e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707028 |

Two-Stage Cascade Framework for Multimodality Face Anti-Spoofing

N.Radha, S.Krishnapriya

Assistant Professor, Department of CSE, Annapoorana Engineering College (Autonomous), Sankari Main Road, Periya Seeragapadi, Tamil Nadu, India

PG Student, Department of CSE, Annapoorana Engineering College (Autonomous), Sankari Main Road, Periya Seeragapadi, Tamil Nadu, India

ABSTRACT: Face presentation attacks (FPA), also known as face spoofing, and have brought increasing concerns to the public through various malicious applications, such as financial fraud and privacy leakage. Therefore, safeguarding face recognition systems against FPA is of utmost importance. Although existing learning-based face anti-spoofing (FAS) models can achieve outstanding detection performance, they lack generalization capability and suffer significant performance drops in unforeseen environments. Many methodologies seek to use auxiliary modality data (*e.g*., depth and infrared maps) during the presentation attack detection (PAD) to address this limitation. However, these methods can be limited since (1) they require specific sensors such as depth and infrared cameras for data capture, which are rarely available on commodity mobile devices, and (2) they cannot work properly in practical scenarios when either modality is missing or of poor quality. In this paper, we devise an accurate and robust MultiModal Mobile Face Anti-Spoofing system named M3FAS to overcome the issues above. The primary innovation of this work lies in the following aspects: (1) To achieve robust PAD, our system combines visual and auditory modalities using three commonly available sensors: camera, speaker, and microphone; (2) We design a novel two-branch neural network with three hierarchical feature aggregation modules to perform cross-modal feature fusion; (3). We propose a multi-head training strategy, allowing the model to output predictions from the vision, acoustic, and fusion heads, resulting in a more flexible PAD. Extensive experiments have demonstrated the accuracy, robustness, and flexibility of M3FAS under various challenging experimental settings

KEY WORDS: Mobile Sensing, Face Anti-Spoofing, and Multimodal Network.

I. INTRODUCTION

1.1 OVERVIEW

The face is one of the most salient and stable biometrics, it often relies on various kinds of interactive AI systems, and has been widely used in many crowd gathering and sensitive areas [1-3] such as attendance registration, security surveillance, etc. In despite of successful applications in many types of face authentication scenarios, most of existing face recognition systems are easily spoofed by presentation attacks (PAs) ranging from a 2D printing attack or a vivid 3Dmask attack [4, 5]. For example, with the help of silicone or latex masks, users easily portray another identity or obfuscate their identity for entertainment purposes. However, such masks have been treated as criminal tools to deceive automatic face recognition systems. Therefore, it is important to distinguish a real face and a fake face for face recognition and authentications systems. In general, a robust face recognition system can cope with variants of face states, such as face partial occlusion, the change of face expression, etc. On the contrary, variant face presentations should be strictly restricted on face anti-spoofing tasks, and an entire frontal face presentation is required. More importantly, an advanced face anti-spoofing model needs to show strong discriminability on intra-dataset with prior defined face knowledge, and performs well on interdataset with unknown faces.

Most popular face anti-spoofing methods extract generalized liveness feature under binary supervision. In practice, a well discriminative feature map is composed by structure clues, texture clues, depth clues, material clues, etc. There are many structure and texture distinctions between original and recaptured images. Here we present two obvious distinctions, the first is light reflection. Typically, fake face materials are much flat and smooth than real faces, easily cause specular reflection, especially under active infrared light spectral. Secondly, Moire´ Pattern based image is formed due to superimposing of the gratings and can be extracted by traditional feature descriptors such as Local Binary Pattern (LBP), Histograms of Oriented Gradients (HOG), Difference of Gaussian (DOG), Speeded Up Robust

Features (SURF), etc. In addition, face anti-spoofing models are also improved by considering the influence from image blur, distortion and noise, etc.

