An Authentication Approach for Lossy Surveillance Videos

HimaBindu U\textsuperscript{1} Prakasha S\textsuperscript{2}
\textsuperscript{1}PG Student \textsuperscript{2}Assistant Professor
\textsuperscript{1,2}Department of Information Science & Engineering
\textsuperscript{1,2}RNSIT, Bangalore, India

Abstract— Nowadays, there is increased use of surveillance cameras for security purpose in public or in other organizations (banks, airports, traffic etc.). The videos that are captured from these cameras need to be secured, as the video content may contain confidential information or any other activities that need to be analyzed. For analyzing the video content, videos are need to be transmitted across the internet for different authorities. But, due to wireless transmissions there would be packet loss or any change in the video content. Different authentication methods are used to authenticate videos, but these techniques sacrifice video distinctiveness. Hence we describe a novel approach for securing these videos, which is called video fingerprint generation and matching technique. Our proposed approach ensures trust for videos that are transmitted via internet. Experimental results have shown concise and accurate video fingerprint generation.

Key words: Authentication, digital signature, video surveillance, video analysis.

I. INTRODUCTION

Due to the increase use of surveillance cameras, different generation of cameras is being used for security purposes. Previously the camera data was transmitted over dedicated channels. As the usage of internet has made the things simpler, data is being transmitted via internet. Since this camera are being used by multiple authorities for video analysis or some other purposes, there is a lot of security risks involved while transmitting the data across various authorities (e.g., police, fire, traffic). One risk involved while transmitting the data is that, the data which is sent from the source may not be the same as that of the received one at the destination, due to the lack of security for this data. Therefore authentication mechanism is very much essential to ensure trust in the video content.

The method that can eliminate the above mentioned threat would be data stream encryption. The authorized camera, in possession of a secret key, would encrypt the data, and the authorized receiver, in possession of the same secret key, would decrypt. Integrity of the data would be protected by secrecy of the key. The data should be made available to various authorities. As the public cameras are increasingly being used made open to the general public for reasons like providing access to information on traffic congestion, viewing public entertainment venues, crowd-sourcing the security task, etc. Since it would be a security vulnerability to distribute decryption keys across agency bounds, and since this is impractical for the general public, encrypting camera streams is not an option.

Digital signature authentication is a cryptographic alternative that assures that each and every received byte is the same as from the authorized capture device. An advantage of digital signature authentication over symmetric (single key) encryption is that it obviates the challenge of key distribution.

Digital signature encryption and decryption use two separate keys. The private key is owned and safeguarded by the signing entity, in this case the camera. A public key can be widely distributed, since it is not secret to the general public (if they wish to authenticate) as they can all share the same public key. However, there is a problem using straightforward digital signature authentication for the video transmission protocols often used in surveillance (RTP/UDP broadcast), and especially over wireless channels. Bytes can change due to packet loss, transcoding, and transrating (scaling), so the data at capture and receiver ends may not be byte-for-byte equivalent. The goal is to describe a method to authenticate camera data transmitted over lossy channels.

Authentication of inexact signals is not a new problem. The term, “robust hashing”, is used for determining if a multimedia or biometric query signal is perceptually equivalent to the original, [2]. This combines two concepts. One is of a hash function, whose result is a highly distinctive and concise digest of data. The one-way hash is used to reduce a data file of arbitrary length to a much smaller fixed length, or digest, which in turn can be used for quickly comparing different data files. And “robust” relaxes the exactness requirement of matching hashes by allowing matches of “perceptually close”, but inexact, data. A common way to produce a robust hash is to extract invariant features from a signal that tend to maintain their value despite incidental—but not malicious—distortion.

Surveillance video often undergoes video processing close to its capture source to determine higher level features such as motion or deviation from normal activity. This information in-turn can reduce bandwidth need, which is an important savings for networks of thousands of camera as these features can help reduce the need to send full video. The authentication technique used here is that, generation of video fingerprint and matching of that fingerprint, which is short but accurate time series graph obtained by extracting the frames from the captured video content along with the digital signature technique.

II. RELATED WORK

Media authentication methods can be categorized as: stream or packet-based [1], [2], and content-based [3], [4]. Stream based authentication is to directly authenticate at the stream or packet level. The system security can be mathematically proven, as it is based on conventional data security approaches. In Stream based method each block is digitally signed. As long as the number of lost packets is less than a threshold, all received packets can be authenticated. However, this scheme has high computational overhead. It also suffers from a high receiver delay because the receiver
has to wait for a minimum number of the received packets for authentication. These methods can offer quantifiably strong security, however they are usually designed for a specific communications protocol or compression scheme, they are not generally robust to video changes other than the changes designed for, and they do not account for transcoding, where authentication means must be transferred across different coding schemes.

In contrast to stream-based methods which authenticate at the data level, content-based methods authenticate at the level of features, and are designed to do so with qualities of persistence and perceptual invariance. These are calculated by frame, and then assembled into sequences to provide authentication.

