

Video Temporing Detection Assessment in No-Reference Mode using Machine Learning

Gajanand Sharma, Dr. Bright Keswani, Dr. Dinesh Goyal

Abstract-Information System Security is an integral part of almost all modern communication system. The requirement for ensuring security varies with the application under consideration. Some applications require encryption so that no-one, except the intended receiver, should have the access of the digital data. This form of information hiding comes under cryptography. The encryption can be combined with covert communication in which the encrypted file is embedded using suitable techniques in some other media so that no-one could think of its presence. One the other hand, there are applications in which we need copyright protection. One large class of applications is the one in which some form of watermarking is required so as to later provide a proof of ownership of the document. All the three techniques; viz, Cryptography, Steganography and Watermarking shares common attributes and are related with Information System Security. Despite of the purpose which differentiates Steganography and Watermarking, the tools and techniques used to change the cover media are equally applicable to both these techniques. In this paper, video tempering detection techniques are proposed for detection of spatial and temporal tempering in the video frames of mpeg encoded video. The tempering detection has been investigated under No Reference Mode. In this mode, correlation properties between image segments in adjacent frames and statistical feature set is used to train machine learning classifier for tempering detection and the type and location of tempering. Particular emphasis has been given on MPEG-2 (Moving Picture Expert Group-2) as it is a widely accepted standard. Simulation of the proposed techniques is done using MATLAB and R. Standard Quality metrics are derived and it turns out that the proposed techniques outperform those of the benchmark techniques.

Index Terms- Video Temporal, Tempering Detection, No-Reference Mode, Machine Learning, MSE, PSNR.

I. INTRODUCTION

A. Digital Video Forensics

The widespread availability of the Internet, coupled with the easy availability of video and image capturing devices, such as low-price cameras, digital camcorders and CCTVs

Manuscript revised June 9, 2019 and published on July 10, 2019

Gajanand Sharma, *Research Scholar, Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur, INDIA.*

Dr. Bright Keswani, *Professor, Department of Computer Application, Suresh Gyan Vihar University, Jaipur, INDIA.*

Dr. Dinesh Goyal, *Professor, Department of Computer Science and Engineering, Poornima Institute of Technology, Jaipur, INDIA.*

cheaper editing software, meant that it is getting easier for ordinary people with nefarious aims to have access to digital doctoring tools in order to modify images and videos. This in turn means that the images and videos that are seen in mass media such as television, popular Internet websites such as YouTube, may have been tampered and the adage “a picture speaks a thousand words” while still holding true – may now have a hidden and subverted meaning, i.e., their authenticity can no longer always be taken for granted.

Yet, at the same time, organizations and countries continue to rely heavily on the authenticity of images and videos. For example, an increasing number of governments and organizations are employing the use of CCTVs because these devices are not only getting cheaper but also are being heavily relied on as useful and necessary tools in surveillance. A report by Urbaneye Project published in June 2002 estimated that there were possibly 500,000 CCTVs in London [1]. Other countries too, such as European countries like France, Italy, etc. and even Asian countries such as Singapore are also using CCTVs to prevent and detect violence and theft.

B. Problem Statement

Video Tempering Detection, in Spatial and Temporal Domain is investigated in this dissertation. Spatial Tempering refers to the tempering made in individual frames whereas temporal tempering refers to the dropping of frames or the change in the sequence of frames by duplicating or appending additional frames in the frame sequence. All these cases are investigated under Full and No reference schemes, which equals to the case of Blind and Non-Blind Techniques for Watermarking Detection or Steganalysis. When there is no information available about the “to be examined video”, verification of the authenticity of the “to be examined video” (which may or may not be a tampered video) is the most challenging and difficult task. Under such circumstances, detection of tampering (here we are concerned about temporal tampering) is usually performed by examining the video for any clue or artifacts left by forger in form of inconsistency or abnormality across the video frames in the “to be examined video”. Additionally, a video watermarking technique is proposed for raw video frames of the original video. This technique is proposed using DWT and SVD Transform. The watermark data is embedded in the RGB planes of I frames of the mpeg which are encoded in JPG image, and used as a means to implement robust video watermarking scheme.

C. Research Approach

As stated previously, techniques are preened in this thesis for Video Tempering Detection in both the Full and No Reference Modes of analysis. By the term Video Tempering, there is the possibility of one or all of the actions stated below:

1. Individual Frame Drop

2. A Sequence of Frames Dropping
3. Inclusion of Additional Frames in the Video Sequence.
4. Copying of Same frames redundantly in video.
5. Changing Frame Sequences
6. Changing the content of the frames
7. Combining one or
8. more of the above Stated Operations

In all the above cases, in No Reference Mode, correlation analysis is required to be made. In this mode, a feature set based on correlation parameters is required to be constructed to train a machine learning classifier.

II. PROPOSED WORK

A. Spatial Video Tempering Detection Technique under No reference Scheme:

By the term “No-Reference”, we mean that the original authentic video is not available for the reference. The full-reference schemes are easy to implement as only a frame-by-frame comparison is required to be made between the video under consideration and the referenced video. This is demonstrated in test video of squirrel in the Chapter 4. The No-Reference scheme is complex to implement for frame tempering either in spatial, temporal or spatio-temporal domain. The Tempering Detection in mpeg-2 video in Spatial Domain is proposed under No-reference model through RS analysis as indicated in the proposed algorithm.

