國 立 中 央 大 學

通 訊 工 程 學 系

碩 士 論 文

基於特徵值空間分解之影像認證系統

Image Authentication System based on Eigenvalue Decomposition

指導教授：張 寶 基 博士

研 究 生：梁 凱 雯

中華民國九十三年七月

國立中央大學圖書館
碩博士論文電子檔授權書

(93 年 5 月最新修正版)

本授權書所授權之論文全文電子檔，為本人於國立中央大學，撰寫之碩/博士學位論文。(以下請擇一勾選)

(v) **同意** (立即開放)

(　) **同意** (一年後開放)，原因是：_____

(　) **同意** (二年後開放)，原因是：_____

(　) **不同意**，原因是：_____

以非專屬、無償授權國立中央大學圖書館與國家圖書館，基於推動讀者間「資源共享、互惠合作」之理念，於回饋社會與學術研究之目的，得不限地域、時間與次數，以紙本、微縮、光碟及其它各種方法將上列論文收錄、重製、公開陳列、與發行，或再授權他人以各種方法重製與利用，並得將數位化之上列論文與論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

研究生簽名: 梁凱雯

論文名稱: 基於特徵值空間分解之影像認證系統

指導教授姓名：張寶基

系所 ：通訊工程 所 □ 博士 ☑ 碩士班

學號：91523053

日期：民國 93 年 7 月 7 日

備註：

1. 本授權書請填寫並**親筆**簽名後，裝訂於各紙本論文封面後之次頁（全文電子檔內之授權書簽名，可用電腦打字代替）。
2. 請加印一份單張之授權書，填寫並親筆簽名後，於辦理離校時交圖書館（以統一代轉寄給國家圖書館）。
3. 讀者基於個人非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。

國立中央大學碩士班研究生

論文指導教授推薦書

通訊工程 學系/研究所 梁凱雯 研究生

所提之論文

基於特徵值空間分解之影像認證系統

係由本人指導撰述，同意提付審查。

指導教授 張寶基 (簽章)

九十三 年 七 月 六 日

國立中央大學碩士班研究生

論文口試委員審定書

通訊工程 學系/研究所 梁凱雯 研究生

所提之論文

基於特徵值空間分解之影像認證系統

經本委員會審議，認定符合碩士資格標準。

學位考試委員會召集人 張隆紋

委　　員 張寶基

呂俊賢

范國清

許秋婷 （簽章）

中 華 民 國　93　年　7　月　6　日

基於特徵值空間分解之影像認證系統

摘要

在本篇論文中，我們提出基於特徵值空間分解影像認證系統，為期達到對一定範圍內之失真容忍度及區域性認證，此系統以區塊為單位計算影像係數值之統計特性，將區塊內之影像信號作特徵空間之分解，擷取其特徵值作為代表各區塊之數位簽章的來源，加密後傳送，接收端將解密後的特徵值用以認證所收到之影像資訊。

我們將區塊先分成大小相同之子區塊，每一子區塊視為一隨機變數，各子區塊的大小作隨機變數之信號觀測空間以計算其期望值，進而對得到之信號空間作特徵值分解，取其最重要之一到數個特徵值作為一區塊之特徵代表。將取出之特徵代表經過量化後得到對應於各區塊之數位簽章，在認證端計算特徵之方式均與傳送端相同，唯在認證端的量化過程中在各量化區間的邊界處設一模糊區域用以容忍影像區塊之特徵值之小幅度變化。

實驗結果證明此系統以區塊為單位並以特徵值擷取作為數位簽章的方法對於經過常見之影像處理後所造成的失真能達到一定的容忍度，並對於有意義之影像竄改能有高偵測率。

Image Authentication System based on Eigenvalue Decomposition

Abstract

In this thesis, we propose an image authentication system based on eigenvalue decomposition. In order to be incidental distortion tolerant and localization dominating, the feature is extracted on block based.

The feature extracted from each block is depended on the statistical characteristic inside one block. Each block is first divided into sub-blocks with the same size. Each sub-block is considered as an observation signal space of a random variable. Each block has random variables that match the number of sub-blocks. The eigenvalue decomposition operation is done within one image block. The feature is produced from the dominated eigenvalues. The feature is then quantized to generate the final signature for each block.

In the receiver end, the feature of one image block is extracted from the same process in the transmission end, but it is quantized with a neutral zone. The neutral zone can tolerate the small change of the feature caused by the incidental distortions.

The simulation results show that the proposed system works well in tolerating some kinds of image processing and achieves high detection accuracy.

誌　謝

口試前幾天的一個下午，意外的看見全景的彩虹，據說看見全虹的人會有很好的運氣，坐在這裡打字的我，此刻是真心的相信了。

最感謝的是張寶基老師，接納我從零開始的學習背景，給予我機會充實自己，引領我進入一個更寬廣的知識領域，不僅在學術研究，更在立身處世之上為我立下一個完美的典範，典範無法模仿，卻在我心中時時提醒自己，還要，還要更努力。再來感謝大德學長不厭其煩在我的研究生活之中每一個跌跌撞撞的時刻拉我一把，給予我最多的包容與提攜。幸運的我在應當辛苦萬分的研究路上遇見了最好的夥伴，康瑋，逸隆，宗紘，培智，茗光與意瑄，感謝個個身懷絕技的你們從未嫌棄我的遲鈍與散漫，願意與我齊肩並進。也謝謝可愛的學弟們默默的在背後加油打氣。

最多的支持與關心來自我最親愛的家人，我年過八十卻還不忘在口試當天幫我注意衣著的好婆婆，媽咪和爸爸，遠在對岸的小舅，從四歲到二十二歲的弟弟妹妹們，你們支撐了我所有的勇氣。還有像彩虹一樣意外出現的凡碩，卻在我沮喪和失意的時候，給了我最多的陪伴與鼓勵。

人說如果沒有天空，又怎麼會有星星的出現，只是如果沒有星星，那這個天空又怎麼會這麼漂亮？把我的人生舞台當作一片天空來看，那所有的你們，就是那最耀眼的星星群，照亮了我的生命。

最後，僅以本論文獻給所有關心我與我所關心的人。

Contents

List of Figures

Chapter 1 Introduction

This chapter will introduce the motivation of this thesis and the related researches in the image authentication area. The organizations of the thesis will also be given in section 1.3.

