Poster: Video Chat Scam Detection Leveraging Screen Light Reflection

Hongbo Liu*, Zhihua Li†, Yucheng Xie*, Ruizhe Jiang*,

Yan Wang†, Xiaonan Guo*, Yingying Chen§

*Indiana University Purdue University Indianapolis, Indianapolis, IN 46202, USA

†Binghamton University, Binghamton, NY 13902, USA

§WINLAB, Rutgers University, North Brunswick, NJ 08902, USA

hl45@iupui.edu,zli191@binghamton.edu,yx11@iupui.edu,ruizjian@iu.edu,yanwang@binghamton.edu,
xg6@iu.edu,yingying.chen@rutgers.edu

ABSTRACT

The rapid advancement of social media and communication technology enables video chat to become an important and convenient way of daily communication. However, such convenience also makes personal video clips easily obtained and exploited by malicious users who launch scam attacks. Existing studies only deal with the attacks that use fabricated facial masks, while the liveness detection that targets the playback attacks using a virtual camera is still elusive. In this work, we develop a novel video chat liveness detection system, which can track the weak light changes reflected off the skin of a human face leveraging chromatic eigenspace differences. We design an inconspicuous challenge frame with minimal intervention to the video chat and develop a robust anomaly frame detector to verify the liveness of remote user in a video chat session. Furthermore, we propose a resilient defense strategy to defeat both naive and intelligent playback attacks leveraging spatial and temporal verification. The evaluation results show that our system can achieve accurate and robust liveness detection with the accuracy and false detection rate as high as 97.7% (94.8%) and 1% (1.6%) on smartphones (laptops), respectively.

CCS CONCEPTS

• **Security and privacy** → *Usability in security and privacy*;
• **Human-centered computing** → *Ubiquitous and mobile computing systems and tools*;

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MobiCom '19, October 21–25, 2019, Los Cabos, Mexico

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6169-9/19/10.
https://doi.org/10.1145/3300061.3343403

Figure 1: A video scam attacker uses a pre-recorded video to impersonate a user in a video chat.

KEYWORDS

Video Chat Scam, Liveness Detection, Chromatic Eigenspace

ACM Reference Format:
Hongbo Liu*, Zhihua Li†, Yucheng Xie*, Ruizhe Jiang*,, Yan Wang†, Xiaonan Guo*, Yingying Chen§. 2019. Poster: Video Chat Scam Detection Leveraging Screen Light Reflection. In *The 25th Annual International Conference on Mobile Computing and Networking (MobiCom '19), October 21–25, 2019, Los Cabos, Mexico.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3300061.3343403

1 INTRODUCTION

Due to the rapid development of social media and communication technology, recent years have witnessed video chat gradually becoming a convenient and indispensable means for people's daily communication. However, such convenience also makes personal images and videos easily obtained and exploited by malicious users to launch impersonation scam attacks as shown in Figure 1. The attacker usually obtains video footages of victims from social media or a stolen smartphone and invites the victim (i.e., victim's relative or friend) to engage in an appealingly genuine video chat with a muted voice using the stolen video footage. If the victims are convinced, the attackers will claim to run into some financial difficulties or emergencies and ask for money, which would result in irreparable economic damage for the victims. Similarly, there have been online romance scams [1] that reach out to the victims on their social media accounts (e.g., Facebook and WhatsApp) and lure the victims into performing obscene acts in a live video chat while the victims never actually chat with the attacker but a pre-recorded video of someone else. All these video scams are usually premeditated, organized crimes that steal millions, potentially billions, of dollars from people over the internet.

Intuitively, video scam attacks may be thwarted by requesting the person in chatting with to respond in accordance with some specific challenges (e.g., blinking, reading words or numbers aloud, head movements, etc.). However, the short video playback used for impersonation attacks may end before the victims are aware of its malicious intent, and the attackers also usually ignore or reject the challenges with reasonable excuses (e.g., broken microphone), which reassures the victims that this is a live video conversation. Existing methods [2] benefit from the explosive advancement of image processing and machine learning techniques, can detect media-based facial forgery or impersonation attack leveraging fabricated 2D/3D facial masks. However, if the attacker impersonates someone by playing a prerecorded video through a virtual camera, existing approaches, even human eyes, are failing to verify the liveness of people appearing in the video chat window.