The Lossless capacity reflects the fact that the LSB plane – even though it looks random– is related to the other bit planes and the method is based on the fact that the content of each bit plane of an image is correlated with the remaining bit planes. In RS Analysis the image is partitioned into groups of pixels (Regular and Singular groups) of a fixed shape depending upon specific properties. Each group classified as ‘regular’ or ‘singular’ depending on whether the pixel noise within the group (as measured by the mean absolute value of the differences between adjacent pixels) is increased or decreased after flipping the LSBs of a fixed set of pixels within each group (the pattern of pixels to flip is called the ‘mask’). The classification is repeated for a dual type of flipping. The proportion of regular and singular groups form curves quadratic in the amount of alteration induced in the original image file. It turns out that the proportion of the Regular blocks increases substantially as compared to the singular blocks as a result of tempering in the original image. Using this assumption, some estimate can be drawn about the amount of tempering which has been done in the original image file. The estimate can be accurate (often within 1%) most of the time. This can be illustrated empirically using a set of natural images and artificial images for the count of regular and singular blocks.

In Fig 1 (a) and (b) respectively, natural and artificial images are shown.

Definition, a natural image is one which is clicked from camera. On the other hand, an artificial image is the image obtained from the combination of two or more natural images with additional artifacts like clipart etc.

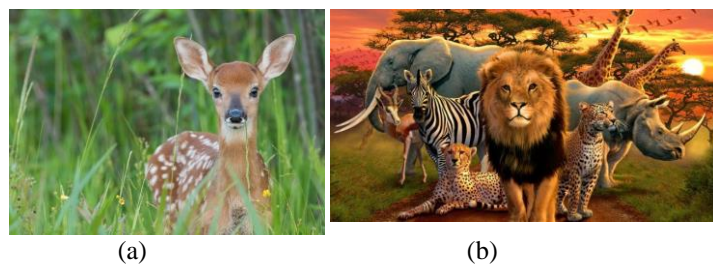


Fig 1 (a) Natural Image and (b) Artificial Image

Consider 8X8 segments (shown in red rectangle) of the above two images as shown in figure 2

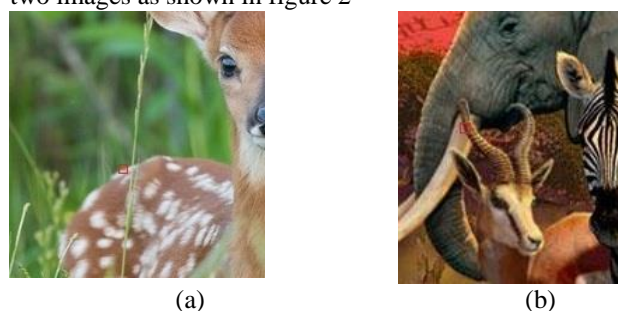


Fig 2 Identified 8X8 block (Red Square) of images

The coordinates of the top left corner of the image are (0,0). In image (a), the coordinates of the top left corner of the 8X8 square is (316, 246). In the second image, the coordinates are (125,210). Here, the first quantity refers to the X coordinate and the second refers to Y coordinate. The images shown are resized for the purpose of inclusion herein; however, the original images are included in the CD ROM enclosed at the back-cover.

For the sake of illustration and demonstration, we have considered Red Component of the image. The image matrix so obtained, for the natural image is:

| | | | | | |
|---|----|----|----|-----|-----|
| 74 | 75 | 78 | 84 | 83 | 87 |
| 79 | 81 | 84 | 85 | 82 | 83 |
| 85 | 86 | 88 | 88 | 84 | 83 |
| 91 | 93 | 90 | 92 | 87 | 85 |
| 94 | 97 | 92 | 92 | 88 | 88 |
| 89 | 92 | 93 | 94 | 91 | 93 |
| 88 | 91 | 91 | 94 | 95 | 101 |
| 86 | 88 | 88 | 97 | 102 | 111 |
| For the artificial image, the aforesaid red square matrix is: | | | | | |
| 66 | 63 | 73 | 82 | 72 | 50 |
| 75 | 71 | 86 | 85 | 63 | 34 |
| 82 | 83 | 96 | 87 | 58 | 30 |
| 88 | 90 | 92 | 76 | 49 | 25 |
| 92 | 96 | 88 | 62 | 25 | 21 |
| 94 | 94 | 81 | 46 | 9 | 23 |
| 88 | 86 | 70 | 37 | 7 | 30 |
| 86 | 77 | 59 | 29 | 8 | 42 |

A zig-zag scan is performed, taking the differences at each step and adding the absolute values of all the differences. This value is termed as the correlation index. A small value of correlation index indicates correlated pixels. A large value indicates uncorrelated pixels. For the first image, the correlation index is:
 $|74-75| + |75-79| + |79-85| + |85-81| + |81-78| + |78-84| + |84-84| + |84-86| + |86-91| + |91-93| + |93-88| + |88-85| + |85-88| + |88-90| + |90-92|$
 $=1+4+6+4+3+6+0+2+5+2+5+3+3+2+2$
 $=48$

For the second image, the artificial one, the corresponding value is

$|66-63| + |63-75| + |75-82| + |82-71| + |71-73| + |73-82| + |82-86| + |86-83| + |83-88| + |88-90| + |90-96| + |96-85| + |85-87| + |87-92| + |92-76|$
 $=3+8+7+3+2+9+4+3+5+2+6+11+2+5+16$
 $=86$

This value is almost twice as big as the natural image. Thus, even for small image segments, this can be accomplished using Regular and Singular analysis that the natural images have a high degree of correlation between the pixel values as compared to the artificial images. It is this artifact that can be used to differentiate between spatially tempered images considering the 'I' frames which are being encoded as the JPEG images.

