Spring 5-31-2013

Detection of video frame insertion based on constraint of human visual perception

Lu Zheng
New Jersey Institute of Technology

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

DETECTION OF VIDEO FRAME INSERTION BASED ON CONSTRAINT OF HUMAN VISUAL PERCEPTION

by
Lu Zheng

Recently, due to availability of inexpensive and easily-operable multimedia tools, digital multimedia technology has experienced drastic advancements. At the same time, video forgery becomes much easier and makes more difficult to validate the video content. Consequently, the origin and integrity of video can no longer be taken for granted. A methodology is developed that is capable of detecting the video frame insertion based on the constraint of human visual perception. The main idea is based on the so-called differential sensitivity. That is, that the variation of brightness of neighboring video frames has some constraint. First, the video sequence is partitioned into short and overlapping sub-sequences. Second, the ratio of the temporal variation of brightness calculated at the beginning and the ending frames of each sub-sequence is computed and compared with a threshold to determine the approximate location of the video frame insertion. Third, a procedure is conducted to determine the exact location of the insertion. The success of simulation works on more than 200 video sequences. The precision rate of detection is about 94.09%, and the precision rate of detecting location of frame insertion is 84.88% on testing database

DETECTION OF VIDEO FRAME INSERTION BASED ON CONSTRAINT OF HUMAN VISUAL PERCEPTION

by
Lu Zheng

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Electrical Engineering

Department of Electrical and Computer Engineering

May 2013

APPROVAL PAGE

DETECTION OF VIDEO FRAME INSERTION BASED ON CONSTRAINT OF HUMAN VISUAL PERCEPTION

Lu Zheng

Dr. Yun-Qing Shi, Thesis Advisor Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Sui-Hoi Edwin Hou, Committee Member
Date
Associate Professor of Electrical and Computer Engineering, NJIT

Dr. Tan-Feng Sun, Committee Member Date
Associate Professor of Electronic Information and Electrical Engineering, SJTU

BIOGRAPHICAL SKETCH

Author: Lu Zheng

Degree: Master of Science

Date: May 2013

Undergraduate and Graduate Education:

- Master of Science in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2013

- Bachelor of Science in Electronic Information Engineering,
  Zhengzhou University, Zhengzhou, Henan, P. R. China, 2011

Major: Electrical Engineering

To my beloved family, teachers and friends

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my thesis advisor, Dr.Yun-Qing Shi, who gave me the chance to join his research group. He guided the road and helped me start on the path on digital multimedia processing. He gave me constant support and encouragement throughout the whole year of my study. I am also grateful to Dr. Tan-Feng Sun for assistance and suggestions throughout my thesis, who was always available for my questions and always knew where to look for the answers to them. Special thanks are given to Dr. Sui-Hoi Edwin Hou for participating in my committee. Thank you so much again for your kindness, patience and professionalism. I also appreciate the support and help from Da-Wen Xu and Jing-Yu Ye in Dr. Shi research group.

I also want to thank to all my friends, especially to Li Li and Yang Zhang, for always listening and giving me words of encouragement and not letting me give up. To all my beloved family, all my special thanks are for their endless love and belief in me.

Most of all, I am fully indebted to Dr. Misra, my graduate advisor, for guiding me to reach my dream step by step.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

Recently, due to availability of inexpensive and easily-operable multimedia devices, digital multimedia technology has experienced enormous improvement, video forgery technology becomes much more sophisticated and makes more difficult to validate the multimedia content. As a consequence, the origin and integrity of the multimedia can no longer be taken for granted. With these reasons, video forensics is becoming increasing important, especially when the digital video content is used for legal support. For example, digital video evidence is an integral part of the investigation on the judicial process. If some important details are erased from the recorded video without any perception, many accused may change their pleas from ‘‘guilty’’ to ‘‘not guilty’’

1.1 Video Forgery

Video forgery is a technique that alternates or damages the type or the content of a video for some purpose so that the origin and integrity are lost. Usually, comparing the original video with the new one, these changes are subtle and unnoticeable to the naked eye. Basically, there are two types of video forgery as follows:

(1) Frame-based video forgery

Frame-based video forgery technology focuses on alternating the whole frame in videos. It can be divided into two types: video frame insertion and video frame deletion. As for video frame insertion, it means to duplicate a sequence of frames from one video and insert them to another part of the same video or a different video. To detect this kind of

video forgery in the same video, Wang and Farid [1] proposed a technique based on correlation coefficient to detect duplication. These will be discussed in Chapter 2 in details.

As for video frame deletion, it means to delete a sequence of frames from the original video. As known, most of the video signals are represented from 24 to 30 frames/s. That is, that naked eye isn't sensitive enough when video content is tempered by only a few frames, like one or three frames. And it is impossible to reach the purpose of erasing a certain object in the videos by only deleting a single frame. Thus, the common video deletion forgery is to delete a sequence of frame and utilize some digital processing tools or algorithms to make more naturally.

(2) Content-based video forgery

Unlike the frame based video forgery, it aims to alternate or erase some parts of the frame. It can also be called copy-move video forgery with two different types: intra-frame forgery and inter-frame forgery [2]. An intra-frame region copy-move forgery is defined as duplicating some parts of the frame and copy to other places in the same frame in order to cover some original objects. Inter-frame copy-move forgery is to duplicate part of frames in other frames and paste in different frames. Wang and Farid [1], as we mentioned before, also introduce an effective algorithm for detection the copy-move forgery in videos.

1.2 Video Forensics

As video forgery technology has experienced enormous improvement, video forensics plays a crucial role in digital video processing. Although a numerous of researcher have proposed many effective methodologies and solutions on digital forensics, most of the researchers are aiming to analysis images instead of video. It seems that these image processing techniques can be applied to video content. In spite of the fact that digital video

is made of a sequence of frames which can be treated as an image, video forensics is much more complex because of its unique characteristic. That means the type of the encoding of digital video is different and the degree of the lossy compression is significant as a result of much higher level of the correlation or similarity between neighboring frames [2, 3].

As known, digital video consists of a sequence of frame. That is, when taking videos, images are taken first. The processes of taking video usually go through these steps: distorted by optical lenses, merged by RGB Color Filter Array, storing pixel values on the internal CCD/CMOS array, processed by the in-camera software and encoding the frames by using MPEG-x or H.26x codes or 3GP codes [2, 4]. As for the physical mechanism and digital image processing in these steps, there will be more or less footprints in this process which have some certain or unique characteristics. Thus, these characteristics can be utilized to find the type of device or algorithm used for the test video. Also, it can be used for detection of reproduction of videos [4, 5]. While, as for the copy-move video detection, usually, the way to solve this kind of problem is to analysis the suspect or useful information in the frame itself instead of focusing on the characteristics of the device or the encoding format [2].

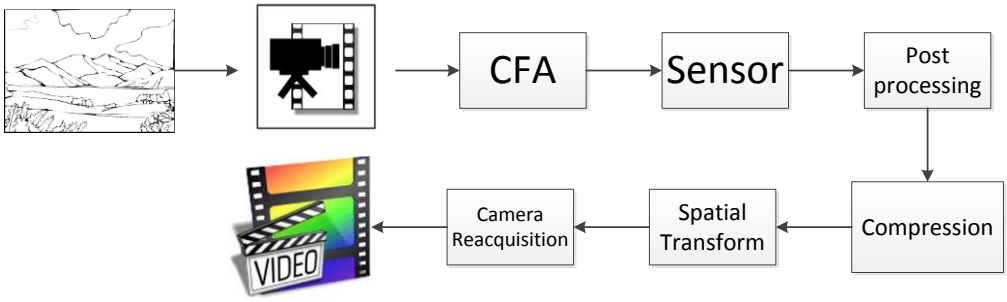


Figure 1.1 Video acquisition pipelines.