1.1 Motivation

The incoming digital world changes the way that people exchange information. The information distributed and stored in digital form is becoming a common practice. Because of the rapid user growth and technological development of the Internet, the digitized media content is becoming more and more important as the days move on.

However, due to the popularity of the Internet and characteristics of digital signals, circumstantial problems are also on the rise. Through the Internet, people get and distribute information easier than the past. And for digital contents, the manipulations and modifications are easy to be done. The various concepts of protecting the digital contents induce researches over different issues.

Several considerations were put on the protection of digital images, IPR (Intelligent Property Right), authority and integrity of images. The digital watermarking techniques were developed to protect the IPR and the public and private key algorithm is commonly used to verify the authority in transmission.

Another category, which this thesis works on, is the content authentication techniques that prevent the modifications of contents by any illegal party. The digital contents inclusive of the digital image, video and audio contents suffer from the fact that any user could edit them using multimedia software easily. The modifications to the content will cause the misunderstanding of the original messages. The user may also manipulate or modify these contents for illegal uses. The importance of the integrity of digital contents grows by days. According to different applications, the forgery tolerances are different. Sometimes the exactly accuracy is asked for situations such as an evidence dependent usage or medical usage. But in other cases, only the parts of the contents or the rough of the main idea of the message is concerned. Different considerations influence the design of the authentication system.

This thesis works on the image authentication under considerations of image integrity, incidental distortions tolerance and localization.

The image integrity concerns about whether the image content had been modified. The modifications may be totally or partially done to the whole image. The modifications may or may not change the principal of the image. For example, the noise liked change will not influence the information expressions of an image, but an object replacement or positions exchange of the items in an image will confuse the original subjectivity. The users who receive the image after the modifications will have no idea about the origin. The misusage may be introduced

from the misunderstanding. For this reason, any legal user through the communication channel should have the ability to authenticate the received image to verify the integrity of this image.

As far as the importance of the integrity is concerned, the tamper detection plays the main role of the image authentication process. However, due to the characteristics of digital image, the digital images will undergo several image processing operations most of the time for applications and storage requirements. The common image processing operations such as lossless compression, image enhancement, filtering and blurring will not change the meaning of the original image although they can still cause some distortions. When our authentication system is designed to apply to detect the meaningful modifications that change the idea of the image, the distortions caused by the common image processing operations should not influence the authentication results. Therefore, in this kind of scheme, the characteristics of distortions caused by operations done to the image should be considered while designing. The image processing operations are often associated with the transformations or quantization steps for different purposes. Under those linear or nonlinear operations, the statistic characteristics may be changed a little to maintain the perspective of view. The authentication system could be designed according to the characteristic movements to tolerate the incidental distortions.

In this thesis, the eigenvalue decomposition is taken as the based

characteristics. The eigenvalue computes the statistics of signals that represents the projection quantities of the orthogonal vector spaces. The maximum eigenvalue is taken as the energy compaction of image signals. The common signal processing operations preserving the perceptual of an image will only influence the eigenvalue slightly. The variant level of the eigenvalue is used to distinguish the modifications between the image processing operations and malicious tampering that change the meaning of the original image content.

For the localization ability, the block based concept is the simplest way to achieve our goal. Our work is performed on this basis also. But in the block based concern with the incidental distortion tolerance, the accuracy of the localization and the tolerance of distortion in one block have been the trade-off related situation. The system proposed in this thesis is to reach more accuracy in the localization ability and effect less in the tolerance of incidental distortion.

1.2 Image Authentication Related Researches

There have been tremendous enthusiasms in the image authentication researching area since the late 1990's decade. Many techniques were proposed to authenticate digital images based on the different concepts of view. The various schemes concerned different characteristics and applications of images. Whether spatial or frequency domain are used to analyze the authentication performance, and both the signature based and watermark based approach are proposed.

In [1], Wong proposed a watermark-based and block based authentication system with the public key encryption that was taken as a basic idea in later works. Wong used the hash function to short the information of one image block into a fixed length code and insert it back to the image. This authentication system can achieve the localization ability at the certain accuracy the same as the block dimension. The independent block based scheme had been referenced frequently in the other schemes with different signature generation and embedding. [2] proposed the absolutely localization authentication and the authentication result correspond to every single pixel of image. Although the above two works achieved the localization capability, there are still limitations in them. First, the independent block design suffered from the counterfeit attack proposed by [3]. [4] extended the Wong's work to prevent the counterfeit attack, and the scheme proposed in [4] can further extend to various block based scheme. On the other hand, [1] and [2] can not

tolerate any image processing operation even only the slight difference in these images. [5] proposed a watermark based scheme in the wavelet domain. The wavelet coefficients are quantized and combined to the watermark bits in the selected frequency subbands. The proposed technique could roughly imply the operations that the authenticating image had been undergone. Since the limitations in the above works restricted the applications of an image authentication system, Lin and Chang worked on the JPEG compression tolerant scheme [6]. In [6], they found the characteristic of the DCT coefficients that kept invariant after the quantization step in the compression process. [7] moved the work to the wavelet domain. [7] recorded the positions between coefficients in the low and high frequency bands when the relations were fit to a certain situation, called a structural scheme. And the similar idea to [5], [8] designed the mean quantization scheme to tolerant the incidental distortions. The content based scheme with incidental distortions tolerant were proposed continuing in [9][10][11], both the signature based and watermark based approach were taken as the framework. The authentications of color images were presented in [12][13]. The principles of the color image authentication are not far from the gray level images. The difference lies in the feature extraction and embedding domains. The most recently work further combine the coding techniques and configured under the JPEG2000 standard [14].