Given an Image I, we partition the image into 8X8 blocks. For each of the block, the correlation coefficient is computed as outline above. In the next step, each partition is subjected to positive and negative flipping operation as shown:

Positive Flipping : $0 \rightleftharpoons 1, 2 \rightleftharpoons 3, 4 \rightleftharpoons 5, 6 \rightleftharpoons 7, \dots, 254 \rightleftharpoons 255$

Negative Flipping: $-1 \rightleftharpoons 0, 1 \rightleftharpoons 2, 3 \rightleftharpoons 4, 5 \rightleftharpoons 6, \dots, 253 \rightleftharpoons 254$

The correlation index is computed by using a series of flipping masks over the image and the average value is computed. If the correlation index after flipping exceeds the original value of the correlation index before flipping, the Block is termed as Regular Block, other-wise it is called singular block.

Corresponding to positive and negative flipping, there are four types of Blocks, namely;

R^+ ; Regular Blocks obtained after positive flipping

R^- ; Regular Blocks obtained after negative flipping

S^+ ; Singular Blocks obtained after positive flipping

S^- ; Singular Blocks obtained after negative flipping

We hereby propose the following algorithm for analysis of spatially tempered image under no reference scheme.

Algorithm:

1. Obtain the I Frame sequence through decompression of the mpeg-2 bit stream.
2. The I Frame is encoded as the jpeg image. Transform the color space from YCbCr to RGB.
3. Separate the image into Red, Green and Blue Planes.
4. Create Random Flipping masks to flip the LSB bit planes.
5. Obtain the number of number of Regular and Singular Blocks over standard 8x8 partitions of the image. This step is to be performed over all the three color planes of the image in RGB space. A regular block is one in which the pixel noise in the frame partition is increased when some tempering is made in the spatial

domain. The number of such blocks is increased greatly when some tempering is made in the spatial domain. On the other hand, a singular block is one in which the pixel noise in the frame partition get reduced. The number of such blocks is greatly reduced when some tempering is made to the frame in the spatial domain. A set of partition sizes for the entire frame can be constructed based on the overall aspect ratio of the video.

6. A set of pixels, known as a flipping mask, is flipped to obtain a possibly doctored image. Apply a series of flipping masks to obtain flipped LSB variants of the given image.
7. If the difference in the number of R and S blocks is well beyond threshold, then the presence of tempering can be stated with high probability.

The system model for the proposed technique of Spatial Tempering Detection is illustrated in figure 3.

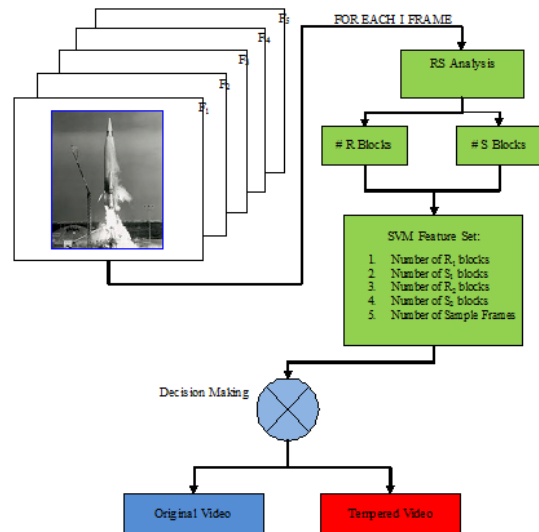


Fig 3 Classifier for Original and Spatially Tempered Video
 Consider the tempering of the frame in the spatial domain in which tempering has been made in the image to sharpen the contrast. The difference in the original and the tempered image is depicted as shown in fig 5.4.



Fig. 4 Edge improvement difference to enhance the contrast of the image

However, at times, not only the brightness or the contrast of the image is changed; the objects in the image are changed from some malicious intent. For this analysis, the I frames in the original video are analyzed for Regular and Singular Blocks. Consider the two images shown in fig 3.5. Both these images correspond to some I frame in the video.

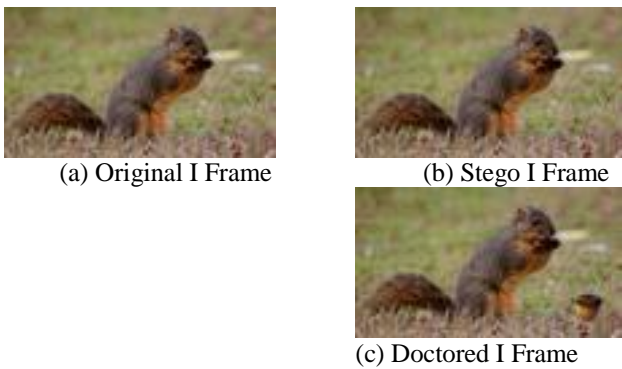


Fig. 5 Spatial Tempering Example

In the first image, the original video frame is shown. In the second image, a spatially tempered video frame is shown in which there, a nut is lying behind the squirrel. The spatial tempering in the video frame can be identified in the following categories as shown in table 1.

TABLE 1
CLASSIFICATION OF SPATIAL TEMPERING IN VIDEO FRAMES

| Spatial Tempering Category | Description |
|--|---|
| Tempering in the Original Video Frame in the LSB Bit Planes to Embed secret Message. | The spatial tempering detection in this case is called steganalysis. The image is partitioned into blocks and each is labeled as Regular or Singular under different masks and flipping operations. An estimation of the count of regular and singular blocks classifies a given frame image as natural or stego image. |
| Tempering in the Original Video Frame to include other objects | Such a video is referred to as a doctored video. The resultant image is an artificial image and can be classified using the Entropy, Hue and Saturation values which can be used as inputs to a machine learning classifier. |

III. RESULTS

A. Tempering in the Video to Change LSB of I Frames

The first category of spatial tempering can be detected using a correlation test called Regular Singular Analysis over the image frame.

