12-2013

A Study on Countermeasures against Steganography: an Active Warden Approach

Qilin Qi
University of Nebraska-Lincoln, qqi2@unl.edu

A STUDY ON COUNTERMEASURES AGAINST STEGANOGRAPHY: AN ACTIVE WARDEN APPROACH

by

Qilin Qi

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Master of Science

Major: Telecommunications Engineering

Under the Supervision of Professor Dongming Peng and Professor Yaoqing Yang

Lincoln, Nebraska

December, 2013

A STUDY ON COUNTERMEASURES AGAINST STEGANOGRAPHY: AN ACTIVE WARDEN APPROACH

Qilin Qi, M.S.

University of Nebraska, 2013

Adviser: Dongming Peng and Yaoqing Yang

Digital steganography is a method used for hiding information in digital images. It can be used for secure communication. There have been many robust digital steganography methods invented in recent decades. The steganographic message can be inserted in multimedia cover signal such as audio, image and video. However, this technique also may be used by malicious users to transmit dangerous information through the Internet beyond the control of security agencies. How to detect and/or block potentially dangerous information transmission on the Internet through billions of multimedia files while not affecting innocent multimedia communications becomes a challenging problem. Existing steganalysis methods or steganography attacking methods which are mostly passive methods cannot be used for analyzing a large volume of digital images in a short time. In addition those passive methods also cannot be generic enough to defeat various steganographic algorithms on the Internet.

In this paper, we propose an active attacking model to defeat the rising threat of steganography. The active protection mechanism is proved to be more effective to protect the integrity of the multimedia data. Based on the active attacking model, a steganography attacking method which is not limited by the types of the steganography methods is proposed. The proposed method can process the digital multimedia data to remove the potential dangerous hidden information while keeping the digital data in a high visual quality. This attack method is based on a proposed transform called Discrete Spring Transform. Some implementations of the Discrete Spring Transform in audio, image and video signals are proposed. The proposed transform causes that the numerical values of the image to be changed dramatically and then the hidden information is not able to be recovered, while at the same time the visual image quality can be maintained. This method is a generic approach for multimedia signals and contains theoretical advantages over similar methods. Our experimental results have demonstrated that the quality of the multimedia signal can be guaranteed while the stego-data are considerably destroyed.

To Mingrun, my parents and grandparents.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Dr. Dongming Peng and Dr. Yaoqing Yang for the continuous support of my Master study and research, for his patience, motivation, enthusiasm, and immense knowledge.

Besides my advisors, I would like to thank the rest of my thesis committee: Dr. Hamid Sharif, Dr. Michael Hampel for their encouragement, insightful comments, and hard questions.

Last but not the least, I would like to thank my family: my parents Zengfan Qi and Hua Guo, for giving birth to me at the first place and supporting me spiritually throughout my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Digital steganography [3] is a technique used to embed and transmit hidden information in multimedia cover media in a secret way. Regarding the conventional text encryption technique, there are many kinds of mature, systematic and well-defined cryptanalysis algorithms which can be used to attack the encrypted information [4]. In the worst cases, the brute-force searching can be used to attack nearly all kinds of text encryption methods by taking advantage of state of the art super computers. In addition, though some text encryption methods are difficult to be compromised, use of the encryption itself also will invoke suspicions. However, digital steganography that hides information in image or multimedia carriers in an invisible way will draw no extra attention among billions of images over Internet [5]. Therefore, Image steganography is potential for various communication applications in order to improve communication security. The basic idea behind the digital steganography is invisible digital watermarking. Due to the redundancy of the multimedia signal, it provides a large capacity for embedding hidden information. Some common steganography methods are implemented in the spatial or frequency domain of the multimedia signal. By slightly and meticulously altering the multimedia cover signal, the intended hidden information can be embedded in the cover media. This slight numerical change will not cause noticeable attention to human auditory system for audio or human visual system for image and video. The hidden information can be hidden in a very high order statistical domain where it is extremely difficult to be detected. Therefore, digital steganographic techniques will not only carry some secret information as the conventional encryption does but also will keep the cover media perceptually unchanged. Besides the difficulty to decode the hidden information without the knowledge of the key, it is even difficult or impossible to detect whether hidden information exists or not. This advantage over the conventional encryption methods provides a second layer of protection to the hidden information.

However, it also leads to another security concern that how to prevent the illegal or malicious information transmission through the Internet. In some sensitive network scenario, the encryption is prohibited in order to prevent the uncontrolled information transmission. The steganography methods will make this problem more complicated because it is difficult to know if the steganography methods are applied in a multimedia file. By this technology, terrorists or attackers are able to transmit illegal or dangerous information with each other through the Internet out of any control. There is already some evidence that terrorists tried to use image steganography to transmit their attack plans and training manuals through Internet websites such as Ebay[4]. Nowadays, there are billions of audio, image, and video files uploading, transmitting and downloading on the Internet in a second, so this security problem has become very critical to authorities and researchers.

For the purpose of hiding information in multimedia signal, digital steganography method must have some properties. First of all, the hidden information should be imperceptible. It not only means the hidden information cannot be perceived directly by observing the carriers but also means the visual artifacts of the cover media caused by the hidden information cannot be easily recognized. Secondly, the hidden information must be able to be retrieved by the intended receiver who has the knowledge of the steganography decoding method and key. The redundancy in the cover media makes these requirements possible. Furthermore, the latest digital steganography algorithms are also designed to be as robust as they can to resist detecting and attacking.

As mentioned above, in case of digital steganography abuse by terrorists for illegal purpose, it is worthwhile to investigate the countermeasures of the digital steganography. Currently, one of a large category of techniques against digital steganography is steganalysis [5] which is a kind of methods trying to detect and crack the hidden information. Steganalysis is mainly focused on cracking what the information is hidden in the cover media. Most of the current steganalysis methods are only effective for a certain kind of image steganography method because they found and used some unique properties of their target steganography methods. Since the mechanism of the steganography methods vary, it is hard to find a generic steganalysis method which works on all the types of steganography methods. Therefore an important weakness for such steganalysis methods is that the types of the steganography method have to be known and well studied in advanced. It is not practical in the Internet environment which is a realtime scenario exchanging billions of multimedia data in a while. It is hard even impossible to know whether steganography is used and which types of image steganography methods are applied. So these methods are only proper for analyzing a certain group of suspicious targets rather than monitoring the general Internet multimedia streams.

Recently, a merging kind of steganalysis method is proposed, called universal steganalysis which is claimed to be adaptive to multiple kinds of digital steganography methods. However, these methods are based on the pattern classification technology which needs a learning process. It is impossible to finish the learning process because of the lack of training information. Obviously, universal methods are not able to classify and detect image streams on the realtime Internet as well. Furthermore, though more advanced and sophisticated steganalysis methods can be invented in future, much more advanced robust steganography method can also be invented accordingly. In all the limitation of the steganalysis method is that it is built on a passive way. This kind of method is not able to actively discover the steganography information. So this passive security protection method cannot be a desired omnipotent solution.

The best defense is a good offense. In fact, in most of the cases, the Internet security only requires to block the harmful information spreading rather than knowing what the information it is. Then, the active attacking-based method will be more effective. This kind of attacking -based method is not interested in what information is hidden in the carrier but just working on destroying the hidden information while keeping a minimum tolerable distortion involved to the cover media.

The difference between the passive and active attacking methods can be described by the classic prisoner problem. Two accomplices Alice and Bob are arrested in separate cells. They plan to escape by working together, however they are not able to communicate directly. They have to exchange their messages through their warden Wendy. So the messages exchanged by Wendy must not to be suspicious to her otherwise they will be put into a high isolated prison where no one can escape. From the warden's point of view, Wendy have to satisfy their rights to communicate innocent information while monitor the illegal one. A passive

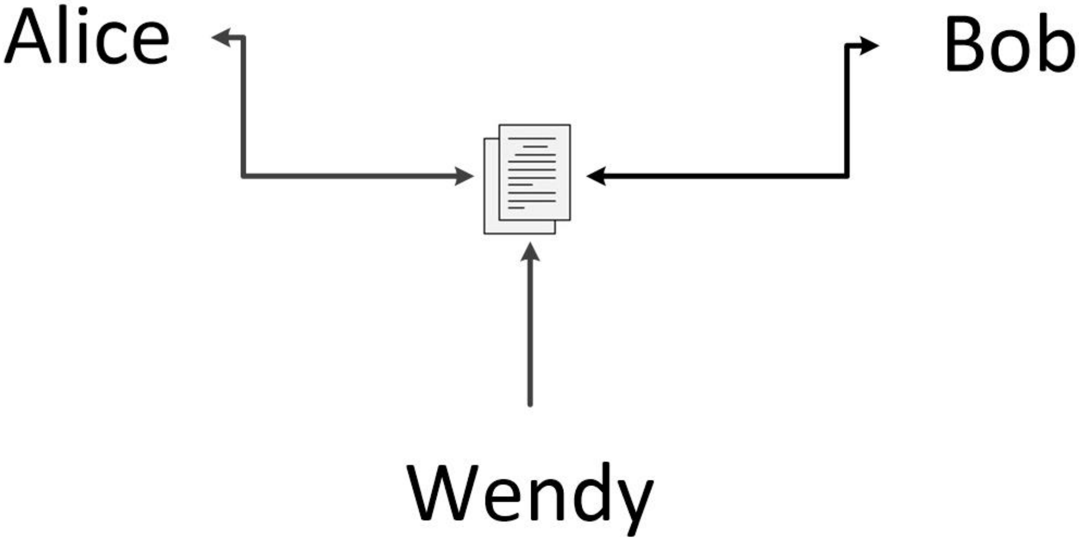

Figure 1.1: Passive Warden Prisoner Problem

warden will not alter the messages initiated by Alice or Bob. She will help to pass it while spying on the suspicious messages. This process is illustrated in Figure1.1. The possibility to preventing the escape plan exchange is solely depended on Wendy's own judgment. The warden's capability to detect the suspicious message is very significant in this place. As a result, it is not very stable to rely on warden's own capability. Alice and Bob may also use different ways to exchange their escape plan. The warden is not able to know all of their ways. It is obvious to be very passive for warden in this scenario.

An active warden, on the contrary, is allowed to modify the message being sent to another prisoner. Specially, a warden can paraphrase the message sent by one prisoner to another. By this way, the straight mean of the message is delivered and the implicit message is most likely to be removed. A mild modification such as replacing words with their synonyms will not change the semantic content of

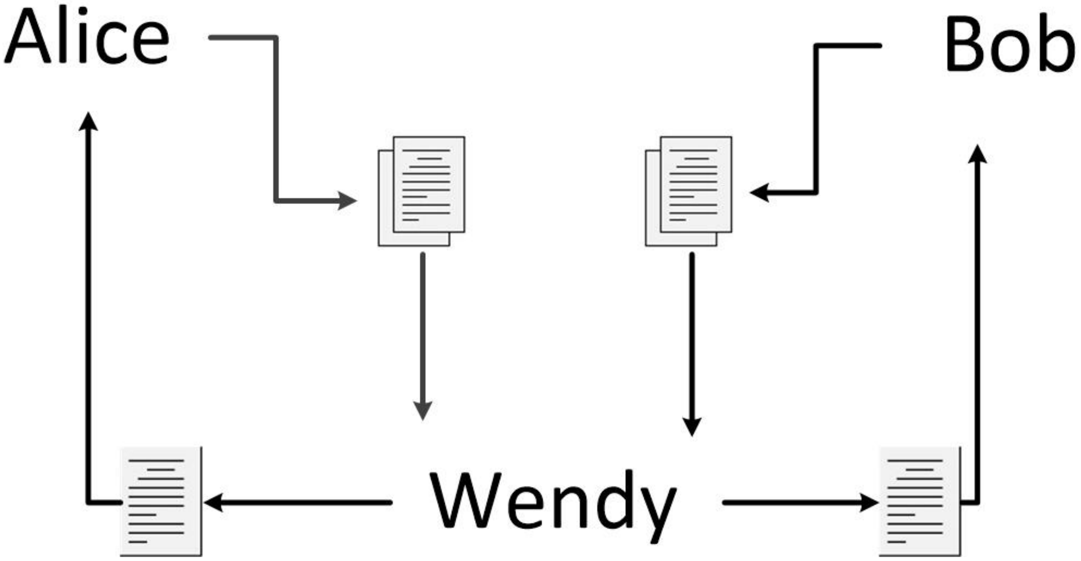

Figure 1.2: Active Warden Prisoner Problem

the text message. This active strategy can be illustrated in Figure1.2. As shown in Figure1.2, Wendy sends the modified text messages to the other prisoner and the hidden information is gone in the modified messages.

This active strategy works perfect for the Internet environment where it is impossible to know the potential steganography existence and the types of the steganography methods. Such a universal active method can be used in a blind basis. Whatever there is steganography or not in the multimedia data, after the universal attack the original cover media keep the perceptual quality and the hidden messages(if have) is removed.

In order to invent a universal active attacking method, we must have the applied scheme independent to the specific properties of the steganography methods. This attacking method can only take advantage of the nature of human perceptual system. In this paper, we propose a way to mostly distort the numerical values of the cover media while keep the perceptual quality in a high level. It is possible because the visual and auditory perception by human is not fully in coherence with the numerical values and there is a gap between them [18]. There are some kinds of distortions which will heavily change the numerical values while less affect the image human visual perceptual evaluation. By involving this kind of distortions the visual information will be kept and the hidden information will be destroyed in larger sense of possibility since the numerical values of the host images have been significantly changed.

In order to find such visually unnoticeable distortion, we are motivated by the print-scan process. When a digital image is printed out into a paper and scanned back into digital, most of the pixel values are significantly changed and most of robust digital steganography methods are not resilient to these digital-analog-digital process. By virtualizing two kinds of distortions which print-scan process caused, we propose an effective feature-independent active attacking method which includes a new transform in image called Discrete Spring Transform and geometrized of the pixel value of the image. The state-of-art printer can output the image in a very high perceptual quality although multiple types of distortions are caused in this digital to analog process. The reason the high quality can be kept is that the distortion caused by the printer is relatively less noticeable for human visual system. An example of those distortions are the distortion caused by the deviation of the ink jet head. Due to the mechanical accuracy, every ink dot is printed in the paper with a small position deviation. This deviation is small and caused by many uncountable random factors. This randomness makes the deviation irreversible, therefore the potential hidden message cannot be recovered easily. This process can be expressed in a 1-D model as

$$a(t) \longrightarrow a(t^{'}) \tag{1.1}$$

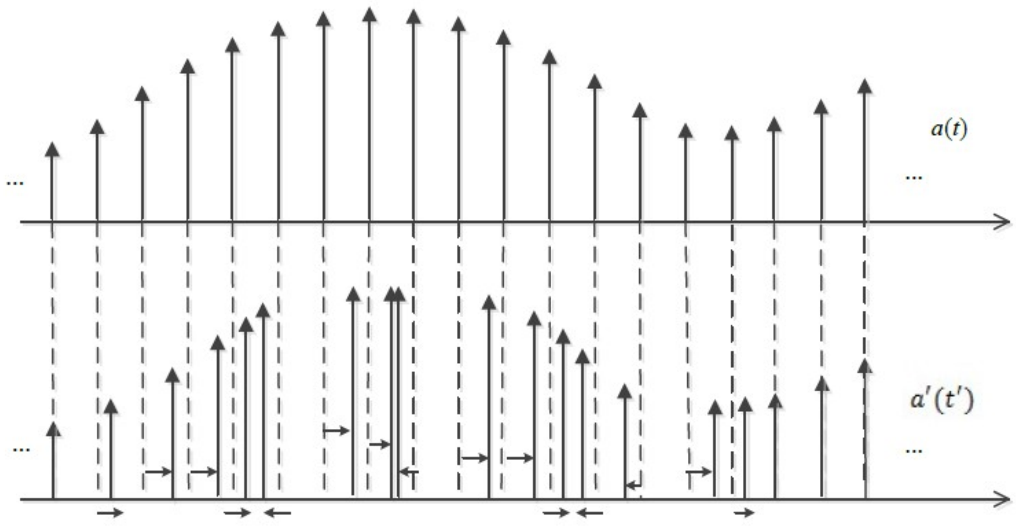

Figure 1.3: Printer Random Deviation

It is also shown in Figure1.3.

