Synthesis Lectures on Computer Vision

SYNTHESIS
COLLECTION OF TECHNOLOGY

Jun Wan · Guodong Guo ·
Sergio Escalera · Hugo Jair Escalante ·
Stan Z. Li

Advances in Face Presentation Attack Detection

Second Edition

Springer

Synthesis Lectures on Computer Vision

This series publishes on topics pertaining to computer vision and pattern recognition. The scope follows the purview of premier computer science conferences, and includes the science of scene reconstruction, event detection, video tracking, object recognition, 3D pose estimation, learning, indexing, motion estimation, and image restoration. As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner. As a technological discipline, computer vision seeks to apply its theories and models for the construction of computer vision systems, such as those in self-driving cars/navigation systems, medical image analysis, and industrial robots.

Jun Wan · Guodong Guo · Sergio Escalera ·
Hugo Jair Escalante · Stan Z. Li

Advances in Face Presentation Attack Detection

Second Edition

Springer

Jun Wan
State Key Laboratory of Multimodal Artificial Intelligence Systems
Institute of Automation, Chinese Academy of Sciences
Beijing, China

Sergio Escalera
Department of Mathematics and Informatics
University of Barcelona and Computer Vision Center
Barcelona, Spain

Stan Z. Li
AI Lab
Westlake University
Hangzhou, China

Guodong Guo
Department of Computer Science and Electrical Engineering
West Virginia University
Morgantown, China

Hugo Jair Escalante
Department of Computer Science
Instituto Nacional de Astrofísica, Optica y Electrónica
Puebla, Mexico

ISSN 2153-1056 ISSN 2153-1064 (electronic)
Synthesis Lectures on Computer Vision
ISBN 978-3-031-32905-0 ISBN 978-3-031-32906-7 (eBook)
https://doi.org/10.1007/978-3-031-32906-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The field of biometric face recognition has achieved great success in the last few years, especially with the great progress in deep learning. Face recognition systems have been widely used in diverse applications, such as mobile phone unlocking, security supervision systems in railway or subway stations, and other access control systems. However, as promising as face recognition is, there also exist potential flaws that should be addressed in practice. For instance, user photos can be easily found on social networks and used to spoof face recognition systems. These face presentation attacks make authentication systems vulnerable.

Therefore, face anti-spoofing technology is important to protect sensitive data, such as user's identity, and privacy in smartphones and similar devices. In this context, we organized a series of the face anti-spoofing workshops and competitions around face presentation attack detection at CVPR 2019, CVPR 2020, and ICCV 2021. We focused on different topics on face anti-spoofing challenges in different years, such as multi-modal face presentation attack detection at CVPR 2019, cross-ethnicity face anti-spoofing recognition at CVPR 2020, and 3D high-Fidelity mask face presentation attack detection at ICCV 2021, where these topics are quite relevant and are motivated by real applications.

This book presents a comprehensive review of solutions developed by challenge participants of the face anti-spoofing challenges. The motivation behind organizing such a competition and a brief review of the state of the art are provided. The datasets associated with the challenges are introduced, and the results of the challenge are analyzed. Finally, research opportunities are outlined. This book provides, in a single source, a compilation that summarizes the state of the art in this critical subject; we foresee the book will become a reference for researchers and practitioners on face recognition.

This book would not be possible without the support of many people involved in the aforementioned events. In particular, we would like to thank all participants in the face anti-spoofing challenges, who provided us the abundant material, especially for the top three winning teams. We would like to thank Ajian Liu, Benjia Zhou, and Jun Li, who helped us prepare the book materials and proof the whole book. Also, we would like to thank Springer publishers for working with us in producing this manuscript.