1.3 Organization

In this introduction chapter, the motivation of this thesis was described. In the followed, chapter 2 will give an overview of content authentication inclusive of the requirements, applications and techniques. The proposed image authentication system will be presented in chapter 3. Chapter 4 will illustrate the experiment results and performance analysis of this system. Finally, chapter 5 will conclude this thesis.

Chapter 2 Content Authentication Overview

The overview of content authentication will be given in this section, and the applications and requirements of an authentication system will also be introduced.

Authentication is used in the secret communication, which allowed the participants in the communication to verify the characteristics of the received content. The characteristics that should be concerned in a communication could include the transmitter identity, access legality and content authentication that verify the integrity of content. Each of the consideration will have different design to verify its own authenticity that will be discussed as follows:

2.1 Transmission Identity and Access Legality

2.1.1 Transmission Identity

In a secret communication, the participants should authenticate the identity of the transmission. And the transmitter should have no chance to deny the originality of the content. The most common way to the goal is the asymmetric key; the private and public key system shown in figure 1.In the private and public key system, the original owner encrypts content using the private key, which is keeping unknown by others. The other participants in the receiver end will use the public key to decrypt

the content. In the asymmetric system, the owner of the private key could generate its personal public key that is used to decrypt but the others who get the public key could not know the private key by any way. In this circumstance, any public key will correspond to only one private key. Whenever a legal user uses a public key to decrypt a message successfully, he can be sure of the originality of this message and the identity of the transmitter.

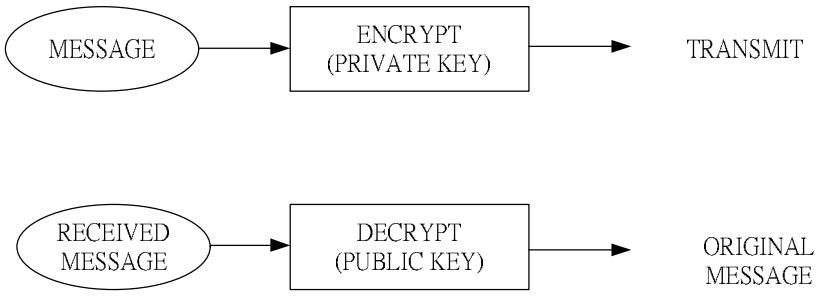


Figure 2.1 Asymmetric key scheme, the messages are encrypted with private key and decrypted with public key.

2.1.2 Access Legality

In another concept contrary to the transmitter identity, the access admission is another consideration in the secret communication. As mentioned above, the asymmetric key algorithm is also the solution for this issue. After a private key encrypted the message, the one who encrypted this message will distribute the public key only to the legal user

in this communication. It can protect the message from the others who are not allowed to read the message.

2.1.3 Content Authentication

The digital media system provides the sophisticated processing framework. All kinds of digital media are easy to be edited using some simple software. In many circumstances, alterations to content serve legitimate purposes. However, in other cases, the changes may be intentionally malicious or may inadvertently affect the interpretation of the content. For example, an inadvertent change to a medical image results in a misdiagnosis, whereas malicious tampering of photographic evidence in a criminal trial can result in either a wrong conviction or acquittal. Thus, in applications for which we must be certain the content has not been modified, there is a need for verification or authentication of the integrity of the content.

Specifically, researches are interested in the method for answering several questions. The first, an authenticator wants to know if the content has been altered in any way whatsoever. And the second problem is the level of the alteration. And further more, the parts of the alteration are concerned. Finally, the restoration of the covered content is also under consideration. Both the watermark and digital signature methods are proposed for the content authentication for solving the above questions.

2.2 Requirements and Applications of Content Authentication

In terms of content authentication mechanisms, the basic idea is to verify the integrity of contents that are distributed in an insecure channel. The requirements of the authentication processing are different from the applications. The authentication system can be designed based on various situations such as exact authentication, semi-exactly authentication and localized authentication. We will describe them in details in the following paragraphs.

2.2.1 Exact Authentication

The most basic authentication task is to verify that the content has not been altered at all. The word “exact” expresses this kind of content authentication, which implies that the authentication operation should obtain the total confidence to the whole message content. Even one bit of the message change will disturb the authentication result. This kind of design consideration is due to some applications such as images used in medical or military purposes. Images used for those applications can not be any different from the original because the image content will confuse the understanding of the meaning.

The approaches for the exact authentication are involved in fragile

watermark and digital signatures.

- **Fragile watermark**

The fragile watermark is simply a mark likely to become undetectable after the embedding content is modified. If a very fragile watermark is detected in the authenticated content, it can be inferred that the content has probably not been altered since the watermark was embedded.

A simple example of the fragile watermark is the least-significant-bit watermark. The watermark is embedded in the LSB plane such that any slight modification destroys the watermark. Although fragile watermarks indicate that the content has not been modified, the use of predefined patterns cannot guarantee that no one has intentionally tampered with the content. This is because an adversary can easily forge a fragile watermark if the embedded watermark pattern is not dependent on the covered content. In the case of the LSB watermark, forgery is a simple matter of copying the least significant bits from the authentic content to the tampered cover content. The proposed LSB watermark method [1] generates the watermark bits from the original image content except the least-significant-plane and inserts the watermark bits back to the LSB plane.