Under the help of depth sensors, such as time-of-flight, structure light, stereo cameras, etc., point clouds of objects can be directly constructed to prevent 2D-based fake face attacks, such as flatten screens, papers, etc. Most existing studies resort to a face detector to obtain face landmarks, and then reconstruct 3D face models based on the stereo vision. Motivated by these ideas, many auxiliary depth supervised face anti-spoofing models have been developed. Intuitively, the face-like depth can be regressed from images of real faces, whereas the none face depth is regressed from images of fake faces such as printing style, replaying style, etc.

1.2. Face antispoofing (FAS)

Face antispoofing (FAS) is attracting increasing attention from researchers because of its important role in preventing facial recognition systems from face spoofing attacks. With the advancement of various new convolutional neural network structures and the construction of various face antispoofing databases, the deep learning-based face antispoofing algorithm has become the main method in the FAS field. However, the generalization performance of the current multimodal face antispoofing algorithm is poor, and the recognition performance of the model on different datasets is quite different. Therefore, we design a multimodal face antispoofing framework based on a multifeature transformer (MFViT) and multirank fusion (MRF). First, we use a vision transformer structure, MFViT, for multimodal face antispoofing and a combination of modalities to capture the distinguishing characteristics in each modality. Second, we design a multidimensional multimodal fusion module, MRF, according to the various modal fusion characteristics obtained by the MFViT to fuse modal information in different dimensions more effectively. Evaluation results indicate that framework we designed achieves an average classification error rate (ACER) of 1.61% on the CASIA-SURF dataset and an ACER of 6.5% on the CASIA-SURF CeFA dataset.

1.3 Problem Statement

In recent years, face recognition based identity authentication systems [1], [2] are popular. However, similar to other biometric modalities [3], [4], security risks hide in the system. Many authentication systems can’t judge whether faces are captured from authorized clients or from presentation attacks.

There are various presentation attacks, for example, prints, photographs, videos displayed on screens and 3D models such as face masks [5]. Images or videos of an authorized user can be easily obtained from Internet or by portable cameras. 2D fake faces are cheap to make, but 3D masks are expensive to build and are rare in real applications. Hence in this paper, we focus on 2D presentation attacks including prints, photos and videos. As shown in Fig. 1, telling real faces is difficult even for humans. Consequently, robust presentation attack detection (PAD) methods are needed.

II. LITERATURE SURVEY

1. Data-Fusion-Based Two-Stage Cascade Framework for Multimodality Face Anti-Spoofing

Author- Weihua Liu, Xiaokang Wei

Year-2021

Existing face anti-spoofing models using deep learning for multi-modality data suffer from low generalization in the case of using variety of presentation attacks such as 2D printing and high-precision 3D face masks. One of the main reasons is that the non-linearity of multi-spectral information used to preserve the intrinsic attributes between a real and a fake face are not well extracted. To address this issue, we propose a multi-modility data based two-stage cascade framework for face anti-spoofing. The proposed framework has two advantages. Firstly, we design a two-stage cascade architecture that can selectively fuse low-level and high-level features from different modalities to improve feature representation. Secondly, we use multi-modality data to construct a distance-free spectral on RGB and infrared (IR) to augment the non-linearity of data. The presented data fusion strategy is different from popular fusion approaches, since it can strengthen discrimination ability of network models on physical attribute features than identity structure features under certain constraints. In addition, a multi-scale patch based weighted fine-tuning strategy is designed to learn each specific local face region. Experimental results show that the proposed framework achieves better performance than other state-of-the-art methods on both benchmark datasets and self-established datasets, especially on multi-material masks spoofing.

2. Face Anti-Spoofing Using Deep Learning
Author- GOPALA KRISHNAN K
Year-2022
Face recognition has been widely researched and achieved great success in a variety of applications because the human face saves the most extravagant data for perceiving people.. Face spoofing attacks continue to pose a threat to modern face recognition systems. Recent researchers had proposed their research in this area but many existing methods for face anti-spoofing have been degraded by illuminations. It inspires us to create an illumination-invariant anti-spoofing method. Our proposed method focused on a Fusion-based approach by fusing two complementary images that is RGB images and Multi-Scale Retinex (MSR) images. The RGB images contain detailed facial information and it is sensitive to illumination. The MSR image is invariant to illumination but contains less facial information. MSR images can effectively capture high-frequency information that is discriminative for face spoofing detection. Both the two complementary images are fused by the weighted pixel averaging method and it is fed to the convolutional neural network for the classification of real and spoof faces. Our proposed framework is trained and validated by the standard CASIA_SURF dataset and achieves improved results compared to the existing work.