Currently, there are two types of forgery or tampering detection: active detection or watermarking and passive detection or blind detection. As for the active detection, the owner of video or image embedded digital watermarking into the publishing video or image in order to identify the ownership of the copyright. The embedded message often stands for the identity information of the owner, the video or image version or some specific character [4]. However, there are several drawbacks using watermarking:

(1) The protected videos or image should be pre-embedded before published.

(2) The robustness of the watermarking is a crucial factor of the quality of the watermarking. It is fragile when facing to all kinds of attacks. Once the watermarking is damaged, the function of protection will be lost immediately.

(3) As for the imperceptible watermarking, it is impossible to detect the watermarking if the source information of watermarking is lost.

As for passive detection, also called blind detection, there is no need to embed watermarking into the protected files. On the contrary, the intrinsic feature of the video itself can be used for forgery detection. As for the digital forgery video, some techniques about digital image forgery detection can be applied to the video forgery detection in separate frame. However, digital video not only contains a large amount of information in spatial domain, but also are full of features in the temporal domain on which is different from digital image processing.

When it comes to malicious tampering for video in the temporal domain, such as deleting or inserting frame and deleting or inserting a short video sequence, the image forgery detection technology cannot be applied to these fields. One of the most important works for the video forgery passive detection is the video feature selection and extraction.

Under the influence hardware and software factors of digital multimedia device, the scene and the content captured by video, the output information usually contains some special statistical feature which can be extracted to compare so as to detect whether the video has been tampered or not. Due to the diversity and complexity of the forgery techniques, more technical analysis is needed. And the limited accessibility of the original video content will lead to the passive forgery video detection more attractive and prevalent.

CHAPTER 2

RELATED WORK

Since videos cannot always be recorded with watermarking, passive forgery video detection is more attractive and prevalent. In recent years, more and more approaches have been developed for passive forgery video detection. Kurosawa et al. [6] proposed using the non-uniformity of the dark current of CCD chips for camcorder identification. Hsu et al. [7] proposed a technique for locating forged regions in a video using correlation of noise residue at block level. In [8], Mondaini et al. proposed a related technique based on sensor pattern noise [9].

With the detection of the frame insertion or frame duplication, Wang and Farid [1] developed a method to detect frame duplication based on the correlation coefficient. Chao et al [10] introduced an approach based on optical flow. And these two methods will be discussed in details in this chapter.

2.1 Exposing Digital Forgeries in Video By Detecting Duplication

As for the frame duplication in a video sequence, first, partition the whole video sequence into several overlapping sub-sequences. Then, use the temporal and spatial correlations to each overlapping sub-sequence. As for the temporal correlations, compare correlation coefficient to calculate the variation across the sub-sequence. That is, if the value of the correlation coefficient is close to 1, it means there is little variation. If the value is near to -1, it means the variation between these two sub-sequences is big [1].

When it comes to the spatial correlations, the spatial correlation of frames is computed in each sub-sequence. At first, the frame is partitioned into m non-overlapping

blocks. After that, the correlation coefficient is computed between two blocks and compared with the specific threshold in order to detect duplicated frames in the test video.

```
FRAMEDUP(f(x, y, t))
 1  ▷ f(x, y, t): video sequence of length N
 2
 3  ▷ n: sub-sequence length
 4  ▷ γ_m: minimum temporal correlation threshold
 5  ▷ γ_t: temporal correlation threshold
 6  ▷ γ_s: spatial correlation threshold
 7
 8  for τ = 1 : N − (n − 1)
 9       do S_τ = {f(x, y, t + τ) | t ∈ [0, n − 1]}
10          build T_τ ▷ temporal correlation
11
12  for τ_1 = 1 : N − (2n − 1)
13       do for τ_2 = τ_1 + n : N − (n − 1)
14               do if (min(T_τ1) > γ_m & C(T_τ1, T_τ2) > γ_t)
15                  build B_τ1,k ▷ spatial correlation
16                  build B_τ2,k ▷ spatial correlation
17                  if ( C(B_τ1,k, B_τ2,k) > γ_s ) ▷ ∀k
18                       do ▷ Frame Duplication at τ_1
```

Figure 2.1 Pseudo-code for detecting frame duplication Source: [1]

2.1.1 Correlation Coefficient

Correlation Coefficient, shorted for Pearson Product-moment Correlation Coefficient (PPMC or PPC), is widely used in the digital image processing especially image comparison, such as image registration, object recognition and disparity measurement [11]. PMCC is a way to evaluate the correlation of the two variables. The equation of the PMCC is defined as the follows [11]:

$$r = \frac{\sum_i (x_i - x_m)(y_i - y_m)}{\sqrt{\sum_i (x_i - x_m)^2}\sqrt{(y_i - y_m)^2}} \tag{2.1}$$

In the Equation (2.1), x and y are the two variables, x_m and y_m are the mean value of the x and y separately. When applying to the image processing, x and y are referred to the

intensity of the i^{th} and y^{th} pixel in the two compared images. And x_m and y_m refer to the corresponding mean intensity of the each image. The value of r which refers to the correlation coefficient ranges from -1 and 1. If r reaches to -1, it means that these two images are entirely different or anti-correlated [11]. If the value is close to 1, it means these two images are nearly the same.

Although PMCC can effectively decrease the calculation difficulties of image comparison by reducing the two-dimensional images into a single scalar, some drawbacks and limitations about PMCC need to be taken into account. For example, computationally intensive, that is, that PMCC is too sensitive to the image skewing (as shown in Figure 2.2), pincushioning (as shown in Figure 2.3) and vignetting (as shown in Figure 2.4) [11]. Another problem is that when it comes to compare the difference these two images, which the intensity of all the different pixels are the mean intensity of the image, it is impossible to detect the difference by using PMCC method. In addition, the over-complexity of the test image will effect on the constancy of the results, such as too much fine detail and over-sensitivity to pixel noise [11].

Figure 2.2 Image skewing.
Source: Image (left) http://www.ee.columbia.edu/~eleft/e6830-Spring96/samples/images/lenna.gif accessed April 3, 2013

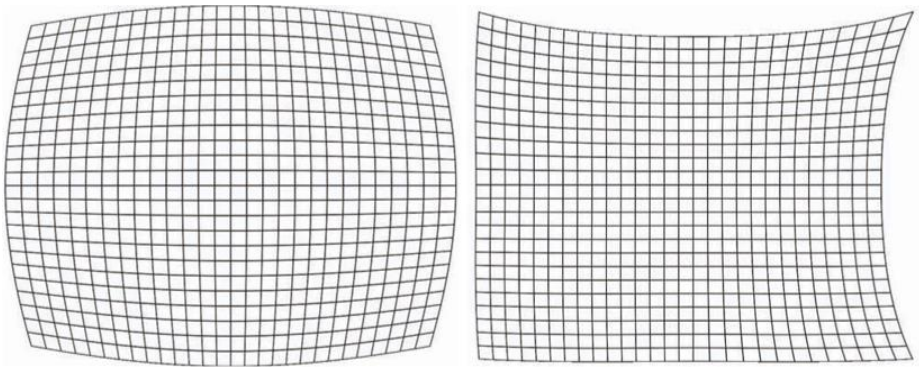


Figure 2.3 Image pincushioning.
Source: http://ieeexplore.ieee.org/ieee_pilot/articles/06/ttg2009061307/figures.html accessed April 3, 2013

Figure 2.4 Image vignetting.
Source: Image (left) http://www.ee.columbia.edu/~eleft/e6830-Spring96/samples/images/lenna.gif accessed April 3, 2013

2.1.2 Result

As for the uncompressed video, nearly 84.2% of the duplicated frames are detected taken from a stationary camera and the false positive is 0.03. When it comes to using the video taken from a moving camera, the result of detection becomes 100% with 0 false positives.

