2012), and image quality based. Face anti-spoofing methods based on motion exploits vitality or liveness sign in the face images such as eye blinking, lips movement, temperature, and optical-flow (Kim, Yoo & Choi, 2011; Kollreider, Fronthaler, Faraj & Bigun, 2007; Anjos, Chakka & Marcel, 2013). These techniques utilized the fact that motion patterns between a 2D planer face images and 3D real face images are different. The texture-based approaches compute either global texture features like image quality (Galbally, Marcel & Fierrez, 2014b; Wen, Han & Jain, 2015) or local texture features (Boulkenafet, Komulainen & Hadid, 2016; Määttä, Hadid & Pietikäinen, 2011; Määttä et al., 2012) such as LBP (Chingovska, Anjos & Marcel, 2012) and HOG (Yang, Lei, Yi & Li, 2015), using independent quantization or joint quantization schemes (Gragnaniello, Poggi, Sansone & Verdoliva, 2015). These techniques were proved to be quite robust in the detection of different types of face presentation attacks. There is a vast literature available for monocular camera-based face spoofing detection using CNN classifiers. These techniques either perform liveness detection at the frame-level (De Souza, Da Silva Santos, Pires, Marana & Papa, 2017; Jourabloo, Liu & Liu, 2018; Li et al., 2016; Menotti et al., 2015; Nguyen et al., 2018) or video-level (Lakshminarayana, Narayan, Napp, Setlur & Govindaraju, 2017; Xu, Li & Deng, 2015) on 2D face anti-spoofing databases like as NUAA database (Tan, Li, Liu & Jiang, 2010), Idiap Replay Attack database (Chingovska et al., 2012), and CASIA (Zhang et al., 2012) database.

The disparity or depth of face obtained from the stereo camera is a vital cue to detect face PA produced from planer mediums (e.g., print and tablet face PA). For example, Wang, Yang, Lei, Liao and Li (2013) used a 3D face structure recovery method by utilizing face images captured from a single camera in different views. According to their method, they map the 2D landmark points of the face on 3D frontal face structure to calculate the 3D structure for live faces and PA. However, the 3D face structure recovery method using 2D face images is computationally expensive and required multiple stages of camera calibration and facial structure refinement. Song, Zhao, Fang and Lin (2019), first obtained facial landmark points using a calibrated pair of cameras. These facial landmark points were then registered with a template frontal face precomputed using a stereo camera. The final landmarks were then used as a feature vector for face liveness detection. In Atoum, Liu, Jourabloo and Liu (2017), the face-depth and face-patches like eyes, nose, mouth, and eyebrows were utilized with CNN for face liveness detection. Two CNN were utilized for achieving the face liveness detection task: one for classification of face-patch and second for obtaining the depth map from the input face image. The output depth map is then fed to an SVM classifier, and score-level fusion strategy was used to improve the face liveness detection rate further. Although obtaining remarkable accuracy on intra-database face liveness detection, no results were reported for cross-database face liveness detection. Related work in Wang, Nian, Li, Meng and Wang (2017) utilized texture features from CNN and depth maps obtained from kinetic sensors for the detection of the live face and PA.

In Liu, Jourabloo and Liu (2018), the authors utilized the estimated 3D information calculated using 3D Morphable model with remote photoplethysmography (rPPG) signals for face liveness detection. However, obtaining a stable rPPG signal is time-dependent, and it introduces additional latency. In Di Martino, Qiu, Nagenalli and Sapiro (2018), the authors utilized the disparity information between the two binocular images and flashlight to detect spoofing attacks. However, the authors only utilized a fixed number of images captured using controlled conditions. Similar work in Sun et al. (2016) fused the 2D and 3D features for face liveness detection. They evaluated their method based on only two types of face PA, i.e., printed and tablet PA.

3. Methodology

In contrast to computing disparity between dual-camera images prior to training a CNN classifier, the proposed method takes advantage of simultaneously learning the dynamic disparity-maps and training a CNN classifier in an end-to-end fashion. Additionally, the proposed dynamic disparity-maps represents multi-channel disparity features that represent a broad range of features compared to the fixed features-based disparity. Further, feeding the learned dynamic disparity-maps enhances the performance of CNN classifier against unknown face PA, both in controlled and adverse conditions, which are explained in the following sections.

3.1. Preprocessing

Before feeding the stereo face images to the CNN, the stereo face images were first registered and normalized. To register the two stereo images, we utilized a single landmark point matching technique. Given two images captured by two cameras that are horizontally aligned, we first detect the face area and 68 facial landmarks point using Dlib Histogram of Oriented Gradient (HOG) based face detector (DLib Face Detector, 2019). After obtaining the face and landmark from both images, the image captured from the left camera is registered with the corresponding image captured from the right camera based on the corner landmark location of the right eye in both the images as shown in Fig. 2.

Let suppose P_{re}^{rc} denotes the landmark location of the right eye-corner in the right camera face image and P_{re}^{lc} denotes the landmark location of the right eye-corner in the left camera face image. Then, the rigid transformation T between P_{re}^{rc} and P_{re}^{lc} can be represented by Eqs. (1)–(5). Since both left and right cameras are identical and aligned horizontally, therefore only the translation component between the two images based on the landmark points P_{re}^{rc} and P_{re}^{lc} is computed. In Eq. (2), I represent 3×3 identity matrix. The image captured by the right camera is considered as a reference frame, and the image captured by the left camera is translated by tx_{diff} in the x direction and ty_{diff} in the y direction respectively based on landmark locations of the right-eye corner in both face images.

$$P_{re}^{rc} = TP_{re}^{lc} \tag{1}$$

$$T = [I] + t[I] \tag{2}$$

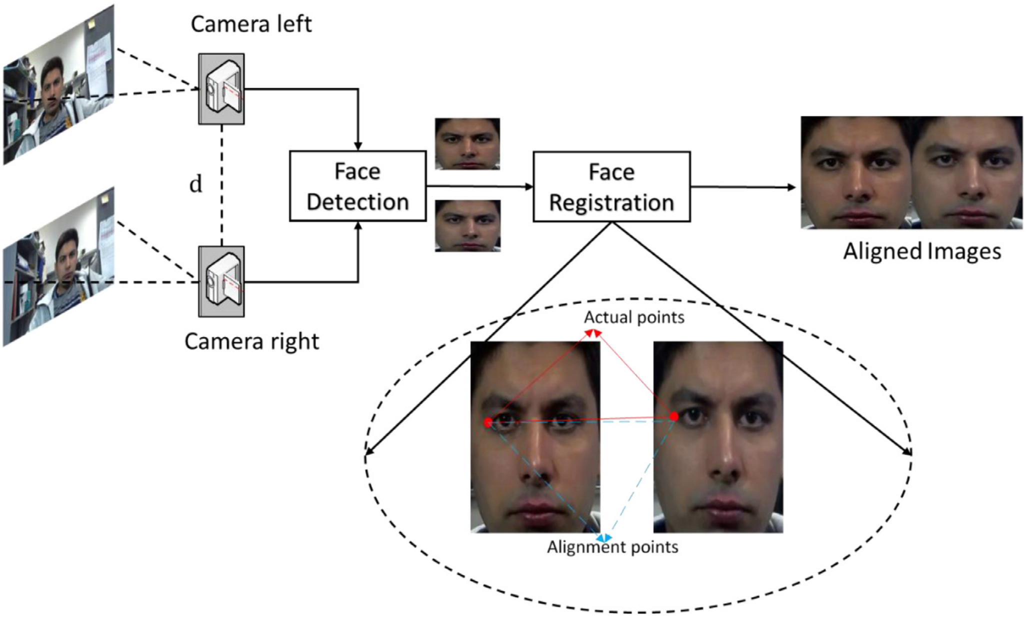


Fig. 2. Stereo face generation using face alignment. The image is captured by the left and right camera, followed by face detection and landmark estimation. The right eye corner landmark point in both cameras frame is selected as the key feature point, and accordingly, the left camera frame is adjusted (only translation in x-direction and y-direction) according to the right camera frame to obtain an aligned image.