- **Digital signature**

The digital signature used in the authentication system is designed to be associated with the whole content when the exact authentication is expected. The signature will be generated from the features of the authenticated content that whenever the modifications happened, the signature is changed to be invalid. The one-way hash function is often used to generate the digital signature of an authenticated content. The difference between the digital signature and fragile watermark is embedding procedure. The digital signature is transmitted as the side information with the authenticated content instead of embedding it back to the content. The related consideration is that the fragile watermark embedding also causes distortions to the original content. It will impact the principle of the exact authentication. The digital signature approach solves this problem while the length of the side information is another discussed issue.

2.2.2 Selective- Authentication

In the real Internet environment, the digital contents are commonly processed by several kinds of operations in order to increase the utilization of the digital media. Common digital signal processing includes compression, low pass filter and etc. These kinds of operations

will not change the meaning of the content but the representation will be changed, which are called incidental distortion or content-preserving modification. It is reasonable to distinguish the incidental distortion or content preserving modification from the other modifications that change the original meaning of the content called malicious modification or content-changing modifications. Under this circumstance, the authentication system should be designed to tolerate the incidental distortion while the malicious modification will not pass the authentication process.

The two approaches described in section 2.2.1 can be used again in the selective authentication scheme. The design of a fragile watermark or a digital signature becomes adaptive sensitive to the distortions. This means that a fragile watermark becomes semi-fragile or selective-fragile to distortions in the content. A certain distortions that are heavier than the predestining level will make the watermark become undetectable. Otherwise the watermark will survive in distortions happened. The main concern in the digital signature is similar to the fragile watermark when it is used in the selective authentication system. The signature is expected to be less sensitive comparing to the signature in the exact authentication system. The signature generation in a selective authentication system is often designed to extract the feature that is not influenced by the considered distortions. Only when the undesired distortions that are not considered in the designing end happen, the signature becomes invalid.

2.2.3Localized Authentication

According to the authentication results, the authenticator can decide whether the authenticated content is available or not. Furthermore, in some cases, not only the pass or not in the authentication was concerned, but also the modification regions. It is useful to know the parts of the modification because the whole message can have much information. Even though the parts of message are modified, there are still other parts of the content that contain the correct information. The indication of the modification regions could be used to estimate the motivation of forgery and the remaining parts can be taken as reference still.

The ability of the localization authentication system can be accurate from the pixel wise to the block wise. The block-based scheme authenticates the content based on the block and returns the result of each block. That distortion within a block will fail the authentication result. Even the absolutely accuracy of a localization authentication system can be achieved by the pixel wise scheme. The authentication result depends on every pixel. The drawback is that the selective authentication can not be reached when every pixel is concerned.

Different kinds of applications require various designs of authentication systems as discussed above.

In this thesis, our concern is focus on the digital images. An image authentication system based on the eigenvalue decomposition that will

consider the tolerance of incidental distortion and the localized functionality is proposed in this thesis. The techniques of the image authentication will be introduced in the next section chapter 3.

Chapter 3 Image Authentication Techniques

3.1 Categories of Image Authentication Techniques

To verify the image contents, the methods can be classified into digital signature-based and watermark-based approaches.

3.1.1 Digital Signature Based Approach

Traditional digital signature is a short digital message that is used to sign a digital content to prevent illegal manipulation or other usages without the author's permission. The signatures are encrypted before transmission using the private key as described in Sec.(previous section). Similarly, in the design of an image authentication system, a digital signature is a set of features extracted from the original image and stored as a file, which will be used to authenticate later. The signature based approach authentication procedures are shown in Figure 2.2.

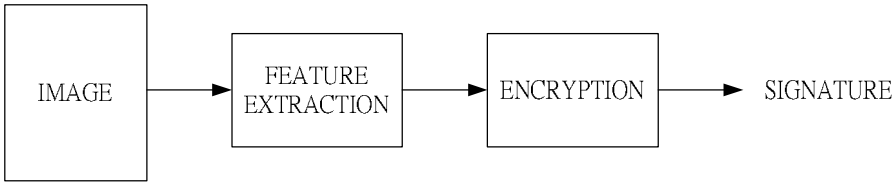


Figure 3.1 The generation of the digital signature

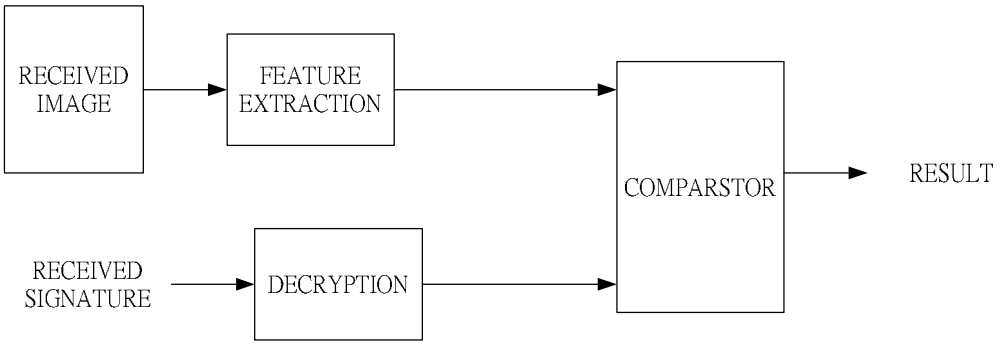


Figure 3.2 The verification in authentication system

According to the requirements and applications described in chapter2, when a digital signature is used to authenticate an image, the characteristics formed the signature should be taken from several concepts. For example, when the authentication system is designed to be the exact authentication, the signature must be related to the whole image and destroyed when any modification happened. On the other hand, in the case of incidental distortion, a digital signature used to authenticate the content-preserving images is expected to have selective sensitivity which means that the designed signature is sensitive to the malicious modification while keeps undestroyed under incidental distortions. To generate this kind of signature, the features used as the signature must have characteristics that have been unchanged under common image processing operations but changed after malicious modifications. Furthermore, if a digital signature is used to verify the content with localized ability, the design of this signature should contain the partial information of the content. For an image to be authenticated, the signature generation process could be done partially in the divided