3. Fusion Methods for Face Presentation Attack Detection
Author- FaseelaAbdullakutty, Pamela Johnston, EyadElyan
Year-2022
Face presentation attacks (PA) are a serious threat to face recognition (FR) applications. These attacks are easy to execute and difficult to detect. An attack can be carried out simply by presenting a video, photo, or mask to the camera. The literature shows that both modern, pre-trained, deep learning-based methods, and traditional hand-crafted, feature-engineered methods have been effective in detecting PAs. However, the question remains as to whether features learned in existing, deep neural networks sufficiently encompass traditional, low-level features in order to achieve optimal performance on PA detection tasks. In this paper, we present a simple feature-fusion method that integrates features extracted by using pre-trained, deep learning models with more traditional colour and texture features. Extensive experiments clearly show the benefit of enriching the feature space to improve detection rates by using three common public datasets, namely CASIA, Replay Attack, and SiW. This work opens future research to improve face presentation attack detection by exploring new characterizing features and fusion strategies.

4.Deep Learning for Face Anti-Spoofing: A Survey
Author- **Muhammad Irfan Khalid,**Jawaid Iqbal, Ahmad Alturki
Year-2021
Face anti-spoofing (FAS) has lately attracted increasing attention due to its vital role in securing face recognition systems from presentation attacks (PAs). As more and more realistic PAs with novel types spring up, traditional FAS methods based on handcrafted features become unreliable due to their limited representation capacity. With the emergence of large-scale academic datasets in the recent decade, deep learning based FAS achieves remarkable performance and dominates this area. However, existing reviews in this field mainly focus on the handcrafted features, which are outdated and uninspiring for the progress of FAS community. In this paper, to stimulate future research, we present the first comprehensive review of recent advances in deep learning based FAS. It covers several novel and insightful components: 1) besides supervision with binary label (e.g., '0' for bonafide vs. '1' for PAs), we also investigate recent methods with pixel-wise supervision (e.g., pseudo depth map); 2) in addition to traditional intra-dataset evaluation, we collect and analyze the latest methods specially designed for domain generalization and open-set FAS; and 3) besides commercial RGB camera, we summarize the deep learning applications under multi-modal (e.g., depth and infrared) or specialized (e.g., light field and flash) sensors. We conclude this survey by emphasizing current open issues and highlighting potential prospects.

5. M3FAS: An Accurate and Robust MultiModal Mobile Face Anti-Spoofing System
Author- Chenqi Kong, KexinZheng, Yibing Liu, Shiqi Wang
Year-2023
Face presentation attacks (FPA), also known as face spoofing, have brought increasing concerns to the public through various malicious applications, such as financial fraud and privacy leakage. Therefore, safeguarding face recognition systems against FPA is of utmost importance. Although existing learning-based face anti-spoofing (FAS) models can achieve outstanding detection performance, they lack generalization capability and suffer significant performance drops in unforeseen environments. Many methodologies seek to use auxiliary modality data (*e.g.*, depth and infrared maps) during the presentation attack detection (PAD) to address this limitation. However, these methods can be limited since (1) they require specific sensors such as depth and infrared cameras for data capture, which are rarely available on commodity mobile devices, and (2) they cannot work properly in practical scenarios when either modality is missing or

of poor quality. In this paper, we devise an accurate and robust MultiModal Mobile Face Anti-Spoofing system named M3FAS to overcome the issues above. The primary innovation of this work lies in the following aspects: (1) To achieve robust PAD, our system combines visual and auditory modalities using three commonly available sensors: camera, speaker, and microphone; (2) We design a novel two-branch neural network with three hierarchical feature aggregation modules to perform cross-modal feature fusion; (3). We propose a multi-head training strategy, allowing the model to output predictions from the vision, acoustic, and fusion heads, resulting in a more flexible PAD. Extensive experiments have demonstrated the accuracy, robustness, and flexibility of M3FAS under various challenging experimental settings.