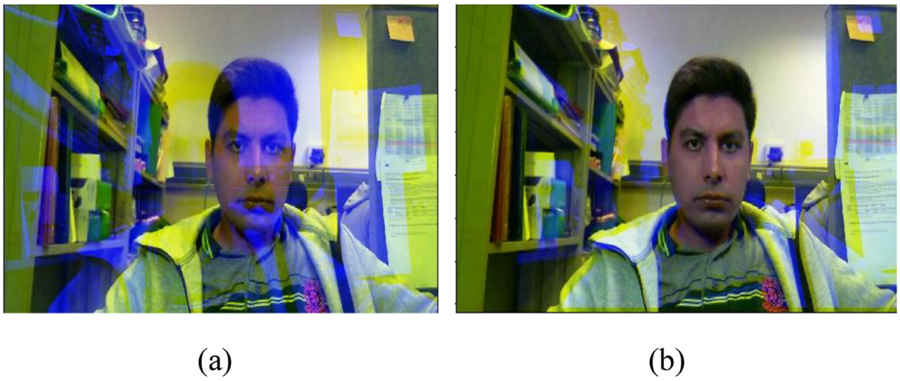


Fig. 3. Result of single point matching. (a) Image before alignment: The blue channel of the left camera image is visible. (b) The image after alignment using single point matching.

$$t = \begin{bmatrix} tx_{diff} \\ ty_{diff} \\ 1 \end{bmatrix} \tag{3}$$

$$tx_{diff} = tx^{rc}_{re} - tx^{lc}_{re} \tag{4}$$

$$ty_{diff} = ty^{rc}_{re} - ty^{lc}_{re} \tag{5}$$

Fig. 3 shows the result of using the above procedure, where the blue channel of the left-camera image is superimposed on the right camera image. As can be seen in Fig. 3(a), the blue channel of the left camera is at some distance d from the right-camera image before tthe transformation. Fig. 3(b) shows the result after the transformation of the left-camera image using the translation of the landmark points. Although there are multiple methods for stereo-face alignment and matching, we found that the proposed method of stereo-face alignment using single point matching is sufficient for face liveness detection system. It should be noted that stereo-matching is a broad field in computer vision, and many efficient stereo-matching techniques have been proposed in recent years. However, the proposed stereo-face alignment is only utilized to align the facial region in the RGB color space. Additionally, we do not perform the disparity computation at the image-level as usually performed in stereo-matching techniques; instead, we learn the Dynamic Disparity-Maps from the deep features within the CNN. Therefore, although the proposed stereo-face alignment technique is simple, yet it is sufficient and adequate for the proposed task, i.e., face liveness detection.

3.2. Dynamic disparity-maps

After stereo face alignment, we input the left-face image and the right-face image captured by the stereo camera to the CNN to first learn the dynamic disparity-maps using the proposed disparity layer. The dynamic disparity-maps are learned by utilizing Square Disparity (SD), Absolute Disparity (AD), Feature Multiplication (FM) and Approximate Disparity (APD) between the convolutional feature maps.

$$AD_k(u, v, k) = \sigma\left(|F_r(u, v, k) - F_l(u, v, k)|\right) \tag{6}$$

$$SD_k(u, v, k) = \sigma\left((F_r(u, v, k) - F_l(u, v, k))^2\right) \tag{7}$$

$$FM_k(u,v,k) = \sigma\left(F_r(u,v,k) \ \circ F_l(u,v,k)\right) \tag{8}$$

$$APD_k(u,v,k) = \sigma\left(\frac{F_r(u,v,k) - F_l(u,v,k)}{\frac{\partial F_l(u,v,k)}{\partial u}}\right) \tag{9}$$

In Eqs. (6)–(9), the symbol σ is the sigmoid activation function defined as: $\sigma(x) = \frac{1}{1+e^{-x}}$, $F_r(u,v,\ k)$ represents the kth convolutional feature-maps learned by convolution layer from the right camera $I_r(u,v)$ face image, $F_l(u,v,\ k)$ represents the kth convolutional feature-map learned by convolution layer from the left camera $I_l(u,v)$ face image, AD_k, SD_k, FM_k *and* APD_k are the disparity-maps learned for kth convolutional feature-map and $(u,\ v)$ are the spatial size of the respective kth convolutional feature-map. The learned dynamic disparity-maps are then fed to the rest of the convolution layers to learn the discriminative features for face liveness detection. The proposed CNN is guided by these learned dynamic disparity-maps as compared to stereo face images and hand-crafted disparity proposed by Di Martino et al. (2018). The partial derivative in the denominator of Eq. (9) is computed by using depth-wise convolution (Chollet, 2017) with the following gradient operator:

$$R_x = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{10}$$

In general, the proposed dynamic disparity-maps can be combined to generate new types of dynamic disparity maps. For example, the proposed dynamic disparity-maps can be added, subtracted, and concatenated together. However, in the present case, we only utilized one type of dynamic disparity-maps at a time, in the disparity layer, to evaluate its effectiveness in CNN for face liveness detection.

3.3. CNN architecture

Table 1 shows the proposed CNN architecture, which consists of 17 convolutional layers except for the disparity layer. The disparity layer at the top of CNN architecture is used to learn dynamic disparity-maps from the stereo face images. The disparity layer output 8 feature maps which are fed to next convolutional layer. Each convolution layer in our proposed CNN consists of a 3 × 3 kernel, followed by Batch-Normalization (BN) and Rectified Linear Unit (RELU). We also use a dropout rate of 0.2 after each max-pooling layer and an additional l_2 regularization of 0.0005 in convolution layers. We also used the concatenation layers denoted as F_i, where i indicate the concatenation layer index, to concatenate the output of the intermediate convolutional layers, as shown in Table 1. After F_5 concatenation layer, all the concatenation layers are mapped to 2 feature-maps using 1 × 1 convolutional layers that result in 10 feature maps of the same size that are concatenated again by F_6 concatenation layer. The output of F_6 concatenation layer is then given to the Global Average Pooling (GAP) layer that averages each feature map and output a 10-element feature vector. This 10-element feature vector is then given to a 2-way softmax classifier for classification of the input stereo face image as live face or face PA. Since, there is no learning in the GAP layer, the gradient computation from the soft-max layer is available to each layer in the CNN network. As a result, there are multiple path for the gradient to flow back to the CNN during back-propagation stage thus minimizing the vanishing gradient problem as suggested in Simonyan and Zisserman (2014), He, Zhang, Ren and Sun (2016) and Huang, Liu, Van Der Maaten and Weinberger (2017).

Table 1
Architecture of proposed CNN.