This heuristic model can be further abstracted as a Discrete Spring Transform. This transform can be implemented in audio, image and video signals in order to remove the hidden message by various steganography methods. It is called Spring Transform because it performs like a spring being stretched. The spring is stretched or pressed in every spot, but the stretch rate is different from spot to spot. This variable stretch rate is continuously changed. This concept can be applied in the 1-D signal and multi-dimensional signal. An important implementation issue in the Spring Transform is that how to let the stretch rate continuously change in a digital signal. In this paper we propose a block-based method for audio signal, and interpolation and geometrized method for the image signal. The image method can be extended into the video signal by applying the Discrete Spring Transform to the time frame as well. The experimental results in this thesis shown the implemented methods effectively removed the steganographic information in spatial and frequency domain for audio, image and video signal. Meanwhile, the cover media maintains an acceptable perceptual quality under our evaluation methods.

The spring transform has two important features. The first is that it will not distort the perceptual quality of the multimedia signal. Secondly, it will greatly change the numerical value of the multimedia signal. These two features enable an active warden approach for the steganography attacking and provide security for the Internet security especially for the multimedia applications.

Chapter 2

Background

Steganography comes from Greek origin and means ”concealed writing”. Historically, hidden message is hidden in cover text, image. The hidden message can be carried between lines of a letter by invisible ink. Generally the steganography is a very broad conception for hidden message techniques. For example In 480 BC a Greek by the name of Demaratus warned Spartans warning about an incoming war by sending a message using the method described as follows

> As the danger of discovery was great, there was only one way in which he could contrive to get the message through: this was by scraping the wax off a pair of wooden folding tables, writing on the wood underneath what Xerxes intended to do, and then covering the message over with the wax again. In this way the tablets, being apparently blank, would cause no trouble with the guards along the road.

One significant advantage of steganography over cryptography is that the cover media used to convey hidden message looks the same as the innocent one. As a result, no additional attention will be drawed. Sometimes even the cryptography is strong enough to be not compromised, the use of cryptography is suspicious or illegal in some circumstances. Steganography is a way to conceal the fact that the hidden message is sent and the content of the hidden message as well.

In this thesis, we only focus on digital steganography. The modern steganography methods conceal hidden message in digital multimedia files such as text, image, music and video. The digital multimedia signal provides a larger capacity for steganographic messages. The boosting wideband Internet and wireless technologies make more and more multimedia applications possible. People today cannot live without the Internet and social networks. People upload, share, view and download music, image and video from their social network accounts. Those huge amounts of multimedia streaming on the Internet are capable to convey lots of hidden messages.

However those multimedia data can be easily compromised. In fact, billions of network attacks happen accompanying with the normal network activities. A instance of the attack associated with the steganography is that the attacker hijacks the online multimedia data as the cover media to transmit hidden malicious messages. It is difficult to notice because the steganography will not change what the multimedia content looks like. The hidden messages embedded will be recognized as network noise as it usually has. The normal cryptography will not prevent this kind of attack. The attack may add the hidden messages onto the cryptography cover media. The hidden messages can still be decoded correctly with the key known. Even if the malicious information is captured, it is not able to be tracked. The owner of the cover media may not be the party who adding the hidden messages. So a very critical security problem arises by the development of the steganogrphy methods. It is also reasonable to believe the steganography tends to be abused because its properties. In order to protect the secret information, the normal cryptography can be used by general public and government. There is no special interest for innocent people to use steganography to protect their privacy or conduct security communication. The motivation of use of the steganography is that people do not want the third part to know there is something important in it. It is of the interests of the terrorists.

Consequently, the countermeasures against steganography become a popular topic on research. Historically the countermeasures against steganography are not regarded as important as steganography. The research of the countermeasures is usually used as a test bench for the steganograph research. Nowadays the security concerns discussed above stimulate people to focus on this topic. This thesis will discuss the existing countermeasures against the steganography. A common weakness of the existing methods is that they may not work in a generic basis or work without the nature of the steganography. An active method is proposed in this thesis to resolve this problem.

2.1 Motivation

The countermeasures of the steganography mainly focus on how to prevent the transmission of the illegal messages encoded by the steganograph. A straightforward solution for this problem is a passive method. This passive method monitors the communication channel periodically. It will detect if a steganography is used in the monitored cover media. If detected, a further classification is needed to tell what steganography method is used. Based on the type of the steganography method, a respective steganalysis method will be applied to decode the steganography messages. This method cannot deal with an unknown steganography method because the steganalysis method has to be designed accordingly. Meanwhile the computational complexity is very high for decoding the steganographic information. It is not very useful for the realtime scenario in the Internet where multiple steganography methods may be used

In this problem, the content of the steganographic message may not always be very important. If the sender of the steganographic message is clearly known, then the content of the message may be very intellectually meaningful. Whereas on the general Internet application, actually most of the content is innocent. Though steganographic messages are embedded, the content of the hidden message may not be of interest. So in order to increase the security of the network, it is good enough to just simply block the transmission of the steganographic messages without knowing what it is. A selective filter is desired that the innocent information can be passed while the hidden message would be blocked. This blocking means it is unable to be decoded by the intended receiver. The design of this filter depends on the properties of the cover media and the steganographic message. It may be thought to be impossible because it is difficult to only change the hidden message while keeping the cover media intact. It is true if it is required that the numerical must be intact . However in most of the cases the numerical value is not as much of interest as the perceptual results for the multimedia cover media. So it is possible to find a way to change the numerical value of the cover media while keep the perceptual effect. This thesis will focus on finding a way to achieve this task.

2.2 Related Works

2.2.1 Steganography Methods

In the audio steganography, the hidden message is covered by audio signal [6] such as speech and music. In some of the circumstances, the detail of the audio signal is not important, so by slightly altering the audio signal, a large amount of steganographic messages can be embedded. Some main audio steganography algorithms can be classified as follows,

- phase coding [6]
- spread spectrum[7, 8]
- quantization index modulation[9]
- echo hiding[10]
- patchwork[6]

Recently, the audio steganography focuses on how to improve the robustness of the hidden messages, some improved steganography [11, 12, 13, 14, 15] are able to resist a lot of different noises and distortions. A main important distortion against the audio steganography is the time scale modification. Most of the steganography methods cannot survive from the time scale modification. Some TSM-robust audio steganography are proposed recently [16, 17, 18].

Image steganography is a well-studied topic[5]. Some typical digital watermarking methods can be used as steganography as well[19, 20, 21]. Some wavelet-based steganography methods are proposed[22, 23, 24, 25]. Recently, some steganography methods[26, 27, 28, 29] are proposed to be robust to the Rotation, Scaling and Translation. Another direction for the image watermarking is the reversible watermarking[30, 31, 32, 33, 34, 35]. Some steganography methods can be applied in image-based steganography and video-based steganography. These techniques [36, 37, 38] are hiding message in the images, while it can be easily applied for the video signal by repeatedly inserting the hidden messages in every frame of the video. There is no hidden message between frames. So these methods are mainly classified as the 2-D steganography methods.

Some steganography methods are designed solely for the video steganography, the motion and time frames are used to hide messages. So this method can be simply adjusted from the 2-D steganography methods. It is classified as 3-D steganography.

The motion-based steganography alters the motion frame to hide steganographic messages. The [39, 40, 41, 42]. This method explored the capacities in the motion vector of the video signal. When combining with the image-based 2-D steganography and the time/motion-based video steganography, it is called multi-dimensional video steganography. It will simultaneously hide information in 2D and 3D margins. In Figure**??**, a multi-dimensional steganography scheme is illustrated. The image-based steganography is encoded in the I-Frame, while the motion vector-based steganographic message is hidden in P-Frame as well. It provides a larger capacity for the steganographic messages and increases the difficulty for the countermeasures.

2.2.2 Countermeasures against Steganography Methods

One of a large category of techniques against steganography is steganalysis [5] which are methods trying to detect and crack the hidden information. Steganalysis is mainly focused on cracking what the information is hidden in a stego image. Most of the current steganalysis methods are only effective for a certain kind of image steganography method[43, 44, 45, 46, 47, 48, 49, 50]. For all these steganalysis methods, the most significant weakness is that the types of the steganography method have to be known in advanced. It is not practical in Internet environment among billions of images. It is hard even impossible to know whether steganography is used in an image and which types of image steganography methods are applied. So these methods are only proper for analyzing a certain group of suspicious images rather than monitoring the general Internet image streams. Recently, a merging kind of steganalysis method is proposed, called universal steganalysis which is adaptive to multiple kinds of image steganography methods [51, 52]. However, these methods are based on the pattern classification technology which needs a learning process. It is impossible to finish the learning process because of the lack of training information. Obviously, universal methods are not able to classify and detect image streams on Internet as well. Furthermore, though more advanced and sophisticated steganalysis methods can be invented in future, much more advanced robust steganography method can also be invented. So this passive security protection method cannot be a desired omnipotent solution. The best defense is a good offense. In fact, in most of the cases, the Internet security only requires to block the harmful information spreading rather than knowing what the information it is. Then, the active attacking-based method will be more effective. This kind of attacking-based method is not interested in what information is hidden in the carrier but just working on destroying the hidden information while keeping as minimum distortion as it can to the host images. There are some kinds of attacking methods [53, 54, 55, 56, 57]. All these methods are only effective for the certain kinds of image steganography methods too. This method is always working by recognizing the unique features taken to the host images by a steganography method where the normal images do not have. In addition research on attacking-based steganography security method is still an uncultivated area where it did not draw deserved attentions. In the Internet environment it desires that there is a universal active method which can be used in the images without the knowledge of what type of steganography methods are. Whatever there is steganography or not in the images, after the universal attack the original host image should keep the high visual quality and the hidden image they may have is removed. The image-based steganography attacking methods can be applied to the video steaganography as long as the hidden message is embedded in the I-Frame. For the motion vector-based steganography methods, those attacking are void because the motion vector is a unique medium for video signal. Some motion vector specific attacking methods are [58, 59], all of which are passive methods as we classified above.

In order to invent a universal active attacking method, we must have the scheme independent to the specific properties of the steganography methods. This attacking method can only take advantage of the nature of human visual system. In this paper, we try to find a way to mostly distort the numerical values of the image while keep the visual quality in a high level. It is possible because the visual perception by human visual system is not fully in coherence with the numerical values and there is a gap between them [60, 61]. There are some kinds of distortions which will heavily change the numerical values while less affect the image human visual perceptual evaluation. By involving this kind of distortions the visual information will be kept and the hidden information will be destroyed in larger sense of possibility since the numerical values of the host images have been significantly changed. Some of our preliminary works on the generic steganography attacking are proposed in[62, 1, 2, 63, 64, 65, 66]

Chapter 3

Discrete Spring Transform:A novel Perceptual Invariant Approach

3.1 Introduction

Most of the digital steganography methods take advantage of the margin between the numerical value and visual perception of the multimedia carriers. In other words, the steganographic messages are embedded in the carriers by involving some slight distortions which are non-significant for human perception system. The steganographic capacity of carrier derives from the defect of the biological perceptual capability. For instance, a true color image pixel costs 24 bits to store. Therefore, a pixel varies in $2^2$4 different colors which is far more than the number of colors human visual system is able to differentiate. As a result, a bit steganographic message can be inserted into the pixel by changing the least significant bit of this pixel. The slightly color change caused by the change of the least significant bit of the pixel is not able to draw any attention from the human visual system.

Although the knowledge to the mechanism of human perceptual system is very limited up to date, there are some empirical facts about it. In general, human visual system may pay more attentions on some certain kinds of changes. Meanwhile, some specific areas in an image may also draw more attentions to human visual system. These unbalanced attention phenomenons also happen in human auditory system. For a pixel in an image, the affect to human visual system caused by a numerical change can be quantified as a set of functions,

$$\mathcal{A} = f_k(x, y, d) \tag{3.1}$$

where $\mathcal{A}$ is the quantified affect to human visual system, the model of f_k varies by different types of change involved in the image. The affect is related to the position of the pixel (x, y) and the numerical difference d caused by this change. The objective of steganography methods can be expressed as an optimization problem,

$$\begin{aligned} \text{maximize} \quad & d & (3.2) \\ \text{subject to} \quad & f_k(x, y, d) \leq T & (3.3) \end{aligned}$$

where T is a threshold for maximal tolerable visual distortion. Steganography methods try to find some types of distortion involved in some specific area of the image to obtain the maximized capacity while make the perceptual distortion is tolerable for human visual system.

In order to make the generic steganographic attacking method works on varies types of steganography algorithms, the attacking method should not lay on any specific property of a certain type of steganography methods. However, as discussed above, most of the steganography methods take advantage of the margin existing between the human perceptual system and numerical values, a countermeasure also can take advantage of it. A further distortion can be manually involved by the countermeasure to let the numerical values of the cover media further changed. By this way, the steganographic message hidden in the cover media may be largely destroyed because the hidden information is highly depends on the intact of the numerical values in the cover media. Meanwhile, this distortion can be controlled to cause minimal human noticeable change to the cover media.

In this chapter, a Discrete Spring Transform is proposed to achieve this goal. This transform can greatly change the numerical values of the cover media while maintain a high perceptual quality for the cover media. In order to reveal the mechanism of the Discrete Spring Transform, the features of how human visual system perceives visual signals should be further investigated.

3.2 Less Significant Perceptual Difference and Less Significant Perceptual Area

In image and video applications, instead of how to draw more attentions to human visual system, we are more interested in how to draw less attentions to human visual system while applying a transform to the image or video. There are two problems in visual perception. First where draw fewer attentions to human visual system. Second what transform draw fewer attentions to human visual system.