As for the compressed video, this method compares the result by using the MPEG compression with a bit rate of 3, 6 or 9 Mbps. The results are as follows:

Table 2.1 Result of Detecting Frame Duplication

	detection			false positive		
video	3	6	9	3	6	9
stationary	87.9%	84.8%	84.4%	0.06	0.0	0.0
moving	86.8%	99.0%	100.0%	0.0	0.0	0.0

Source: [1]

2.2 Inter-Frame Forgery Model Detection Scheme Based on Optical Flow Consistency

Chao et al. [10] proposed a passive detection methodology for the Inter-frame forgery for video content based on optical flow consistency. The optical flow consistency will change if the given video has been tampered by frame insertion or deletion.

Lucas and Kanade [12] proposed an effective image registration technique named Optical flow (The Lucas Kanade optical flow) in 1981 which had been widely used for many kinds of digital image processing, such as layered motion, mosaic construction and face coding [10, 13]. Optic flow (as shown in Figure 2.5) is the apparent visual motion that you experience as you move through the world and it is based on the spatial intensity gradient information [14]. In Chao's paper, they make use of optical flow into video processing.

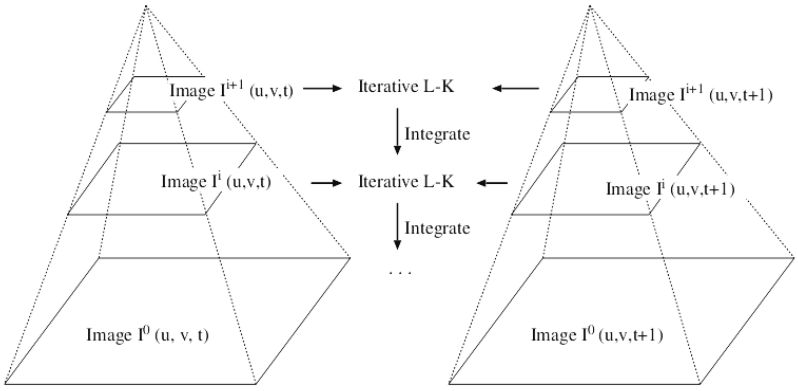


Figure 2.5 Procedure for computing Optical Flow by the Lucas–Kanade method with pyramids. Source: [14]

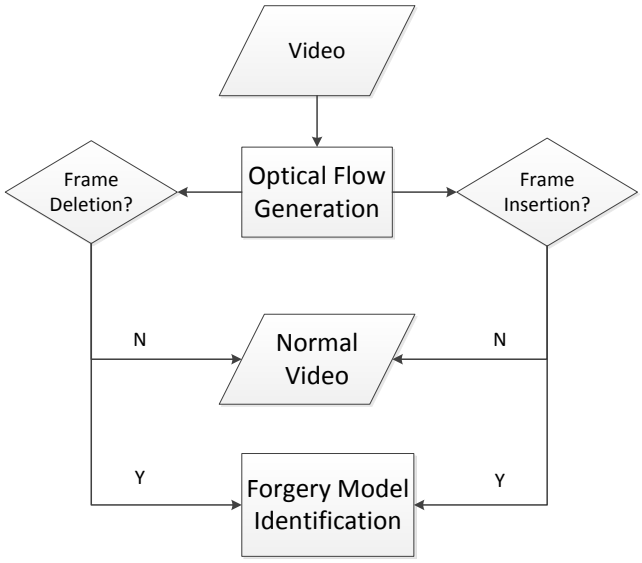


Figure 2.6 Procedure for inter-frame forgery model detection.

With the insertion forgery, this method uses a window as rough detection in to initially validate whether the test video has been tampered or not. If the test video is suspected to be tampered, the binary searching scheme will be used in further detection.

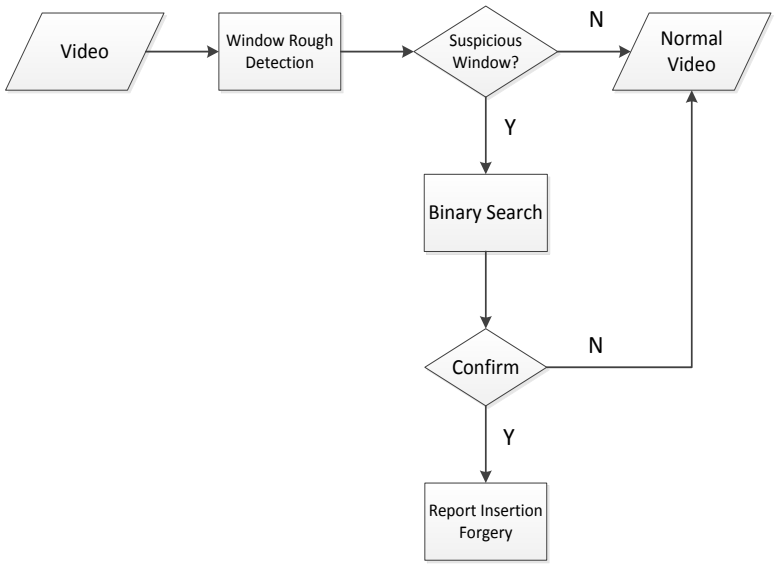


Figure 2.7 Procedure for frame insertion forgery detection.

As for the deletion model, the optical flow and double adaptive threshold is applied to detect forgery [14]. Because the difference of optical flow in frame deletion video is much smaller than frame insertion video, Chao et al. compute the optical flow between all the adjacent frames.

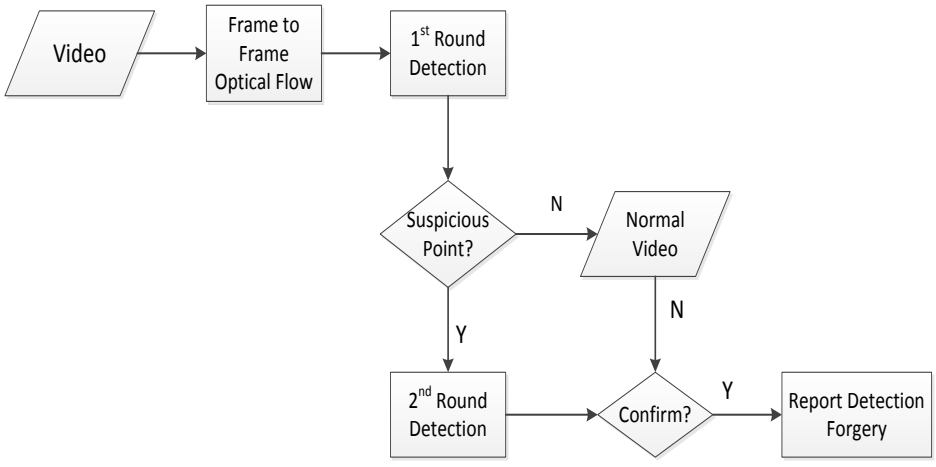


Figure 2.8 Procedure for frame deletion forgery detection.

In this experiment, 1,3000 frame insertion videos are tested. The frame insertion detection recall rate reaches 95.43% and the precision rate reaches 95.34%. As for the frame deletion, the result is a little worse than frame insertion which the recall rate and precision rate are both lower than 90% [10].

CHAPTER 3

SHORT-TEMPORAL VARIATION OF THE BRIGHTNESS

In this work, an effective method for detecting video frame insertion is proposed. It is based on the short-temporal variation of the brightness in video sequence. The main idea is that the consistency of the ratio of the short-temporal variation of the brightness (STVB) in equal time intervals will be disturbed in frame insertion video. This method can not only detect whether the video is tampered or not, but also can detect the location of the frame insertion.

3.1 Human Visual Perception

In Victorian times, when the video camera was not invented, there was a popular toy named thaumatrope as the Figure 3.1 shown. A card with different pictures on each side is attached to two pieces of string, just like in the Figure 3.1. When holding the two strings and spinning them quickly, a new picture which merges two pictures on each side and the bird appeared to be in the cage because of the persistence of vision phenomenon [15, 16].

