FACE DETECTION & RECOGNITION IN ORGANIC VIDEO A COMPARATIVE STUDY FOR SPORT CELEBRITIES DATABASE

K. ASHWINI[1],NENAVATH BHASKAR NAYAK[2], BATTINI VINEELA[3], MUKUNDU SUNAYANA[4], KANDULA NIKHILVARMA[5]

[1]Assistant Professor, Department of CSE, Malla Reddy College of Engineering Hyderabad, TS, India.

[2,3,4,5] UG students, Department of CSE, Malla Reddy College of Engineering Hyderabad, TS, India.

ABSTRACT:

The face image is the most accessible biometric modality which is used for highly accurate face recognition systems, while it is vulnerable to many different types of presentation attacks. Face anti-spoofing is a very critical step before feeding the face image to biometric systems. In this paper, we propose a novel two-stream CNN-based approach for face anti-spoofing, by extracting the local features and holistic depth maps from the face images. The local features facilitate CNN to discriminate the spoof patches independent of the spatial face areas. On the other hand, holistic depth map examine whether the input image has a face-like depth. Extensive experiments are conducted on the challenging databases (CASIA-FASD, MSU-USSA, and Replay Attack), with comparison to the state of the art.

Keywords: Detection, semantic segmentation, color detection, shape.

1. INTRODUCTION

Biometrics utilize physiological, such as fingerprint, face, and iris, or behavioral characteristics, such as typing rhythm and gait, to uniquely identify or authenticate an individual. As biometric systems are widely used in real-world applications including mobile phone authentication and access control, biometric spoof, or Presentation Attack (PA) are becoming a larger threat, where a spoofed biometric sample is presented to the biometric system and attempted to be authenticated. Since face is the most accessible biometric modality, there have been many different types of PAs for faces including print attack, replay attack, 3D masks, etc. As a result, conventional face recognition systems can be very vulnerable to such PAs. In order to develop a face recognition system that is invulnerable to various types of PAs, there is an increasing demand on designing a robust face anti-spoofing (or PA detection) system to classify a face sample as live or spoof before recognizing its identity. Previous approaches to tackle face anti-spoofing can be categorized in three groups. The first is the texture-based methods, which discover discriminative texture characteristics unique to various attack mediums. Due to a lack of an explicit correlation between pixel intensities and different types of attacks, extracting robust texture features is challenging. The second is the motionbased methods that aim at classifying face videos based on detecting movements of facial parts, e.g., eye blinking and lip movements. These methods are suitable for static attacks, but not dynamic attacks such as replay or mask attacks. The third is image quality and reflectance-based methods, which design features to capture the superimposed illumination and noise information to the spoof images.

OVER VIEW:

The historically influential works in anti-spoofing area contains four major approaches. One, is the texture-based methods which incorporate some hand-crafted features such as HoG, and LBP followed by traditional classifiers such as SVM to perform the task. The temporal-based methods, on the other hand, either use the facial motion patterns (e.g., eye blinking) or involve the movements between face and the background and employ methods such as the optical flow to track the movement of face in order to discriminate real faces from the fake ones. Some 3D structure-based methods have also been developed which either extract depth information from 2D images, or they analyze the 3D shape information being recorded with 3D sensors and then compare the 3D model of the input sample with that of a genuine face. This method, however requires specific 3D devices which are not easily available and should be costly. Finally, the rPPG (Remote Photoplethysmography) methods extract pulse signal from facial videos without contacting any skin. Nevertheless, all these systems are highly vulnerable against the fake face attacks and masks, and may not cope with these attacks without the auxiliary data assistance, such as depth information, IR. In recent years, the deep learning based methods have been pervasively used for many detection and recognition tasks, as well as anti-spoofing.