A sample image mask for Bit plane flipping can be depicted as shown in Figure 6.



Fig. 6 Image Mask

The black color indicates a 0 whereas white color indicates a 1. Either 0 or 1 corresponds to a negative or a positive flipping operation. Therefore, corresponding to each mask, we may have two type of flipping operations. The image may be partitioned into non overlapping segments of various sizes. The count of the image partitions is equal to image size divided by the partition size. Each petition is classified as Regular or Singular depending upon whether the discrimination function is greater or smaller than its initial value.

The tabulation of the Regular and Singular Blocks, on the tempered image shown above, corresponding to the illustrated flipping mask, in the block size ranging from 2X2 to 16X16 pixels is depicted as shown in table 2

TABLE 2
TABULATION OF R AND S BLOCKS AFTER SPATIAL TEMPERING

| Block Size | Total Block s | R+ Block s | S+ Block s | R- Block s | S- Block s |
|------------|---------------|------------|------------|------------|------------|
| 4 | 1152 | 841 | 311 | 841 | 311 |
| 9 | 512 | 369 | 143 | 358 | 154 |
| 16 | 288 | 204 | 84 | 216 | 72 |
| 36 | 128 | 93 | 35 | 95 | 33 |
| 64 | 72 | 52 | 20 | 51 | 21 |
| 144 | 32 | 22 | 10 | 23 | 9 |
| 256 | 18 | 12 | 6 | 13 | 5 |

The Histogram of RS Steg-analysis on the frame is depicted as shown in figure 7

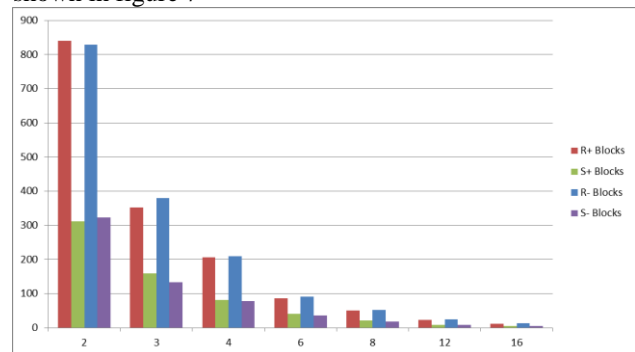


Fig.7 Number of Regular and Singular Blocks Corresponding to the Mask shown above and Flipping Operations

Horizontal axis shows the block size whereas vertical axis shows the count of the blocks. It is evident from the count of blocks that the count of R blocks is significantly above the count of S blocks. This is likely because of the fact that the correlation of the pixel values becomes distorted as the bit planes are changed by the embedding messages.

B. Tempering in the Original Video Frame to include other objects: An image which is modified to include some other object is called a doctored image. Such an image is called an artificial or synthetic image. As mentioned in Chapter 3, the feature set is required to be extracted from an image to classify it into natural image or artificial image. The feature set is

comprises of the mean histogram value, the Hue and the saturation index. The histogram of the natural Image I shown in figure 5) are shown in figure 10 and 11.

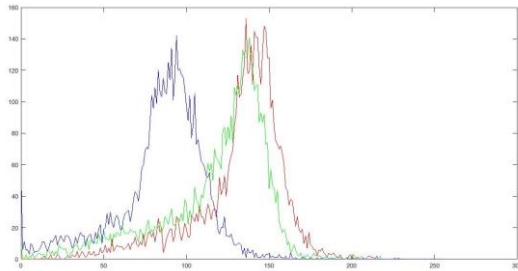


Fig. 8 Histogram of Red, Blue and Green in Natural Image

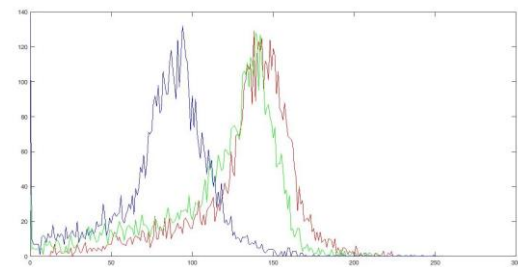


Fig. 9 Histogram of Red, Blue and Green in Doctored Image
The Hue, Saturation and Value illustration of the natural Image and Artificial Image of the sample frame from squirrel video are shown in figure 10. In the artificial image, the same video frame is tempered to include a nut with the squirrel.



Hue Saturation Value

Fig. 10 Hue, Saturation and Value for Natural Image



Hue Saturation Value

Fig. 11 Hue, Saturation and Value for Doctored Frame (Squirrel with a nut).

The mean value of Hue and Saturation in Natural and Artificial Image are shown in Table 3

TABLE 3

TABULATION OF FEATURE SET VALUES FOR SVM

| Parameter | Natural Image | | | Artificial Image | | |
|-----------|---------------|--------|--------|------------------|--------|--------|
| | R | G | B | R | G | B |
| Entropy | 6.5399 | 6.7443 | 6.5723 | 6.7626 | 6.9227 | 6.6944 |
| Mean Hue | 0.1210 | | | 0.1222 | | |

| | | |
|-----------------|--------|--------|
| Mean Saturation | 0.3793 | 0.4107 |
| Mean Value | 0.5194 | 0.5225 |

Over a Dataset of 50 images in the 2 second video of squirrel, the confusion matrix of the SVM Classifier is prepared as indicated below:

TRAINING DATA : 50 samples

TEST DATA : 25 samples

TABLE 4

CONFUSION MATRIX DEFINITION

| | | |
|---------------------------|--------------------|-------------------|
| ORIGINALITY→ DETECTION | TEMPERED YES | TEMPERED NO |
| ↓ | | |
| TEMPERED YES | TRUE POSITIVES | FALSE POSTIVES |
| TEMPERED NO | FALSE NEGATIVES | TRUE NEGATIVES |

Out of a sample of 25 test frames in which 20 are spatial tempered and 5 are original in the raw video, the confusion matrix of the SVM gives the following output.