The human brain can retain an image for a fraction of a second longer than the eye actually sees it. This phenomenon is called the persistence of vision effect. It means, when seeing a fast moving object, the human eye can continue to retain the image for about 0.1 to 0.4 second after the disappearance of the objects [16, 17]. As known, video is made of a sequence of individual still frames. With the effect of the persistence of vision, the human visual can process 10 to 12 separate images per second, perceiving them individually [17]. In addition, Due to these perception thresholds, the frame rate of the video is usually set as 24 fps for common film, 25 fps for PAL television standard and 30 fps for NTSC video

standard in order to eliminate the feeling of discontinuity and the blink between two frames.

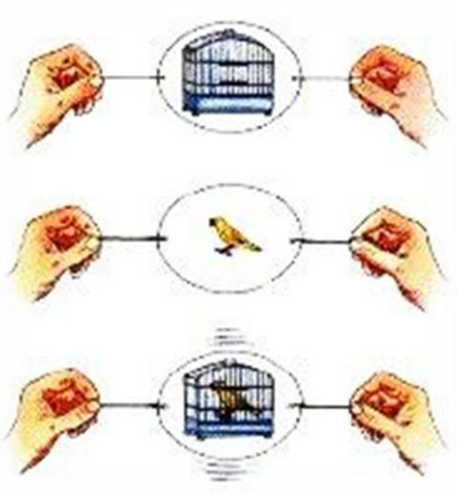


Figure 3.1 Application of persistence of vision phenomenon. Source: [15]

3.2 Weber's Law

Before discussing the new theory, a famous theory named Weber's Law is needed to be mentioned ahead.

Weber's Law describes the perceptual difference between increment stimulus and original stimulus. It can be utilized for detecting the changes of weights, brightness or length. For example, when you lift a thing with a weight of 500 grams, suppose that the weight of 50 grams can only just be distinguished from that of 500 grams. It is impossible for you to notice the additional weights after adding 50 grams to a weight of 5 kilograms thing. And for this time, only adding 500 grams can one notice the variance. The equation of the Weber's Law is as follows:

$$\mathrm{K} = \frac{\Delta I}{I} \tag{3.1}$$

where I is the original stimulus and ΔI is variation of I. K is the so-called Weber's rate. Weber's Law predicts a linear relationship between the increment threshold and the background intensity [18].

Weber's law describes a constant ratio between original stimulus and the changed stimulus under the premise that other factors have no obviously variation. When it is applied to explain the video sequence, it can be described like that, the video is made of a sequence of individual still frames with coherence in some degree, at least, this kind of consistency can pull the wool over the human naked eye without noticing the inconsistency between frames. With each frame, it can be treated as a stimulus in the Weber's Law. As for any two frames with the same time intervals, the ratio the variation between these two frames is approximately equal. Moreover, all the assumptions above are based on the fact that this kind of video is taken by a stationary camera in a consistent scene without sudden changing.

3.3 Short-Temporal Variation of Brightness

As discussed above, a new feature called Short-Temporal Variation of the Brightness (STVB) in consistent video sequence is proposed. As known, digital video not only contains a large amount of information in spatial domain, but also is full of features in the temporal domain. In the temporal domain, the correlation of the every two adjacent frames is very high. That is, that the corresponding variation is very low. As mentioned before, because of the persistence of vision phenomenon on human eye, the image can be retained in the human visual system for about 0.1 to 0.4 seconds after its disappearance [17]. In one video, if two frames with a 0.4 seconds time interval has a very low correlation; it seems to be weird when people see it, just like seeing separate still images. So, two frames with

some short time intervals such as 0.4 seconds must have little variation in order to guarantee the consistency of the video content. As for the frame rate with 24 fps, the minimum value of time interval is as follows:

$$24 \text{ fps} \times 0.4 \text{ seconds} = 9.6 \text{ frames} \approx 10 \text{ frames} \tag{3.2}$$

which means every two frames with an interval of approximately 10 frames, they still have the correlation in some degree. On the other hand, it means that the variation of every two frames with an interval of 10 frames has a very low variation without obvious perception by the human naked eye. This variation refers to the brightness or the gray value of each pixel in the frame. Similar to Weber's Law, the ratio of the variation can be defined as the follows:

$$\mathrm{R} = \frac{\Delta \mathrm{B}}{\mathrm{B}} \tag{3.3}$$

where B is the brightness or the gray value of each pixel in one frame and ΔB is the variation value. R is referred to the ratio of the STVB of pixels between two video frames at the same position.

Thus, as for an original normal video, the ratio of STVB between two frames with a certain time interval is usually a constant. However, this consistency will be disturbed in the frame insertion video.

3.4 3σ Rule

Now, as known above, the ratio of STVB will be disturbed by the frame insertion video. However, how to judge or validate whether the video has been tampered or not is needed to be solved. With further study, 3σ Rule is found to solve this problem.

3σ Rule is the common criteria of gross error detection. Its basic principle is random error subordinated to the normal distribution, then the absolute value of error mainly concentrated in the vicinity of the its medium [18]. It can be expressed as follows:

$$P(-3\sigma < z - \mu < 3\sigma) = 0.9974 \tag{3.4}$$

where z~N (μ, σ^2), σ is the standard deviation of z. The Equation (3.4) implies that data can be treated as the gross error that is more than3σ.

When applying the 3σ Rule to our method, if the ratio of STVB is more than 3σ, where σ refers to the standard deviation of a sequence of ratio of STVB derived from every two frames with the equal time interval in a video, this value can be treated as a gross error. That means these two frames are no longer subordinated to consistency of those frames with equal time intervals. Thus, between these two frames, those frames must contain the insertion frame and original frame. In this way, the approximate location of the insertion frame will be initially found.

CHAPTER 4

DETECTION OF VIDEO FRAME INSERTION BASED ON CONSTRAINT OF HUMAN VISUAL PERCEPTION

As mentioned above, this method only focuses on the meaningful frame insertion video. And it is based on several assumptions as the follows:

(1) Each test video sequence is one shot video sequence taken by the stationary video camera.

(2) Each forgery video has only one type forgery: frame insertion forgery.

(3) In frame insertion video, the number of the insertion frame is more than 10.

(4) Each frame insertion video only has one time frame insertion.

Based on the above assumptions, our proposal is described in details as follows:

Given a test video, parse it into a continuous sequence of frames. As for each frame, partition it into 4×4 blocks, denoted as $B= \{b_1, b_2, b_3 \ldots b_i \ldots b_{16}\}$. Then, partition the full-length video sequence into short overlapping sub-sequences group with the length of 15, $G= \{g_1, g_2, g_3 \ldots g_j \ldots g_n\}$, which n ($1<j<n$) is defined as the total number of the sub-sequence group. The way to partition in details is as follows:

(1) 1^{st} sub-sequence group: 1^{st} frame to 15^{th} frame;

(2) 2^{nd} sub-sequence group: 11^{th} frame to 25^{th} frame

(3) 3^{rd} sub-sequence group: 21^{st} frame to 35^{th} frame

……

(j) j^{th} sub-sequence group: w^{th} frame to $(w+15)^{th}$ frame, where $w= [(j-1) *10+1]$.