Even though the human visual system is a very complicated biological system which is still limited known by human some empirical observations about human visual system are very useful for our research. Generally, it is believed the local geometric transform will draw fewer attentions to human visual system. The local geometric transform can be expressed as

$$I^{'}(x^{'},y^{'},1) = \begin{pmatrix} t_{11}(x,y) & t_{12}(x,y) & 0 \\ t_{21}(x,y) & t_{22}(x,y) & 0 \\ t_{31}(x,y) & t_{32}(x,y) & 1 \end{pmatrix} I(x,y,1)$$

The locality of this transform makes the pixel is projected in different direction and distance from the original pixel. The direction and distance is not identical so that the transform is not able to generate a global transform trend for the image. In a micro view, every pixel is changed, however this change is not consistent in the macro view. As a result, this change will not draw large attentions to human visual system. On the other hand, human visual system is not sensitive to the location projection in the image as well. In all this localized transformation can be a candidate for the proposed transform to be less attractive to human visual system.

Human visual system pays fewer attentions to the plain area in the image compared to the edge area in the image in terms of the content of the image. It is straightforward because the edge area contains more information capacity. On the contrary, the plain area only reflects a few information because of lack of change. Consequently, a change in the plain area will draw fewer attentions than a change in the edge area considering the changes are comparable in terms of the means squares. In the frequency domain, it is well-known that human visual system discards high frequency components of the image. This is also the fundamental for JPEG image compression. Provided the image is not compressed, the change of the high frequency is also makes less affection on the entire image.

In the audio signal, human auditory system has some similar properties. The experiment results shown a localized time scale change in the audio signal will negligibly affect the audio quality. It is obvious that for a normal music or speech where thousands of frames are played in one second, a simple change in a few frames may not be recognized by human auditory system. The proposed Discrete Spring Transform will take advantages of those properties to perform transformation and apply the transformation in selected areas. It should emphasize the transforms discussed here will only less affect the human perceptual quality of the image or audio, the numerical value of the multimedia signal will still be dramatically changed. This is just a fundamental to apply this transformation to the active warden attacking methods. As a result, the normal objective mean square error based methods will not be fair enough to evaluate the signal quality of the multimedia signal.

3.3 Prototype of Discrete Spring Transform

A preliminary work for the prototype of Discrete Spring Transform is presented in [1]. Let's consider the 1-D Discrete Spring Transform. For a physical spring the stretch rate is different in different spot. This can be considered as a time-variable scaling process for an audio 1-D signal. The signal $a(t)$ is stretched by a stretch rate function $r(t)$. In t_0, it is scaled as

$$a(t_0^{'}) = a(t_0) + r(t_0)\Delta t \tag{3.4}$$

This process is shown in Figure3.1, we can see the stretch rate varies spot by spot. If the rate function $r(t)$ is randomized, the transformed signal is difficult to reverse. Next let's see how the local stretch rate affects their neighbors. For a finite length

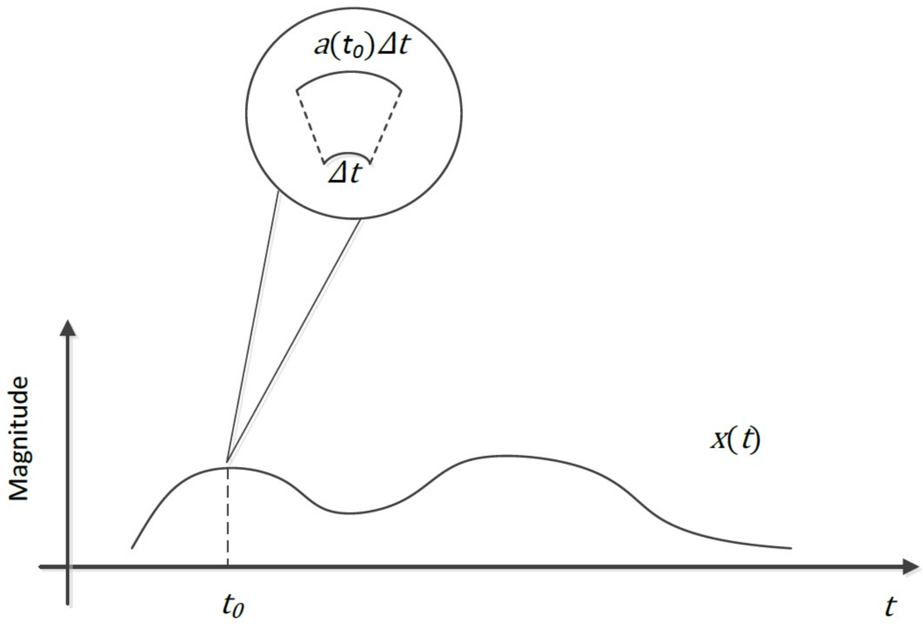

Figure 3.1: Time-variable Scaling Rate[1]

continuous 1-D signal $a(t), 0 \leq t \leq T$,if it is scaled in a constant rate r then the scaled signal can be expressed as

$$a^{'}(t) = a(t/r), \quad 0 \leq t \leq rT \tag{3.5}$$

It can be reversed by

$$a(t) = a^{'}(rt), \quad 0 \leq t \leq T \tag{3.6}$$

Now consider the signal is scaled by two scale rates, $a(t), 0 \leq t \leq t_1$ is scaled by r_1, and the other part of $a(t), t1 \leq t \leq T$ is scaled by r_2, and $r_1 \neq r_2$. Then the scaled signal can be expressed as

$$a^{'}(t) = \begin{cases} a(t/r_1) & 0 < t \leq r_1 t_1 \\ a(t/r_2) & r_1 t_1 < t < (r_1 - r_2)t_1 + r_2 T \end{cases} \tag{3.7}$$

By observation, the reverse relation is expressed as

$$a(t) = \begin{cases} a^{'}(tr_1) & 0 < t \leq t_1 \\ a^{'}(tr_2 + t_1(r_1 - r_2)) & t_1 < t < T \end{cases} \tag{3.8}$$

The second part of the signal is not only depends on its own scaling rate r_2 but also depends on the previous scaling rate r_1. In the case where the signal is scaled by multiple scale rate, this relation can be further investigated. For the signal $a(t)$ which separated by $(t_0, t_1, t_2, ..., t_{n-1}, t_n), (t_0 = 0, t_n = T)$and scaled by $r_1, r_2, ..., r_n$. The scaled signal can be expressed as

$$a^{'}(t) = \begin{cases} a(t/r_1) & 0 < t \leq r_1 t_1 \\ a(t/r_2) & r_1 t_1 < t < (r_1 - r_2)t_1 + r_2 T \\ ... & ... \\ a(t/r_i) & \sum_{k=1}^{i-1} r_k(t_k - t_{k-1}) < t < \sum_{k=1}^{i} r_k(t_k - t_{k-1}) \\ ... & ... \\ a(t/r_n) & \sum_{k=1}^{n-1} r_k(t_k - t_{k-1}) < t < \sum_{k=1}^{n} r_k(t_k - t_{k-1}) \end{cases} \tag{3.9}$$

It can be reversed by

$$a(t) = \begin{cases} a^{'}(tr_1) & 0 < t \leq t_1 \\ a^{'}(tr_2 + t_1 < t < (r_1 - r_2)) & t_1 < t < t_2 \\ \ldots & \ldots \\ a^{'}(tr_i + \sum_{k=1}^{i-1} r_k(t_k - t_{k-1}) - r_i t_{i-1}) & t_{i-1} < t < t_i \\ \ldots & \ldots \\ a^{'}(tr_n + \sum_{k=1}^{n-1} r_k(t_k - t_{k-1}) - r_n t_{n-1}) & t_{n-1} < t < t_n \end{cases} \tag{3.10}$$

From the above observations, it shown the signal reverse depends on all the previous scaling rate $r(t)$. If n goes to infinite, every point is not able to be recovered because it depends infinite previous scaling rate. So this transform becomes a one-way transform. It is very useful when applying for the countermeasures against the steganography because it is desirable to make the attacked signal unrecoverable. In fact, this is an advantage over the time scale modification which can be recovered by some synchronization techniques. In order to make the signal not change too much, the scale rate $r(t)$ should be a function close to 1. The maximum magnitude of the difference caused by the scale function is expressed as

$$max(|r(t) - 1|) \tag{3.11}$$

which is called the maximum scaling range. The property of the scaling function conforms to the physical spring where every spot is stretched in a different rate while the every spot interact its neighbors. One of the important advantages is that though the signal is scaled in various signal the order of the signal is kept. That is for the original signal if $t_1 < t_2$ it must be $t_1^{'} < t_2^{'}$ for the transformed signal. It is a very crucial property to maintain the perceptual quality of the audio signal in steganography attacking process.

The proposed transform is a reasonable tool against the steganography. But the implementation of the spring transform cannot be done in continuous form in digital multimedia signal. The implementation of the spring transform is discussed in next two chapters.

Chapter 4

Proposed Active Warden Attack against Audio Steganography Methods

4.1 Introduction

In the audio signal, human auditory system is not sensitive to the time scale modification. The proposed attacking method is based on the 1-D discrete spring transform[1]. The transform will scale the audio sequences with a time variant stretch rate. In order to make the transformation irreversible, the stretch rate is a random series. Since the stretch rate is small and not identical, the overall affection of the audio signal is negligible. The advantages of this method are that it need not to detect whether steganographic message is embedded in the audio signal or not. Moreover this is a generic method where the types of the steganography methods are not important in terms of the performance of the attacking methods. It is obvious that the smaller the stretch rate is the less distortion will involve. On the other hand, larger stretch rate has a better capability to remove the steganographic message while a larger distortion will be involved as well. Therefore it is desirable to dynamically adjust the stretch rate to meet the requirement for the audio quality and the security standards. Also the stretch rate can be adjusted according to the location and the nature of the audio. The stretch rate must be strictly controlled in some important section of the audio whereas a larger stretch rate is allowable in some transitional section of the audio. Another advantage for this method is that even though the stretch rate is small the steganographic message still tends to be nonrecoverable because the stretch rate changes very quickly. It is difficult even impossible to capture the stretch rate.

The proposed Spring Transform is not able to implement in the digital audio signal, because the stretch rate cannot change point to point in discrete signal. So a block-based Discrete Spring Transform is proposed. This block-based method will be used to attack the steganography hidden in the audio signal.

4.2 Block-based Discrete Spring Transform and Audio Steganography Attacking

As shown in Figure4.1 a block-based Spring Transform is implemented in the audio steganography attacking. The stretch rate is fixed in a certain block which contains a few time series. The stretch rate is dynamic from one block to another. For the audio signal, $a[n], n = 0, ..., N-1$, it is divided into K blocks where the index of the blocks are $B_1[N_0, N_1], B_2[N_1, N_2], ..., B_k[N_{k-1}, N_k]$ and $N_0 = 0, N_k = N-1$. Number of samples in each block is $n_1, n_2..., n_k, n_i = N_i - N_{i-1}$. The number of each block is not the same and irrelevant to each other. The maximum block size

$\max n_i$ need to be small enough to make the stretch negligible. The stretch rate cane be expressed as

$$r[n], n = 1, 2, ..., k \tag{4.1}$$

Then the block-based Discrete Spring Transform can be expressed as

$$a^{'}[n] = f(a[n]) \; n = 0, 1, ..., N^{'} - 1 \tag{4.2}$$

In each block the scaling can be implemented by the interpolation or resampling. The interpolation function can be expressed as a convolution

$$\hat{a}_i(t) = a_i[n] * w(x) \tag{4.3}$$

$$\hat{a}_i(t) = \sum_{k=N_{i-1}}^{k=N_i} a_i[k] w(t-k). \tag{4.4}$$

where

$$w(x) = \begin{cases} 1 & x = 0 \\ 3\frac{\sin(\pi x)\sin\left(\frac{\pi x}{3}\right)}{\pi^2 x^2} & 0 < |x| < 3 \\ 0 & otherwise \end{cases} \tag{4.5}$$

is the window function of the Lanczos interpolation. The interpolated signal$\hat{a}_i(n)$ is a continuous signal of the signal $a[n]$ in block B_i. After the interpolation, the scaling is implemented based on the resampling of the continuous signal $\hat{a}_i(n)$ with a new stretch rate $r(i)Fs$,where Fs is the original sampling rate of the audio signal The resampling process is shown as

$$a_i^{'}[n] = \hat{a}_i[N_{i-1} + (n - N_{i-1}^{'})\frac{1}{r(i)Fs}] \quad n = N_{i-1}^{'}...N_i^{'} \tag{4.6}$$

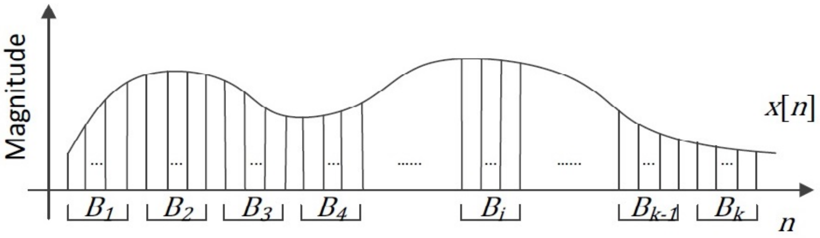

Figure 4.1: Block-based Discrete Spring Transform[1]

where N'_{i-1} and N'_i is the adjusted boundaries of the signal block. It can be obtained by

$$N'_i = N'_{i-1} + r[i]n_i \tag{4.7}$$

and $N'_0 = 0$.

Finally, the entire transformation is expressed as

$$a'[n] = f(a[n]) = \begin{cases} a'_1[n] & 0 < n \leq N'_1 \\ a'_2[n] & N'_1 < n \leq N'_2 \\ \ldots & \ldots \\ a'_i[n] & N'_{i-1} < n \leq N'_i \\ \ldots & \ldots \\ a'_k[n] & N'_{k-1} < n \leq N'_k \end{cases} \tag{4.8}$$

As mentioned above, the stretch rate is a random series so that it is difficult to be followed. Though it makes the signal practically non-reversible, the possibility still exists if the stretch rate is known. Unlike the prototype of the spring transform where the stretch rate is infinite, the stretch rate function in the block-based discrete spring transform is finite. So in finite blocks the signal can be recovered by scaling with a stretch function

$$c[k] = 1/r[k] \tag{4.9}$$

An additional adjustment is conducted in each block to make the stretch rate totally nonreversible. For block B_i, this adjustment is expressed as

$$a_i^{'}[n] = \begin{cases} \hat{a}_i[\frac{n}{r_i^{'}(1)Fs}] & n = 1 \\ \hat{a}_i[\frac{n}{r_i^{'}(2)Fs}] & n = 2 \\ \quad ... & ... \\ \hat{a}_i[\frac{n}{r_i^{'}(j)Fs}] & n = j \\ \quad ... & ... \\ \hat{a}_i[\frac{n}{r_i^{'}(n_i^{'})Fs}] & n = n_i^{'} \end{cases} \tag{4.10}$$

This process resampled the audio signal in each block with an associate time stretch function $r_i^{'}(k)$. It makes the signal unable to recover even $r[n]$ and $r_i^{'}(k)$ are explicitly known.

