International Journal of Trend in Scientific Research and Development (IJTSRD)
International Open Access Journal
ISSN No: 2456 - 6470 | www.ijtsrd.com | Volume - 2 | Issue – 3

Video Liveness Verification

Cyrus Deboo, Shubham Kshatriya, Rajat Bhat
BE Computer, MET, Bhujbal Knowledge City,
Adgaon, Nashik, Maharashtra, India

ABSTRACT

The ubiquitous and connected nature of camera loaded mobile devices has greatly estimated the value and importance of visual information they capture. Today, sending videos from camera phones uploaded by unknown users is relevant on news networks, and banking customers expect to be able to deposit checks using mobile devices. In this paper we represent Movee, a system that addresses the fundamental question of whether the visual stream exchange by a user has been captured live on a mobile device, and has not been tampered with by an adversary. Movee leverages the mobile device motion sensors and the inherent user movements during the shooting of the video. Movee exploits the observation that the movement of the scene recorded on the video stream should be related to the movement of the device simultaneously captured by the accelerometer. the last decade e-lecturing has become more and more popular. We model the distribution of correlation of temporal noise residue in a forged video as a Gaussian mixture model (GMM). We propose a twostep scheme to estimate the model parameters. Consequently, a Bayesian classifier is used to find the optimal threshold value based on the estimated parameters.

Keywords: *Lecture videos, automatic video indexing, content-based video search, lecture video archives*

I. INTRODUCTION

Digital video has become a popular storage and exchange medium due to the rapid development in recording technology, improved video compression techniques and high-speed networks in the last few years. Therefore audio visual recordings are used more and more frequently in e-lecturing systems. A number of universities and research institutions are taking the chance to record their lectures and represent them online for students to access independent of time and location. As a result, there has been a large growth in the amount of multimedia data on the Web. Therefore, for a user it is nearly impossible to find desired videos without a finding function within a video archive. Even when the user has get related video data, it is still difficult most of the time for him to judge whether a video is useful by only flash at the title and other international metadata which are often brief and high level. Moreover, the requested information may be covered in only a few minutes, the user might thus want to find the piece of information he requires without viewing the complete video. In response to the ubiquitous and connected nature of mobile and wearable devices, industries such as utilities, insurance, banking, retail, and broadcast news have started to trust visual information gleaned from or created using mobile devices. Mobile apps utilize mobile and wearable device cameras for purposes varying from authentication to location verification, tracking, witnessing, and remote assistance. Today, one can deposit a check using a mobile phone, and videos from mobile phones uploaded by unknown users are shown on broadcast news to a national audience. The correlation measurement between the reference pattern noise image and pattern noise image is used here. The sensor pattern noise has also been used for scanner model identification and tampering detection of scanned images [8]. In [8], in addition to camera source identification, sensor pattern noise was first utilized for image forgery detection. This method proposed an accurate pattern noise extraction scheme. The above methods [6][7][8] need to pre-collect a

number of images captured from specific video cameras to extract the sensor pattern noise of the cameras. Besides, it is difficult to extract sensor pattern noise from a video without a extensive variety of video contents.

II.EXISTING SYSTEM

Information retrieval in the multimedia-based learning domain is an active and integrative research area. Video texts, spoken language, community tagging, manual annotations, video actions, or gestures of speakers can act as the source to open up the content of lectures.

A. Slide Video Segmentation

F. Chang, C [13].Video browsing can be achieve by segmenting video into representative key frames. The chosen key frames can provide a visual guideline for navigation in the lecture video portal. Moreover, video segmentation and key-frame selection is also often armed as a preprocessing for other analysis tasks such as video OCR, visual concept revelation, etc. Choosing a sufficient segmentation method is based on the definition of "video segment" and usually depends on the brand of the video. In the lecture video domain, the video sequence of an individual lecture topic or subtopic is often considered as a video segment. This can be roughly resolve by analyzing the materialistic scope of lecture slides. Many methods (as, e.g., [11]) make use of global pixel-level-differencing metrics for capturing slide transitions. A disadvantage of this kind of method is that the salt and pepper noise of video signal can affect the segmentation accuracy. After analyzing the content of lecture slides, we realize that the major content as, e.g., text lines, figures, tables, etc., can be considered as Connected Components (CCs). We therefore initiate to use CC instead of pixel as the basis element for the differencing analysis. We call it component-level-differencing metric.