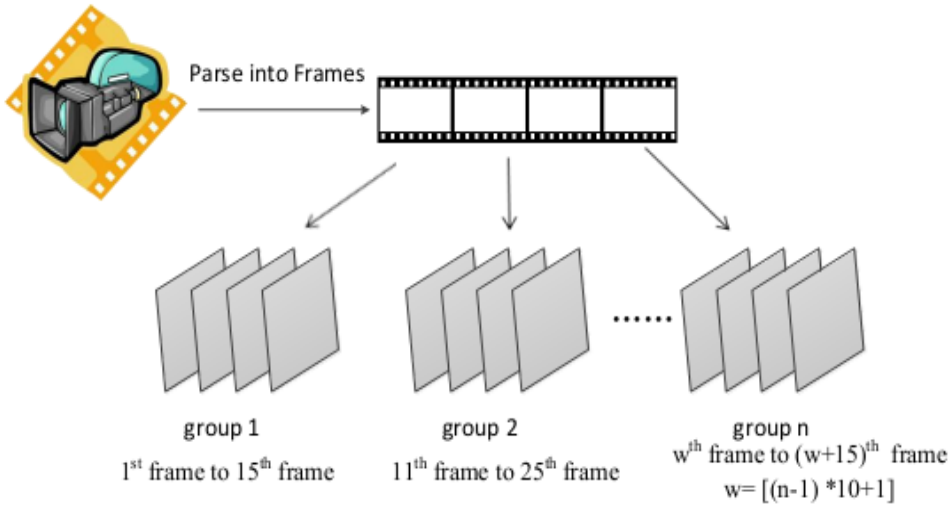


Figure 4.1 Partition whole video sequence into short overlapping sub-sequence.

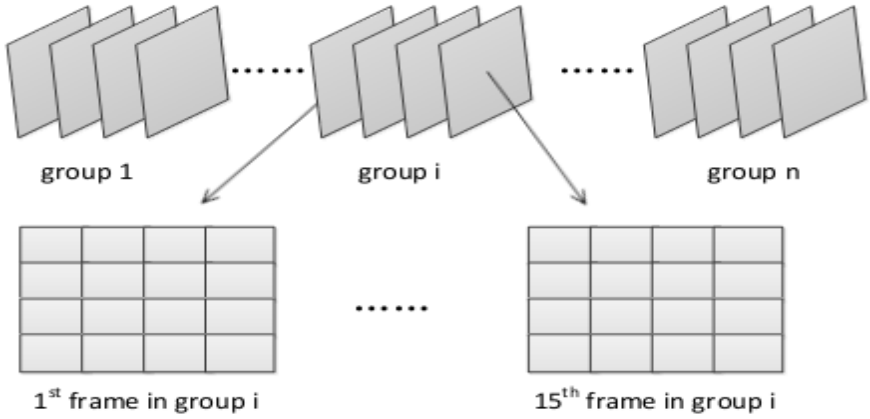


Figure 4.2 Partition each frame into blocks of 4×4.

As for the each sub-sequence group, take the j^{th} sub-sequence group as an example, extract the first frame (w^{th} frame) and the last frame ($(w+15)^{th}$ frame) in the current sub-sequence group. According to the Equation (3.3), compute the ratio of STVB between the each block in the first frame and the corresponding block in the last frame in each

sub-sequence group. Then extract the average value of the all the ration of STVB in the current sub-sequence.

$$R_b = \frac{\Delta p}{P_{ave}} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{Pf_{ij} - Pl_{ij}}{P_{ave}} \tag{4.1}$$

where, R_b represents the ratio of STVB between the each block in the first frame and the corresponding block in the last frame in each sub-sequence group. Δp is defined as the variation of the gray value of pixels in the corresponding blocks. Pf_{ij} and Pl_{ij} represent the gray value in each pixel of the current block in the first frame and the last frame of the current sub-sequence group respectively. M and N represent the number of the pixel in row and column in each block. P_{ave} is the average gray value of pixels in the current block of the first frame in each sub-sequence group and the equation of the P_{ave} can be defined as follows:

$$P_{ave} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} Pf_{ij} \tag{4.2}$$

As the R_b of each block in the corresponding two frames has been calculated in each group, calculate the average of these series $R_{b,}$ value defined as R_f, in each sub-sequence:

$$R_f = \frac{1}{16} \sum_{i=1}^{16} R_{bi} \tag{4.3}$$

In this way, a series R_f is calculated in the whole-length video sequence as the shown in the Figure 4.4. Among a series value of R_f, there are two obvious peak points. As mentioned before, the ratio of STVB is near to a constant in a normal video and this consistency will be disturbed in frame insertion video. As for these two peak points in

Figure 4.4, the corresponding sub-sequences to these two peak points must contain the original frame and the insertion frame. Although it is easy to determine the peak point from the figure, how to set a threshold to find it is still needed. With further study, 3σ Rule is found to solve this problem.

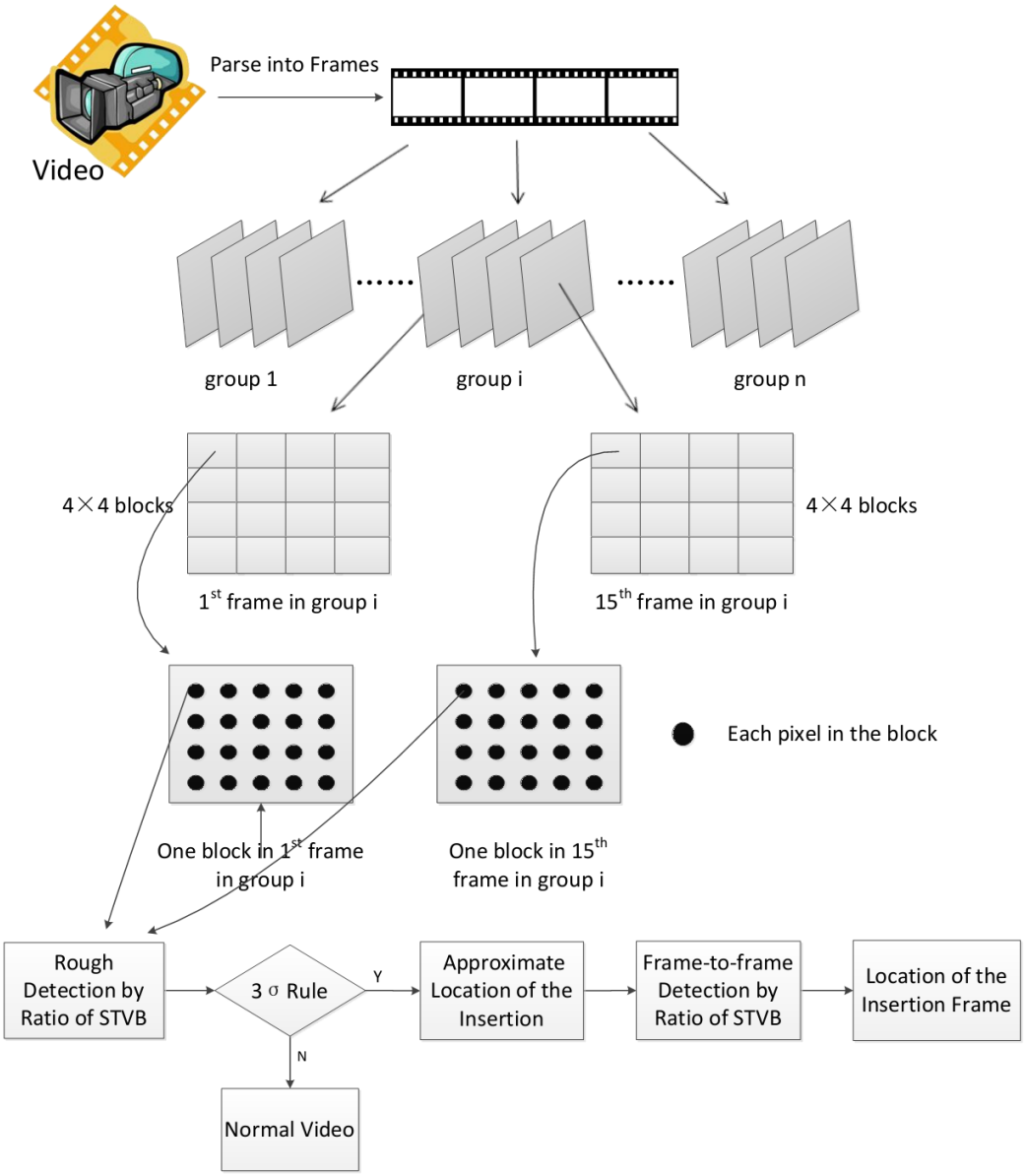


Figure 4.3 Procedure for the detection of video frame insertion based on constraint of human visual perception.

