Journal of Information Security and Applications 41 (2018) 159–169

Contents lists available at ScienceDirect

Journal of Information Security and Applications

journal homepage: www.elsevier.com/locate/jisa

Video authentication using spatio temporal relationship for tampering detection

Sowmya K.N.[a,*], H.R. Chennamma[b], Lalitha Rangarajan[c]

[a] Department of ISE, JSS Academy of Technical Education, Bangalore, Karnataka, India
[b] Department of MCA, Sri Jayachamarajendra College of Engineering, JSS Science & Technology University, Mysuru, Karnataka, India
[c] Department of Studies in Computer Science, University of Mysore, Mysuru, Karnataka, India

ARTICLE INFO

Article history:

Keywords:
Video tampering detection
Video authentication
Digital signature
Video forensics
Video fingerprint
Video hashing

ABSTRACT

This paper discusses a novel approach to detect inter frame and intra frame video forgery using content based signature. A novel technique called the "Spatio Temporal Triad Feature Relationship" (STTFR) is employed to generate a unique content based signature – value for any given video sequence. The proposed STTFR algorithm aims to verify video integrity through the creation of a 128 bit message digest from the input video of variable length that will be unique to that video and acts as a fingerprint. Change in the video sequence, either at the spatial or at the temporal level will result in a different fingerprint than the one obtained originally. The knowledge of the signature will not enable any person/entity to recreate the original video as the signature is generated by combining spatial and temporal fingerprints in an orderly and systematic approach. We have verified our technique with standard datasets and found accurate results.

1. Introduction

With the advancement in technology, new and innovative ways are being used by the new-age criminals. The increase in crime rate requires the law enforcers to be a step-ahead with the electronic surveillance devices. The data captured by such devices can become the primary source of evidence in any legal matters with devices like CCTV becoming ubiquitous for surveillance, security and integrity. The obtained digital data poses a new challenge as the video and data analytics are performed on the surveillance footages. Video Forensics is the scientific study of video sequences in legal matters, to help establish the authenticity and genuineness. The primary goal of video forensics is to preserve the digital evidence in its original form and establishing the integrity of the video as recorded by the sources like CCTV DVR, digital cameras, smartphone camera, etc. The visual data obtained through these devices is lossy in nature due to compression and are vulnerable for tampering through powerful editing tools. The forensic techniques adopted to prove video authenticity needs to be reliable and fool-proof, to be admissible by the legal bodies. The proposed approach is one step towards it.

2. Related works

2.1. Review

Intentional tampering of the digital evidences at spatial, temporal or spatio-temporal mode, are commonly done to mislead investigations. This could lead to wrong judgements, interpretations and chaos in the investigation process. In order to detect traces of such tampering in videos, multiple approaches are adopted. Video tampering detection approaches can be predominantly classified into passive and active [1].

An untampered video remains consistent after some post processing operations while the natural characteristics of the video are exploited when there is a premeditated tampering and it is observed keenly to detect traces of tampering in passive forensics. A forged video containing tampered region artefact disturbs the statistical patterns which need to be efficiently identified using sophisticated algorithms. Detection of these anomalies must be carried out in a scientific way. Spatial tampering is identified by considering modifications at pixel level like splicing, copy move forgery, object insertion, deletion [2], object modification [3] format based operations like compression or blocking [4–6], camera based tampering like demosaicing, chromatic aberrations and sensor noise [7–9]. Physical and geometrical based tampering detection properties such as light direction, environment, and principal point are also considered [10,11]. Temporal tampering like frame

* Corresponding author.
E-mail addresses: kn_sowmya@rediffmail.com (S. K.N.), hrchennamma@sjce.ac.in (H.R. Chennamma), lali85arun@yahoo.co.in (L. Rangarajan).

https://doi.org/10.1016/j.jisa.2018.07.002
2214-2126/

insertion, deletion, shuffling are detected through various statistical disturbance and codec's, GOP structures [12–15] and frame rate up conversions FRUC [16,17]. Combination of spatial and temporal approach needs to be considered for detecting spatio temporal tampering. Unfortunately there is no unified standard approach available to identify all types of tampering efficiently in a passive loom. Passive forensic techniques are mainly based on detection of the inherent pattern, pattern removal, and pattern insertion, often these can be addressed through pattern recognition techniques [18].

In active forensics, watermark or digital signature will be embedded in the video and the extraction of the same helps in verifying its integrity. Diverse watermark algorithms exist and the degree of their retrieval helps to affirm the originality of the video content. Video watermarks can be classified based on factors such as their characteristics and embedding domain [19]. Fragile watermark techniques [20–23] are sensitive to modifications and semi fragile watermarks [21,24,25] are less sensitive to content preserving classical modifications such as compression, changing brightness, saving file in different file formats etc. However these watermarks are sensitive to content alterations. Multiple approaches based on spatial and frequency domains like DCT, DWT, DFT [26–32] and compressed domain watermarking [33–35] belong to fragile or semi fragile nature. Often, watermark or signature is generated and embedded at the time of video creation. However, occasionally, in absence of sophisticated devices watermarking can happen at a later stage. Retrieval of the same watermark or digital signature helps in establishing credibility and authenticity of video.