regions of the whole image. Thus each region has its own signature. Whenever parts of the image are forged, the signatures of the modification regions will be undistinguished and the results will indicate the localization. Under this consideration the signature corresponding to each region will survive while other regions have been modified. The simplest way to achieve the localization is to divide the image into small regions independently and generate the individual signature to each. Thus the individual signature can be verified separately without influencing each other. But this concept will have its drawback of counterfeit attack which uses a vector quantization method to forge watermarked image confusing the authentication result. This attack will be further discussed in the next section. The direct idea to fight the attack is to set the divided regions in overlapping patterns. It means that the individual signature is accurately related to not only one region itself but also still other regions involved. In the overlapped design, the localization accuracy will be sacrificed because when there is a modification in only one region, it influences not only this signature, but the other signatures that took information in this region in the generation process will also be influenced as well.

3.1.2 Watermark-Based Approach

The digital watermark technique is one of the categories in data

hiding that is used to protect the IPIR of digital content. The so-called ‘digital watermark’ means to embed (i.e. hiding) invisible digital information into original sources. The goal is to identify the authority or ownership from the hiding information. By extracting the watermark from the distributed content, the author’s information could be identified. In this kind of applications, the embedded watermark label within the source contents are spread over the Internet and undergone various operations which could be viewed as attacks to the watermark. So the watermark technique must be able to against the attacks in the environment basically. This requirement is called the robustness, and there are other requirements such as transparency and low false positive/negative rate considered in any watermark technique design.

For an image authentication system, the watermark-based approached method means to embed watermark into original image and verify the authenticity by the extracted watermark. The major difference between digital signature and watermark approach is that the watermark approached method will produce distortions in the embedding process.

In this approach to authentication, the embed watermark is designed in different concept from the original watermark embedding. In authentication applications, the watermark used to verify the integrity of images should be destroyed after modification. This kind of watermark is called fragile watermark. Similar to the digital signature based design; applications influence the watermark embedding design. When the

system is expected to tolerate the incidental distortion, the semi-fragile watermark is used. The semi-fragile watermark is meant to be less fragile than the fragile watermarks such that only certain distortions will destroy the watermark. If there are only slight modifications or incidental distortions caused by common image processing operations, the watermark remains available. As to the localization ability, the extracted watermark should be able to indicate the regions that have been modified.

3.2 Techniques in Image Authentication

The authentication system can be designed using different techniques. Techniques taken in the signature and watermark based approach can be classified into quantization-based, relation-based and feature points-based and etc. Both the spatial and frequency domains are selected depending on the considerations for different applications. Usually the spatial domain design is used to detect the exact modifications. Any micro forgery will cause the authentication to become false. As to the frequency domain, the authentication system done in the frequency domain is willing to tolerate the common image processing operations. Because the processing operations are often operated in the frequency domain, authentication systems are designed under the circumstance to reach the ability of selective-fragile.

3.2.1 Feature Point-based

The feature point-based technique exacted the feature from the original image to form the signature or watermark that are transmitted or embedded back to the original image. Because the features represented the image often contain large amount of bits. The more information the feature include, the more accuracy will be achieved by authenticating the features. A one-way hash function is widely used to shorten the features into a fixed length such that both the transmission security and embedding capacity can be increased. Addition to fix the feature length, the one way hash function map input bits to a single output, thus when the input bits have any difference, the output will change instantaneously. [1] hashed the 7 MSB of the pixels in one block and combined the output bits with the watermark bits. And the exclusive or results are later embedded back to the LSB of the pixels in this block itself. The verification process computed the output of the 7 MSB using hash function and compared it with the bits extracted from the LSB. The embedding and verification are both done to each block individually to achieve the localization ability. [12] extracted the LSBs in the red and green domain of an color image and inserted the hash code combined with the watermark into the LSB plane of the blue domain. Based on the characteristic of hash function, even a single bits error disturbed the authentication results.

3.2.2 Quantization-based

There are researches that work on the concept of quantizing the coefficients into a quantized interval when the watermark bit is encoded. [5] designed the quantization-based watermark scheme that is adaptive sensitive to the different subbands. They defined a tamper assessment function (TAF), which is the ratio of the number of the tampered coefficients to the total number coefficients in a specific subband, in order to measure the degree of tampering. [8] quantize the mean of a group pixels instead of the individual pixel.

3.2.3 Relation-based

The relation based scheme found the invariant relations between coefficients in the transform domain. The invariant relations will be unchanged under the incidental distortions. [6] drives the relations between two DCT coefficients of two distinct blocks in the same position. This relation will remain unchanged after JPEG compression even if the recompression had been done. In [7], the wavelet coefficients relationships called structure between subbands are utilized to be the features. Both the two works can tolerate certain level of the incidental distortions.

Chapter 4 Image Authentication System Based on Eigenvalue Decomposition

This chapter will describe the whole system architecture proposed in this thesis in detail. The requirements and principal considerations of an image authentication system were presented in chapter 3. This chapter will introduce each part of the design considerations. The reasons for the feature extraction will also be explained.

4.1 System Architecture

In chapter1, the desired functionalities of the designed image authentication system were mentioned in the paragraph of motivations. An image authentication system that is suitable for the distribution over practical Internet environment must first have the ability of distinguishing the incidental distortions and malicious tampering. Because that the common image processing operations have been done most of time. Further more the localization capability is also desired to increasing the utilization of the authentication results.