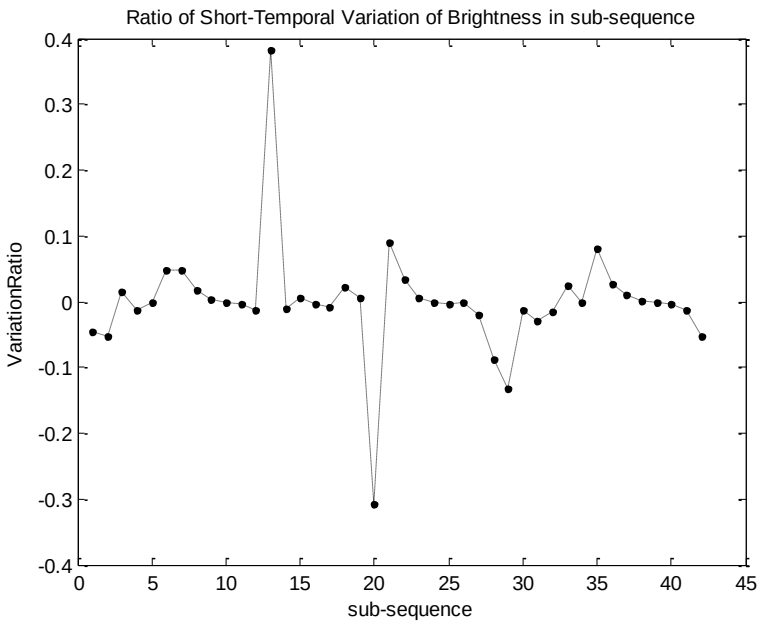


Figure 4.4 The ratio of STVB in each sub-sequence.

Before using the 3σ Rule to the R_f, to validate the whether it subordinated to Normal distribution is needed. The results are as the follows:

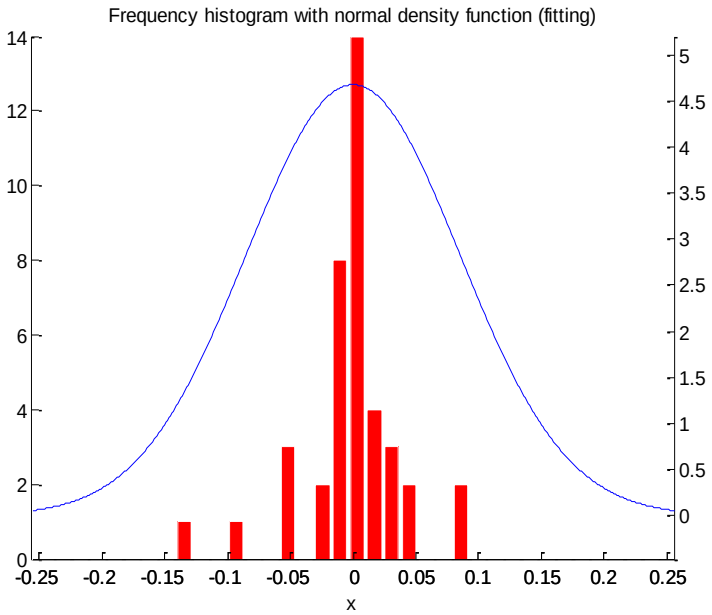


Figure 4.5 The frequency histogram of the ratio of STVB.

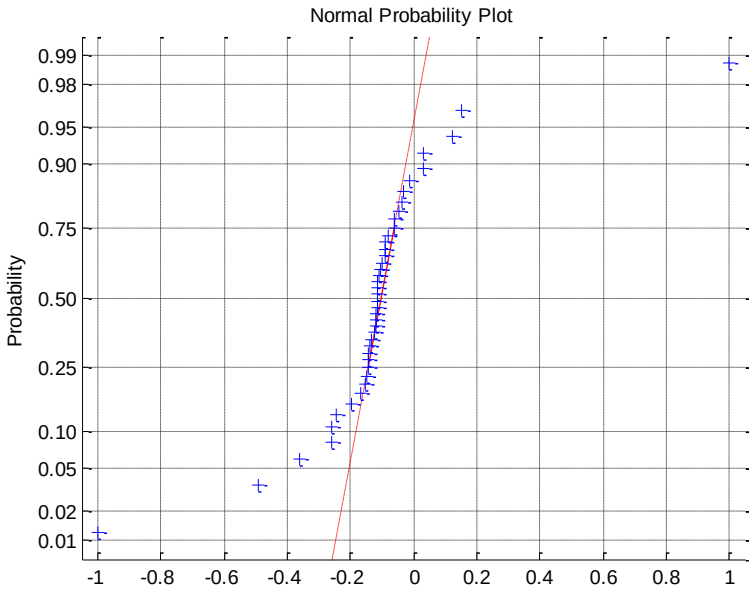


Figure 4.6 The Normal Probability of the ratio of STVB.

From the Figure 4.5, it shows that the series of R_f is subordinated to normal distribution. And the 3σ Rule can be applied to the RTVB sequence. From the Figure 4.6, it is obvious that most R_f are close to 0, only few values have large numerical deviation which can be treated as the gross error. Usually, the number of the gross error is two which imply that the corresponding sub-sequence frames of these two values contain both insertion frames and original frames. And 3σ Rule can be applied to the RTVB sequence.

After validation, calculate the standard deviation of series $R_f = \{R_{f1}, R_{f2}, R_{f3} \ldots \ldots R_{fn}\}$ in the in the whole-length video sequence:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} R_{fi} \tag{4.4}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(R_{fi} - u)^2} \tag{4.5}$$

where μ is the mean value of the R_f, σ is the standard deviation of series R_f.

According to the 3σ Rule, if the value of R_f is more than 3σ, this value can be treated as a gross error. If none of R_f is more than 3σ, the test video is assumed as a normal video. If the number of the gross error is more than two, the test video is frame insertion video. As for the frame insertion video, the ratio of STVB at the beginning and the end of the insertion frame has a greater volatility than other normal frames. Just like the Figure 4.4 showing, two distinct peak points occur in the whole series of the R_f. Thus, it is easy to conclude that the corresponding sub-sequence group two these two peak points must contain the insertion frame and original frame.

Since the location of two maximum peak points are found, it is easily to figure out the corresponding sub-sequences. As known, each sub-sequence has 15 frames, which contains the normal frames and insertion frames. That means, more algorithms are needed to find the accurate location of the frame insertion.

So, the ratio of STVB frame to frame is computed to find the accurate position. First, select the first frame, fi, as a reference position in the corresponding sub-sequence to the peak point. Then, select 40 adjacent frames of which 20 frames is ahead of fi and the other 20 frames is behind. After that, calculate the ration of STVB of each two adjacent frames by utilizing the Equation (4.1) and Equation (4.3). In this way, two new series of ratio of STVB frame to frame are calculated, each of which is corresponding to a peak point. The results are shown in Figure 4.7.

From the Figure 4.7, all the values except one are nearly equal to 0. As known, the variation between two adjacent frames in non-tampered frame sequence is extremely low which implies that the ratio of STVB is near to 0. Only two corresponding frames are from

different video frame subsequences can lead to the sudden volatility. Thus, select the maximum value of each result and find the corresponding frames to this result. The location of frame is accurate location of the frame insertion.