Beijing, China	Jun Wan
Morgantown, USA	Guodong Guo
Barcelona, Spain	Sergio Escalera
Puebla, Mexico	Hugo Jair Escalante
Hangzhou, China	Stan Z. Li

Contents

Face Anti-spoofing Progress Driven by Academic Challenges

1

1.1 Introduction

Face anti-spoofing is essential to prevent face recognition systems from a security breach. Progress in this field has been largely motivated by the availability of face anti-spoofing benchmark datasets that resemble realistic scenarios. Despite research advances in recent years have been considerable, there are several limitations on the existing face anti-spoofing benchmarks, these include:

- The limited number of subjects ($\leq$170) and data modalities ($\leq$2) considered, which hinder further development of the academic community;
- The existence of ethnic biases, whose importance has been extensively documented in other face recognition tasks;
- The fact that most benchmarks target 2D information, while the few existing 3D mask-based benchmarks consider low-quality facial masks and very small number of samples.

Through the organization of academic challenges and associated workshops, the ChaLearn Face Anti-Spoofing challenge series has contributed with resources, evaluation protocols and dissemination forums for advancing research on facial presentation attack detection. Summarizing, our work in recent years has impacted into the above problems in turn, as follows:

- To facilitate face anti-spoofing research for the scientific community, we introduced a large-scale multi-modal dataset, namely **CASIA-SURF**. This dataset is the largest publicly available dataset for face anti-spoofing in terms of both subjects and modalities. Specifically, it consists of $1,000$ subjects with $21,000$ videos and each sample was recorded in 3 modalities (i.e., RGB, Depth and IR). Moreover, we presented a novel

J. Wan et al., *Advances in Face Presentation Attack Detection*, Synthesis Lectures on Computer Vision, https://doi.org/10.1007/978-3-031-32906-7_1

multi-modal multi-scale fusion method as a strong baseline, namely **Multi-scale SEF**, which performs feature re-weighting to select the more informative channel features while suppressing the less useful ones for each modality across different scales.

- To study the ethnic bias for face anti-spoofing, we introduced the CASIA-SURF Cross-ethnicity Face Anti-spoofing dataset, **CeFA**, covering 3 ethnicities, 3 modalities, 1, 607 subjects, and 2D plus 3D attack types. Then, we proposed a novel multi-modal fusion method as a strong baseline to alleviate the ethnic bias, namely **PSMM-Net**, which uses a partially shared fusion strategy to learn complementary information from multiple modalities.
- To bridge the gap to real-world applications, we introduced a large-scale High-Fidelity Mask dataset, namely **HiFiMask**. Specifically, a total of 54, 600 videos are recorded from 75 subjects with 225 realistic masks by 7 new kinds of sensors. Along with the dataset, we proposed a novel Contrastive Context-aware Learning framework, namely **CCL**. CCL is a new training methodology for supervised PAD tasks, which is able to learn by leveraging rich contexts accurately (e.g., subjects, mask material and lighting) among pairs of live faces and high-fidelity mask attacks.

The remainder of this chapter introduces the problem of face presentation attack detection (PAD), and describes the need for research on this topic. Also, we outline the main limitations of research practices, in particular in terms of benchmarking. Then we describe our efforts in trying to stimulate the community to target the PAD task.

1.2 Face Presentation Attack Detection

1.2.1 Formulation of the Problem

With the development of science and technology and the advent of the information age, part of the identity authentication systems based on traditional methods, such as signature and seal, are being gradually replaced by biometric recognition systems. A biometric recognition system collects human innate physiological characteristics, analyzes them and finally determines the real identity of the subject under analysis. In recent years, as an important branch of biometric recognition, face recognition has attracted extensive attention because of its unique advantages such as intuitive, natural, real-time non-intrusive, and contactless.

The research of face recognition can be traced back to the late 1960s [1–3]. Since the 1990s, with the rapid improvement of computer hardware performance, face recognition has gradually become one of the important branches in the field of computer vision and biometric recognition. There have been a series of studies represented by the characteristic face recognition algorithms [4, 5] proposed by Kirby and Turk. Since 2000, the U.S. Department of defense has organized evaluation (face recognition vendor test, FRVT) for face recognition service providers. So far, FRVT2000, FRVT2002, FRVT2006, FRVT2010, FRVT2013,