In order to achieve the localization ability, we divide the original image first and do the feature extraction based on the eigenvalue decomposition independently in each block, as shown in figure 4.1. The eigenvalues produced from the autocorrelation matrix of an image block

will be used to characterize this block. The design consideration is based on taking the image pixels as random variables and constructing the observation data space by the subblock composition. Therefore, the eigenvalues of the autocorrelation matrix of an image block will reflect the intensities of the image pixels in this block and correlations between them.

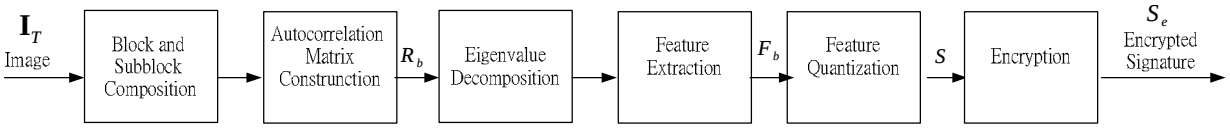


Figure 4.1 the signature generation process

4.1.1 Transmission End

In the transmission end, the image is first divided into blocks and subblocks. The image pixels are regarded as random variables. The autocorrelation matrix is constructed by taking each subblock as an observation data space of a random variable. The autocorrelation matrix then is eigenvalue decomposed to produce the eigenvalues that will be used to characterize this block. A block feature is extracted from the dominated eigenvalues. Signature of this block will be generated by quantizing the feature. All the signatures of image blocks will be combined and encrypted before transmitting.

The proposed system extracts the feature based on the eigenvalue

decomposition because the eigenvalues from the autocorrelation matrix of an image block inflect the intensities and correlations of the image pixels in this block. The elements in the autocorrelation matrix have the information of the average magnitudes of image pixels through a subblock and covariance of subblocks. Then the values of the eigenvalues indicate the intensities of pixels and the distributions of the eigenvalues express the activity of one block. The theoretical explain will be given in detail in the next sections.

4.1.2 Authentication End

In the authentication end, the feature extraction process is the same with the transmission end. The eigenvalues produced from the autocorrelation matrix of an image block is assumed to be a robust characteristic of the image block under distortions by the common image processing. This assumption is based on the knowledge that the most image processing operations keeping the meaning of the image will not change the mean of the image pixels heavily and maintain the most structure of the image block in most of time. The values and distributions of the eigenvalues of the autocorrelation matrix of an image block will not change a lot when the incidental distortions have happened in this block. In the authentication process, the authenticator calculates the feature and compares it to the decrypted signature with a

neutral zone consideration. This operation means that the feature is quantized with a neutral zone. The neutral zone is designed to tolerate the small variant of the feature caused by the incidental distortions. Figure 4.2 expresses the authentication process.

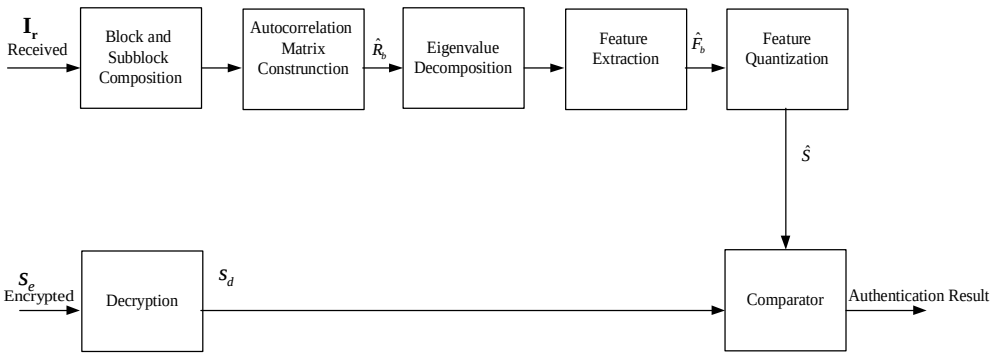


Figure 4.2 Authentication Process

4.1.3 Decision of Authentication Result

The final decision of the authentication result will be determined by a threshold and distribution judgment. The predefined threshold is set to ignore a few number of mistake blocks. This is because the incidental distortions might still disturb the features of some special blocks. To determine the authentication result, the percentage of the mistake blocks will be compared with the threshold. When the percentage of the mistake blocks is less than the threshold, the authentication result will be deemed to be successful. In the contrary, mistake percentage that is larger than the threshold will make the authentication result fail. The distribution of the

positions of mistake blocks is also under consideration. This judgment is based on the point of view that incidental distortions influence the feature extraction of blocks randomly while the meaningful tampering will be some kind of patterns of regions or objects. The spreading distribution of the mistake blocks indicates the mistakes are caused by the incidental distortions. The specific patterns of the authentication results by the block groups show the meanings of the meaningful tampering. The figure 4.3 illustrates the decision of the authentication result.

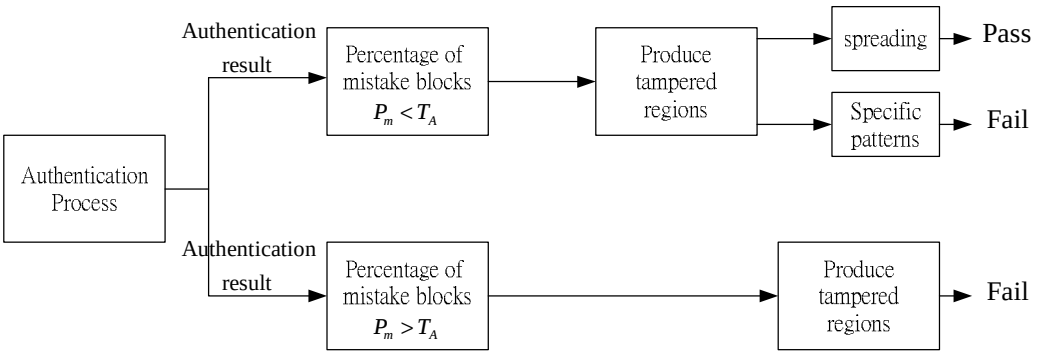


Figure 4.3 Authentication result decision

Figure 4.1 and 4.2 show the whole system architecture of the proposed system. The details of each block in the figures will be further explained in the later sections.