Content based signature algorithms are primarily used for integrity verification and also used for identification of video, stored in databases. The video hashing scheme helps to identify the intentional, malevolent content altering attacks, as the hash computation of the forged/tampered video will yield a different hash when compared to the original authentic video. Few of the video hashing schemes are discussed here [36] have developed a video hashing based on radial projections of the image pixels passing through the centre characterized by angular orientations. Variance of such pixel luminance is found and normalized. 40 low frequency DCT coefficients from the feature vector are chosen for RASH and a 320-bit image digest is obtained by quantizing each coefficient on 8 bit. Noise and frame dropping affect the key frame and hence the hash value too [37] have developed hash functions based on Fourier transforms which are invariant to 2D affine transformations and converted to polar coordinates. Normalized key dependent pseudo random numbers are computed using either weighted sum or secret key of polar coordinates. They undergo post processing operations and the hash values thus generated is permuted based on permutation table with the help of a key. Authors have addressed both security and robustness in their work and their approach can detect forgery attack such as copy paste, mild filtering, compression and rotations [38] have constructed short fingerprint based on signal processing operations. They have considered the lowest 64 3D-DCT coefficient frequency transformation values as bases for hash function. In order to address security needs, a random parameter is introduced in 3D-RBT approach where the 3D-DCT coefficients are selected randomly. The problem with selection of few low frequency coefficients is that the significant multiple low frequency coefficients which are of high importance is lost. If all the 3D-DCT coefficients selected are attacked then 3D-RBT approach fails to meet its objective and its search space is limited too [39] have considered centroid of gradients which are based on differences of pixels having least confrontation for content based fingerprint generation. They have used their approach for identifying video in a database search. A threshold based on squared Euclidean distance is used to find the fingerprint match in the database. It is sensitive to geometrical transformations like rotation and cropping resulting in performance degradation [40] have used object driven visual attention regions for generating video fingerprints which are invariant to content preserving distortions. Region of interest represented by salient maps based on colour, intensity and orientation are normalized before reducing their dimension to a preferred stage and transformed to a binary vector which acts as a fingerprint. Their approach helps to effectively identify video in the database although they are sensitive to orientations of overlays within frames [41] divide the video into separate shots and perform signal processing operation by applying 2D-DCT on each frame and generate the hash value based on those values which occur most frequently within the shots [42] have developed a fingerprint algorithm called 3D-LBT. It uses AdaBoost based machine learning approach for feature selection of 64 3D-DCT coefficient values. Learned 3D-DCT coefficients vary depending on the video category ensuring security. Weighted distance concept adopted, helps improve the performance [43] have considered combination of histogram based approach along with DCT for video hashing. Initially, edge oriented histogram is found for all frames and sampled to bins. After finding the EOH descriptor vector, 1D-DCT is applied along the temporal axis from which a subset of first few columns are chosen and randomly rearranged to compute hash of the given video. Their hashing scheme provides good differentiating power and also robust against content preserving attacks [44] have considered scene frame features along with video ID as fingerprint and are stored in Bag-words as a hash key. To verify authenticity inverted index file IIF based algorithm is used and the performance is found to be satisfactory [45] have developed a video hashing system for content identification and authentication of short video clips. Differential luminance block mean (DLBM) is extracted in vertical, horizontal and temporal directions. DCT/DST based signal normalization is done to produce signal coefficient pair. Digital signals are calibrated to exhibit sensitivity to malicious manipulations and to resolve synchronization issues. The hash value thus obtained is robust to noise and low pass filtering. The identification and authentication threshold is based on Neyman–Pearson criterion and is used in such a way that the missed similarity detection probability is minimized.

Most of the hashing schemes found on videos have considered DCT as feature vectors and they are sensitive to block based operations. Robust techniques which need to detect content altering modifications and flexible to content preserving modifications is the requirement of the day.

3. Proposed digital signature approach

The content based signature approach proposed in this paper meets high level fidelity and is sensitive to malicious attacks and recompression. This authentication approach generates a content dependent signature, adopting a hierarchical model for a given video sequence by using the local descriptors which are strongly coupled. Human perception remains same for consecutive frames, where there is no much motion but the fingerprint of such frames remains distinct. Storage space occupied by the video varies with its format. Normally videos are stored in a compressed format to facilitate effective storage. In order to perform any tampering on such videos, individual frames of the video need to be edited and then recompressed. Irrespective of the video type and size, the proposed system is successful in content based signature generation.

Initially the given video sequence or clip is broken down to frames. Then the content of each frame is uniquely represented using local keypoints obtained. The keypoints which are affine to scale invariant transformations like SIFT [46], SURF [47], HOG [48], Harris [49] and MSER [50] are considered in our work. The positions of such extracted key points are considered as features and