2. RELATED STUDY

Most of the prior face anti-spoofing work, as one of our key observations, apply SVM on hand-crafted features. While Convolutional Neural Network (CNN) exhibits its superior performance in many computer vision tasks, there are only a few CNN-based methods for face anti-spoofing. Existing CNN methods typically use CNN for learning representations, which will be further classified by SVM. In our view, further utilizing CNN in multiple ways, such as end-to-end training and learning with additional supervision, is a viable option for solving face anti-spoofing problems. On one hand, with an increasing variety of sensing environments and PAs, it is not desirable to have a hand-crafted feature to cover all attacks. On the other hand, we need CNN to learn a robust feature from the data. With the growing numbers of face spoofing databases, CNN is known to be able to leverage the larger amount of training data, and learn generalizable information to discriminate live vs. spoof samples. Following this perspective, as shown in Figure 1, this paper proposes a novel two-stream CNN-based face anti spoofing method, for print and replay attacks. The proposed method extracts the local features and holistic depth maps from face images. Here the local features are extracted from random patches within the face region, while the depth features leverage the whole face, and describe the live face as a 3D object but the spoof face as a flat plain (assuming PAs include print attack and replay attack). Since face spoofing datasets contain videos with different qualities, combining the local and holistic features has two benefits: First, utilizing the local patches help to learn spoof patterns independent of spatial face areas. Second, holistic depth maps ensure the input live sample has a face-like depth. Hence, we use two CNNs to learn local and holistic features respectively. The first CNN is end-to-end trained, and assign a score to each randomly extracted patch from a face image. We assign the face image with the average of scores. The second CNN estimates the depth map of the face image and provide the face image with a liveness score based on estimated depth map. The fusion of the scores of both CNNs lead to the final estimated class of lives vs. spoof.

3. PROPOSED SYSTEM

The problem of face anti-spoofing could be cast as a binary classification problem, which attempts to discriminate between real and fake images. However, the amount of fake samples is normally dominant vs. the real ones due to the enormous types of attacks and variations of the fake images within each type which could be given to the system. Hence, the system is likely to be exposed to the imbalanced training data. In order to gather the required data for antispoofing purposes, there are issues which hinder the clean data preparation. For instance, the background person who passes by or the portraits in the background could easily leak into the data if no pre-processing is performed. Correspondingly, these outliers have to be thrown away. The functional flow graph of the proposed system for appropriate and reliable data preparation is depicted.

Spoofing detection is actually a binary (real vs. fake face) classification problem. In deep learning era, a natural solution of this task is to feed the input RGB images to a carefully designed CNN with classification loss (softmax and cross entropy loss) for end-to-end training. This CNN-based framework has been widely investigated by [2]. Despite the strong nonlinear feature learning capacity of deep learning, the performance of anti-spoofing degrades when the input images are captured by different devices, under different lighting, etc. In this work, we aim to train a CNN which generalizes better to various environments, mainly various lightings. The RGB images are sensitive to illumination variations yet cover very detailed facial texture information. Motivated by extensive research of (single-scale and multi-scale) Retinex image, we find the Retinex (we use Multi-Scale Retinex - MSR in this work) image is invariant to illumination yet loses minor facial texture. Thus, in this work, we propose a two-stream CNN (TSCNN) which trains two separate CNNs accepting RGB images and MSR images as input respectively. To effectively fuse RGB feature and MSR feature, we propose an attention based fusion method.

ANALYSIS:

A. Benchmark Database

In this subsection, to assess the effectiveness of our proposed anti-spoofing technique, an experimental evaluation on the CASIA Face Anti-Spoofing Database, the REPLAYATTACK database and the OULU database is provided. These three datasets consist of real client accesses and different types of attacks, which are captured in different imaging qualities with different cameras. In the following paragraphs, we will have a brief introduction of the databases.

1) The CASIA Face Anti-Spoofing Database (CASIA FASD): The CASIA Face Anti-Spoofing Database is divided into the training set consisted of 20 subjects and the test set containing 30 individuals (see, Fig.3). The fake faces were made by capturing the genuine faces. Three different cameras are used in this database to collect the videos with various imaging qualities: low, normal, and high. In addition, the individuals were asked to blink and not to keep still in the videos to collect abundant frames for detection. Three types of face attacks were designed as follows:

1) Warped Photo Attack: A high resolution (1920 _ 1080) image, which is recorded by a Sony NEX-5 camera, was used to print a photo. The attacker simulates the facial motion by warps the photo in a warped photo attack.