Towards this end, we propose a low-cost video chat liveness detection system for various video chat terminals (e.g., smartphones and computers). In particular, our system examines the light reflected off the skin of human faces to verify the liveness of a remote user in a video chat automatically, as facial skin reflects more light from the screen than other objects in the scene of video chat due to its close distance to the screen. Our system is low-cost and easy to integrated into existing video chat terminals because it only requires a screen and a camera, which are already used in the video chat. Different from existing solutions, our system leverages the chromatic eigenspace difference to capture the minute changes of the light reflected off the human face, enabling robust video liveness detection under various practical scenarios with complex environmental light conditions, head movements, and non-stationary video background.

In a video chat, the camera of both remote and local users continuously capture the users' faces and send them to each others' screen in the form of video frames. To verify the liveness of the remote user, the local user customizes some video frames captured by the local camera with a special light pattern. The light pattern works as a *challenge* that will be displayed at the screen of the remote user and projected onto the remote user's face. The pattern of the light reflected from the user's face will be captured by the remote user's camera and sent back to the local user as a *response* along with other normal video chat frames. Thus, our system can detect the video liveness by examining the change of the light intensity without requiring active participation of the remote user.

We summarize the main contributions of this work as follows:

- We devise a non-invasive, low-cost and light-weight liveness detection system, which can be easily integrated into existing video chat applications without additional devices.

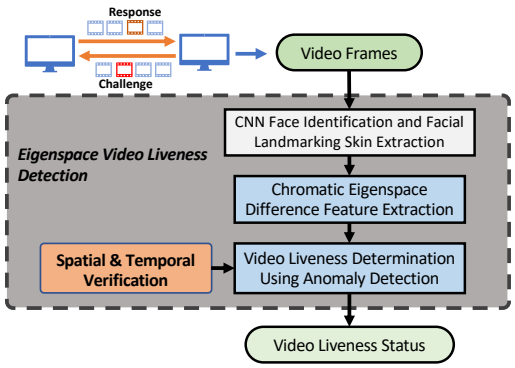


Figure 2: Overview of Video Chat Scam Detection System.

- We extensively explore the light reflected off human skin and design an inconspicuous challenge that can minimize the interference to the users' viewing experience in video chat.
- We propose resilient defense strategies that leverage the spatial and temporal verification on the light intensity changes in video chat frames to defend our system against types of attacks.
- We build a prototype video chat application with the integration of our liveness detection system and conduct extensive experiments on both laptop and smartphone platforms.

2 APPROACH OVERVIEW

2.1 Attacking Scenarios

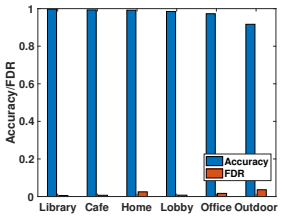
Naive Playback Attack. The naive attacker can either just play the pre-recorded video clip in front of the camera (**Naive-A1**) or stream the pre-recorded video through a virtual camera to emulate a live video chat (**Naive-A2**).

Intelligent Playback Attack. The intelligent attacker has the capability to process the video frames from the user and modify the video frames that are sent to the user. Therefore, the attacker can detect the challenges embedded in the video frames and modify the frame to emulate a valid response to the detected challenge (**Intelli-A**).

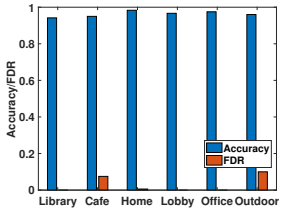
2.2 System Overview

The architecture of our system is shown in Figure 2. The system first sends the challenge to the remote chatting end, which plays the challenge on its screen and sends the video frames captured by its camera back to the system. For each receiving frame, the system first performs the *Face Identification using Convolutional Neural Network* to locate the human face in the frame by using a pre-trained convolutional neural network model. Then to further boost the detection accuracy, we employ the *Facial-landmark-based Skin Extraction* to exclude the non-skin parts on the identified face area and extract the skin-related pixels.