4.2 The Signature Generation Process

As shown in the figure 4.1, the whole signature generation process before the signatures are encrypted includes the block and subblock composition, autocorrelation matrix construction, feature extraction, and feature quantization steps. Each step will be described as follows.

4.2.1 Block and Subblock Composition

The image pixels are regarded as random variables in this authentication system. The statistical characteristics are computed in a determined observation data space that is constructed by the subblock composition. The original image is first divided into blocks with size of 8 by 8. In each 8-by-8 block, the 64 pixels in the block will be further divided into 2-by-2 subblocks as shown in figure 4.3. Pixels x_{11}, x_{12}, x_{13} and x_{14} form a subblock.

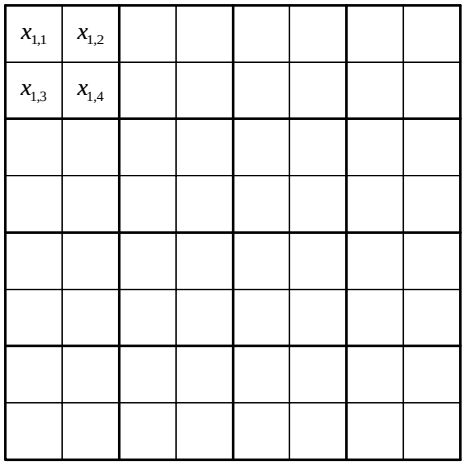


Figure 4.4 An image block structure contains 16 subblocks.

A subblock contains 4 pixels that will be regarded as the observation data of one random variable. Therefore, an 8-by-8 block consists of 16 random variables, where each random variable contains 4 observation data. The 4 neighbor pixels are taken as the observed data because the 4 adjacent pixels in the most of images are most stochastic similar. The statistical characteristic of one random variable can be estimated probably from them. The correlations of those random variables considered from the subblocks can reflect the activity of this image block. The more the subblocks are correlated, the smoother this image block is. This is a very directly point from the perspective of view. The behaviors of the random variables express the activities of image pixels. To analysis these random variables in an image block, the autocorrelation matrix of them will be eigenvalue decomposed. The autocorrelation matrix will be constructed in the next section.

4.2.2 Autocorrelation Matrix Construction

The autocorrelation matrix in conventional random process is defined as eq(1), Let a $N \times 1$ random vector x consisting of n random variables , which is denoted $\mathbf{x} = [x_1, x_2 \cdots x_N]^T$, the autocorrelation matrix $\mathbf{R}_{N \times N}$ of x is

$$\mathbf{R}_{N \times N} = E\left[\mathbf{x}\mathbf{x}^T\right] \text{.....(1)}$$

The autocorrelation matrix is the expectation of the outer product of

x and its transport $\mathbf{x}^T$, $\mathbf{R}_{N\times N}$ can be expressed by the elements of x as eq(2)

$$\mathbf{R}_{N\times N} = E\begin{bmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_N \\ x_2x_1 & x_2^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ x_Nx_1 & \cdots & \cdots & x_N^2 \end{bmatrix} \quad \text{.....(2)}$$

The diagonal elements of $\mathbf{R}_{n\times n}$ are the square of each element in x and the other terms are the cross terms that can display the related stochastic information between the random variables. Since the autocorrelation is the given form

$$\mathbf{R}_{N\times N} = E\left[\mathbf{x}\mathbf{x}^T\right] = E\left[(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)^T\right] + \boldsymbol{\mu}_x\boldsymbol{\mu}_x^T \quad \text{.....(3)}$$

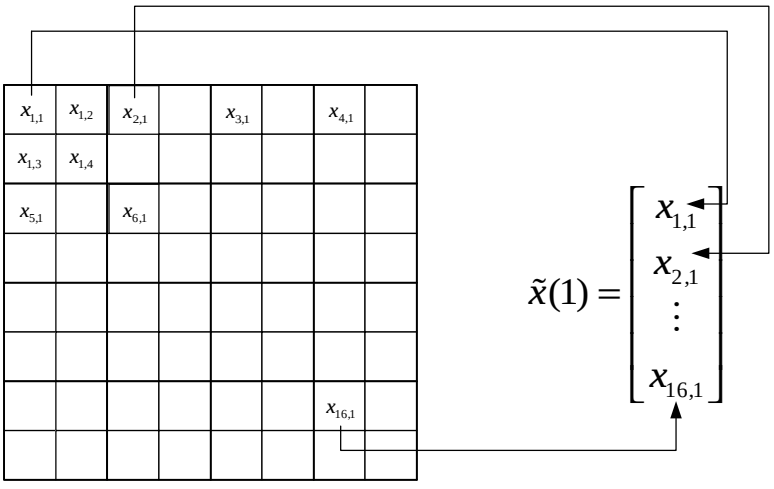
where the $\boldsymbol{\mu}_x$ is the mean vector of the random vector x.

$$\boldsymbol{\mu}_x = \begin{bmatrix} m_{x_1} \\ m_{x_2} \\ \vdots \\ m_{x_N} \end{bmatrix}, \quad \boldsymbol{\mu}_{x_i} \text{ is the mean of random variable } \mathbf{x}_i$$

(3) has the information of power and average terms. The power information has the physical meaning with the deviation of those random variables. In one image block, the deviation of the random variables in them implies the variance of the image pixels in it, or in other words, the frequency performance of this block. On the other concept, the average terms of the autocorrelation matrix are the average values of the image

pixels within an observation data space. More specifically, each diagonal element in the autocorrelation matrix is square of the average value of image pixels in a subblock. Since the autocorrelation matrix contains the information of the image pixels, it will be analyzed later by the eigenvalue decomposition. The values of the random variables in the image block will be indicated by the magnitudes of the eigenvalues and the crosscorrelation behaviors of those random variables are reflected to the distribution of the eigenvalues.