Input image = $[I_r(u,v),\ I_l(u,v)]$			
Layer name	Kernel size	Output channel	Input
Disparity layer	3 × 3	8	$[I_l(u,v),\ I_r(u,v)]$
Conv_1	3 × 3	16	*Disparity layer*
F1 = concatenate [disparity layer, Conv_1]			
Max-pool_1	2 × 2	24	F1
Conv_2	3 × 3	32	Max-pool_1
Conv_3	3 × 3	32	Conv_2
F2 = concatenate [Conv_2, Conv_3]			
Max-pool_2	2 × 2	64	Max-pool_2
Conv_4	3 × 3	64	Conv_3
Conv_5	3 × 3	64	Conv_4
Conv_6	3 × 3	64	Conv_5
F3 = concatenate [Conv_4, Conv_5, Conv_6]			
Max-pool_3	2 × 2	192	F3
Conv_7	3 × 3	128	Max-pool_3
Conv_8	3 × 3	128	Conv_7
Conv_9	3 × 3	128	Conv_8
F4 = concatenate [Conv_7, Conv_8, Conv_9]			
Max-pool_4	2 × 2	384	F4
Conv_10	3 × 3	256	Max-pool_4
Conv_11	3 × 3	256	Conv_10
Conv_12	3 × 3	256	Conv_11
F5 = concatenate [Conv_10, Conv_11, Conv_12]			
Conv_13	1 × 1	2	F1
Conv_14	1 × 1	2	F2
Conv_15	1 × 1	2	F3
Conv_16	1 × 1	2	F4
Conv_17	1 × 1	2	F5
F6 = concatenate [F1, F2, F3, F4, F5]			
GAP	–	10	F6
Fc1	10	2	GAP
2-way soft-max			

Conv → Convolution, Fc→Fully-connected layer.

3.4. Training

We train the proposed CNN for 20 epochs using Stochastic Gradient Descent (SGD) algorithm with an initial learning rate of 0.01 and a momentum of 0.9. A factor of 0.1 is used to reduce the learning rate after 10 and 15 epochs. The batch size is set to 32 following the work in Masters and Luschi (2018). For pre-processing, we normalize the input face images before feeding it to the proposed model. The total time taken by the proposed model for training on 1080 GTX TI GPU is 30 min considering the size of the proposed stereo camera-based face database.

4. Experimental results and discussions

This section details and discusses the experimental setup and results obtained using the proposed stereo face liveness detection system. As there is no public stereo face anti-spoofing database available, we evaluated the performance of the proposed model on our own stereo face anti-spoofing database using various test case scenarios.

4.1. Stereo face anti-spoofing database development

Since there is no publicly available database for studying stereo video-based face anti-spoofing, we created our own video-based stereo face anti-spoofing database with 50 subjects on four types of attacks: printed attack, cut-photo attack, mobile phone attack, and tablet attack. The database has been collected using Logitech HD Webcam C525 with three video resolutions: 320 × 240 representing low resolution, 640 × 480 representing the normal resolution and 1280 × 720 representing high-resolution images. For generating face PA, we utilized four types of PA mediums: printed photo, cut-photo, mobile screen, and tablet screen. For generating

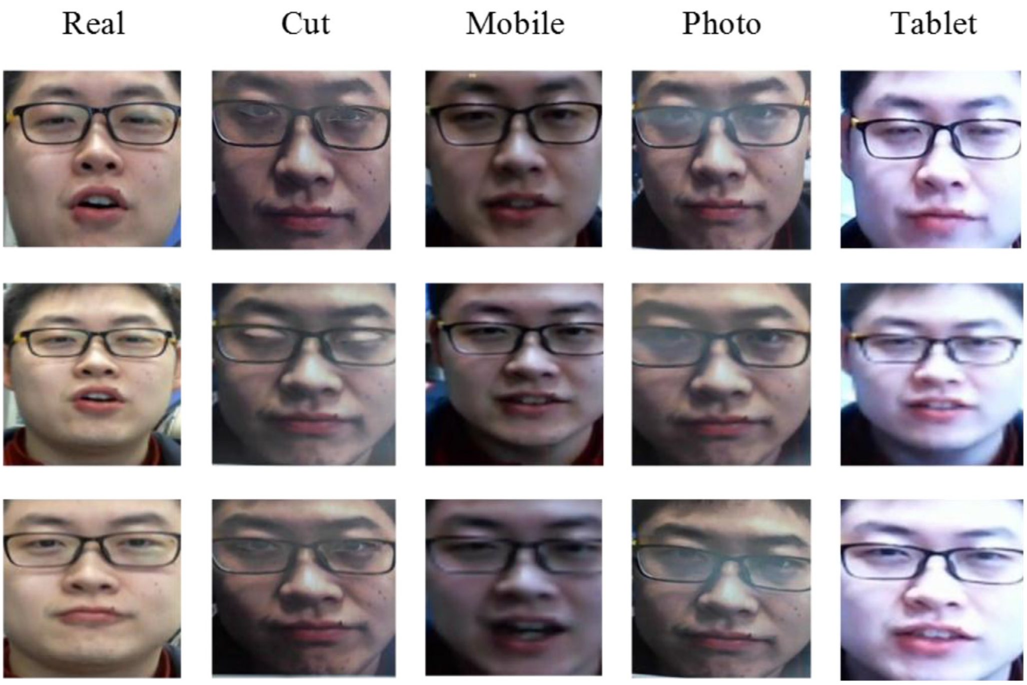


Fig. 4. Real face and face PA samples in the proposed stereo camera face anti-spoofing database: Top row: High-quality samples. Middle row: Normal-quality samples. Bottom row: Low-quality Samples.

Table 2
Distribution of training and testing set in proposed stereo face anti-spoofing database.

	Training set	Testing Set	Total
Subjects	40	10	50
PA video samples (left camera + right camera)	960	240	1200
Live video samples (left camera + right camera)	240	60	300
Total video samples (Live + Attack)	1200	300	1500

a print photo attack, we took a high definition photo of the frontal face and printed it on a fine A4 sheet of paper. For generating the cut-photo attack, the same printed photo was utilized, however with eyes section removed for imitating eye-blinking attack. The mobile PA and tablet PA represent a replay-based PA, in which the video of a person is played in front of the face liveness detection system. The mobile PA and tablet PA are more realistic and high-quality compared to the photo PA and cut-photo PA. Also, the loss of high-frequency details in mobile PA and tablet PA is lower as compared to photo PA and tablet PA. For generating mobile PA and tablet PA, we utilized google phone Pixel XL and Samsung tablet. Fig. 4 shows examples of a real face and corresponding face PA. Each video in our database is 20 s long and contain 700 frames on average. Table 2 shows the distribution of the training set and the testing set. The total number of participants in this database were 50 with various ethnicity. For training a face anti-spoofing algorithm, the training set contains 40 subjects, while the testing set contains the rest of the subjects. This type of arrangement is particularly useful for training CNN that require more data for training.

4.2. Evaluation protocol

For performance evaluation, we use the Bona fide Presentation Classification Error rate (BPCER) and Attack Presentation Classification Error Rate (APCER). The Average Classification Error Rate (ACER) is the average of BPCER and APCER (ISO/IEC JTC 1/SC 37 Biometrics, 2016). The BPCER can be defined as the proportion of Bona fide samples incorrectly classified, by a biometric face anti-spoofing system, as face PA. The APCER can be defined as the proportion of face PA samples incorrectly classified by a biometric face anti-spoofing system as genuine. In practice, the APCER is calculated individually for each face PA type, and the face PA type with maximum APCER is considered as the APCER of the whole biometric face anti-spoofing system in the worst-case scenario (Boulkenafet et al., 2018). The BPCER, APCER, and ACER can be defined as follows:

$$BPCER = \frac{1}{N_G}\sum_{l=0}^{m-1} S_l^x, \;\; x = G_g \tag{11}$$

$$APCER_{PA} = \frac{1}{N_{PA}}\sum_{l=0}^{m-1} G_l^x, \;\; x = S_g \tag{12}$$

$$ACER = \frac{BPCER + APCER}{2} \tag{13}$$

In Eq. (11), S_l^x represents that the l^{th}Bona fide image sample of a user x having groundtruth label G_g, presented to a biometric face anti-spoofing system, is incorrectly classified as face PA . N_G represents the total number of Bona fide samples in the database. Conversely in Eq. (12), G_l^x represents that the l^{th} face PA sample of a user x having groundtruth label S_g presented to a biometric face anti-spoofing system, is incorrectly classified as a Bona fide sample. N_{PA} represents the total number of PA samples.