TABLE 5

CONFUSION MATRIX FOR SAMPLE VIDEO

| | | |
|---------------------------|--------------|-------------|
| ORIGINALITY→ DETECTION | TEMPERED YES | TEMPERED NO |
| ↓ | | |
| TEMPERED YES | 18 | 3 |
| TEMPERED NO | 2 | 2 |

It is clearly evident from the confusion matrix that the proposed scheme performs well for the classification of tempered and the original video frames. Only 18 out of 20 tempered video frames are successfully identified, thereby providing an accuracy of $18/20 = 90$ percent in the classification process.

C. Proposed Schemes for Dropped Frame Identification

(Temporal Tempering Detection) under No Reference

condition: The No Reference scheme for dropped frame identification is based on the structure of the I, P and B Frames in the mpeg stream. When one or more frames have been removed from the mpeg stream, the resulting structure of the mpeg stream gets distorted. This is because the specification of the number of P and B Frames is provided in the Group of Pictures (GOP) Header. This header structure can be analyzed for the number of P and B frames which are in the stream and rendered before another I frame followed by a set of P and B Frames. MPEG-1 video encoder block diagram is shown in Figure 5.12. Input Video is first converted from RGB to YUV format. If the given frame is I frame, it will be simply compressed as JPEG compression. This means that I frame is passed through DCT, quantization and then entropy coding is done. If it is P frame or B frame then motion estimation is done with respect to reference frame and prediction error and then motion vector is coded.

The detection of the frame drop under no reference scheme uses machine learning classifier to be trained on a feature set representing the entropy of the pixel blocks in the adjacent

frames. An SVM based classifier is proposed to classify temporally tempered video. This classifier is required to be trained over a training dataset to be created using a set of similarly encoded video, with explicitly generated positive and negative examples. We have used such dataset from sample videos from the web resource <https://media.xiph.org/video/derf/>. Proposed frame tempering detection uses Entropy, Hue, Saturation and Value differences between adjacent frames as the feature set for SVM classifier for detection of spatial tempering. The sample videos investigated to train the SVM classifier are indicated as shown in table 6.

TABLE 6
SAMPLE VIDEOS CONSIDERED FOR TEMPORAL TEMPERING DETECTION



akiyo (4:3 | 300 frames)
| [readme](#))

Source: http://meru.cecs.missouri.edu/free_download/videos/, <ftp://sotka.cs.tut.fi/cost/Ossi/sequences/>, <http://trace.eas.asu.edu/yuv/>

Download: [[CIF](#) (44 MB)
| [QCIF](#) (11 MB)]



bowing (4:3 | 300 frames)
| [Copyright](#))

Source: <ftp://ftp3.itu.int/video-site/sequences/Bowing/> ([login required](#))

Download: [[CIF](#) (44 MB)
| [QCIF](#) (11 MB)]



bridge_close (4:3 | 2001 frames)

Source: <http://trace.eas.asu.edu/yuv/>

Download: [[CIF](#) (291 MB)
| [QCIF](#) (73 MB)]



bridge_far (4:3 | 2101 frames)

Source: <http://trace.eas.asu.edu/yuv/>

Download: [[CIF](#) (305 MB)
| [QCIF](#) (77 MB)]



bus (4:3 | 150 frames)

Source: <ftp://ftp.tnt.uni-hannover.de/pub/svc/testsequences/>, <http://trace.eas.asu.edu/yuv/>

Download: [[CIF](#) (22 MB) | [15 fps QCIF](#) (2.8 MB)]



carphone (4:3 | 382 frames)
Source: <http://ise.stanford.edu/Video/carphone.qcif.gz>, <ftp://sotka.cs.tut.fi/cost/Ossi/seq/>, <http://trace.eas.asu.edu/yuv/>
Download: [[QCIF](#) (14 MB)]

A sample of 120 such video are considered, from the URL as specified, for the purpose of temporal tempering detection.

We propose the extended feature set, as compared to the set of 4 features proposed by Manish Kr. Thet. al. [7]. The complete set of 5 features for SVM training, including the four used in [10] are:

1. Entropy: A measure of randomness between the image pixels in adjacent frames.
2. Average Object Area (AOA) ; Between two adjacent frames of the video under investigation. It is computed as:

$$A(i) = \frac{1}{n} \sum_{i=1}^n area(i)$$

Where, area(i) is the area of the area of the i^{th} object of the binary difference frame and n is the total number of objects in the binary difference frame.

3. Mean Squared Error
4. Count of Displaced Blocks

This feature represents the count of the 8X8 blocks which are not present in the subsequent block.

5. Harris Corner Point Feature;

This feature is used to find the points of interest in the Frame image and the check whether these important points are present in the subsequent frame or not.

The confusion matrix for the video files under investigation is shown in table 7.