Chapter 5

Proposed Active Warden Attack against Image and Video Steganography Methods

5.1 Print-scan Process Inspired Analog Location Transform

Consider an inkjet printer which is mostly used for photo printing today. Ink droplets are printed on the paper by an inkjet head in order to transform a digital representation of an image onto physical paper. One of the distortions in this process is caused by pixel location deviation. Since the inkjet head is a mechanical structure, it cannot guarantee that every ink droplet which represents a certain pixel can be exactly and uniformly printed at the designated location. This pixel location deviation is an analog change which is hard to be dealt with by steganography algorithms compared to some rotation, scaling, translating and cropping (RSTC) transforms. A steganography algorithm could be designed to survive RSTC transforms, but can hardly survive the print or scan process. Since the human visual system is insensitive to small-scale deviations in the printed image this uncorrected location deviation changes only the pixel values of the image rather than the perception of the image. Thus the small pixel location deviations are not able to create a change noticeable to the human visual system. This pixel location deviation can be formulated as an analog location transform (ALT). For simplicity, we first consider the 1-D case. Consider a 1-D N-points signal where every data point is located in integer points 0,1,2,,N-1 as shown in Fig.1, after the ALT every discrete data point is mapped to a new location as dictated by the ALT. The mapping of the 1-D ALT can be described as

$$a(t) \rightarrow a^{'}(t^{'}) \tag{5.1}$$

and

$$t^{'} = t + \Phi(t) \quad t \in \mathbb{Z} \quad t^{'} \in \mathbb{R} \tag{5.2}$$

where $\Phi(t)$ is the transform function denoting time-variant location deviations. In order to prevent disordering, we define $|\Phi(t)| \leq 0.5$ and $\Phi(0) \geq 0 \quad \Phi(N) \leq 0$. It should be aware that after the ALT, t could mapped into arbitrary continuous location range in $(0, N-1)$.

It is then straightforward to extend to the 2-D situation where an image ALT can be expressed as

$$I(x,y) \rightarrow I^{'}(x^{'},y^{'}) \tag{5.3}$$

where

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Phi_1(x,y) \\ 0 & 1 & \Phi_2(x,y) \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad x,y \in \mathbb{Z} \quad x',y' \in \mathbb{R} \tag{5.4}$$

This transform differs from a conventional affine transform or geometric transform since the transform functions are not constant as the 2-D ALT is a spatial-variant transform. For the same reason as 1-D ALT, the transform functions are confined as $|\Phi_1, \Phi_2| \leq 0$ and $\forall x = 0, y = 0$ we have $0 \leq \Phi_1(x,y), \Phi_2(x,y)$ and $\forall x = M-1, y = N-1$ we have $\Phi_1(x,y), \Phi_2(x,y) \leq 0$. This definition allows the location of each pixel to deviate only in relation to its nearby area. There are various methods that can be used to create the transform functions $\Phi_1(x,y)$ and $\Phi_2(x,y)$. Thus the ALT can be described as an analogy to the random inkjet head mechanical jitter, where the transform functions could be uniformly distributed i.i.d random processes. The probability density function can be expressed as

$$P(\Phi_i(x,y)) = \begin{cases} \dfrac{1}{2\delta_i} & -\delta_i < \Phi_i(x,y) < \delta_i \\ 0 & otherwise \end{cases} \tag{5.5}$$

where $i = 1,2$ and $0 \leq \delta_i < 0.5$ is the maximum range of the deviation. This random method did not well consider the properties of the human visual system and therefore cannot be a very effective method. However this method provides us a reference threshold for the other advanced methods. Obviously, the larger deviation ranges that are used, the larger distortions will be caused. We therefore set a threshold $T_1 = M\frac{\delta_1}{2}$ and $T_2 = N\frac{\delta_2}{2}$ denoting the maximum deviation range in horizontal and vertical directions. This threshold can be found by searching the maximum possible δ_x and δ_y under desired image visual quality requirement and tuning appropriately. These thresholds will then be used for normalizing the transform functions in the following methods.

As mentioned, in order to attack the steganography maximally while maintaining high image quality and preserving the human visual perception of the image the properties of human visual quality have to be taken into account properly. Specifically, in ALT method, larger deviations could be involved in the area where human visual system is not as sensitive. Rather than perceiving the absolute value of the pixel intensity, the human visual system is more sensitive to the contrast and the edge area which will draw more attention than the plain area in an image. In other words, the deviations in plain area will be increased while the distortion in edge area will be suppressed. In order to realize this idea, we utilize two methods, the first being an edge detector and novel curve length method. Sobel edge detection is one of the most simple edge detectors which can extract the edge information of an image. The masks of the Sobel detector in horizontal and vertical directions are

$$M_1 = \begin{Bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{Bmatrix} \tag{5.6}$$

and

$$M_2 = \begin{Bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{Bmatrix} \tag{5.7}$$

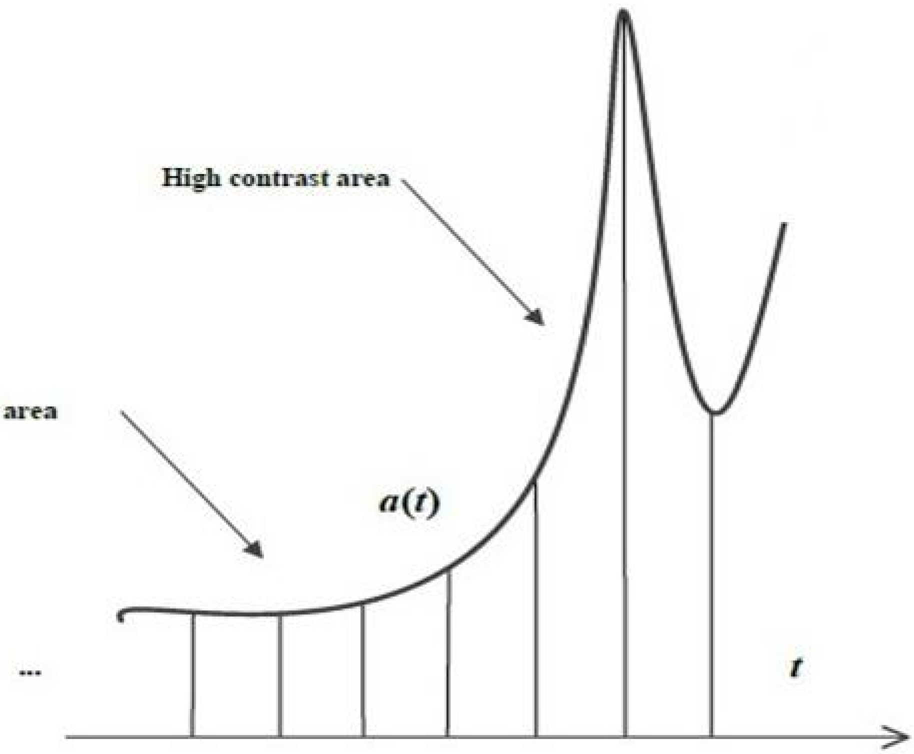

Figure 5.1: Illustration of the Relation Between Curver Length and Pixel Difference

The gradient of image in horizontal and vertical directions can be calculated by per-forming 2-D convolution of the image and the Sobel detector masks in two directions respectively as follows,

$$G_1 = I \times M_1 \tag{5.8}$$

$$G_2 = I \times M_2 \tag{5.9}$$

This operation discriminates the edge area of the image from the plain area because the gradient in edge area is much larger than the plain area. To make less distortions in the edge area, let the transform parameters be inversely proportional G_1, G_2. Remembering that the quality threshold T_i is related to a desirable visual quality level, the deviation functions can be normalized by the threshold to obtain the desired visual quality. So the transform functions in edge detector method can be expressed as

$$\Phi_i(x, y) = A_i(\frac{1}{G_i(x, y)}) \tag{5.10}$$

where

$$A_i = \frac{T_i}{\sum_{p=1}^{M} \sum_{q=1}^{N} \frac{1}{G_i(p, q)}} \tag{5.11}$$

This method exploits the more unnoticeable or plain area in the image. Edge detection can only discriminate an edge from the other areas of the image but cannot reflect progressive and smooth trends in respect to the contrast of the image pixel values. We invent a novel curve length method which can accomplish this task. As shown in Figure5.1 in a 1-D signal, large differences between neighbor signal point values causes longer curve lengths between them if we interpolate the time sequence into a continuous curve. This curve length is thus progressively changing and dependent on the entire signal pattern. Therefore, we can use the length of the curve to model the signal contrasts. In the plain area of the image the curve length would be shorter than the length in the high contrast area. The curve length can then be obtained by integrating the interpolated 2-D signal in horizontal and vertical directions respectively.

The interpolated signal can be expressed as

$$\hat{I}(x,y) = I(x,y) \times W_L(x,y) \tag{5.12}$$

where W_L is the interpolation window kernel, in this paper, the 3-order Lanczos window kernel of 1-D form can be express below as

$$w(x) = \begin{cases} 1 & |x| = 0 \\ 3\dfrac{\sin \pi x \sin \dfrac{\pi x}{3}}{\pi^2 x^2} & 0 < |x| < 3 \\ 0 & 3 \leq |x| \end{cases} \tag{5.13}$$

then the 2-D window kernel is The curve length between $I(k,y)$ and $I(k+1,y)$ in horizontal direction and between $I(x,k)$ and $I(x,k+1)$ in vertical direction are as follows

$$d_1(k,y) = \int_{x=k}^{x=k+1} \left(\frac{d\hat{I}}{x}(x,y)^2 + 1 \right) dx \tag{5.14}$$

$$d_2(x,k) = \int_{y=k}^{y=k+1} \left(\frac{d\hat{I}}{y}(x,y)^2 + 1 \right) dy \tag{5.15}$$

Then the transform function in the curve length method is formulated as

$$\Phi_i(x,y) = B_x \left(\frac{1}{d_i(x,y)} \right) \tag{5.16}$$

where

$$B_x = \frac{T_i}{\sum 1/d_i} \tag{5.17}$$

Now, every pixel of the image is mapped into a new location that may not represent an integer point. The next step is to deter-mine how to scan back this projected ink-jet image into a digital image. In other words, it means reconstructing the pixel value at the integer point based on the ALT-pixels. One way to accomplish this reconstruction is through interpolation which can be expressed as

$$I_r(x,y) = \sum_{v=Y'} \sum_{u=X'} (I^{'}(u,v) W_L(x-u, y-v)) \tag{5.18}$$

where

$$X^{'} = \{x^{'}(x,y) | x^{'}(x,y) = x + \Phi_1(x,y)\} \tag{5.19}$$

$$Y^{'} = \{y^{'}(x,y) | y^{'}(x,y) = y + \Phi_2(x,y)\} \tag{5.20}$$

denote the mapped location coordinator sets of each pixel. However, based on experimental results, this method may not destroy the steganographic information very well. A novel pixel geometrized method is presented in next section to solve this problem. This method introduces another kind of distortion in print-scan process.

5.2 Geometrized Image Reconstruction

Recall the idea by the ink-jet printer, where we assume the ink volume for each pixel is proportional to the pixel value and the ink droplets will spread into a circle ink dot whose area is proportional to the ink volume. So each pixel can be geometrized into a size-variable circle ink dot. We define that every pixel value is equal to the area of the corresponding ink dot as follows

$$I^{'}(x^{'},y^{'}) = \pi r(x,y)^2 \tag{5.21}$$

where r denoting the radius of the ink dot circle. Ideally, the ink dot is centered uniformly in the paper, however, since after the ALT each circle center will have a slightly deviation and will deviate from its center due to the spreading effect, neighbor ink dots would be overlapped. Assume then that a scanner will scan the paper by sliding a square shape scanning window whose size is comparable to the size of the ink dots. Every pixel will be scanned into digital value by sensing the ink amount in a scanning window. In this process, the scanned image pixel value will be distorted by the circle center deviations, the size of the scanning window, and the neighbor pixels. With different size ink circle areas and a firm size scanning window, there will be nine different geometric patterns as shown in Figure5.2. In Figure5.2 it assumed that the horizontal deviation is towards right direction and vertical deviation is towards down direction. As observed, each ink dot will affect up to eight nearby ink dots, meanwhile each ink dot also would be affected by up to eight nearby ink dots. With different size relations between 2, circle radius circle radius r and scanning window edge length d, these nine patterns would occur in different orders, Figure5.2 is only one of these possible orders. Though this will result in many different patterns, it can be formulated using a concise expression to the recovered value. Firstly, define two functions as below

$$Q(x) = [r^2 \arccos \frac{x}{r} - x\sqrt{r^2 - x^2}, 0]^+ \tag{5.22}$$

$$R(x, y) = [\frac{1}{2}(S(x) + S(y)) + xy - \frac{\pi r^2}{4}, 0]^+ \tag{5.23}$$

$Q(x)$ expresses area in a circle divided by a chord. $R(x, y)$ denotes the area in the middle of two intersect chords. Then the ink area in nine nearby scanning windows spread by pixel $I(i, j)$ can be expressed as

$$S^{(i,j)}(i \pm 1, j \pm 1) = R(\frac{d}{2} \mp \Phi_x(x, y), \frac{d}{3} b\Phi_y(x, y)) \tag{5.24}$$

$$S^{(i,j)}(i \pm 1, j) = Q(\frac{d}{2} - \Phi_x(i, j)) - S^{(i,j)}(i \pm 1, j - 1) - S^{(i,j)}(i \pm 1, j + 1) \tag{5.25}$$

$$S^{(i,j)}(i, j \pm 1) = Q(\frac{d}{2} - \Phi_y(i, j)) - S^{(i,j)}(i - 1, j \pm 1) - S^{(i,j)}(i + 1, j \pm 1) \tag{5.26}$$

$$S^{(i,j)}(i, j) = \pi r^2(i, j) - \sum_{p=-1}^{1} \sum_{q=-1}^{1} S^{(i,j)}(i + p, j + q) \tag{5.27}$$