For evaluating the proposed stereo face liveness detection system, we used various test case scenarios. We first designed three protocols, as shown in Table 3. In each protocol, we rule out one face PA type and train the proposed system on the rest of the

Table 3
Evaluation protocols on unseen face PA.

Protocols	Training PA samples	Test PA samples
Protocol 1	Photo, Cut-photo, Mobile	Photo, Cut-photo, Mobile, Tablet
Protocol 2	Photo, Cut-photo, Tablet	
Protocol 3	Photo, Mobile, Tablet	

training set. During the testing stage, we used all the real and face PA in the test set. These protocols are designed to test the performance of the face liveness detection system when an unseen face PA is presented to it. Further, we tested the performance of the proposed Dynamic Disparity-Maps by using the protocols, defined above, in various tests case scenarios, such as *Overall Performance Test, Spoof Face Detection Test, Image Scaling Test, and Blur Test.*

Overall Performance Test: For this case, we grouped the photo face PA and cut-photo face PA into a printed attack category and the mobile face PA and tablet face PA into the video attack category. The attack having highest APCER is selected to be the overall system APCER, and accordingly, the ACER is determined.

Spoof Face Detection Test: For this case, we tested the performance of the proposed method on the individual face PA. This test is done in order to investigate the vulnerability and robustness of the proposed method to different face PA. For this case, the face PA with the highest APCER determines the performance of the proposed system, i.e., the worst-case scenario.

Image Scaling Test: In the image scaling test, we trained our CNN using a fixed size image, and tested it on a different size image. Since different image acquisition devices have a variety of capturing resolutions, the image scaling test can show the performance of the proposed system under varying resolution changes in the image. For the evaluation, we utilized the same setup as used for the overall performance test.

Blur Test: In the blur test, we added Gaussian noise to the samples during testing to check the robustness of the proposed method for face liveness detection in adverse scenarios. For the evaluation, we utilized the same setup as used for the *overall performance test*.

4.3. Overall performance test

We evaluated the face liveness detection performance on *overall performance test* using disparity-maps learned between the convolutional-features by using Eqs. (6)–(9). Table 4 summarizes the performance of the proposed CNN, for each dynamic disparity-map, on the three protocols of the proposed stereo face anti-spoofing database. As can be seen in Table 4, all dynamic disparity-maps provide better performance against printed attacks. In the case of video attacks, we observed that the disparity-maps learned using SD provides better performance by obtaining the %APCER of 0.15 ± 0.08 compared to other learned disparity-maps. However, the disparity-maps learned using APD provides an overall better performance by providing a better trade-off between %APCER and %BPCER compared to AD, SD, and FM, and obtaining %ACER of 0.20 ± 0.20. We further observed that the dynamic disparity-maps learned using FM provides the lowest performance among other learned disparity-maps. Notably, the %BPCER obtained using FM is higher, 5.77 ± 4.57%, compared to AD, SD, and APD. This suggests that the dynamic disparity-maps learned using FM are more biased toward face PA as compared to AD, SD, and APD. It can be further observed in Table 4 that all disparity-maps, except SD, have detected the printed-based face PA (photo and cut-photo) with 100% accuracy as can be seen with 0.00±0.00% APCER, under the Printed column, for AD, FM, and APD in all three protocols. This shows that the proposed dynamic disparity-maps are effective against printed-based face PA.

4.4. Spoof face detection

In order to evaluate the vulnerability and robustness of the proposed to various kind of face PA, *spoof face detection* tests were performed. In this test, the objective was to evaluate the performance of the dynamic-disparity maps learned using various disparity measurements, as explained in Section 3.2. Table 5 reports the results obtained by applying the spoof face detection test using each AD, SD, FM, APD in the disparity layer of the CNN.

It can be observed in Table 5, that the proposed method achieve remarkable performance against printed attacks, i.e., photo PA and cut-photo PA. On the other hand, we found that dynamic disparity-maps computed using SD provides better performance on mobile PA and overall PA by obtaining the lowest %APCER of 0.28 ± 0.19 compared to other dynamic disparity-maps. It can be noted in Table 5 that the dynamic disparity-maps learned using AD, and APD accurately detected tablet-based face PA compared to the dynamic disparity-maps learned using SD and FM. Further observation of the results obtained in Table 5 placed the dynamic disparity-maps learned using APD in the third place after SD and AD. However, since APD provides excellent performance in terms of %APCER and %BPCER in the *Overall Performance Test*, therefore the *performance* of dynamic disparity-maps learned using APD is better compared to AD, SD, and FM.

Table 4
Performance of the proposed method on overall performance test.

Disparity-maps	Protocols	Test				
		Printed	Video	Overall		
		APCER (%)	APCER (%)	APCER (%)	BPCER (%)	ACER (%)
AD	1	0.00	0.21	0.21	0.47	0.21
	2	0.00	0.10	0.10	0.22	0.16
	3	0.00	0.25	0.25	0.50	0.38
	Overall	0.00 ± 0.00	0.19 ± 0.08	0.19 ± 0.08	0.40 ± 0.15	0.25 ± 0.12
SD	1	0.03	0.06	0.06	0.17	0.11
	2	0.00	0.21	0.21	0.33	0.27
	3	0.00	0.18	0.18	0.44	0.31
	Overall	0.01 ± 0.02	**0.15 ± 0.08**	**0.15 ± 0.08**	0.31 ± 0.14	0.23 ± 0.11
FM	1	0.00	5.39	5.39	10.64	8.01
	2	0.00	2.75	2.75	5.08	3.92
	3	0.00	0.75	0.75	1.58	1.17
	Overall	0.00 ± 0.00	2.96 ± 2.32	2.96 ± 2.32	5.77 ± 4.57	4.37 ± 3.44
APD	1	0.00	0.70	0.70	0.14	0.10
	2	0.00	0.58	0.58	0.28	0.43
	3	0.00	0.06	0.06	0.08	0.07
	Overall	**0.00 ± 0.00**	**0.45 ± 0.34**	0.45 ± 0.34	**0.17 ± 0.10**	**0.20 ± 0.20**

Table 5
Performance various dynamic disparity-maps on spoof face detection test.