III. PROPOSED SYSTEM

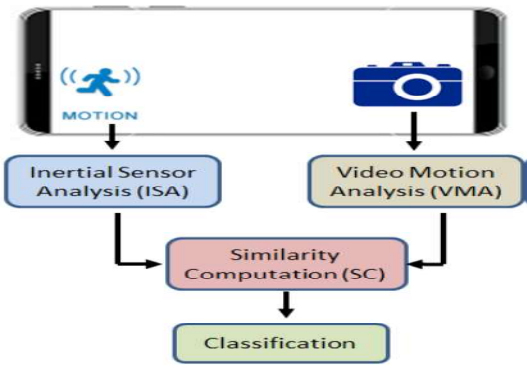

Fig. 1. Movee uses four modules to verify a video stream: the i) Video Motion Analysis (VMA), and the ii) Inertial Sensor Motion Analysis (IMA), produce movement estimations during capture, iii) Similarity Computation

We solve the fundamental question of whether the visual stream uploaded by a user has been captured live on a mobile device, and has not been tampered with by a malicious user attempting to game the system. We refer to this problem as video "liveness" verification. The practical attacks we consider are feeding a previously recorded video through man in the middle software ("Copy-Paste" attack) and pointing the camera to a projection of a video ("Projection" attack). This problem is a cornerstone in a variety of practical applications that use the mobile device camera as a trusted witness. Examples applications include citizen journalism, where people record witnessed events (e.g., public protests, natural or man-made disasters) and share their records with the community at large. Other applications add video based proofs of physical possession of products and prototypes (e.g., for sites like Kick starter [5], Amazon [1] and eBay [3]), and of deposited checks [11], [3] In this paper we represent Movee, a motion sensor based video liveness verification system. Movee edge the pervasive mobile device accelerometers and the fundamental movements of the user's hand and body during the shooting of the video. Movee exploits the intuition that video frames and accelerometer data captured simultaneously will bear certain relations. Specifically, the movement of the scene recorded in the video stream should be related to the movement of the device registered by the accelerometer. We conjecture that such relations are hard to fabricate and emulate. If the data from the accelerometer corroborates the data from the camera, Movee proved that the video stream was genuine, and has been taken by the user pointing the camera to a real scene. 2 In essence, Movee provides CAPTCHA like verifications, by adding the user, through her mobile device, into the visual verification process. However, instead of using the cognitive strength of humans to read visual information, we rely on their innately flawed ability to hold a camera still. Movee can also belook as a visual public notary that stamps an untrusted video stream, with data simultaneously captured from a trusted sensor. This data can later be used to verify the liveness of the video. Previous work has proposed to use audio streams in captured videos to protect against spoofing attacks in biometric authentication. The current verifications use static and

dynamic relations between the recorded voice and the motion of the user's face. In this paper we use the previously unexplored combination of video and accelerometer data to solve a different problem: verify the liveness of the video capture process. Movee consists of four modules, explained in Figure 1. The Video Motion Analysis (VMA) module processes the video stream captured by the camera. It uses video processing techniques to infer the motion of the camera, generating a time-dependent motion vector. VMA is inspired by the process used in image stabilization capable cameras.

III THE MODEL: SYSTEM AND ADVERSARY

We now describe the system and adversary models that we assume in this work.

A. *System Model*

We consider a system that consists of a service provider, e.g. video sharing services such as Vine YouTube or check deposit services The provider offers an interface for subscribers to upload or stream videos they shot on their mobile devices. We assume subscribers own mobile devices adapted with a camera and inertial sensors (i.e., accelerometers). Devices have Internet connectivity, which, for the purpose of this work may be infrequent. Donor need to install an application on their mobile devices, which we henceforth denote as the "client". A subscriber needs to use this client to capture videos. In addition to video, the client simultaneously captures the inertial sensor (accelerometer) stream from the device. The client uploads both the video and the accelerometer streams to the provider. The provider verifies the authenticity of the video by checking the consistency of the two streams. The verification is performed using restricted information: the two streams are from independent sources, but have been captured at the same time on the same device. We assume a system where the problems of constructing trust in the mobile device, operating system and identify drivers, and the mobile client are already addressed. This added for instance a system where a a chain of trust has been developed The chain of trust ensures that the operating system, adding the camera and sensor device drivers, and the installed apps, are trusted and have not been tampered with by an attacker, see e.g., [4], [6]. A discussion of limitations is included in Section VII. In the remainder of the paper we use the terms accelerometer and inertial sensor interchangeably.

B. *Adversary Model*

We assume that the service provider is honest. Users however can be malicious. An adversarial user can tamper with or copy video streams and inertial sensor data. The goal is to fraudulently claim ownership of videos they upload to the provider. Let V be such a video. The adversary can use a trusted device to launch the following attacks, that generate fraudulent videos or fraudulent video and acceleration data:

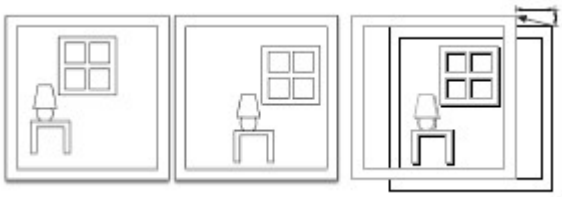

Fig. 2. The Video Motion Analysis module processes each consecutive video frame and finds the motion vector by computing the amount of displacement that common image components have shifted between two frames.