Next, the system performs the *Chromatic Eigenspace Difference Feature Extraction* to derive the chromatic eigenspace



(a) Smartphone



(b) Laptop

Figure 3: Performance of liveness detection under different real-life scenarios when the responder is a smartphone or a laptop.

difference feature, utilizing eigenspace distance in the RGB color space to capture the minute light intensity changes caused by the challenge, which is derived from the light intensity of skin-related pixels of two adjacent video frames. Last, we conduct *Video Liveness Determination Using Anomaly Detection* to identify valid response based on the time series of the eigenspace difference features and determine the liveness of a video chat. Specifically, we adopt Hodrick-Prescott filter [3] to remove the cyclical component and obtain a smoothed-curve representation of the time series. Next, we conduct Median Absolute Deviation (MAD) test, which is a robust measure of variability, to detect the response frames.

Furthermore, in order to defend against the attacks launched at the remote end, we also adopt two defense strategies: (1) *Spatial Verification*, which examines the spatial distribution of light intensity in both skin and non-skin area to detect Naive Playback Attack. Because the valid response should only appear in the human face and no other area during a video chat. (2) *Temporal Verification* can detect attackers that have access to the responder's system and synthesize fake response with a temporal verification scheme, it determines whether the response is legitimate or not (data processing causes time delay) based on the time delay between consecutive frames.

3 PERFORMANCE EVALUATION

We build a prototype on both laptop and smartphone platforms with Python to evaluate our system. Our experiments involve two laptops and three smartphones. We recruit 30 volunteers with different ages and skin colors, including 21 brown, 4 white, 5 dark skin individuals. We use *Accuracy* and *False Detection Rate (FDR)* to evaluate our system performance. Accuracy is defined as the ratio between the number of correctly detected responses and the total number of challenge frames. FDR is defined as the ratio between the number of incorrectly detected responses and the total number of challenge frames.

To validate the scalability of our proposed system, we carry out the experiments under six common real-life environments (i.e., library, coffee store, home, lobby, home, outdoor) and compare the results in Figure 3. For all the indoor

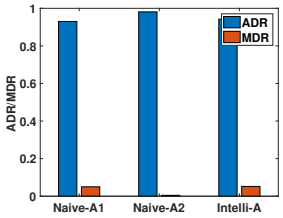


Figure 4: Performance of attack detection.

environments, our system always achieves high average detection accuracy of 94.5% on both smartphone and laptop platforms with less than 2.5% FDR. For outdoor environments, our method still maintains over 90% detection accuracy but relative high FDR of 4% and 10% on smartphones and laptops, respectively.

Then we evaluate the performance of our system's defense mechanism under the naive and intelligent playback attacks. To facilitate the evaluation, we define the *Attack Detection Rate (ADR)* as the ratio between the number of accurately detected attacks and the total number of effective attacks (i.e., the total number of challenges frames). We also define the *Miss Detection Rate (MDR)* as the ratio between the number of incorrectly detected attacks and the total number of effective attacks. We conduct the experiments with each of the three attackers (i.e., Naive-A1, Naive-A2, and Intelli-A) performing attacks on 200 challenges sent in a video chat protected by our system. Note that we use a smartphone to playback a victim's pre-recorded video in front of the camera to launch the Naive-A1. As shown in Figure 4, our system can achieve high accuracy and low miss rate on detecting Naive-A1, Naive-A2, and Intelli-A. In particular, the ADR for detecting the three attackers are 93%, 98%, and 94%, and the MDR for detecting the three attackers are 5%, below 1%, and 5%, respectively. Overall, the results confirm the effectiveness of our defense strategy levering spatial & temporal verifications.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation Grants CNS-1566455, CNS-1815908, CNS-1717356, CNS-1814590, CNS-1820624, CNS-1826647 and ARO Grant W911NF-18-1-0221.

REFERENCES

[1] Romance scams. https://www.fbi.gov/news/stories/romance-scams.

[2] Bao, W., Li, H., Li, N., and Jiang, W. A liveness detection method for face recognition based on optical flow field. In *Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on* (2009), pp. 233–236.

[3] Pedersen, T. M. The hodrick–prescott filter, the slutzky effect, and the distortionary effect of filters. *Journal of economic dynamics and control 25*, 8 (2001), 1081–1101.

[4] Wang, W., Stuijk, S., and De Haan, G. A novel algorithm for remote photoplethysmography: Spatial subspace rotation. *IEEE transactions on biomedical engineering 63*, 9 (2016), 1974–1984.