TABLE 7
CONFUSION MATRIX FOR TEMPORALLY TEMPERED VIDEO

| | Temporally Tempered Prediction | | Temporally Tempered Actual | | Accuracy |
|----------|--------------------------------|----|----------------------------|----|----------|
| | Yes | No | Yes | No | |
| Video 1 | 1 | 0 | 1 | 0 | 1 |
| Video 2 | 1 | 0 | 1 | 0 | 1 |
| Video 3 | 0 | 1 | 1 | 0 | 0 |
| Video 4 | 1 | 0 | 1 | 0 | 1 |
| Video 5 | 1 | 0 | 1 | 0 | 1 |
| Video 6 | 0 | 1 | 1 | 0 | 0 |
| Video 7 | 0 | 1 | 1 | 0 | 0 |
| Video 8 | 0 | 1 | 1 | 0 | 0 |
| Video 9 | 1 | 0 | 1 | 0 | 1 |
| Video 10 | 1 | 0 | 1 | 0 | 1 |
| Video 11 | 1 | 0 | 1 | 0 | 1 |

| | | | | | |
|----------|---|---|---|---|---|
| Video 12 | 1 | 0 | 1 | 0 | 1 |
| Video 13 | 1 | 0 | 1 | 0 | 1 |
| Video 14 | 1 | 0 | 1 | 0 | 1 |
| Video 15 | 0 | 1 | 1 | 0 | 0 |
| Video 16 | 1 | 0 | 1 | 0 | 1 |
| Video 17 | 1 | 0 | 1 | 0 | 1 |
| Video 18 | 0 | 1 | 1 | 0 | 0 |
| Video 19 | 1 | 0 | 1 | 0 | 1 |
| Video 20 | 0 | 1 | 1 | 0 | 0 |
| Video 21 | 1 | 0 | 1 | 0 | 1 |
| Video 22 | 1 | 0 | 1 | 0 | 1 |
| Video 23 | 1 | 0 | 1 | 0 | 1 |
| Video 24 | 1 | 0 | 1 | 0 | 1 |
| Video 25 | 1 | 0 | 1 | 0 | 1 |
| Video 26 | 1 | 0 | 1 | 0 | 1 |
| Video 27 | 0 | 1 | 1 | 0 | 0 |
| Video 28 | 1 | 0 | 1 | 0 | 1 |
| Video 29 | 1 | 0 | 1 | 0 | 1 |
| Video 30 | 0 | 1 | 1 | 0 | 0 |
| Video 31 | 0 | 1 | 1 | 0 | 0 |
| Video 32 | 1 | 0 | 1 | 0 | 1 |
| Video 33 | 1 | 0 | 1 | 0 | 1 |
| Video 34 | 1 | 0 | 1 | 0 | 1 |
| Video 35 | 0 | 1 | 1 | 0 | 0 |
| Video 36 | 1 | 0 | 1 | 0 | 1 |
| Video 37 | 1 | 0 | 1 | 0 | 1 |
| Video 38 | 1 | 0 | 1 | 0 | 1 |
| Video 39 | 1 | 0 | 1 | 0 | 1 |
| Video 40 | 0 | 1 | 1 | 0 | 0 |
| Video 41 | 1 | 0 | 1 | 0 | 1 |
| Video 42 | 1 | 0 | 1 | 0 | 1 |
| Video 43 | 1 | 0 | 1 | 0 | 1 |
| Video 44 | 1 | 0 | 1 | 0 | 1 |
| Video 45 | 1 | 0 | 1 | 0 | 1 |
| Video 46 | 1 | 0 | 1 | 0 | 1 |
| Video 47 | 1 | 0 | 1 | 0 | 1 |
| Video 48 | 0 | 1 | 1 | 0 | 0 |
| Video 49 | 1 | 0 | 1 | 0 | 1 |
| Video 50 | 1 | 0 | 1 | 0 | 1 |
| Video 51 | 1 | 0 | 1 | 0 | 1 |
| Video 52 | 1 | 0 | 1 | 0 | 1 |
| Video 53 | 1 | 0 | 1 | 0 | 1 |
| Video 54 | 1 | 0 | 1 | 0 | 1 |
| Video 55 | 1 | 0 | 1 | 0 | 1 |
| Video 56 | 1 | 0 | 1 | 0 | 1 |
| Video 57 | 1 | 0 | 1 | 0 | 1 |
| Video 58 | 1 | 0 | 1 | 0 | 1 |
| Video 59 | 1 | 0 | 1 | 0 | 1 |
| Video 60 | 1 | 0 | 1 | 0 | 1 |
| Video 61 | 1 | 0 | 1 | 0 | 1 |
| Video 62 | 1 | 0 | 1 | 0 | 1 |
| Video 63 | 1 | 0 | 1 | 0 | 1 |
| Video 64 | 1 | 0 | 1 | 0 | 1 |
| Video 65 | 0 | 1 | 1 | 0 | 0 |
| Video 66 | 1 | 0 | 1 | 0 | 1 |
| Video 67 | 0 | 1 | 1 | 0 | 0 |
| Video 68 | 1 | 0 | 1 | 0 | 1 |
| Video 69 | 0 | 1 | 1 | 0 | 0 |
| Video 70 | 1 | 0 | 1 | 0 | 1 |
| Video 71 | 0 | 1 | 1 | 0 | 0 |
| Video 72 | 1 | 0 | 1 | 0 | 1 |
| Video 73 | 1 | 0 | 1 | 0 | 1 |
| Video 74 | 1 | 0 | 1 | 0 | 1 |
| Video 75 | 1 | 0 | 1 | 0 | 1 |
| Video 76 | 1 | 0 | 1 | 0 | 1 |
| Video 77 | 0 | 1 | 1 | 0 | 0 |
| Video 78 | 1 | 0 | 1 | 0 | 1 |
| Video 79 | 1 | 0 | 1 | 0 | 1 |
| Video 80 | 1 | 0 | 1 | 0 | 1 |
| Video 81 | 1 | 0 | 1 | 0 | 1 |
| Video 82 | 0 | 1 | 1 | 0 | 0 |
| Video 83 | 1 | 0 | 1 | 0 | 1 |
| Video 84 | 1 | 0 | 1 | 0 | 1 |
| Video 85 | 1 | 0 | 1 | 0 | 1 |
| Video 86 | 1 | 0 | 1 | 0 | 1 |
| Video 87 | 1 | 0 | 1 | 0 | 1 |
| Video 88 | 1 | 0 | 1 | 0 | 1 |
| Video 89 | 1 | 0 | 1 | 0 | 1 |
| Video 90 | 0 | 1 | 1 | 0 | 0 |
| Video 91 | 1 | 0 | 1 | 0 | 1 |
| Video 92 | 1 | 0 | 1 | 0 | 1 |
| Video 93 | 0 | 1 | 1 | 0 | 0 |
| Video 94 | 1 | 0 | 1 | 0 | 1 |
| Video 95 | 1 | 0 | 1 | 0 | 1 |
| Video 96 | 1 | 0 | 1 | 0 | 1 |
| Video 97 | 0 | 1 | 1 | 0 | 0 |