Then the recovered pixel value in $I_r(i, j)$ can be expressed as

$$I_r(i, j) \sum_{q=j-1}^{j+1} \sum_{p=i-1}^{i+1} S^{(p,q)}(i, j) \tag{5.28}$$

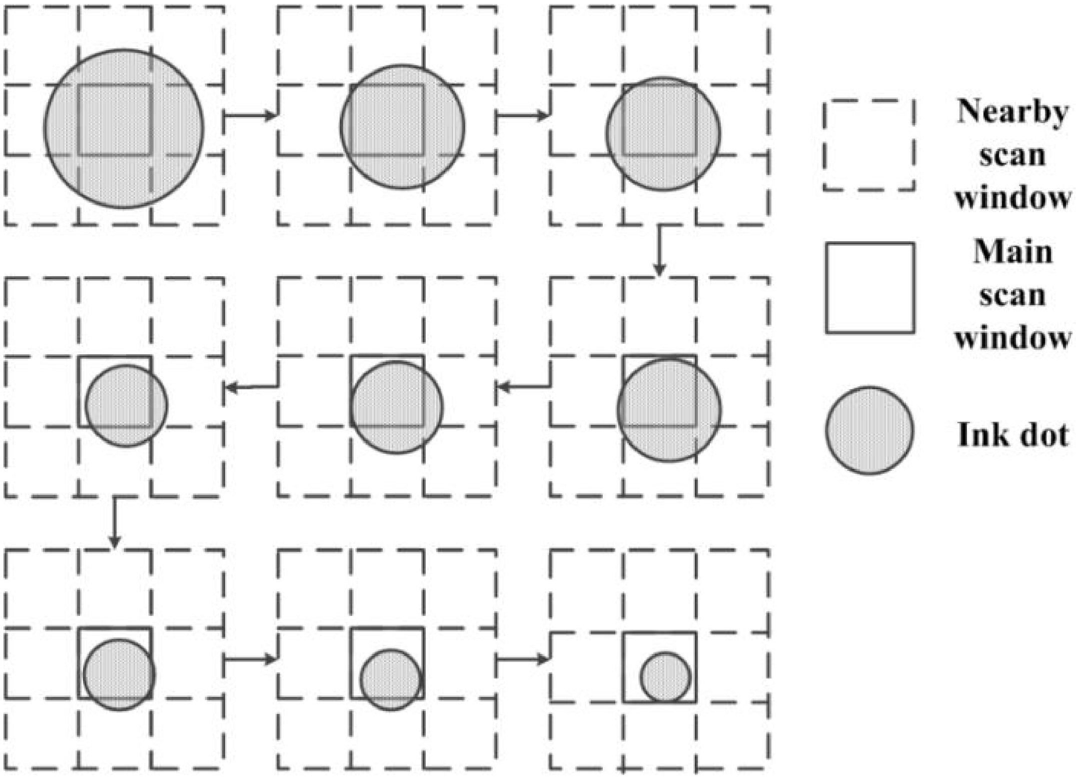

Figure 5.2: Nine Patterns of Scanning Window and Variable-size Ink Dot

5.3 Multi-dimensional Steganography Attack against Video Steganography

The video steganography is implemented in frequency domain. A Discrete Spring Transform extention in frequency domain will be proposed at first.

5.3.1 Discrete Spring Transform in Frequency Domain

In the Frequency DST (FDST), the image $C = c(x, y)$ is transformed in the frequency domain initially which can be expressed as $F(\omega_1, \omega_2)$. The FDST is defined as:

$$F(\omega_1, \omega_2) \xrightarrow{FDST} F^{'}(\omega_1, \omega_2) \tag{5.29}$$

Then, the transformed signal in the time domain is

$$C^{'} = c^{'}(x, y) = IFFT[F^{'}(\omega_1, \omega_2)] \tag{5.30}$$

$C^{'}$ is assumed to be identical to C visually while the hidden message in C is unrecoverable in $C^{'}$. The FDST is mainly conducted in the mid-range Frequency area. Roughly, the mid-frequency range is defined as

$$M_c = \{F(\omega_1, \omega_2) | \gamma_1 < \omega_1 < \gamma_2, \quad \delta_1 < \omega_1 < \delta_2\} \tag{5.31}$$

Specifically, the FDST strength operator is obtained by a bandpass filter denoted as

$$O = A \times f_b(\gamma_1, \gamma_2, \delta_1, \delta_2) \tag{5.32}$$

In this place, A is the original FDST strength. By this way, the FDST can be first conducted in the entire frequency area and then the FDST strength operator is multiplied. Therefore, the affect resulted by the FDST in the low and high frequency portions are filtered by the bandpass filter. The choice of the bandpass filter model and the cut-off frequencies determine the FDST boundary in the image.

In order to let the FDST blocks discretely distributed in the frequency domain where the best attacking performance can be achieved, we define the center set of the blocks as $\Phi = \{\Phi_1, \Phi_2, \Phi_3, , \Phi_N\}$ where N denotes the number of the blocks. The first step of the block generation is the block center selection. These N block centers are randomly selected from the $F(\omega_1, \omega_2)$. Then the block center can be expressed as

$$\phi_i = (\omega_{x_i}, \omega_{y_i}) \tag{5.33}$$

Then the block distance which is defined as the distance between two block centers is expressed as

$$d_{ij} = \sqrt{|y_j - y_i|^2 + |x_j - x_i|^2} \tag{5.34}$$

The block shape is irregular and grows from the center to eight different orientations. The block shape can be expressed by the shape matrix

$$S_i = \begin{Bmatrix} s_i^1 & s_i^2 & s_i^3 \\ s_i^4 & 0 & s_i^5 \\ s_i^6 & s_i^7 & s_i^8 \end{Bmatrix} \tag{5.35}$$

An example of the block shape can be shown in the figure below: The blocks may have some overlaps or locate very close to each other. Based on our observation, the more sparse, the better attacking performance can be achieved. So we want to optimize the attacking performance by solving an optimized problem to make the blocks distribute as sparse as possible. Lets fix the shape matrix and move the block center from $\phi_i \rightarrow \phi_i^{'}$ where

$$\phi_i^{'} = (\omega_{x_i}^{'}, \omega_{y_i}^{'}) \tag{5.36}$$

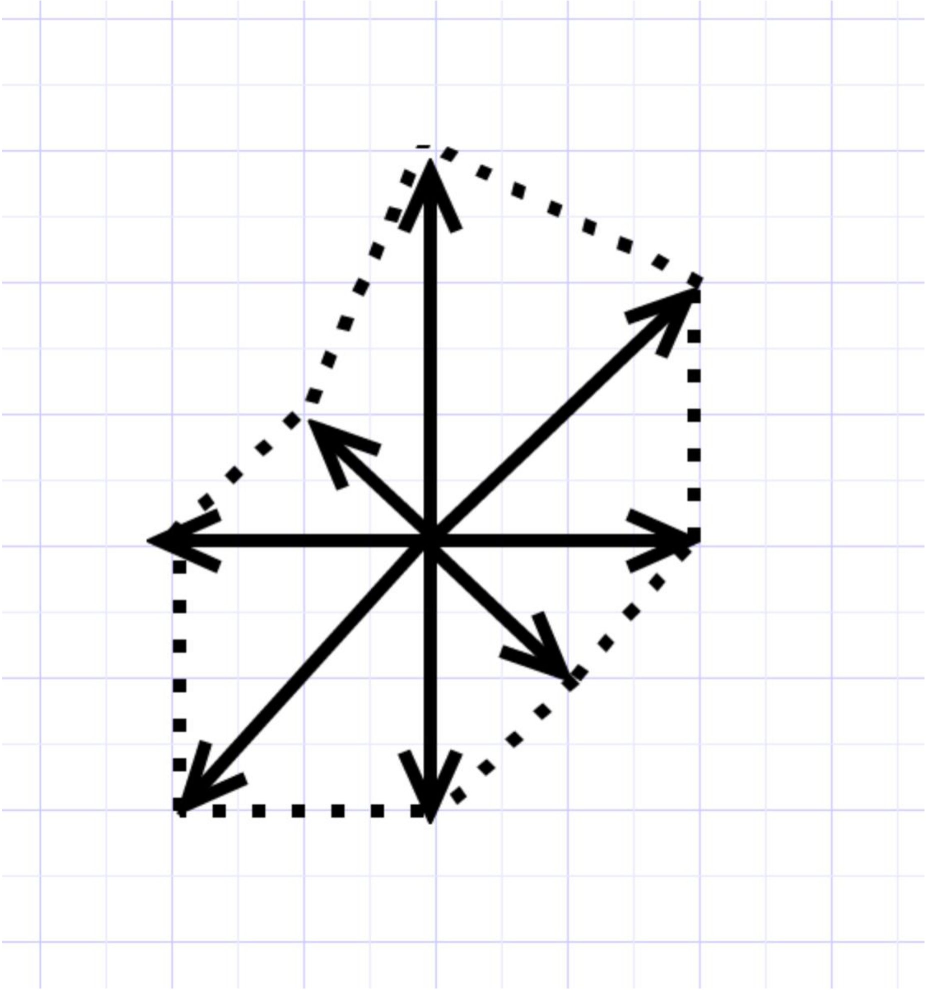

Figure 5.3: An Example of Block Shape in FDST

We define the minimum boundary distance between two blocks as d_{ij}^{b} where it is shown in Figure5.4

The optimization problem can be expressed as

$$\max \sum_{i=0,j=0,i\neq j}^{i=N,j=N} d_{ij}^{b} \tag{5.37}$$

the condition of the optimization problem is that any block must be within the frequency plane.

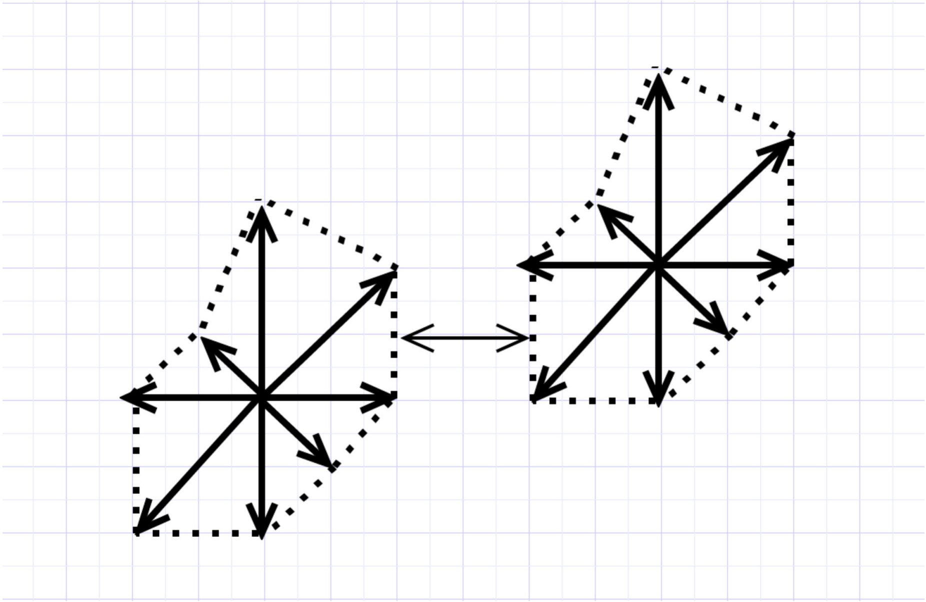

Figure 5.4: Minimum Boundary Distance

In practice, an iterative algorithm is used. We consider two blocks are sparse when the minimum boundary distance is larger than half of the block distance which is represented as

$$d_{ij}^{b} \geq 0.5 d_{ij} \tag{5.38}$$

Firstly, we find the block pairs $[\phi_i, \phi_j]$ whose boundary distance d_{ij}^{b} is smaller than $0.5d_{ij}$. For all those block pairs, we move the block centers towards the opposite directions from each other along the link of those two block centers for one distance unit. Then, we repeat this process until no block pairs can be found. In other words, $d_{ij}^{b} \geq 0.5d_{ij}$ is satisfied for all the blocks. It should note a larger scale rate can be used to define the sparse standard, however more convergence time is required.

The in-block FDST is implemented by the interpolation and resampling.

The interpolated signal can be expressed as

$$F^{'}(\omega_1, \omega_2) = F(\omega_1, \omega_2) * W_L(x, y) \tag{5.39}$$

where $w(x, y)$ is the interpolation window kermal, in this application, the 3-order Lanczos window kernel which 1-D form can be expressed below is used

$$W_L(x, y) = w(x)w(y) \tag{5.40}$$

where

$$w(x) = \begin{cases} 1 & |x| = 0 \\ 3\frac{\sin \pi x \sin \frac{\pi x}{3}}{\pi^2 x^2} & 0 < |x| < 3 \\ 0 & 3 \leq |x| \end{cases} \tag{5.41}$$

After the interpolation, in each block a different resampling rate is adopted.

5.3.2 Video Steganography Attack Extention

The application of the DST in Video signal is presented in Aaron's paper in [2]. Video Signal provides more capacity for the steganography. The steganographic message can be not only embedded in each frame but also can be embedded into the time frame as well.

As a result the Discrete Spring Transform should also be implemented in the time frame so as to remove the steganographic message there.

The basic idea for time-frame Discrete Spring Transform is the same as the 1-D Spring Transform. The time frame is scaled by a time variant rate. This will cause a frame deletion or interpolation in the video signal. When the video is played, a

few frame insertions or missing may not affect the video perceptual quality as the fact in the audio signal.
In each frame of the video signal, it can be treated as the time signal and the proposed image steganography attacking method can be applied.

In order to keep the video duration as identical as the previous signal a part of the frames are interpolated while another part of frames are resampled. The interpolation rate are reciprocal so that the entire number of frames in the video sequences are not changed.

Chapter 6

Experimental Results

6.1 Audio Implementation Simulation Results

The proposed attacking method is similar to the time scale modification. Both methods do not actually remove the steganographic messages but desynchronize the location of the hidden messages so that the messages are not able to be decoded correctly by the intended receiver. One of the advantages over the time scale modification is that our proposed method makes the hidden message more difficult to be recovered by synchronization techniques because of the variability of the stretch rate. As discussed above, the randomness of the stretch rate function makes spring transform a one-way transform. Another advantage over time scale modification is that the cover media can bear a larger stretch rate with the same perceptual quality requirement. It is because the variable stretch rate localizes the changes of the audio signal so that the people are hardly to notice. A larger stretch rate can strengthen the attacking performance to the hidden messages.

The experimental results show our proposed method performs well on some TSM-robust steganography method. It proves the attacking performance of our

Table 6.1: BER of the Hidden Message[1]

	10	**50**	**100**	**500**	**1000**	**2000**
0.02	0.4665	0.3996	0.3292	0.4442	0.4866	0.4911
0.04	**0.5056**	0.4531	**0.5033**	**0.5226**	**0.5045**	0.4989
0.06	0.4688	0.4911	0.4989	**0.5067**	0.4788	0.4810
0.08	0.4754	0.4688	0.4275	**0.5078**	**0.5208**	0.4866
0.1	**0.5379**	**0.5000**	0.4888	**0.5190**	0.4866	0.4710

method outperforms the time scale modification. After the attack, the audio is evaluated by both subjective and objective evaluation methods. It should note that the conventional Signal-to-noise ratio is not fair for our proposed method because our method intends to largely change the numerical value of the audio signal which will cause the decrease of the SNR. Instead, the correlation is used to evaluate the similarity of the attacked audio signal to the original signal.

In the experiments, a TSM-robust audio steganography algorithm[16] is used as the test bench for our proposed attacking methods. Li's method use the point feature to synchronize the audio signal and it achieve a good performance against the time scale modification.