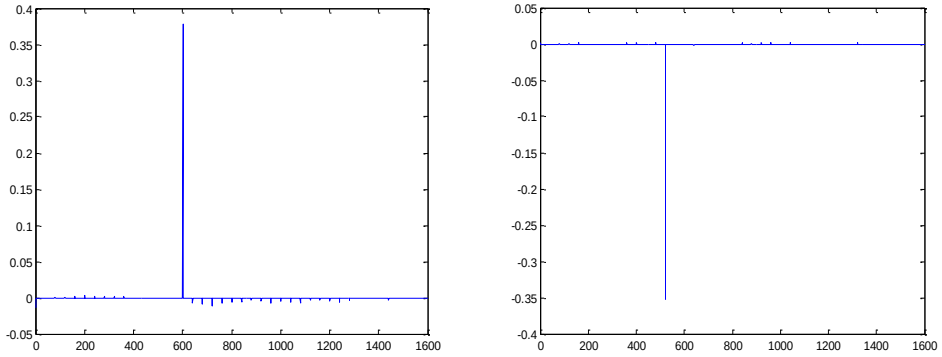


Figure 4.7 Results of the ratio of STVB calculated by frame to frame.

CHAPTER 5

EXPERIMENTAL RESULTS AND CONCLUSION

5.1 Test Video Database

The original videos are from the Recognition of Human Actions Database [19]. The video database, shown in Figure 5.1, contains six types of human actions (walking, jogging, running, boxing, hand waving and hand clapping) in four different scenarios: outdoors *s1*, outdoors with scale variation *s2*, outdoors with different clothes *s3* and indoors *s4* as illustrated below. All sequences were taken over homogeneous backgrounds with a static camera with *25*fps frame rate. The format of all the video sequence is AVI file format [19].

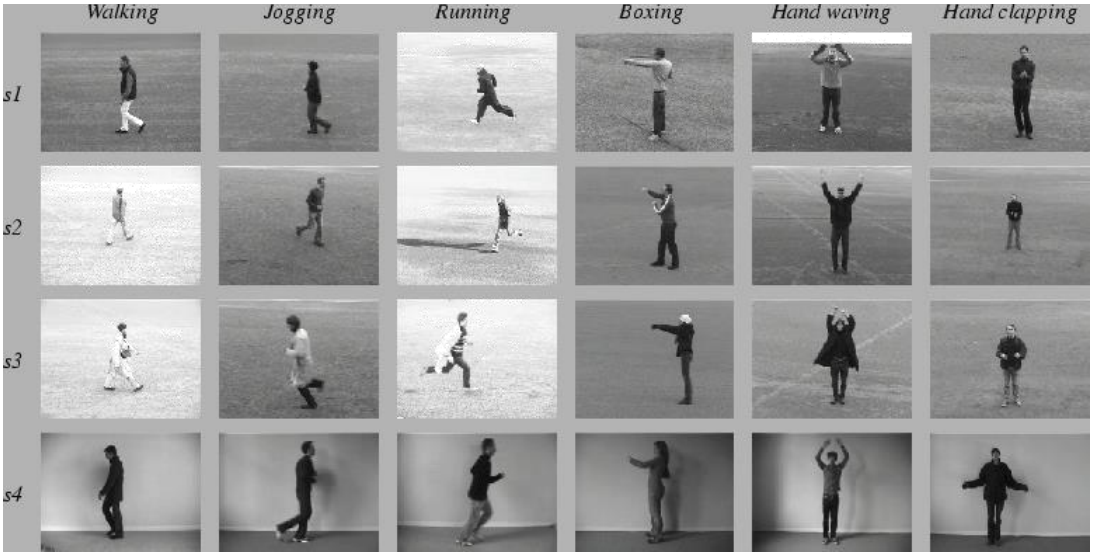


Figure 5.1 Samples of the original videos resource. Source: [19]

The test video database is generated with TRECVID Content Based Copy Detection (CBCD) scripts. CBCD scripts can generate frame insertion videos

automatically with random length. In our approach, 200 frame insertion video sequences and 20 normal video sequences are selected for testing. Each frame is 240×320 pixels in size and the length of each video sequence is range from 375 frames to 625 frames (approximately 15 to 25 seconds).

5.2 Evaluation Standards

To evaluate the detection efficiency, two standards called the recall rate (R_r) and precision rate (R_p) are used. The recall rate is the proportion of correctly detected videos among all tampered videos. The precision rate refers to the percentage of correctly detected video among all the detected videos [10]. The recall rate and the precision rate are defined as follows:

$$R_r = \frac{N_c}{N_c + N_m} \times 100\% \qquad (5.1)$$

$$R_p = \frac{N_c}{N_c + N_f} \times 100\% \qquad (5.2)$$

where N_c is the number of correctly detected video forgeries; N_m is the number of missed video forgeries; N_f is the number of falsely detected video forgeries.

5.3 Results of Frame Insertion Videos

As for validation of the video frame insertion, the recall rate reaches 98.67% and the precision rate reaches 94.09% as shown in Table 5.1. From the results, it shows this algorithm can well detect the video frame insertion. The number of the missed video frame insertion is less than the number of the falsely detected video frame insertion.

Table 5.1 Test Results for Validation of Video Frame Insertion

N_c	N_m	N_f	R_r (100%)	R_p(100%)
223	3	14	98.67%	94.09%

As for detecting the location of the frame insertion, the recall rate reaches 90.62% and the precision rate reaches 84.88% as shown in Table 5.2. From the results, it shows not all forgery videos that have been detected by this algorithm can be found the accurate location of the frame insertion. Because there are some constraints on the test videos which are based on the several assumptions mentioned in Chapter 4. Thus, this algorithm has low level of robustness. That is, if video camera shakes during the process of shooting or the object moves too fast, some values of ratio of STVB will have a big variation, where they are not caused by the frame insertion.

Table 5.2 Test Results for Detecting the Location of the Frame Insertion

N_c	N_m	N_f	R_r (100%)	R_p(100%)
174	18	31	90.62%	84.88%

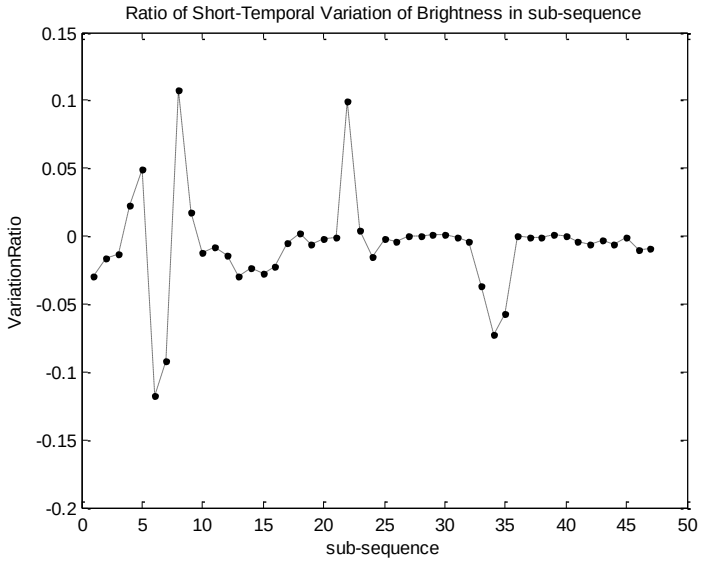
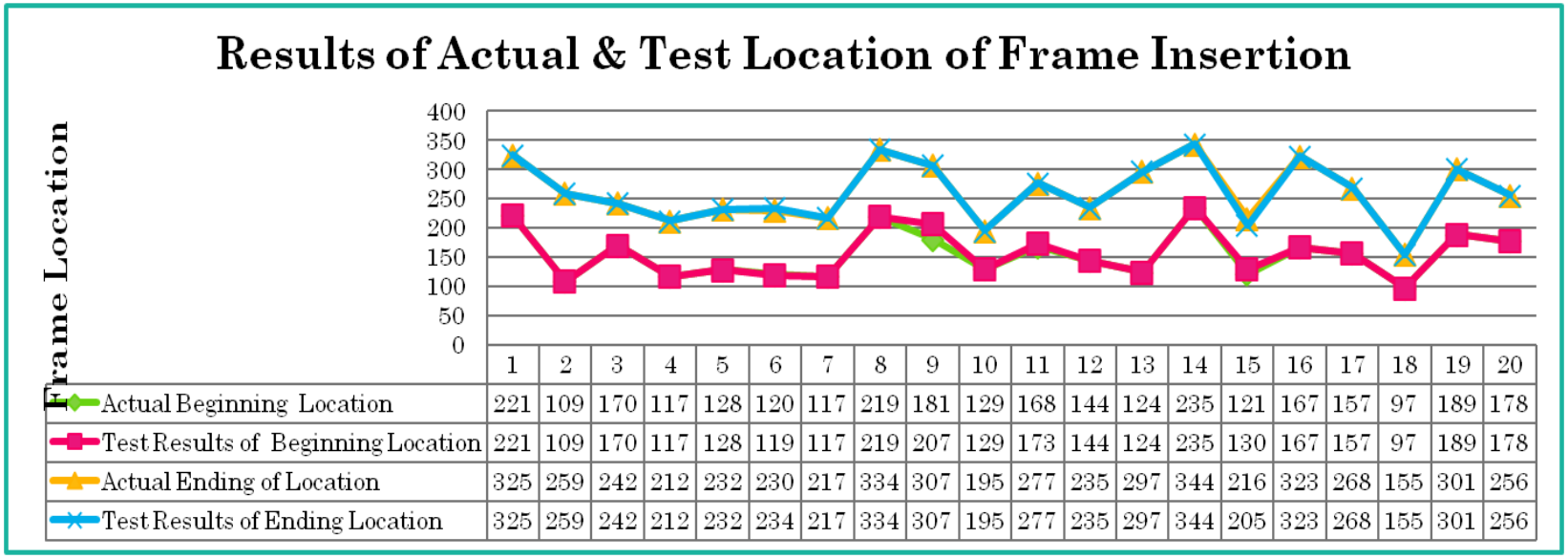