Disparity-maps	Protocols	Test				
		Photo APCER (%)	Cut-photo APCER (%)	Mobile APCER (%)	Tablet APCER (%)	Overall APCER (%)
AD	1	0.00	0.00	0.42	0.00	0.42
	2	0.00	0.00	0.19	0.00	0.19
	3	0.00	0.00	0.50	0.00	0.50
	Overall	**0.00 ± 0.00**	**0.00 ± 0.00**	0.37 ± 0.16	**0.00 ± 0.00**	0.37 ± 0.16
SD	1	0.06	0.00	0.06	0.06	0.06
	2	0.00	0.00	0.42	0.00	0.42
	3	0.00	0.00	0.36	0.00	0.36
	Overall	0.02 ± 0.03	**0.00 ± 0.00**	**0.28 ± 0.19**	0.02 ± 0.03	**0.28 ± 0.19**
FM	1	0.00	0.00	0.28	10.50	10.50
	2	0.00	0.00	5.50	0.00	5.50
	3	0.00	0.00	1.42	0.08	1.42
	Overall	**0.00 ± 0.00**	**0.00 ± 0.00**	2.4±2.75	3.53 ± 6.04	5.81 ± 4.55
APD	1	0.00	0.00	0.14	0.00	0.14
	2	0.00	0.00	1.17	0.00	1.17
	3	0.00	0.00	0.11	0.00	0.11
	Overall	**0.00 ± 0.00**	**0.00 ± 0.00**	0.47 ± 0.60	**0.00 ± 0.00**	0.47 ± 0.60

Table 6
Performance of the proposed approach by up-scaling the stereo-face image by a factor of 2.

Disparity-maps	Protocols	Test				
		Printed	Video	Overall		
		APCER (%)	APCER (%)	APCER (%)	BPCER (%)	ACER (%)
AD	1	0.13	1.90	1.90	3.94	2.92
	2	2.14	1.58	2.14	3.81	2.97
	3	0.60	1.19	1.19	3.53	2.36
	Overall	0.96 ± 1.05	1.56 ± 0.36	1.74 ± 0.50	3.76 ± 0.21	2.75 ± 0.34
SD	1	0.93	0.47	0.93	2.17	1.55
	2	0.03	2.08	2.08	4.03	3.06
	3	0.35	0.93	0.93	2.47	1.70
	Overall	0.44 ± 0.46	1.16 ± 0.83	1.31 ± 0.66	2.89 ± 0.1	2.10 ± 0.83
FM	1	0.00	10.72	10.72	21.25	15.99
	2	0.00	2.81	2.81	5.72	4.26
	3	0.34	4.04	4.04	6.92	5.48
	Overall	0.11 ± 0.20	5.86 ± 4.26	5.86 ± 4.26	11.30 ± 8.64	8.58 ± 6.45
APD	1	0.07	0.76	0.76	1.69	1.22
	2	0.03	1.71	1.71	2.89	2.30
	3	0.12	0.07	0.12	0.28	0.2
	Overall	**0.07 ± 0.05**	**0.85 ± 0.82**	**0.86 ± 0.80**	**1.62 ± 1.31**	**1.24 ± 1.10**

4.5. Image scaling test

To test the performance of the proposed method in adverse conditions, we utilized an image scaling test. In this test, we upsample and down-sample the input stereo-face image by a scale factor of 2 with linear interpolation, during evaluation. Since this process introduces some distortion in the original input images, there is a likelihood that the performance of the proposed system for face liveness detection will be affected. Further, the image scaling test represents the real-world scenarios, as the face bounding-box changes with the variations in the distance between the face and the camera.

Table 6 shows the performance of the proposed method on the test set by up-sampling the input face image by a factor of 2. As can be seen in Table 6, for the up-sampling case, the dynamic disparity-maps learned using APD provide better performance by obtaining the %ACER of 1.24±1.10% compared to AD, SD, and FM. It can be further observed that the performance of the proposed method is affected by the change of resolution in the input image during the test case. In contrast to the rise in the %APCER, we observed that the %BPCER rises significantly. This can be attributed to the distortion introduces in the live samples during linear interpolation operation.

In down-sampling case, as shown in Table 7, the dynamic disparity-maps learned using APD achieved the best overall performance among all other disparity-maps learned using AD, SD, and FM. It can be observed in Table 7, that the SD provides comparatively better performance in detecting face PA by obtaining the overall %APCER of 1.86 ± 0.97% compared to the %APCER of 2.22 ± 0.67% obtained using APD. However, the APD provides significantly lower %BPCER of 12.61 ± 12.37% compared to the%BPCER of 34.34 ± 14.92% obtained using SD. Further, from Fig. 5, it can be verified that the dynamic disparity-maps learned using APD provide better performance AD, SD, and FM in down-sampling. It can also be noted that the dynamic disparity-maps computed using FM do not generalize well in image scaling tests. The face liveness detection in the down-sampling case is more challenging as compared to up-sampling case. In the down-sampling case, the image loses most of the crucial information that can be useful in determining whether the input image has a live face or face PA. As a result, it is evident that the performance of the face liveness detection system will deteriorate when a low resolution or down-sampled image is presented to it.

Table 7
Performance of the proposed approach by down-scaling the stereo-face image by a factor of 2.

Disparity-maps	Protocols	Test				
		Printed	Video	Overall		
		APCER (%)	APCER (%)	APCER (%)	BPCER (%)	ACER (%)
AD	1	0.72	1.42	1.42	40.31	20.86
	2	0.84	2.93	2.93	32.97	17.95
	3	1.49	1.96	1.96	5.97	3.97
	Overall	1.02 ± 0.41	2.10 ± 0.77	2.10 ± 0.77	26.42 ± 18.08	14.26 ± 9.03
SD	1	0.46	0.96	0.96	30.56	15.76
	2	0.15	2.89	2.89	50.78	26.83
	3	0.18	1.72	1.72	21.67	11.69
	Overall	**0.26** ± **0.17**	**1.86** ± **0.97**	**1.86** ± **0.97**	34.34 ± 14.92	18.09 ± 7.84
FM	1	4.94	8.36	8.36	27.28	17.82
	2	1.30	12.78	12.78	27.72	20.25
	3	15.71	0.19	15.71	30.19	22.95
	Overall	7.32 ± 7.50	7.11 ± 6.39	12.28 ± 3.70	28.40 ± 1.57	20.34 ± 2.57
APD	1	2.36	1.58	2.36	5.89	4.13
	2	0.07	2.81	2.81	26.89	14.85
	3	1.49	1.44	1.49	5.06	3.27
	Overall	1.31 ± 1.16	1.94 ± 0.75	2.22 ± 0.67	**12.61** ± **12.37**	**7.42** ± **6.45**

Table 8
Performance of the proposed approach by blurring the face samples using 5 × 5 Gaussian kernel.

Disparity-maps	Protocols	Test				
		Printed	Video	Overall		
		APCER (%)	APCER (%)	APCER (%)	BPCER (%)	ACER (%)
AD	1	0.50	0.82	0.82	2.33	1.58
	2	0.09	0.85	0.85	1.53	1.19
	3	0.53	0.49	0.53	1.92	1.22
	Overall	0.37 ± 0.25	0.72 ± 0.20	0.73 ± 0.18	1.93 ± 0.40	1.33 ± 0.22
SD	1	0.66	0.13	0.66	1.53	1.1
	2	0.06	1.21	1.21	1.33	1.27
	3	0.27	0.85	0.85	2.17	1.51
	Overall	0.33 ± 0.31	0.73 ± 0.55	0.91 ± 0.28	1.68 ± 0.44	1.29 ± 0.21
FM	1	0.00	14.39	14.39	28.81	21.60
	2	0.00	4.57	4.57	8.42	6.49
	3	0.00	3.03	3.03	6.39	4.71
	Overall	0.00 ± 0.00	7.33 ± 6.16	7.33 ± 6.16	14.54 ± 12.40	10.93 ± 9.28
APD	1	0.03	0.19	0.19	0.47	0.33
	2	0.02	1.08	1.08	0.17	0.63
	3	0.02	0.03	0.03	0.08	0.06
	Overall	**0.02** ± **0.01**	**0.43** ± **0.57**	**0.43** ± **0.57**	**0.24** ± **0.20**	**0.34** ± **0.28**