| | | | | | |
|-----------|---|---|---|---|---|
| Video 98 | 1 | 0 | 1 | 0 | 1 |
| Video 99 | 1 | 0 | 1 | 0 | 1 |
| Video 100 | 0 | 1 | 1 | 0 | 0 |
| Video 101 | 1 | 0 | 1 | 0 | 1 |
| Video 102 | 1 | 0 | 1 | 0 | 1 |
| Video 103 | 1 | 0 | 1 | 0 | 1 |
| Video 104 | 1 | 0 | 1 | 0 | 1 |
| Video 105 | 0 | 1 | 1 | 0 | 0 |
| Video 106 | 0 | 1 | 1 | 0 | 0 |
| Video 107 | 1 | 0 | 1 | 0 | 1 |
| Video 108 | 1 | 0 | 1 | 0 | 1 |
| Video 109 | 0 | 1 | 1 | 0 | 0 |
| Video 110 | 0 | 1 | 1 | 0 | 0 |
| Video 111 | 0 | 1 | 1 | 0 | 0 |
| Video 112 | 1 | 0 | 1 | 0 | 1 |
| Video 113 | 1 | 0 | 1 | 0 | 1 |
| Video 114 | 0 | 1 | 1 | 0 | 0 |
| Video 115 | 1 | 0 | 1 | 0 | 1 |
| Video 116 | 1 | 0 | 1 | 0 | 1 |
| Video 117 | 1 | 0 | 1 | 0 | 1 |
| Video 118 | 1 | 0 | 1 | 0 | 1 |
| Video 119 | 1 | 0 | 1 | 0 | 1 |
| Video 120 | 0 | 1 | 1 | 0 | 0 |

IV. CONCLUSION AND FUTURE SCOPE

Digital videos are gaining popularity due to its expressive capabilities, however constrained with its misuse in various areas such as legal, medical, video surveillance, defence, defamation, etc. Recently many of such misuses have been reported in several controversies, where original videos were tampered to create fake or doctored videos. Digital video forensics tools and experts are often involved in detecting tampering with such tampered videos. Due to lack of structured methodologies and effective tools for detection of tampering with videos, the area of digital video forensics is a challenging domain. The schemes proposed in this paper are a small contribution in the area of digital video forensics.

In this paper our major contribution is focused on development of tampering detection schemes to verify the authenticity of a video as well as identification of tampering location in temporally tampered videos. The schemes have been proposed under no reference mode of tampering detection. However, these schemes have been developed with the assumption that any temporal tampering may disturb the temporal redundancy in successive frames, which are usually available in videos. The temporal tampering of frame drop at frame level has been discussed under full reference mode of tampering detection. We have shown that our proposed set of

schemes successfully detected the exact location of dropped frames, if videos were spatio-temporally tampered videos, whereas the watermark based schemes, such as the benchmark schemes are inefficient to identify the exact location of dropped frames in such videos. We have proposed set of features in SVM learning to detect the location of tampering of frame drop in temporally tampered videos and accordingly verify the authenticity of videos. Here, we have shown that, the average classification accuracy with our proposed features are comparatively better than the set of features used so far.

In addition, we have shown that our proposed scheme successfully detected the location of multiple set of dropped frames and accordingly verified the authenticity of temporally tampered videos. We have also proposed a threshold based scheme which successfully identified the location of copied frames in videos and accordingly verified the authenticity of a video. While development of no reference tampering detection techniques we have found and emphasized on to select such features which could trace the temporal tampering with videos, and hence in this thesis come up with such set of features, viz. Entropy and Mean Squared Error (MSE). Based on these features, we have proposed various thresholds along with their recommended values. These thresholds have been used in various schemes and found to locate the temporal tampering of frame drop, etc. However, one may further tune these thresholds to improve the accuracies of various schemes proposed in this thesis. Efforts may also be applied in search of other suitable features and accordingly can be applied in the proposed schemes. On the other hand, in this thesis we have also addressed the issues with common full reference video quality metrics to assess the quality degradation in temporally and spatio-temporally distorted videos.

However, efforts may further applied in search of other parameters which have significant impact on perceived video quality in temporally as well as spatio-temporally distorted videos.