In the content-based category are digital watermarks. A watermark is embedded into a video, and when extracted from the received video can indicate distortions, e.g. A trade-off inherent to watermarking is the boundary between robustness to media channel distortions, and detection of malicious distortions. Some authentication methods use watermarks for transporting a digital signature. Some content-based methods require the authentication to be known by the receiver to authenticate. This is not appropriate for our application where several agencies plus the general public would have to have these keys to authenticate.

III. METHOD

System security plays a vital role in an authentication system. Three modules mainly affect system security: feature extraction, Elliptical Curve cryptography (ECC), and hashing. Therefore, the security performance of the system may be measured in terms of the probability of the system being cracked, i.e., given an video, the probability of finding another video that can pass the signature verification, under the same parameters.

Fig. 1: System Security Illustration in Content Based Scheme

A. Video Fingerprint Generation:

Our method falls under the content-based authentication category for which content is reduced to a video fingerprint. Since motion information is important in surveillance video, we extract motion feature from video and use this as a more concise fingerprint. There is much research on motion feature representation.

Existing systems can achieve speed through the use of global signatures (e.g., color histograms [5], ordinal signature [6]) but these methods sacrifice video distinctiveness. Our proposed method seeks to find a middle ground: tracking local features, called keypoints, allows us to retain more robustness while aggregating across time; and matching as a global signature allows us to achieve speed efficiency.

The video fingerprint we generate for each video seeks to capture trajectories of motion of the most salient features of the video across time. In Fig. 2, fingerprint generation method accepts an input video V represented as a sequence of sampled frames V=\{v_0,v_1,v_2,…v_{s-1}\}. The frames are sampled randomly according to sampling rules agreed between generator and receiver of the fingerprint as shown in fig 3.

Fig. 2: Video Fingerprint Generation Technique

We generate a histogram of orientations of optical flow for each consecutive pair of frames. Salient features are first detected from sampled frame using a local feature detector, FAST [7] as shown in fig 4.

Fig. 3: samples of frames extracted from video content

We then calculate the optical flow of these features from one frame to the next by applying the Lucas–Kanade method [8]. This is done by extracting features similar to histograms of orientations of optical flow (HOOF) [9]. We only retain trajectories with magnitudes within a realistic range (e.g., 3 to 50 pixels for 5 frames/sec sampling rate). Orientation of trajectories is less sensitive to lossy video noise than other metrics such as magnitude of trajectories, which can be changed due to rescaling of original video. We distribute the orientations of trajectories of local features motion into a histogram of angular range bins. Each bin records the number of keypoints that have moved in a given orientation for each pair of consecutive sampled frames. In our implementation, we use 8 bins of 45 degree range each as shown in fig 5.

Fig. 4: Local feature extraction using FAST

Fig. 5: Distribution of orientations of trajectories of local features motion into a histogram of angular range bins.
B. Fingerprint Matching:

Our authentication scheme does not use a robust hash, but instead a robust method for hash-matching [10]. Each frame fingerprint is digitally signed. This hash result cannot be used alone for authentication if there are video distortions, however in addition to the digital signature, we include the video fingerprint in the clear (i.e., unencrypted), the combination of which enables authentication of distorted signals. To authenticate, as shown in Fig. 6, the receiver does three operations: 1) The video fingerprint is hashed $H_1$ (using a public seed), to obtain $H'_1$; 2) The digital signature is decrypted using the public key and the resulting hash, $H'_2$, is compared with $H_3$; 3) The video fingerprint is calculated of the received video, $F_1$, and compared with the received video fingerprint $F_2$. Then the frame is said to be authenticated if they match.

IV. EXPERIMENTAL RESULTS

A. Robustness of Video Fingerprint:

The video distance metric was computed for the lossy video sequences. Fig. 7. The plot of SSIM (Structural Similarity Measure) versus the distance metric suggests a correlation between the two. It should be noted that even though the video is of poor quality at higher error rates and correspondingly lower SSIM, the video is still authentic.

In the case of video transcoding, Fig. 8 suggests that the distance metric is affected more by resolution than by bandwidth changes. Again, the video is authentic in all of these cases.

B. Accuracy of Video Fingerprint:

To understand the accuracy of the proposed video authentication approach, we test on a publicly available video dataset TRECVID 2009 for content-based copy detection [3], which provides 201 untransformed query videos. The queries last from 5 seconds to 2 minutes long.

V. CONCLUSION

Authentication of lossy video is a challenge that will become increasingly important as agencies and the public share video feeds. We have proposed a method to authenticate video whose authentication rate is high for
videos and noise tested. The methods combine a video fingerprint with a cryptographic digital signature. The video fingerprint is sent separately from the video signal, incurring an increase in transmission complexity, and there is a cost of 5–8 seconds latency before authentication results are available.

REFERENCES