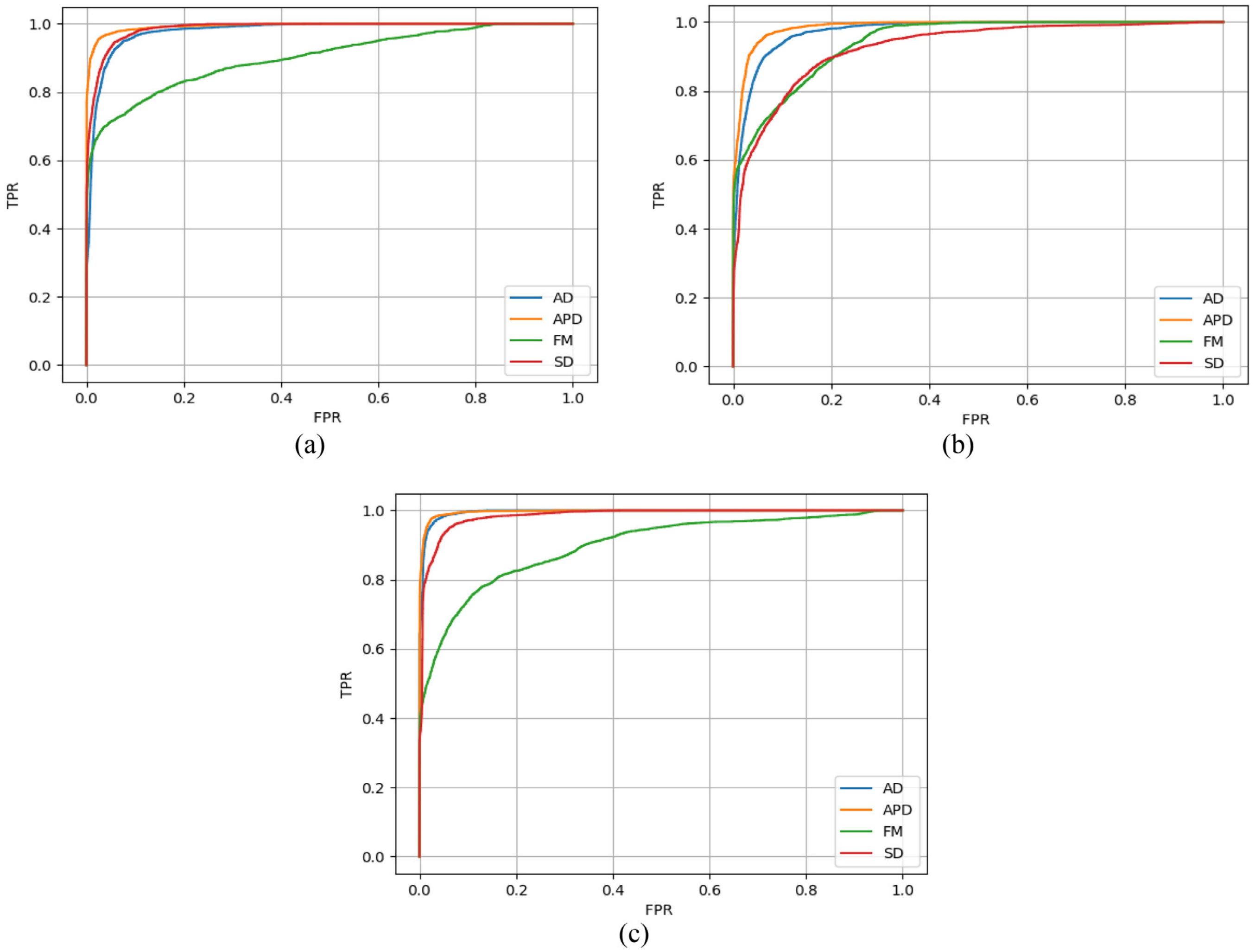
4.6. Blur test

We used the *blur test* to test the performance of the proposed method against unseen face PA when the input face samples are blurred. To obtain blurriness in the input image utilized a Gaussian kernel size of 5 × 5. Table 8 provides the performance of the dynamic disparity-maps learned using AD, SD, FM, and APD. From Table 8, it can be observed that dynamic disparity-maps learned using APD provides better performance compared to disparity-maps learned using other approaches. The APD obtained an overall lower %ACER of 0.34 ± 0.28% compared to other learned dynamic disparity-maps. Further, it can be observed in Table 9, that the learned dynamic disparity-maps are less sensitive to blurriness in the input image compared to an *image scaling test*.

It can be observed from Table 8 that the APD provide better performance on video-based face PA, by obtaining %APCER of 0.43 ± 0.57%, compared to AD, SD, and FM. On the other hand, FM shows the worst performance on video-based face PA, by obtaining % APCER of 7.33 ± 6.16%, compared to AD, SD, and APD. This further shows that the dynamic disparity-maps computed using FM shows high susceptibility to face PA under adverse scenarios. Also, when FM is combined with other dynamic disparity-maps, it may deteriorate the performance of the final classifier.

4.7. Comparison with state-of-the-art approaches

We further compare the performance of the proposed method with the work in Di Martino et al. (2018) and Sun, Huang, and Liu (2016), across all 4 tests. For a fair comparison, we reported the performance across all the three protocols in the proposed testing scenarios. Table 9 shows the overall performance of the proposed approach with other state-of-the-art approaches. It can be seen in Table 9 that the proposed approach using APD disparity maps outperform the state-of-the-art approaches in all 4 testing scenarios, i.e., *Overall Performance Test, Image Scaling Test (Upsampling, Downsampling), and Blur Test*. In Overall Test, the proposed approach obtained a %ACER of 0.20 ± 0.20%. On the Image Scaling Test, Upsampling and Down-sampling, the proposed approach obtained a %ACER of 1.24 ± 1.10% and 7.42 ± 6.45%. Whereas in Blur Test, the proposed method obtained a %ACER of 0.34±0.28% compared to the other methods. It can be noted from Table 9, that the proposed method performed significantly better in the case of video display attacks, compared to previously proposed methods. Additionally, the%BPCER of our proposed approach is significantly lower compared to the work proposed by Di Martino et al. (2018) and Sun et al. (2016). In the down-sampling case, the work proposed by Sun et al. (2016) has obtained a lower %BPCER of 9.10 ± 3.22%



(a)

(b)

(c)

Fig. 5. ROC curve for the various dynamic disparity-maps by down-sampling the stereo face image by a factor of 2 (a) Protocol 1 (b) Protocol 2 (c) Protocol 3.

Table 9
Comparison of the proposed method with conventional approaches in proposed testing scenarios.