REFERENCES

- [1]. Madden, M., "The Audience for Online Video Sharing Sites Shoots Up" <http://fe01.pewinternet.org/Reports/2009/13--The-Audience-for-Online-Video-Sharing-Sites-Shoots-Up.aspx> Accessed 20 January 2014
- [2]. Nielsen, "State of the Media: The Social Media Report," Q3,2011, at http://cn.nielsen.com/documents/Nielsen-Social-Media-Report_FINAL_090911.pdf. Accessed 22 January 2014
- [3]. Dare, L. and Gar, C. S., "Video Sharing: SVEA Training Modules," at <http://www.svea-project.eu>. Accessed 12 January 2014
- [4]. Rocha, A., Scheirer, W., Boulton, T., and Goldenstein, S., "Vision of the unseen: Current trends and challenges in digital image and video forensics," *ACM Computing Surveys*, Vol. 43, No. 4, Article 26, October 2011, pp. 1-42. DOI: 10.1145/1978802.1978805
- [5]. Redi, J.A., Taktak, W., and Dugelay, J.L., "Digital image forensics: a booklet for beginners," *Multimed Tools Appl*, Vol. 51, Issue 1, Jan 2011, pp. 133-162. DOI: 10.1007/s11042-010-0620-1
- [6]. Farid, H., "Image Forgery Detection," *IEEE Signal Processing Magazine*, Vol. 26, Issue 2, March 2009, pp. 16-25

- [7]. Carvalho, T., Riess, C., Angelopoulou, E., Pedrini, H., and Rocha, A., "Exposing Digital Image Forgeries by Illumination Color Classification," *IEEE Trans. on Information Forensics and Security*, Vol. 8, No. 7, July 2013, pp. 1182-1194. DOI: 10.1109/TIFS.2013.2265677

Mahaveer Open University, Kota and Bhoj University, Madhya Pradesh, Suresh Gyan Vihar University, Jaipur etc.

AUTHORS PROFILE



Gajanand Sharma completed his Bachelor's degree in Computer Science & Engineering in year 2005 from JECRC Jaipur (University of Rajasthan). Mr. Sharma joined Rajasthan Technical University to complete his M.Tech. in 2014 with research work on 'Information Security' and Pursuing Ph.D. with specialization in Computer Science & Engineering from Suresh Gyan Vihar University, Jaipur. His areas of research are Image Processing & Security. Mr. Sharma has published 16 research papers

in international publications followed by 9 papers presented in conferences and has guided 11 PG Scholars.

Mr. Sharma is Life time Member, CSI Jaipur Chapter since 2012.(Membership No.-11503681), IAENG since 03/2014 (Membership No.-148248) and GRDS under ADTEL since 04/2017.(Membership No.- ADTEL-M17294).



Dr. Bright Keswani (M.Phil, M.Tech and Ph.D.) honored with 'BEST CITIZEN OF INDIA AWARD- 2013', 'ACADEMIC EXCELLENCE AWARD - 2015', 'SHIKSHA BHUSHAN AWARD - 2009' and 'BHARAT JYOTI AWARD - 2013' for outstanding achievements in the field of Computer Science. Presently Dr. Keswani is working as Professor & Head of the Department of Computer Applications, Principal (Academic Staff College) and Editor-In-Chief (SGVU-Journal of Engineering & Technology) at Suresh Gyan

Vihar University, Jaipur. Dr. Keswani has a long standing of teaching at graduate and postgraduate level for more than 16 years at various Institutions. Dr. Keswani have been taught many courses in M.Tech, MCA, M.Sc., PGDCA, MBA and BCA, BBA Programs in past years. Presently, Dr. Keswani is devoting his maximum time for research in the varied field of Computer Science as a-few students of M.Tech. & Ph.D. is registered under him for research guidance. Apart from regular Teaching and Administrative work, Dr. Keswani is associated with some companies and providing Consultancy on 'Project Planning and Designing' to the team of Software Professionals. To his credit there are a number of research publications which have appeared in leading Journals; some of which have been presented in International/National Conferences and included in Conference Proceedings, leading International/National level Magazines, and in-house Journals of corporate sector. Beside this, Dr. Keswani is associated as Member, Senior Member, Global Member with various International Societies (Technical) of Singapore, Canada, Australia, London, Hong Kong, Belgium and USA. For the year 2011-12, Dr. Keswani has been nominated as 'Honorary Treasurer' of the 'Computer Society of India (CSI)', Jaipur Chapter. Also, Dr. Keswani is working as Reviewer for various International Journal of Computer Application in Engineering, Technology and Science (IJ-CA-ETS) and as a member of the Editorial Board of International Journal of Computer Applications in Engineering, Technology and Sciences (IJ-CA-ETS). As a prolific writer in the arena of Computer Sciences and Information Technology, he penned down a number of books on Operating System (ISBN No: 13/978-81-8496-324-3), E-Banking and Security Transactions (ISBN No: 978-93-80311-07-4), Software Engineering (ISBN No: 81-88870-88-9) and (ISBN No: 13/978-85-8496-312-0), Elementary Computer Applications (ISBN No: 13/978-81-8496-061-7), Introduction to Computer Science (ISBN No: 13/978-81-8496-264-2) and Programming in Visual Basic. Dr. Keswani has authored 'Self Learning Material' of Computer Science for Virdhaman



Dr. Dinesh Goyal has recently joined Poornima Group as Principal & Director, Poornima Institute of Engineering & Technology. He is a renowned academician and a very good researcher. As a leader, he motivates and grows with his team to achieve the desired results. Dr. Goyal completed his Bachelor's degree in Computer Science & Engineering in year 2000 from MBM College of Engineering, Jodhpur. He joined Rajasthan Technical University to complete his M.Tech. in 2012 with research work on 'Information Security

Activities and Societies' and awarded the degree of Ph.D. in 2014 with specialization in Computer Science & Engineering from Suresh Gyan Vihar University, Jaipur.

Dr. Dinesh Goyal, Professor of Research & Academic Experience at Suresh Gyan Vihar University, Jaipur with last 15 year has retired from the post of Director, Centre for Cloud Infrastructure and Security. He has expertise in Information Security, Image Processing and Cloud Computing and has written more than 60 International & National Papers of Good quality. He has guided more than 20 PG Research Scholars and more than 06 Ph.D. Research Scholars.