6.1.1 Attacking Performance

For simplicity, the block size is set to be identical and ranging from 10 to 2000 in this experiment. The sampling rate of the audio signal is 44.1kHz. Stretch rate is a random sequence uniformly distributed in $[1 - \frac{A_{max}}{2}, 1 + \frac{A_{max}}{2}]$. The maximum stretch rate A_{max} varies from 0.02 to 0.1 with the step 0.02. 128 bits hidden messages are embedded in the clips. The attacking performance is shown in Figure6.1, Figure6.2,Figure6.3 and Figure6.4. The attacking result is also shown in Table6.1. The Table6.1 and Figure6.1 present the results from the same audio clip. Overall nearly half result points reach a 0.5 BER which means the decoding results

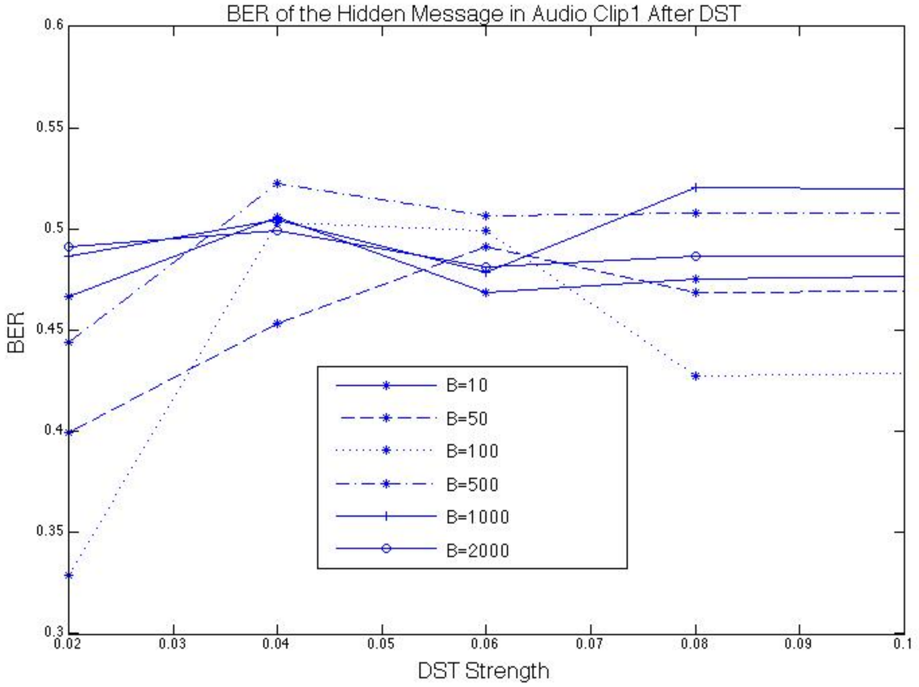

Figure 6.1: BER of Hidden Message in Audio Clip1 after DST[1]

of the attacked hidden message cannot be better than a random guess. According to the results, a medium stretch rate responses a better attacking performance. On other words, the attacking performance will not improve when the stretch rate is higher than a threshold. It is a desirable result because we do not want to make the stretch rate too affect the audio perceptual quality to achieve a better attacking performance. Another observation from the result is that the medium stretch rate achieves a good attacking performance for nearly all block size configuration. So the choice of the block size is not very important when the stretch rate is properly selected. By this way, we can set the block size small to maintain good audio quality.

Similiar results are reported by different audio types in Figure6.2, Figure6.3 and Figure6.4. Figure6.3 and Figure6.4 are male and female speech audio clips.

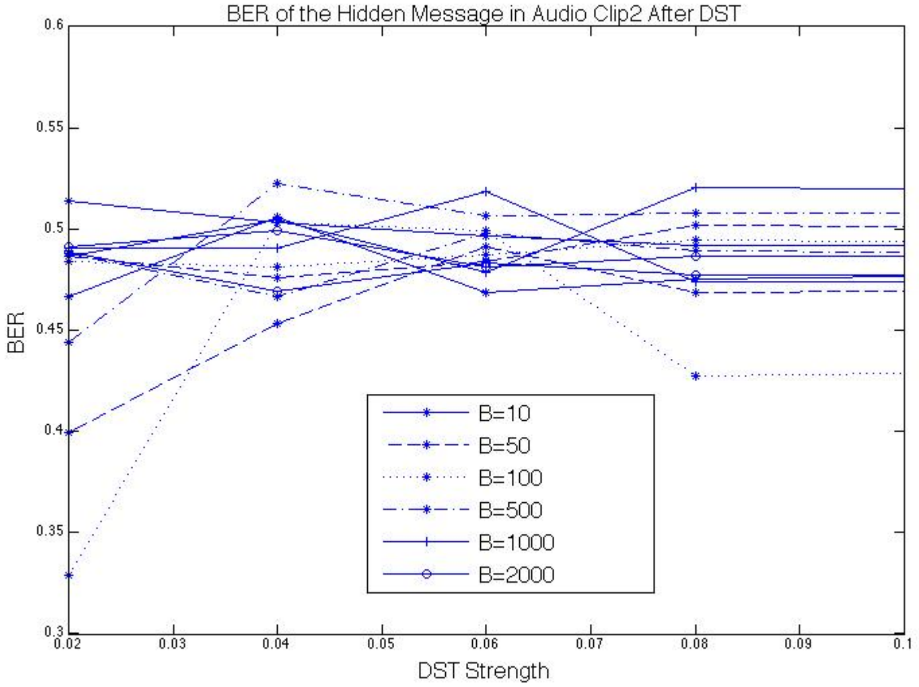

Figure 6.2: BER of Hidden Message in Audio Clip2 after DST[1]

These results show a consistent performance for audio signal. In addition, the speech signal represents a better performance in Figure6.3 and Figure6.4.

6.1.2 Audio Peceptual Quality

6.1.2.1 Objective Method

As mentioned, the SNR is not fair enough to evaluate the audio perceptual quality regarding the audio attacked by our proposed method. For example, for the operation that the audio signal is shifted one sample, the change is too tiny to be noticed however the SNR could be very low because of the position change of every sample. In order to eliminate the fluencies of the position, the cross-correlation is used as a metric to evaluate the audio quality. A higher cross-correlation means

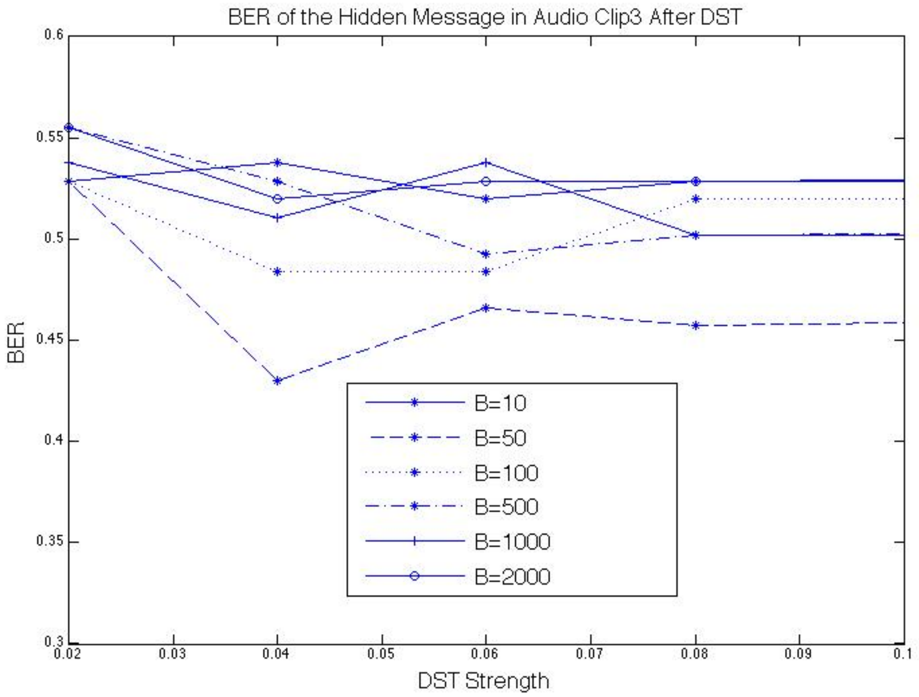

Figure 6.3: BER of Hidden Message in Audio Clip3 after DST[1]

Table 6.2: Cross-correlation between Stego-signal and Attacked signal[1]

	10	**50**	**100**	**500**	**1000**	**2000**
0.02	0.7241	0.6332	0.8616	0.9222	0.8403	0.4180
0.04	0.7923	0.6661	0.6567	0.7817	0.5205	0.3561
0.06	0.5154	0.8400	0.7534	0.6616	0.2690	0.2862
0.08	0.3398	0.6331	0.6768	0.4657	0.1759	0.2233
0.1	0.6853	0.7605	0.3957	0.5363	0.2192	0.1677

more similar the attacked signal compared to the original one. In Table6.2, the cross-correlations are shown in terms of block sizes and stretch rates. Other than the good audio quality shown from the cross-correlation values, Table6.2 also shows the audio quality is relatively independent to the block size and stretch rate. This feature provides us a flexibility to adjust the block size and stretch rate to achieve a better attacking result.

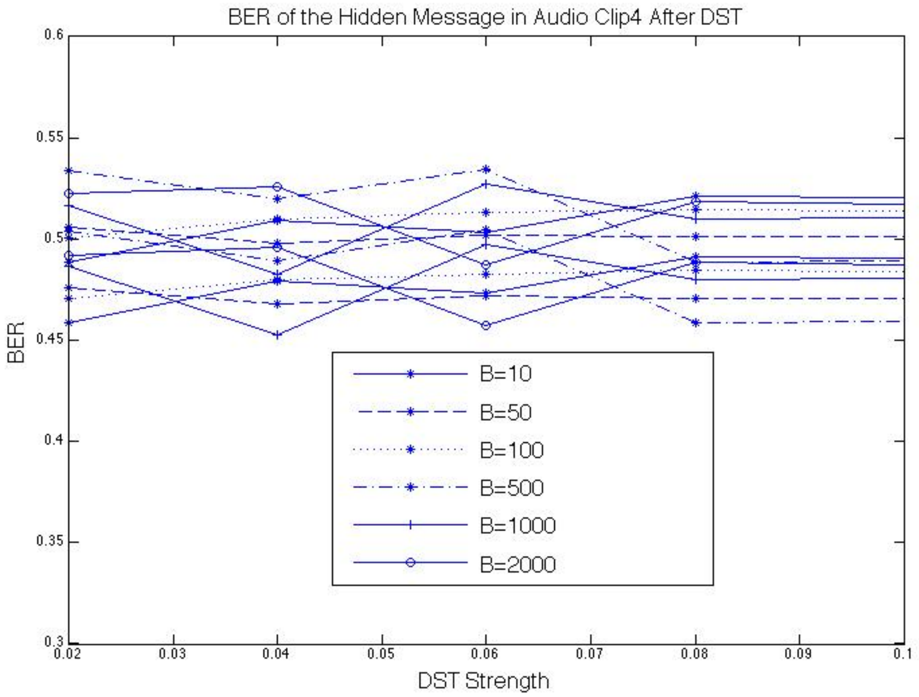

Figure 6.4: BER of Hidden Message in Audio Clip4 after DST[1]

6.1.2.2 Subjective Method

Since our method cultivates the difference between the numerical value and the perceptual quality of the audio signal. The subjective evaluation method seems to be more effective to evaluate the proposed method. A subjective audio quality evaluation based on the SDG (Subjective Diff-Grades) is performed in this experiment. The score is given in Table6.4 based on the criteria in Table6.3. Ten reviewers are assigned the audio clips in terms of different block sizes and stretch rates. As shown in Table6.2, the subjective evaluation shows a similar results in Table6.4 that the block sizes and stretch rates do not affect the audio quality very much. As a result a searching-based method is recommended to find the best audio quality configuration.

Table 6.3: Subjective Diff-Grades[1]

SDG	Description
0.0	imperceptible
-0.1	perceptible, but not annoying
-0.2	slightly annoying
-0.3	annoying
-0.4	very annoying

Table 6.4: Average SDG Grades[1]

	10	**50**	**100**	**500**	**1000**	**2000**
0.02	0.0	0.0	0.0	0.0	0.0	0.0
0.04	0.0	0.0	0.0	0.0	0.0	0.0
0.06	0.0	0.0	0.0	0.0	-0.1	-0.1
0.08	0.0	0.0	0.0	0.0	0.0	-0.1
0.1	0.0	0.0	0.0	0.0	0.0	-0.1

6.2 Image Implementation Simulation Results

6.2.1 Attacking Performance

The spread spectrum steganography is used as test bench to validate our proposed method. In the experiment, 64 bits of information is embedded in the host image by the spread spectrum steganography method. The results show our proposed method works on the steganography in frequency domain as well. We use the curve length based ALT and edge detector based ALT to attack the steganography where the scanning window ranged from 4 to 9. Figure6.5 shows the BER of curve length based method. Figure6.8 shows the BER of edge detector based method. As shown in Figure6.5 and Figure6.8, three different category pictures which are portrait, scenery and still object are tested respectively. Since 64 bits are embedded, when the BER is greater than 0.5 the decoding result of the hidden information will not be better than a random guess since the steganography cannot be obtained with any certainty. The results show, in each method when the scanning window

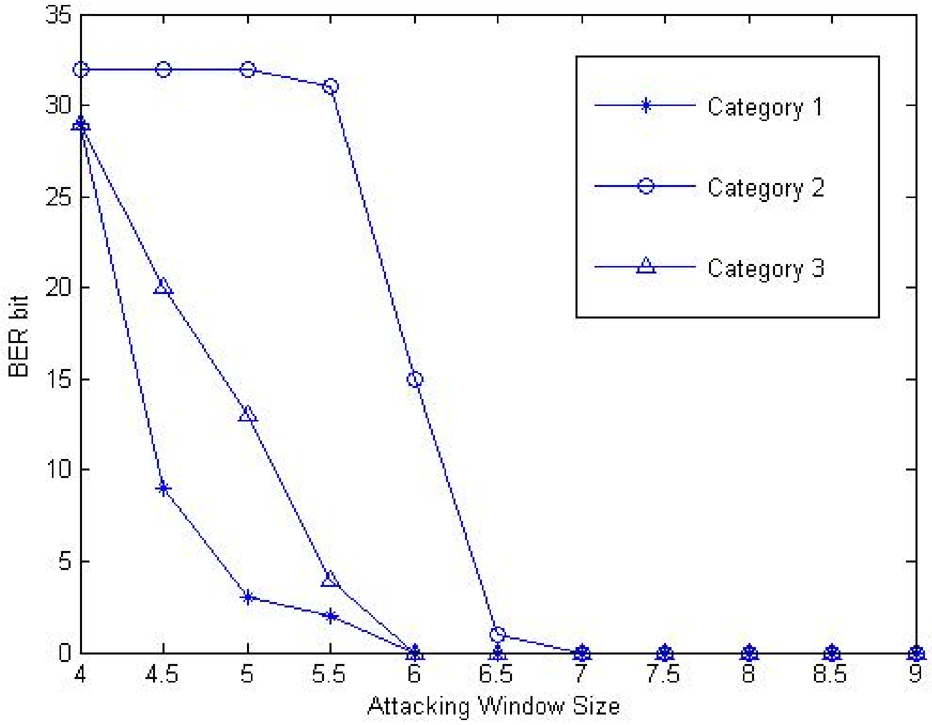

Figure 6.5: Curve Length Method BER

length is 5 that the BER is over 32. Those two kinds of methods do not show much difference in terms of the attacking performance.