Method	Evaluation	Test				
		Printed	Video	Overall		
		APCER (%)	APCER (%)	APCER (%)	BPCER (%)	ACER (%)
Sun et al. (2016)	Overall Test	0.00 ± 0.00	1.91 ± 2.39	1.91 ± 2.39	2.80 ± 3.26	2.35 ± 2.82
	Up-sampling	0.14 ± 0.14	4.77 ± 6.85	4.78 ± 6.84	7.28 ± 12.50	6.03 ± 9.66
	Down-sampling	0.36 ± 0.47	4.82 ± 2.86	4.82 ± 2.86	9.10 ± 3.22	6.96 ± 3.04
	Blur Test	0.21 ± 0.09	6.67 ± 3.89	6.67 ± 3.89	13.68 ± 7.55	10.17 ± 5.72
Di Martino et al. (2018)	Overall Test	0.01 ± 0.01	0.38 ± 0.33	0.38 ± 0.33	0.75 ± 0.65	0.56 ± 0.49
	Up-sampling	7.24 ± 5.01	3.84 ± 2.15	8.77 ± 2.61	21.54 ± 4.72	15.15 ± 3.64
	Down-sampling	6.00 ± 1.98	16.24 ± 2.51	16.24 ± 2.51	44.62 ± 6.07	30.44 ± 4.00
	Blur Test	1.09 ± 0.61	1.82 ± 1.00	2.14 ± 0.57	5.63 ± 0.90	3.89 ± 0.73
Proposed (APD)	Overall Test	**0.00** ± **0.00**	**0.45** ± **0.34**	**0.45** ± **0.34**	**0.17** ± **0.10**	**0.20** ± **0.20**
	Up-sampling	**0.07** ± **0.05**	**0.85** ± **0.82**	**0.86** ± **0.80**	**1.62** ± **1.31**	**1.24** ± **1.10**
	Down-sampling	**1.31** ± **1.16**	**1.94** ± **0.75**	**2.22** ± **0.67**	**12.61** ± **12.37**	**7.42** ± **6.45**
	Blur Test	**0.02** ± **0.01**	**0.43** ± **0.57**	**0.43** ± **0.57**	**0.24** ± **0.20**	**0.34** ± **0.28**

compared to our proposed approach that obtained the %BPCER of 12.61 ± 12.37%, however, the %APCER of their method is higher compared to our proposed approach. As a result, the proposed method have lower false rejection and false acceptance rate compared to previously proposed approaches.

Table 10 shows the performance of the proposed method on the *Spoof face detection test*; It can be seen in Table 10, that the proposed method have low %APCER among all presentation attacks compared to Di Martino et al. (2018) and Sun et al. (2016). Particularly, in case of Mobile and Tablet presentation attacks, the proposed method performed significantly better compared to the previously proposed approaches. In case of Mobile presentation attack, the proposed method obtained the %APCER of 0.47 ± 0.60%, and in case of Tablet presentation attack, the proposed method obtained the %APCER of 0.00 ± 0.00%, which is significantly lower than the methods proposed by Di Martino et al. (2018) and Sun et al. (2016).

Table 10
Comparison of the proposed method with conventional approaches in the spoof face detection test.

	Test				
Method	Photo APCER (%)	Cut-photo APCER (%)	Mobile APCER (%)	Tablet APCER (%)	Overall APCER (%)
Sun et al. (2016)	0.00 ± 0.00	0.00 ± 0.00	0.70 ± 61	3.11 ± 5.39	3.81 ± 4.78
Di Martino et al. (2018)	0.49 ± 0.32	0.22 ± 0.23	2.05 ± 1.10	1.05 ± 1.23	2.37 ± 1.0
Proposed (APD)	**0.00** ± **0.00**	**0.00** ± **0.00**	**0.47** ± **0.60**	**0.00** ± **0.00**	**0.47** ± **0.60**

It can be noted from the results of Tables 9, and 10 that supervising CNN with only stereo RGB images, or disparity-maps (Di Martino et al., 2018; Sun et al., 2016) limits the performance of CNN classifier for face anti-spoofing in adverse scenarios. On the other hand, in the proposed approach, we utilized the dynamic disparity-maps within the proposed CNN that shows improved performance in adverse scenarios. Further, the dynamic disparity-maps were learned as the CNN classifier is trained for face liveness detection. Therefore, the proposed dynamic disparity-maps shows a broader set of disparity features compared to fixed disparity features, as usually obtained from stereo images.

5. Conclusion and future work

In this paper, we proposed a low cost and reliable stereo camera-based face liveness detection method utilizing dynamic disparity-maps learned from convolutional-features with input stereo face images, to supervise a CNN. We further evaluated the various form of disparity-maps learned from convolutional-features of stereo face images, such as Feature Multiplication (FM), the Absolute Disparity (AD), the Square Disparity (SD) and Approximate Disparity (APD). Our experimental results with various challenging case scenarios justify the effectiveness of the proposed method in real-time face liveness detection scenarios. Further, since the depth map estimation using the stereo camera is an active research topic in computer vision, and since current state-of-the-art CNN have been proven to be effective in generating depth and disparity information for a variety of tasks, we will explore these state-of-the-art CNN for stereo-face liveness detection. Also, there is a need to further explore challenging scenarios for testing for stereo-camera based face liveness detection systems that will help to assess the generalization and feasibility of stereo camera-based face liveness detection systems in real-time. For example, the proposed work utilized horizontally aligned stereo camera for face liveness detection. However, further work is needed for the cases, in which it is difficult to align the camera horizontally. Additionally, the proposed stereo face anti-spoofing approach can be combined with other hand-crafted features for further improvement in the performance. Further, other disparity-maps could also be designed to further improve the performance of CNN classifier for face liveness detection.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Internal SRG Funding from the City University of Hong Kong substantially supported the work described in this paper under Project No. 7004430.

References

Abbas, Y., Rehman, U., Po, L., Liu, M., & Zou, Z. (2020). *Perturbing convolutional feature maps with histogram of oriented gradients for face liveness detection*. doi:10.1007/978-3-030-20005-3.

Anjos, A., Chakka, M. M., & Marcel, S. (2013). Motion-based counter-measures to photo attacks in face recognition. *IET Biometrics, 3*(3), 147–158.

Atoum, Y., Liu, Y., Jourabloo, A., & Liu, X. (2017). Face anti-spoofing using patch and depth-based CNNs. In *Proc. IEEE int. joint conf. biometrics (IJCB)* (pp. 319–328). doi:10.1109/BTAS.2017.8272713.

Boulkenafet, Z., Komulainen, J., Akhtar, Z., Benlamoudi, A., Samai, D., & Bekhouche, S. E. (2018). A competition on generalized software-based face presentation attack detection in mobile scenarios. In IEEE international joint conference on biometrics, IJCB 2017, 2018- Janua (pp. 688–696). https://doi.org/10.1109/BTAS.2017.8272758.

Boulkenafet, Zinelabidine, Komulainen, J., & Hadid, A. (2016). Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security, 11*(8), 1818–1830.

Chingovska, I., Anjos, A., & Marcel, S. (2012). On the effectiveness of local binary patterns in face anti-spoofing. In *Biometrics special interest group (BIOSIG), 2012 BIOSIG-proceedings of the international conference of the* (pp. 1–7).

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE conference on computer vision and pattern recognition (CVPR): 7* (pp. 1800–1807). https://doi.org/10.1109/CVPR.2017.195.

de Freitas Pereira, T., Komulainen, J., Anjos, A., De Martino, J. M., Hadid, A., Pietikäinen, M., et al. (2014). Face liveness detection using dynamic texture. *EURASIP Journal on Image and Video Processing, 2014*(1), 2.

De Souza, G. B., Da Silva Santos, D. F., Pires, R. G., Marana, A. N., & Papa, J. P. (2017). Deep texture features for robust face spoofing detection. *IEEE Transactions on Circuits and Systems II: Express Briefs, 64*(12), 1397–1401. https://doi.org/10.1109/TCSII.2017.2764460.