III. EXISTING SYSTEM

Face Anti-Spoofing is profoundly fundamental in the two scholastics and modern fields. Unapproved individuals are attempting to get confirmed through face introduction assaults (PAs, for example, a printed face photo, showing recordings on computerized gadgets, or a 3D veil assault). Along these lines, face introduction assault recognition (facial anti- spoofing) is required, which is the errand of forestalling bogus facial check by utilizing a photograph, video, veil, or an alternate substitute for an approved individual's face. The multimodal (RGB, profundity, and IR) technique dependent on CNN is proposed in this work for antispoofing of face for validation.

DISADVANTAGES

- The technique demonstrated preferable presentation over the single model classifiers. Even though the multi-model demonstrated improved Performance, Feather-Net will present A/B network to decrease the unpredictability.
- This design utilized the combination technique in the Face Anti-mocking Attack identification and accomplished preferred outcomes over multi-model ways.

IV. PROPOSED SYSTEM

The face spoof detection is frequently used with various microscopic biometric modalities. Based on artificial intelligence, The face spoof detection techniques have various phases which include pre-processing, feature extraction and classification. In this paper, the different types of methods used in face spoof detection have been discussed. Various types of feature extraction algorithms are used which can be categorized as textural feature and colour feature extraction algorithms. The different schemes for the face spoof detection have been reviewed which are based on machine learning, deep learning and other general techniques. The schemes have been reviewed on basis of methodology and outcomes

ADVANTAGES

- Applications that require authentication can benefit from the strong and useful solution provided by biometrics.
- Nowadays, academia and industry are paying more and more attention to biometrics authentication thanks to deep learning because of its advancements in security compared to more conventional authentication techniques (such passwords, secret questions, and token codes).
- The most common biometric modalities are voice, iris, face, and fingerprints. Of them, "face" is the most widely used because it doesn't require any extra hardware resource and almost all smartphones come with a front-facing camera.

V. MODULES DESCRIPTION

Image Augmentation There are a few contrasts in the pictures obtained by various gadgets, regardless of whether a similar gadget model is utilized. As appeared in Figure 5. Human eyes can't recognize whether the face has a shape profundity. To diminish the gadget's information contrast, the focal point of genuine face pictures is scaled, as appeared.

MULTIMATERIAL MASKS SPOOFING.

To address this issue, we propose a multimodility data-based two-stage cascade framework for face anti-spoofing. The proposed framework has two advantages. First, we design a two-stage cascade architecture that can selectively fuse low-level and high-level features from different modalities to improve feature representation. Second, we use multimodality data to construct a distance-free spectral on RGB and infrared to augment the nonlinearity of data. The presented data fusion strategy is different from popular fusion approaches, since it can strengthen discrimination ability of network models on physical attribute features than identity structure features under certain constraints. In addition, a multiscale patch-based weighted fine-tuning strategy is designed to learn each specific local face region. The experimental results show that the proposed framework achieves better performance than other state-of-theart methods on both benchmark data sets and self-established data sets, especially on multimaterial masks spoofing.

OBJECT DETECTION AND RECOGNITION

A providing richer spectral information, spectral imaging is far beyond human visual perception ability in the field of object detection and recognition. The researcher's manifest that compared to RGB or monochrome cameras, the utilization of multispectral imaging can enhance spatial heterogeneity that is not easily captured by the human visual system, and thus leads to better face detection and recognition. Clearly, the spectral signature between real and fake faces provides additional spectral–spatial information that is helpful for improving face anti-spoofing.

Multispectral Images

With the demands of upgrading the security level in practical applications, the multispectral technique is used to evaluate the essential attribute of fake and real faces.