6.2.2 Image Perceptual Quality

The image quality is evaluated by PSNR with two attacking methods shown in Figure6.7 and Figure6.8 respectively. A PSNR which is generally greater than 32db is shown in both two figures. The curve length based method is found to be better than the edge detector based method.

A visual result is shown in Figure6.9. Both two results are shown using a window size=6. It is shown no visual difference between the original image.

Since the PSNR can only evaluate the numerical difference of the image we

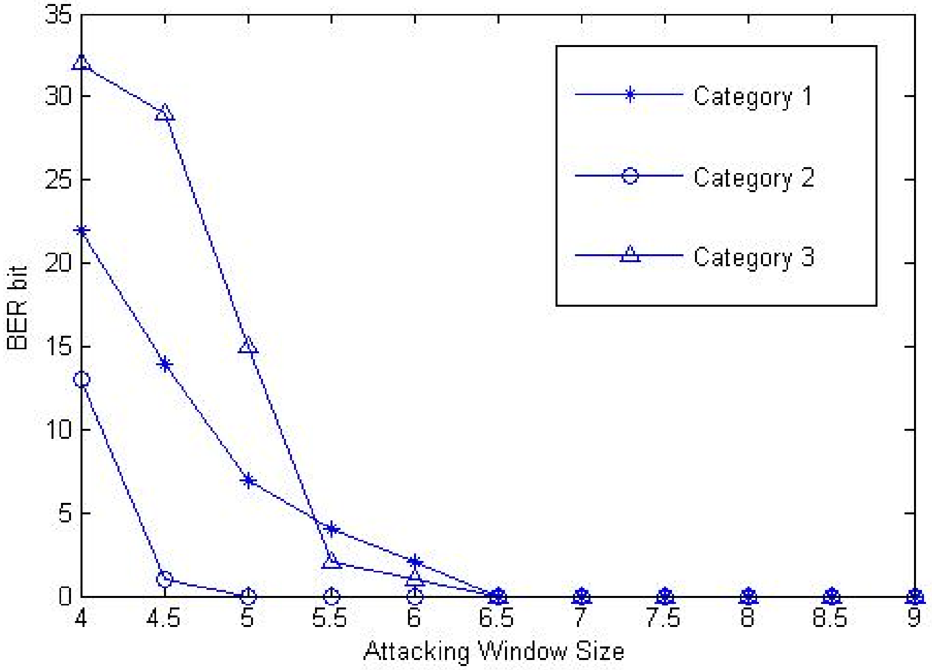

Figure 6.6: Edge Based Method BER

will use a visual quality evaluation method to present the attacking performance. Figure6.10 shows the attacked image of PSNR in 20 dB with our method and the same image with random noise with PSNR in 20 dB. It is evident that at the same PSNR level, the attacked image has a better visual quality. This proves our method exploits the unnoticeable distortions to defeat steganographic data while maintaining high image quality.

6.3 Video Implementation Simulation Results

The results of the video steganography attacking is conducted by Aaron Sharp and the results are also presented in [2].

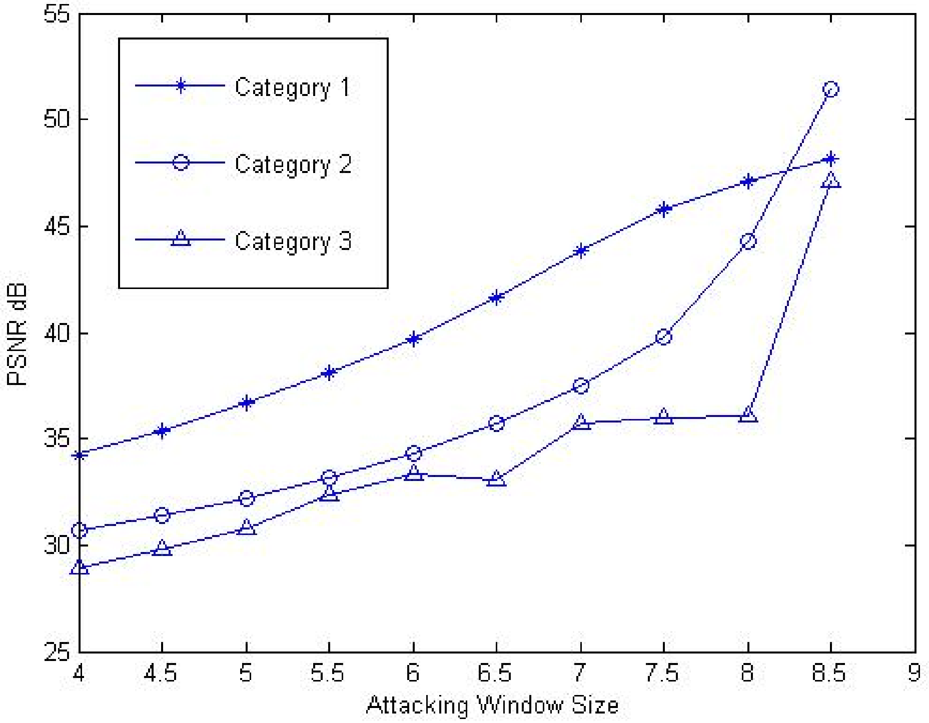

Figure 6.7: Curve Length Method PSNR

6.3.1 Video Attacking Performance

The proposed method is undergoing steganography attack both in image domain and motion vector domain.

6.3.1.1 2D Steganography Attacking

The attacking performance for the steganographic messages hidden in each video frame is shown in Table6.5. The results shows the BER reaches 0.5 which is highlighted in the table.

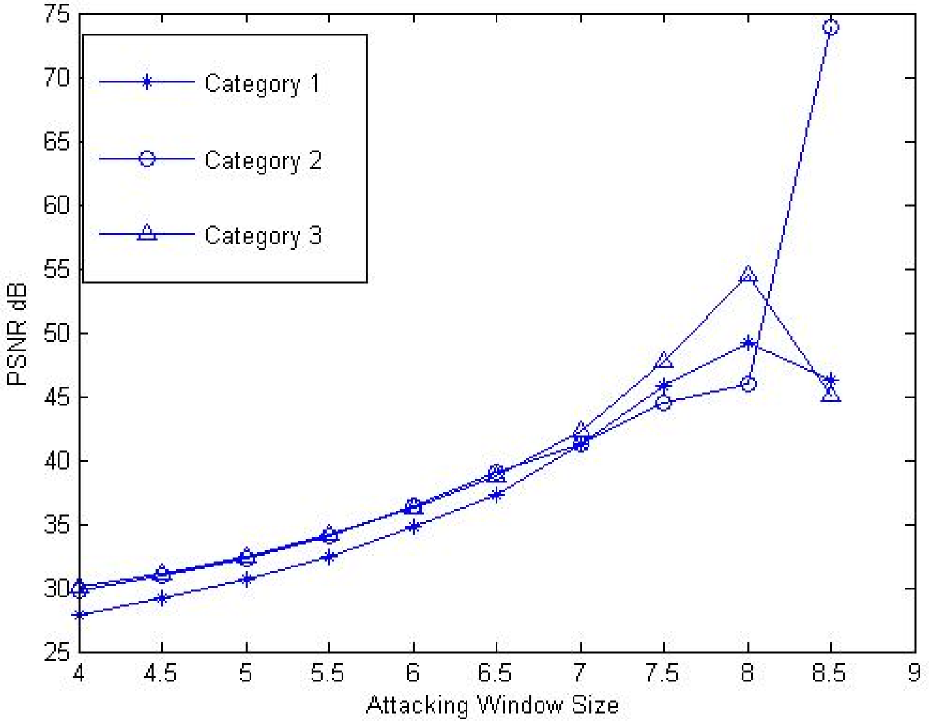

Figure 6.8: Edge Based Method PSNR

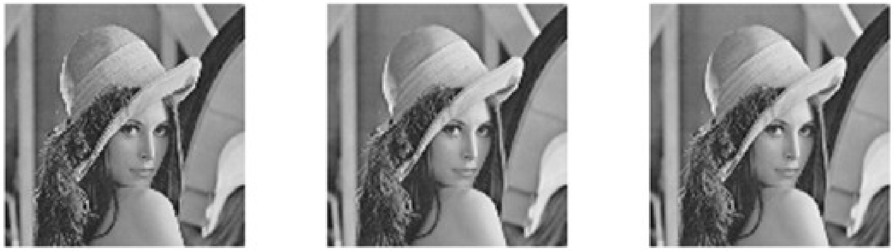

Figure 6.9: Visual Results using Proposed Methods

6.3.1.2 Motion Vector Attacking

Table6.6 shows the attacking performance for the motion vector attack. Like the 2D attack, it also validates the proposed method with the BER greater than 0.5.

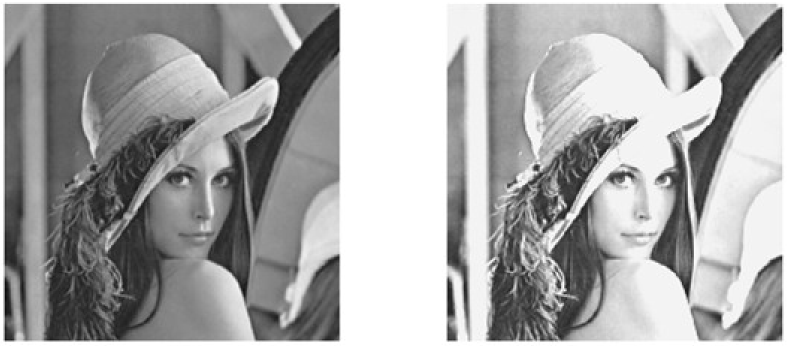

Figure 6.10: Comparison to the Image PSNR=20 dB

		2D Chop (pixels)					
		0	10	20	30	40	50
Time Chop (frames)	0	0.0003	0.1020	0.5351	0.5195	0.5185	0.4611
	5	0.0006	0.1033	0.5364	0.5214	0.5077	0.4531
	10	0.0010	0.1043	0.5418	0.5236	0.5086	0.4557
	15	0.0010	0.1052	0.5402	0.5293	0.5140	0.4534
	20	0.0010	0.1043	0.5450	0.5332	0.5057	0.4544
	25	0.0013	0.1087	0.5421	0.5360	0.5061	0.4585

Table 6.5: 2D Steganography BER[2]

		2D Chop (pixels)					
		0	10	20	30	40	50
Time Chop (frames)	0	0.0000	0.0000	0.3891	0.4713	0.5161	0.5146
	5	0.5197	0.5257	0.5106	0.5338	0.4970	0.5015
	10	0.5423	0.5302	0.5474	0.4945	0.5121	0.5287
	15	0.4990	0.5318	0.5484	0.5378	0.5484	0.5237
	20	0.5302	0.5635	0.5181	0.5297	0.5297	0.5156
	25	0.5514	0.5559	0.5413	0.5398	0.4950	0.5186

Table 6.6: Motion Vector Steganography BER[2]

6.3.2 Video Quality

Table 6.7 shows the PSNR for the video signal after the multi-dimensional steganography attack. It indicates the video quality is acceptable with the PSNR over 30dB is most of the cases. The optimal attack in terms of the video quality is the attack with 20 pixels for 2D chop and 10 frames for time chop where it is highlighted in the table.

		2D Chop (pixels)					
		0	10	20	30	40	50
Time Chop (frames)	0	∞	30.887	29.686	29.339	29.144	29.018
	5	34.250	30.136	29.377	29.137	28.994	28.898
	10	32.238	29.765	29.223	29.036	28.913	28.828
	15	31.470	29.594	29.141	28.979	28.868	28.792
	20	31.090	29.496	29.092	28.943	28.838	28.769
	25	30.864	29.435	29.059	28.919	28.817	28.752

Table 6.7: Motion Vector Steganography PSNR (db)[2]

Chapter 7

Conclusion

Steganography is an emerging technology in this century where the multimedia is everywhere in our daily life. The rapid development of the Internet and wireless technology provides general public much more conveniences and spaces to exchange multimedia data. Many multimedia-based applications such as Youtube, Hulu become popular owing to the dramatically increase of the network bandwidth. The portable devices and wireless technology make the multimedia exchanges happen everywhere in the world. Beside the convenience, a big security concern arises from the multimedia applications. Unlike the text data, multimedia data provides a much more capacity for steganography. It also makes the steganography more robust than ever.

The feature of the steganography facilitates the malicious or illegal use of the steganography. Some effective countermeasures to the steganography are very necessary in this scenario. The existing passive steganography attack methods are proven to be ineffective provided the properties of the Internet environments where multiple steganography methods can be used. It is also not efficient because most of the Internet data exchange is innocent. An active warden attack is needed in this case. This method is based on the aggreement that the network administrator can modify the data on the Internet whereas the content of the data cannot be changed.

A generic attacking method is proposed in this thesis. It can be implemented in audio, image and video steganography applications respectively. The proposed method is built on the fact that the multimedia signal has a gap between the numerical values and the perceptual effects. specifically, some certain types of changes happened in certain areas of the multimedia signal is not very significant in terms of the human perceptual systems. A typical type of change used in this thesis is the location change in image or frequency change in audio. The area chosen to apply this kind of change in image is the plain area where fewer attentions are attracted.

A Discrete Spring Transform is proposed based on this investigation. This transform can be generalized as a perceptual-invariant transform. It means the perceptual effect will not be largely changed by Discrete Spring Transform. However another feature of this transform is that the numerical value can be greatly changed in the meantime. In all, the gap between the numerical value of the multimedia signal and the perceptual effect is exploited by the Spring Transform. It is proven to be effective in steganography attack.

Some implementations respect to different signal formats are proposed in this thesis as well. The implementation considers the requirements of two aspects. First is the quality of the cover media can not be distorted very much. Second is the steganographic message can be effectively removed. The implementations are proven to be effective by the numerical results. It also shows the advantages over some similar schemes. This attack not only works in spatial domain but also on frequency domain.

In future work, perceptual quality evaluation for the image and audio is a very important topic. It will not only be used to justify our results but also further reveal the relations between the perceptual effect and numerical values of the multimedia signal. It is worthwhile to study how large this gap is. This gap makes the steganography possible. This gap also makes the countermeasures possible. In all, this is just a start.