Di Martino, J.M., .Qiu, Q., Nagenalli, T., & Sapiro, G. (2018). *Liveness detection using implicit 3D features*. 1–21. Retrieved from http://arxiv.org/abs/1804.06702

DLib Face Detector. (2019). Retrieved from http://dlib.net/cnn_face_detector.py.html

Erdogmus, N., & Marcel, S. (2013). Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect. In *2013 IEEE sixth international conference on biometrics: Theory, applications and systems (BTAS)* (pp. 1–6). https://doi.org/10.1109/BTAS.2013.6712688.

Galbally, J., Marcel, S., & Fierrez, J. (2014a). Biometric antispoofing methods: A survey in face recognition. *IEEE Access : Practical Innovations, Open Solutions, 2*, 1530–1552.

Galbally, J., Marcel, S., & Fierrez, J. (2014b). Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. *IEEE Transactions on Image Processing, 23*(2), 710–724.

Gragnaniello, D., Poggi, G., Sansone, C., & Verdoliva, L. (2015). An investigation of local descriptors for biometric spoofing detection. *IEEE Transactions on Information Forensics and Security, 10*(4), 849–863.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings - 30th IEEE conference on computer vision and pattern recognition, CVPR 2017, 2017- Janua (pp. 2261–2269). https://doi.org/10.1109/CVPR.2017.243.

ISO/IEC JTC 1/SC 37 biometrics. information technology biometric presentation attack detection part 1: Framework. international organization for standardization, 2016.

Jourabloo, A., Liu, Y., & Liu, X. (2018). Face de-spoofing: Anti-spoofing via noise modeling. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), Computer vision – ECCV 2018. eccv 2018. Lecture notes in computer science, 11217 lncs (pp. 297–315). https://doi.org/10.1007/978-3-030-01261-8_18.

Kim, Y., Yoo, J.-. H., & Choi, K. (2011). A motion and similarity-based fake detection method for biometric face recognition systems. *IEEE Transactions on Consumer Electronics, 57*(2), 756–762.

Kollreider, K., Fronthaler, H., Faraj, M. I., & Bigun, J. (2007). Real-time face detection and motion analysis with application in “liveness” assessment. *IEEE Transactions on Information Forensics and Security, 2*(3), 548–558.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 60*(6), 1097–1105.

Lakshminarayana, N. N., Narayan, N., Napp, N., Setlur, S., & Govindaraju, V. (2017). A discriminative spatio-temporal mapping of face for liveness detection. In *Iden-*

tity, security and behavior analysis (ISBA), 2017 IEEE international conference on (pp. 1–7).

Li, L., Correia, P. L., & Hadid, A. (2018). Face recognition under spoofing attacks: Countermeasures and research directions. *IET Biometrics, 7*(1), 3–14. https://doi.org/10.1049/iet-bmt.2017.0089.

Li, L., Feng, X., Boulkenafet, Z., Xia, Z., Li, M., & Hadid, A. (2016). An original face anti-spoofing approach using partial convolutional neural network. In *Image processing theory tools and applications (IPTA), 2016 6th international conference on* (pp. 1–6).

Liu, Y., Jourabloo, A., & Liu, X. (2018). Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *2018 IEEE/CVF conference on computer vision and pattern recognition* (pp. 389–398). https://doi.org/10.1109/CVPR.2018.00048.

Määttä, J., Hadid, A., & Pietikäinen, M. (2011). Face spoofing detection from single images using micro-texture analysis. In *Biometrics (IJCB), 2011 international joint conference on* (pp. 1–7).

Määttä, J., Hadid, A., & Pietikäinen, M. (2012). Face spoofing detection from single images using texture and local shape analysis. *IET Biometrics, 1*(1), 3–10.

Masters, D., & Luschi, C. (2018). *Revisiting small batch training for deep neural networks* (pp. 1–18) Retrieved from.

Menotti, D., Chiachia, G., Pinto, A., Schwartz, W. R., Pedrini, H., & Falcao, A. X. (2015). Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security, 10*(4), 864–879.

Nguyen, D. T., Pham, T. D., Baek, N. R., & Park, K. R. (2018). Combining deep and handcrafted image features for presentation attack detection in face recognition systems using visible-light camera sensors. *Sensors (Switzerland), 18*(3). https://doi.org/10.3390/s18030699.

Pinto, A., Pedrini, H., Schwartz, W. R., & Rocha, A. (2015). Face spoofing detection through visual codebooks of spectral temporal cubes. *IEEE Transactions on Image Processing, 24*(12), 4726–4740.

Rehman, Y. A. U., Po, L., Liu, M., Zou, Z., Ou, W., & Zhao, Y. (2019). Face liveness detection using convolutional-features fusion of real and deep network generated face images q. *Journal of Visual Communication and Image Representation, 59*, 574–582. https://doi.org/10.1016/j.jvcir.2019.02.014.

Rehman, Y. A. U., Po, L. M., & Liu, M. (2018). LiveNet: Improving features generalization for face liveness detection using convolution neural networks. *Expert Systems with Applications, 108*, 159–169. https://doi.org/10.1016/j.eswa.2018.05.004.

Seo, J., & Chung, I.-. J. (2019). Face liveness detection using thermal face-CNN with external knowledge. *Symmetry, 11*(3), 360. https://doi.org/10.3390/sym11030360.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*. Retrieved from https://arxiv.org/pdf/1409.1556.pdf

Song, L., & Liu, C. (2018). Face liveness detection based on joint analysis of rgb and near-infrared image of faces. *Electronic Imaging, 2018*(10). 373-1-373–376 https://doi.org/10.2352/issn.2470-1173.2018.10.imawm-373 .

Song, X., Zhao, X., Fang, L., & Lin, T. (2019). Discriminative representation combinations for accurate face spoofing detection. *Pattern Recognition, 85*, 220–231. https://doi.org/10.1016/j.patcog.2018.08.019.

Sun, X., Huang, L., & Liu, C. (2016). Dual camera based feature for face spoofing detection. *Communications in Computer and Information Science, 662*, 332–344. https://doi.org/10.1007/978-981-10-3002-4_28.

Sun, X., Huang, L., & Liu, C. (2018). Multimodal face spoofing detection via RGB-D images. In Proceedings - international conference on pattern recognition, 2018-August (pp. 2221–2226). https://doi.org/10.1109/ICPR.2018.8545849.

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). https://doi.org/10.1007/978-3-642-15567-3_37.

Wang, T., Yang, J., Lei, Z., Liao, S., & Li, S. Z. (2013). Face liveness detection using 3D structure recovered from a single camera. In *2013 international conference on biometrics (ICB)* (pp. 1–6). https://doi.org/10.1109/ICB.2013.6612957.

Wang, Y., Nian, F., Li, T., Meng, Z., & Wang, K. (2017). Robust face anti-spoofing with depth information. *Journal of Visual Communication and Image Representation, 49*, 332–337. https://doi.org/10.1016/j.jvcir.2017.09.002.

Wen, D., Han, H., & Jain, A. K. (2015). Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security, 10*(4), 746–761.

Xu, Z., Li, S., & Deng, W. (2015). Learning temporal features using LSTM-CNN architecture for face anti-spoofing. In *Pattern recognition (ACPR), 2015 3rd IAPR Asian conference on* (pp. 141–145).

Yang, J., Lei, Z., Yi, D., & Li, S. Z. (2015). Person-specific face antispoofing with subject domain adaptation. *IEEE Transactions on Information Forensics and Security, 10*(4), 797–809.

Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., & Li, S. Z. (2012). A face antispoofing database with diverse attacks. In *Biometrics (ICB), 2012 5th IAPR international conference on* (pp. 26–31).