2) Cut Photo Attack: The high resolution printed photos are then used for the cut photo attacks. In this scenario, an attacker hides behinds the photo and exhibits eye-blinking through the holes of the eye region, which was cut off before attack. In addition, the attacker put a intact photo behind the cut photo, putting the eye region overlapping from the holes and moving the intact photo up and down slightly to simulate the blinking of the eyes.

3) Video Attack: In this attack, the high resolution videos are displayed on an iPad and captured by a camera.

2) REPLAY-ATTACK Database: The REPLAY-ATTACK Database consists of video recordings of real accesses and attack attempts to 50 clients (see, Fig.4). There are 1200 videos taken by the webcam on a MacBook with the resolution 320 _ 240 under two illumination conditions: 1) controlled condition with a uniform background and light supplied by a fluorescent lamp, 2) adverse condition with non-uniform background and the day-light. For performance evaluation, the data set is divided into three subsets of training (360 videos), development (360 videos), and testing (480 videos). To generate the fake faces, a high resolution videos were taken for each person using a Canon Power Shot camera and an iPhone 3GS camera, under the same illumination conditions. Three types of attacks were designed: (1) Print Attacks: High resolution pictures were printed on A4 paper and recaptured by cameras; (2) Mobile Attacks: High resolution pictures and videos were displayed on the screen of an iPhone 3GS and recaptured by cameras; (3) High Definition Attacks: the pictures and the videos were displayed on the screen of an iPad with resolution of 1024 _ 168.

3) OULU-NPU Database: OULU-NPU face presentation attack database consists of 4950 real access and attack videos that were recorded using front facing cameras of six different mobile phones (see, Fig.). The real videos and attack materials were collected in three sessions with different illumination condition. The attack types considered in the OULU-NPU database are print and video-replay. These attacks were created using two printers (Printer 1 and 2) and two display devices (Display 1 and 2). The videos of the real accesses and attacks, corresponding to the 55 subjects, are divided into three subject disjoint subsets for training, development and testing with 20, 15 and 20 users, respectively.

CASE 1:

Input Image

RGB Face Image

GRAY Face Image

MSR Face Image

Original

5. CONCLUSION

In this paper, we proposed an attention-based two stream convolutional networks for face spoofing detection to distinguish real and fake faces. The proposed approach applies the complementary features (RGB and MSR) extracted via CNN models (MobileNet and ResNet-18) and then employs the attention based fusion method to fuse these two features. The adaptively weighted features contain more discriminative information under various lighting conditions. We evaluated our approaches of face spoofing on three challenging databases, i.e. CASIA-FASD, REPLAY-ATTACK and OULU-NPU, which indicated the competitive performance in both intra-database and inter-database. The experiments of fusion methods show that the attention model can achieve promising results on feature fusion. The cross-database evaluations show the effectiveness of the fusion of RGB and MSR information.

REFERENCES

[1] I. Chingovska, A. Anjos, and S. Marcel. 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, pages 1–7. IEEE, 2012.

[2] A. da Silva Pinto, H. Pedrini, W. Schwartz, and A. Rocha. Video-based face spoofing detection through visual rhythm analysis. In Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on, pages 221–228. IEEE, 2012.

[3] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel. Lbp-top based countermeasure against face spoofing attacks. In Asian Conference on Computer Vision, pages 121– 132. Springer, 2012.

[4] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel. Can face anti-spoofing countermeasures work in a real world scenario? In Biometrics (ICB), 2013 International Conference on, pages 1–8. IEEE, 2013.

[5] L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T. C.-H. Cheung, and K.-W. Cheung. Integration of image quality and motion cues for face anti-

spoofing: A neural network approach. Journal of Visual Communication and Image Representation, 38:451–460, 2016.

[6] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia, pages 675–678. ACM, 2014.

[7] A. Jourabloo and X. Liu. Pose-invariant 3d face alignment. In Proc. International Conference on Computer Vision, Santiago, Chile, December 2015.

[8] A. Jourabloo and X. Liu. Large-pose face alignment via cnnbased dense 3d model fitting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4188–4196, 2016.

[9] A. Jourabloo and X. Liu. Pose-invariant face alignment via cnn-based dense 3d model fitting. International Journal of Computer Vision, pages 1–17, April 2017.

[10] N. Kalchbrenner, E. Grefenstette, and P. Blunsom. A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188, 2014.