Bibliography

[1] Qilin Qi, Aaron Sharp, Dongming Peng, Yaoqing Yang, and Hamid Sharif. An active audio steganography attacking method using discrete spring transform. In *2013 IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 3456–3460, 2013. (document), 2.2.2, 3.3, 3.1, 4.1, 4.1, 6.1, 6.1, 6.2, 6.3, 6.2, 6.4, 6.3, 6.4

[2] A. Sharp, Qilin Qi, Yaoqing Yang, Dongming Peng, and H. Sharif. A novel active warden steganographic attack for next-generation steganography. In *Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International*, pages 1138–1143, 2013. (document), 2.2.2, 5.3.2, 6.3, 6.5, 6.6, 6.7

[3] N. Provos and P. Honeyman. Hide and seek: an introduction to steganography. *Security Privacy, IEEE*, 1(3):32–44, 2003. 1

[4] E.F. Brickell and A.M. Odlyzko. Cryptanalysis: a survey of recent results. *Proceedings of the IEEE*, 76(5):578–593, 1988. 1

[5] Rajarathnam Ch, Mehdi Kharrazi, and Nasir Memon. Image steganography and steganalysis: Concepts and practice. 1, 2.2.1, 2.2.2

[6] W. Bender, D. Gruhl, N. Morimoto, and A. Lu. Techniques for data hiding. *IBM systems journal*, 35(3.4):313–336, 1996. 2.2.1

[7] H.S. Malvar and D.A.F. Florencio. Improved spread spectrum: a new modulation technique for robust watermarking. *Signal Processing, IEEE Transactions on*, 51(4):898 – 905, apr 2003. 2.2.1

[8] I.J. Cox, J. Kilian, F.T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *Image Processing, IEEE Transactions on*, 6(12):1673 –1687, dec 1997. 2.2.1

[9] B. Chen and G.W. Wornell. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. *Information Theory, IEEE Transactions on*, 47(4):1423 –1443, may 2001. 2.2.1

[10] Hyen O Oh, Jong Won Seok, Jin Woo Hong, and Dae Hee Youn. New echo embedding technique for robust and imperceptible audio watermarking. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, volume 3, pages 1341 –1344 vol.3, 2001. 2.2.1

[11] M.A. Akhaee, M.J. Saberian, S. Feizi, and F. Marvasti. Robust audio data hiding using correlated quantization with histogram-based detector. *Multimedia, IEEE Transactions on*, 11(5):834 –842, aug. 2009. 2.2.1

[12] O.T.-C. Chen and Wen-Chih Wu. Highly robust, secure, and perceptual-quality echo hiding scheme. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(3):629 –638, march 2008. 2.2.1

[13] Xiao-Ming Chen, G. Doerr, M. Arnold, and P.G. Baum. Efficient coherent phase quantization for audio watermarking. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 1844 –1847, may 2011. 2.2.1

[14] Kaliappan Gopalan. Audio steganography using bit modification. In *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*, volume 1, pages I–629. IEEE, 2003. 2.2.1

[15] Mohamed A Ahmed, Miss Laiha Mat Kiah, BB Zaidan, and AA Zaidan. A novel embedding method to increase capacity and robustness of low-bit encoding audio steganography technique using noise gate software logic algorithm. *Journal of Applied Science*, 10(1):59–64, 2010. 2.2.1

[16] Wei Li, Xiangyang Xue, and Peizhong Lu. Localized audio watermarking technique robust against time-scale modification. *Multimedia, IEEE Transactions on*, 8(1):60 – 69, feb. 2006. 2.2.1, 6.1

[17] Xiang-Yang Wang and Hong Zhao. A novel synchronization invariant audio watermarking scheme based on dwt and dct. *Signal Processing, IEEE Transactions on*, 54(12):4835 –4840, dec. 2006. 2.2.1

[18] Shaoquan Wu, Jiwu Huang, Daren Huang, and Y.Q. Shi. Efficiently self-synchronized audio watermarking for assured audio data transmission. *Broadcasting, IEEE Transactions on*, 51(1):69 – 76, march 2005. 2.2.1

[19] Wu Xiangsheng and Yin Dicheng. A new watermarking algorithm for withstanding geometric attacks based on image content. In *Image and Graphics, 2009. ICIG '09. Fifth International Conference on*, pages 57–62, 2009. 2.2.1

[20] Ingemar Cox, Matthew Miller, Jeffrey Bloom, Jessica Fridrich, and Ton Kalker. *Digital watermarking and steganography*. Morgan Kaufmann, 2007. 2.2.1

[21] Abbas Cheddad, Joan Condell, Kevin Curran, and Paul Mc Kevitt. Digital image steganography: Survey and analysis of current methods. *Signal Processing*, 90(3):727–752, 2010. 2.2.1

[22] Yiwei Wang, John F Doherty, and Robert E Van Dyck. A wavelet-based watermarking algorithm for ownership verification of digital images. *Image Processing, IEEE Transactions on*, 11(2):77–88, 2002. 2.2.1

[23] Ming-Shing Hsieh, Din-Chang Tseng, and Yong-Huai Huang. Hiding digital watermarks using multiresolution wavelet transform. *Industrial Electronics, IEEE Transactions on*, 48(5):875–882, 2001. 2.2.1

[24] Jong Ryul Kim and Young Shik Moon. A robust wavelet-based digital watermarking using level-adaptive thresholding. In *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, volume 2, pages 226–230. IEEE, 1999. 2.2.1

[25] Zhao Dawei, Chen Guanrong, and Liu Wenbo. A chaos-based robust wavelet-domain watermarking algorithm. *Chaos, Solitons & Fractals*, 22(1):47–54, 2004. 2.2.1

[26] Dong Zheng, Yan Liu, Jiying Zhao, and Abdulmotaleb El Saddik. A survey of rst invariant image watermarking algorithms. *ACM Computing Surveys (CSUR)*, 39(2):5, 2007. 2.2.1

[27] Dong Zheng, Jiying Zhao, and Abdulmotaleb El Saddik. Rst-invariant digital image watermarking based on log-polar mapping and phase correlation. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(8):753–765, 2003. 2.2.1

[28] Min Wu, Matt L Miller, Jeffrey A Bloom, and Ingemar J Cox. A rotation, scale and translation resilient public watermark. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 4, pages 2065–vol. IEEE, 1999. 2.2.1

[29] Hyung Shin Kim and Heung-Kyu Lee. Invariant image watermark using zernike moments. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(8):766–775, 2003. 2.2.1

[30] Xiang Wang, Xiaolong Li, Bin Yang, and Zongming Guo. Efficient generalized integer transform for reversible watermarking. *Signal Processing Letters, IEEE*, 17(6):567–570, 2010. 2.2.1

[31] Shaowei Weng, Yao Zhao, Jeng-Shyang Pan, and Rongrong Ni. Reversible watermarking based on invariability and adjustment on pixel pairs. *Signal Processing Letters, IEEE*, 15:721–724, 2008. 2.2.1

[32] Adnan M Alattar. Reversible watermark using difference expansion of triplets. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, volume 1, pages I–501. IEEE, 2003. 2.2.1

[33] Adnan M Alattar. Reversible watermark using the difference expansion of a generalized integer transform. *Image Processing, IEEE Transactions on*, 13(8):1147–1156, 2004. 2.2.1

[34] Adnan M Alattar. Reversible watermark using difference expansion of quads. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 3, pages iii–377. IEEE, 2004. 2.2.1

[35] JinHa Hwang, JongWeon Kim, and JongUk Choi. A reversible watermarking based on histogram shifting. In *Digital Watermarking*, pages 348–361. Springer, 2006. 2.2.1

[36] ShengDun Hu and U. KinTak. A novel video steganography based on non-uniform rectangular partition. In *Computational Science and Engineering (CSE), 2011 IEEE 14th International Conference on*, pages 57 –61, aug. 2011. 2.2.1

[37] Bin Liu, Fenlin Liu, Chunfang Yang, and Yifeng Sun. Secure steganography in compressed video bitstreams. In *Availability, Reliability and Security, 2008. ARES 08. Third International Conference on*, pages 1382 –1387, march 2008. 2.2.1

[38] K. Raghavendra and K.R. Chetan. A blind and robust watermarking scheme with scrambled watermark for video authentication. In *Internet Multimedia Services Architecture and Applications (IMSAA), 2009 IEEE International Conference on*, pages 1 –6, dec. 2009. 2.2.1

[39] A.T. Sharp, J. Devaney, and A.E. Steiner. Digital video authentication with motion vector watermarking. In *Signal Processing and Communication Systems (ICSPCS), 2010 4th International Conference on*, pages 1 –4, dec. 2010. 2.2.1

[40] N. Mohaghegh and O. Fatemi. H.264 copyright protection with motion vector watermarking. In *Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on*, pages 1384 –1389, july 2008. 2.2.1

[41] Zina Liu, Huaqing Liang, Xinxin Niu, and YixianYang. A robust video watermarking in motion vectors. In *Signal Processing, 2004. Proceedings. ICSP '04. 2004 7th International Conference on*, volume 3, pages 2358 – 2361 vol.3, aug.-4 sept. 2004. 2.2.1

[42] A. Ceddillo-Hernandez, M. Nakano-Miyatake, L. Rojas-Cardenas, and H. Perez-Meana. Robust video watermarking using perceptual information and motion vector. In *Circuits and Systems, 2007. NEWCAS 2007. IEEE Northeast Workshop on*, pages 811 –814, aug. 2007. 2.2.1

[43] Sorina Dumitrescu, Xiaolin Wu, and Zhe Wang. Detection of lsb steganography via sample pair analysis. In FabienA.P. Petitcolas, editor, *Information Hiding*, volume 2578 of *Lecture Notes in Computer Science*, pages 355–372. Springer Berlin Heidelberg, 2003. 2.2.2

[44] Jessica Fridrich. Feature-based steganalysis for jpeg images and its implications for future design of steganographic schemes. In *in Proc. Inf. Hiding Workshop, Springer LNCS*, pages 67–81. 2.2.2

[45] J. Fridrich, M. Goljan, and Rui Du. Detecting lsb steganography in color, and gray-scale images. *MultiMedia, IEEE*, 8(4):22–28, 2001. 2.2.2

[46] Jessica Fridrich, Miroslav Goljan, and Rui Du. Reliable detection of lsb steganography in color and grayscale images. In *Proceedings of the 2001 Workshop on Multimedia and Security: New Challenges*, MM&#38;Sec '01, pages 27–30, New York, NY, USA, 2001. ACM. 2.2.2

[47] Bin Li, Yanmei Fang, and Jiwu Huang. Steganalysis of multiple-base notational system steganography. *Signal Processing Letters, IEEE*, 15:493–496, 2008. 2.2.2

[48] Zhuo Li, Kuijun Lu, Xianting Zeng, and Xuezeng Pan. Feature-based steganalysis for jpeg images. In *Digital Image Processing, 2009 International Conference on*, pages 76–80, 2009. 2.2.2

[49] Siwei Lyu and Hany Farid. Detecting hidden messages using higher-order statistics and support vector machines. In *In 5th International Workshop on Information Hiding*, pages 340–354. Springer-Verlag, 2002. 2.2.2

[50] Hafiz Malik, K. P. Subbalakshmi, and Rajarathnam Chandramouli. Steganalysis of gim-based data hiding using kernel density estimation. In *Proceedings of the 9th Workshop on Multimedia & Security*, MM&#38;Sec '07, pages 149–160, New York, NY, USA, 2007. ACM. 2.2.2

[51] Andreas Westfeld. Generic adoption of spatial steganalysis to transformed domain. In *Information Hiding*, pages 161–177. Springer, 2008. 2.2.2

[52] Yun Q Shi, Guorong Xuan, Chengyun Yang, Jianjiong Gao, Zhenping Zhang, Peiqi Chai, Dekun Zou, Chunhua Chen, and Wen Chen. Effective steganalysis based on statistical moments of wavelet characteristic function. In *Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on*, volume 1, pages 768–773. IEEE, 2005. 2.2.2

[53] Darko Kirovski and Fabien A.P. Petitcolas. Replacement attack on arbitrary watermarking systems. In *ACM Workshop on Digital Rights Management*, pages 177–189. SpringerVerlag, 2003. 2.2.2

[54] D. Kirovski and Fabien A P Petitcolas. Blind pattern matching attack on watermarking systems. *Signal Processing, IEEE Transactions on*, 51(4):1045–1053, 2003. 2.2.2

[55] A.H. Taherinia and M. Jamzad. A new adaptive watermarking attack in wavelet domain. In *Multimedia, Signal Processing and Communication Technologies, 2009. IMPACT '09. International*, pages 320–323, 2009. 2.2.2

[56] S. Voloshynovskiy, S. Pereira, V. Iquise, and T. Pun. Attack modelling: Towards a second generation watermarking benchmark. In *Signal Processing, Special Issue on Information Theoretic Issues in Digital Watermarking*, pages 1177–1214, 2001. 2.2.2

[57] Joseph A O'Sullivan, Pierre Moulin, and J Mark Ettinger. Information theoretic analysis of steganography. In *Information Theory, 1998. Proceedings. 1998 IEEE International Symposium on*, page 297. IEEE, 1998. 2.2.2

[58] Chengqian Zhang, Yuting Su, and Chuntian Zhang. A new video steganalysis algorithm against motion vector steganography. In *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on*, pages 1 –4, oct. 2008. 2.2.2

[59] Yuting Su, Chengqian Zhang, and Chuntian Zhang. A video steganalytic algorithm against motion-vector-based steganography. *Signal Process.*, 91(8):1901–1909, August 2011. 2.2.2

[60] D.C. Hood and M.A. Finkelstein. *Handbook of perception and human performance*, chapter Sensitivity to light. Wiley-Interscience, 1986. 2.2.2

[61] Michael P. Eckert and Andrew P. Bradley. Perceptual quality metrics applied to still image compression. *Signal Processing*, 70(3):177 – 200, 1998. 2.2.2

[62] Aaron Sharp, Qilin Qi, Dongming Peng, Yaoqing Yang, and Hamid Sharif. A video steganography attack using multi-dimensional discrete spring transform. *IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 2013. 2.2.2

[63] Aaron Sharp, Qilin Qi, Dongming Peng, Yaoqing Yang, and Hamid Sharif. Frequency domain discrete spring transform: A novel frequency domain steganographic attack. *Submitted to 9th IEEE/IET InternationalCommunication Systems, Networks and Digital Signal Processing CSNDSP*, 2014. 2.2.2

[64] Qilin Qi, Aaron Sharp, Dongming Peng, Yaoqing Yang, and Hamid Sharif. Steganography attack based on discrete spring transform and image geometrization. *To be submitted*, 2013. 2.2.2

[65] Qilin Qi, Aaron Sharp, Dongming Peng, Yaoqing Yang, and Hamid Sharif. Image steganography attack based on discrete spring transform and image geometrization. *To be submitted*, 2013. 2.2.2

[66] A. Sharp, Qilin Qi, Yaoqing Yang, Dongming Peng, and H. Sharif. Discrete spring transform (dst) based active warden attack for next-generation steganography. *under the review of Wireless Communications and Mobile Computing, Wiley*, Oct. 2013. 2.2.2