Algorithm Video Segmentation

Consider the frames which are converted from the video by using this code

Let F(k) be the kth frames in the given video, where k will takes the values from k=1,2,3,….n, the shot boundary detection algorithm for the above frames can be explained as follows

Step 1: Split the given frames into block with m rows & n Columns B(i,j,k) stands for the block at (i ,j) in the given frame.
Step 2: Computing the histogram matching difference between the neighboring blocks in consecutive frames for a video sequence H(i ,j,k) and H(i ,j,k+1) stands for the histogram of blocks at (i,j) in the kth and (k+1) th frames respectively the block difference is measured by using the flowing equation Where DB = block difference
Step 3: Evaluate the histogram difference between the two consecutive frames by Where Wij is the weight of the block at (i ,j)
Step 4: calculating the threshold by the use of mean and standard variance of histogram which are differ over the whole video sequence and is different for different kind of information extracted. Mean and standard variance can be calculating by using the following equations.
Step 5: Calculating the total number of frames
Copy-Paste attack. Copy V and output it.

Projection attack. Point the camera of the device over a projection of the target video. Output the result.

Random movement attack. Move the device in a random direction, and capture the resulting acceleration data. Output the video V and the captured acceleration stream.

Direction sync attack. Use the video to infer the dominant motion direction of V . Use the device to capture an acceleration sample that encodes the same motion direction. Output V and the acceleration sample.

Cluster attack. Capture a dataset of videos and associated sensor streams. Use a clustering algorithm (e.g., K-means [12]) to group the videos based on their movement (i.e., the values of video features extracted by the VMA module, see Section III-A and Section III-D). Assign V to the cluster containing videos whose movement is closest to V . Randomly choose one of the videos and associated sensor streams in the cluster, (V ′,A′). Output (V , A′).

Replay attack. Study the target video V . Then, holding a mobile device that captures acceleration data, emulate the movements observed in V . Let A′ be the acceleration data captured by the device during this process. Output (V,A′).

CONCLUSIONS

In this paper we have introduced the concept of "liveness" analysis, of verifying that a video has been shot live on a mobile device. We have proposed Movee, a system that relies on the accelerometer sensors ubiquitously deployed on most recent mobile devices to verify the liveness of a simultaneously captured video stream. We have implemented Movee, and, through extensive experiments, we have shown that (i) it is efficient in differentiating fraudulent and genuine videos and (ii) imposes reasonable overheads on the server. In future work we intend to integrate more sensors (e.g., gyroscope), as well as the use of MonoSLAM [12] as an alternative VMA implementation to improve accuracy.

REFERENCES

1) E. Leeuwis, M. Federico, and M. Cettolo, "Language modelling and transcription of the ted corpus lectures," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., 2003, pp. 232–235.
2) D. Lee and G. G. Lee, "A korean spoken document retrieval system for lecture search," in Proc. ACM Special Interest Group Inf. Retrieval Searching Spontaneous Conversational Speech Workshop, 2008.
3) J. Glass, T. J. Hazen, L. Hetherington, and C. Wang, "Analysis and processing of lecture audio data: Preliminary investigations," in Proc. HLT-NAACL Workshop Interdisciplinary Approaches Speech Indexing Retrieval, 2004, pp. 9–12.
4) A. Haubold and J. R. Kender, "Augmented segmentation and visualization for presentation videos," in Proc. 13th Annu. ACM Int. Conf. Multimedia, 2005, pp. 51–60.
5) W. H€urst, T. Kreuzer, and M. Wiesenh€utter, "A qualitative study towards using large vocabulary automatic speech realization to index recorded presentations for search and access over the web," in Proc. IADIS Int. Conf. WWW/Internet, 2002, pp. 135–143.
6) C. Munteanu, G. Penn, R. Baecker, and Y. C. Zhang, "Automatic speech realization for webcasts: How good is good enough and what to do when it isn't," in Proc. 8th Int. Conf. Multimodal Interfaces, 2006.
7) G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Inf. Process. Manage., vol. 24, no. 5, pp. 513– 523, 1988.
8) G. Salton, A. Wong, and C. S. Yang. (Nov. 1975). A vector space model for automatic indexing, Commun. ACM, 18(11), pp. 613–620, [Online]. Available: http://doi.acm.org/10.1145/361219.361220
9) T.-C. Pong, F. Wang, and C.-W. Ngo, "Structuring low-quality videotaped lectures for cross-reference browsing by video text analysis," J. Pattern Recog., vol. 41, no. 10, pp. 3257–3269, 2008..
10) M. Grcar, D. Mladenic, and P. Kese, "Semi-automatic categorization of videos on videolectures.net," in Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases, 2009, pp. 730–733.
11) J. Adcock, M. Cooper, L. Denoue, and H. Pirsiavash, "Talkminer: A lecture webcast search engine," in Proc. ACM Int. Conf. Multimedia, 2010, pp. 241–250.
12) A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(6):1052–1067, June 2007.
13) F. Chang, C.-J. Chen, and C.-J. Lu, "A linear-time component labeling algorithm using contour tracing technique," Comput. Vis. Image Understanding, vol. 93, no. 2, pp. 206–220, Jan. 2004