VI. CONCLUSION

In this section, we propose an end-to-end strategy to extract the fusion information from aforementioned data sources by designing a two-stage cascade network. Specifically, two convolutional networks with similar architecture are designed and a cascade framework for face anti-spoofing is presented. For the first stage of the framework, the multipreprocessed depth faces as the input of D-Net are employed to discriminate significant spoofing attacks, such as printed photos, replayed videos, etc. Before the training, depth faces are preprocessed by three ways, including depth normalization, depth face scale embedding, and normal orientation embedding. For our design, if the predication score from this model is greater than 0.5, the model output is "fake face." Otherwise, the second stage, namely, M-Net is implemented. In the M-Net stage, the fusion multimodality formation of RGB and IR, i.e., stack, summation, and difference, is fed into M-Net for further resisting more 3-D mask attacks.

VII. FUTURE WORK

In this work, we have studied the task of face anti-spoofing for preventing both 2-D and 3-D face attacks under several identification verification scenarios. For this task, we have developed a two-stage cascade framework to extract both face reflectance features and multilevel of face texture features by considering the data nonlinearity fusion strategy and network skip-connection architecture. The experimental results show that the proposed anti-spoofing framework can prevent diversity of face attacking forms, such as dim light, realistic face camouflage, static or motion pattern, etc. Furthermore, the proposed model shows strong generalization ability on PAs since it fuses features from coarse to fine network levels and utilizes the nonlinearity of multimodality information. For future works, we will establish a more pervasive face spoofing data set to analyze the generalization ability of the proposed framework. Moreover, the proposed cascade strategy can also be extended toward other tasks of biometric modality attack detection, such as print attack in iris and palm.

REFERENCES

1. T. Shen, Y. Huang, and Z. Tong, "Facebagnet: bag-of-local-features model for multi-modal face anti-spoofing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, June 2019.View at: Google Scholar
2. Z. Yu, Y. Qin, X. Li et al., "Multi-modal face anti-spoofing based on central difference networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 650-651, Seattle, WA, USA, June 2020.View at: Publisher Site | Google Scholar
3. Q. Yang, X. Zhu, J. K. Fwu et al., "PipeNet: selective modal pipeline of fusion network for multi-modal face anti-spoofing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 644-645, Seattle, WA, USA, June 2020.View at: Publisher Site | Google Scholar
4. J. Määttä, A. Hadid, and M. Pietikäinen, "face spoofing detection from single images using micro-texture analysis," in Proceedings of the 2011 International Joint Conference on Biometrics, pp. 1–7, Washington, DC, USA, October 2011.View at: Publisher Site | Google Scholar
5. J. Komulainen, A. Hadid, and M. Pietikäinen, "Context based face anti-spoofing," in Proceedings of the 2013 IEEE Sixth International Conference on Biometrics, pp. 1–8, Arlington, VA, USA, September 2013.View at: Publisher Site | Google Scholar

6. K. Patel, H. Han, and A. K. Jain, "Secure face unlock: spoof detection on smartphones," IEEE Transactions on Information Forensics and Security, vol. 11, no. 10, pp. 2268–2283, 2016.View at: Publisher Site | Google Scholar
7. Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in Proceedings of the 2015 IEEE International Conference on Image Processing, pp. 2636–2640, Quebec City, Canada, September 2015.View at: Publisher Site | Google Scholar
8. J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of fourier spectra," in Biometric Technology For Human Identification, pp. 296–303, International Society for Optics and Photonics, Bellingham, WA, USA, 2004.View at: Google Scholar
9. G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, pp. 1–8, Rio de Janeiro, Brazil, October 2007.View at: Publisher Site | Google Scholar
10. K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in "l" a," IEEE Transactions on Information Forensics and Security, vol. 2, no. 3, pp. 548–558, 2007.View at: Publisher Site | Google Scholar
11. J. Komulainen, A. Hadid, and M. Pietikäinen, "Face spoofing detection using dynamic texture," in Proceedings of the Asian Conference on Computer Vision, pp. 146–157, Daejeon, Korea, November 2012.View at: Google Scholar
12. W. Bao, H. Li, N. Li, and J. Wei, "A liveness detection method for face recognition based on optical flow field," in Proceedings of the 2009 International Conference On Image Analysis And Signal Processing, pp. 233–236, Linhai, China, 2009.View at: Publisher Site | Google Scholar

INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com