Figure 5.2 Results of the ratio of STVB in each sub-sequence.

As shown in Figure 5.2, this result is obtained from one test video of which the video camera shook during the first 4 seconds. It can be seen clearly that there several obvious peak points at the beginning of this series of ratios, R_f. In this case, the accuracy of detecting location of the frame insertion will be undermined.

In the Figure 5.3, it is the comparison results between the actual location of the frame insertion and test results for that location in 30 test frame insertion videos. From the results, most test results match the actual values and this algorithm achieves good performance in detecting the location of video frame insertion.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Actual Beginning Location	221	109	170	117	128	120	117	219	181	129	168	144	124	235	121	167	157	97	189	178
Test Results of Beginning Location	221	109	170	117	128	119	117	219	207	129	173	144	124	235	130	167	157	97	189	178
Actual Ending of Location	325	259	242	212	232	230	217	334	307	195	277	235	297	344	216	323	268	155	301	256
Test Results of Ending Location	325	259	242	212	232	234	217	334	307	195	277	235	297	344	205	323	268	155	301	256

Figure 5.3 Results of detecting the location of frame insertion in 30 test frame insertion videos.

5.4 Conclusion

In this work, a novel feature called short-temporal variation of the brightness (STVB) is proposed and an algorithm for video frame insertion detection based on this new feature is developed. The constraint of the variation of brightness of neighboring video frames will be undermined in frame insertion videos. In this work, the ratio of the STVB is calculated in each sub-sequence and compared with a threshold in order to validate the frame insertion video. Then, the ratio of STVB frame to frame is conducted to determine the exact location of the insertion. This algorithm can not only identify whether the test video is tempered by frame insertion or not, but also well determine the location of the frame insertion. Experiment shows that the recall rate of detection video frame insertion reaches 98.67% and the precision rate of it reaches 94.09%. As for the detecting the location of the frame insertion, the recall rate reaches 90.62% and the precision rate reaches 84.88%. Future work will be focus on improve quality of the robustness when detecting the location of the video frame insertion and improve its recall rate and precision rate.

REFERENCES

[1] Wang, W., & Farid, H. (2007, September). Exposing digital forgeries in video by detecting duplication. *In Proceedings of the 9th workshop on Multimedia & security (pp. 35-42).* ACM.

[2] Milani, S., Fontani, M., Bestagini, P., Barni, M., Piva, A., Tagliasacchi, M., & Tubaro, S. (2012). An overview on video forensics. *APSIPA Transactions on Signal and Information Processing, 1(1). APSIPA.*

[3] Swaminathan, A., Wu, M., & Liu, K. R. (2008). Digital image forensics via intrinsic fingerprints. *Information Forensics and Security,* IEEE Transactions on, 3(1), 101-117. IEEE.

[4] Hsu, C. C., Hung, T. Y., Lin, C. W., & Hsu, C. T. (2008, October). Video forgery detection using correlation of noise residue. *In Multimedia Signal Processing, 2008 IEEE 10th Workshop on (pp. 170-174).* IEEE.

[5] Wang, W., & Farid, H. (2008, January). Detecting re-projected video. *In Information Hiding (pp. 72-86)*. Springer Berlin, Heidelberg, Germany.

[6] Kurosawa, K., Kuroki, K., & Saitoh, N. (1999). CCD fingerprint method-identification of a video camera from videotaped images. In *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on* (Vol. 3, pp. 537-540). IEEE.

[7] Hsu, C. C., Hung, T. Y., Lin, C. W., & Hsu, C. T. (2008, October). Video forgery detection using correlation of noise residue. In *Multimedia Signal Processing, 2008 IEEE 10th Workshop on* (pp. 170-174). IEEE.

[8] Mondaini, N., Caldelli, R., Piva, A., Barni, M., & Cappellini, V. (2007, February). Detection of malevolent changes in digital video for forensic applications. In *Electronic Imaging 2007* (pp. 65050T-65050T). International Society for Optics and Photonics.

[9] Wang, W. (2009). *Digital video forensics* (Doctoral dissertation, Dartmouth College Hanover, New Hampshire, U.S.).

[10] Juan Chao, Xinghao Jiang, Tanfeng Sun. (2012). A Novel Video Inter-frame Forgery Model Detection Scheme based on Optical Flow Consistency. *11th International Workshop on Digital-forensics and Watermarking*. IWDW.

[11] Yen, K., & Johnston, R. G. (1996). The ineffectiveness of the correlation coefficient for image comparisons. *Los Alamos National Laboratory internal report*. Accessed http://jps.anl.gov/vol.2/3-Correlation.pdf April 8, 2013

[12] Lucas, B. D., & Kanade, T. (1981, April). An iterative image registration technique with an application to stereo vision. *In Proceedings of the 7th international joint conference on Artificial intelligence.*

[13] Hsu, S., Anandan, P., & Peleg, S. (1994, October). Accurate computation of optical flow by using layered motion representations. *In Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on (Vol. 1, pp. 743-746).* IEEE.

[14] Ohnishi, N., & Imiya, A. (2008). Independent component analysis of optical flow for robot navigation. *Neurocomputing,71(10), 2140-2163.*

[15] Internet: http://www.webanswers.com/misc/what-is-a-thaumatrope-bff664 accessed April 5, 2013

[16] Internet: http://bizarrelabs.com/persist.htm accessed April 5, 2013

[17] Read, P., & Meyer, M. P. (2000). Restoration of motion picture film. *Butterworth-Heinemann.* Oxford, U.K..

[18] Tang, G., & Qin, A. (2008, November). ECG de-noising based on empirical mode decomposition. *In Young Computer Scientists, 2008. ICYCS 2008. The 9th International Conference for (pp. 903-906).* IEEE.

[19] Internet: http://www.nada.kth.se/cvap/actions/ accessed April 8, 2013