To construct the autocorrelation matrix in the authentication system designed in this thesis, the pixels $x_{n,i}$ that are in the same position in a subblock x_n, n=1,2,3……16, are arranged to form an observation vectors of a random vector with dimension 16x1 as shown in Figure4.5.



$$, where \quad \tilde{x}(i) = \begin{bmatrix} x_{1,i} \\ x_{2,i} \\ \vdots \\ x_{16,i} \end{bmatrix} \qquad i = 1, 2, 3, 4$$

Figure 4.5 Observation vector mapping

i denotes the positions in a subblock.

And the autocorrelation matrix of an image block can be estimated as the average of outer product of $\tilde{\mathbf{x}}(i)$.

$$\hat{\mathbf{R}}_{16\times16} = \frac{1}{4}\left[\tilde{\mathbf{x}}(1)\tilde{\mathbf{x}}(1)^T + \cdots + \tilde{\mathbf{x}}(4)\tilde{\mathbf{x}}(4)^T\right] \ldots\ldots\ldots\ldots\ldots\ldots(4)$$

The autocorrelation matrix of these random variables interprets the magnitudes of random variables themselves and the relations between them. The characteristics involved with pixel values and the autocorrelation matrix as mentioned above demonstrates relations between pixels of an image block. To analyze the autocorrelation matrix, the magnitudes and the correlations of the 16 random variables can be observed by the eigenvalue decomposition that will be introduced in the next.

4.2.3 Eigenvalue Decomposition

After the autocorrelation matrix was constructed, it will be eigenvalue decomposed.

For a given matrix, the eigenvalue decomposition of $\mathbf{R}_{N\times N}$ can be expressed as eq (3)

$$\mathbf{R}_{N\times N} = \mathbf{E}_\mathbf{s}\mathbf{\Lambda}_\mathbf{s}\mathbf{E}_\mathbf{s} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5)$$

where $\mathbf{E}_\mathbf{s} = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_N]$, which is the $N\times N$ matrix constructed by the eigenvectors $\mathbf{v}_n$ corresponding to the eigenvalue λ_n. And $\mathbf{\Lambda}_\mathbf{s}$ is the

diagonal matrix with the elements respect to the eigenvalues λ_n.

The $\mathbf{R}_{N\times N}$ can also be represented as

$$\mathbf{R}_{N\times N} = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^T + \cdots + \lambda_N \mathbf{v}_N \mathbf{v}_N^T \quad \text{.....(6)}$$

with λ_n are arranged in the descending order that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$$

and $\mathbf{v}_n$ is the corresponding eigenvector of λ_n

The $\mathbf{v}_n \mathbf{v}_n^T$ is the projection matrix that spans the η_{th} vector space. The corresponding eigenvalue λ_n is regarded as the projection quantity of this projection space. These vector spaces are orthogonal and can be regarded as the coordinates in the N dimension space. The each eigenvalue is the quantity in each coordinate and expresses the correlation in each projection direction.

In the image block, the autocorrelation matrix of this block $\hat{\mathbf{R}}_{16\times16}$ can be expressed by eignevalue decomposition as .

$$\hat{\mathbf{R}}_{16\times16} = \mathbf{E}_x \mathbf{\Lambda}_x \mathbf{E}_x \quad \text{.....(7)}$$

where the $\mathbf{\Lambda}_x$ is the 16x16 diagonal matrix that the diagonal elements λ_n, n =1,2,…16, is the eigenvalues of the autocorrelation matrix of the image block. In 4.2.2, eq(3) defined the $\hat{\mathbf{R}}_{16\times16}$ to be estimated by the average of outer product of $\tilde{\mathbf{x}}(i)$. According to the observation data size, each random variable is observed by 4 samples, the autocorrelation matrix of an image block will be a matrix with the rank equal to 4. The resulted eigenvalues will only have 4 nonzero λ_n.The 4 nonzero eigenvalues will later been analyzed to characterize each image

block.

The eigenvalues indicate the projection quantity of each corresponding projection space. In the image signals, the eigenvalues produced from the autocorrelation matrix expressed the correlation of the original image pixels in each orthogonal coordinate. The information of the image block implied in the autocorrelation matrix was described in the last section. The eigenvalues produced from the autocorrelation matrix will further be analyzed to explain that information, which are the intensities and correlation of the image pixels within an image block.

In the theoretical of view in the linear algebra, diagonal elements of an autocorrelation matrix $\mathbf{R}_{N\times N}$ are the square term of the random variables. In the image block, the nonzero eigenvlues of $\hat{\mathbf{R}}_{16\times 16}$ are related to the element as eq(7)

$$\sum_{j=1}^{4} l_j = tr(\tilde{\mathbf{R}}_{16\times 16}) = \sum_{i=1}^{16} x_{ii}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

l_j denote the nonzero l_n

In eq(7), The sum of the eigenvalues equals to the trace of the autocorrelation matrix. The trace means the sum of the diagonal elements of a matrix. In the autocorrelation matrix of an image block, the diagonal elements are the square terms of the average values of image pixels in an observation data space, which is the subblock. The relation demonstrates the eigenvalues are associated with the pixel values in the image block.

In addition, each cross term of the autocorrelation matrix also influences the eigenvalues. The influence will react to the distribution of the nonzero eigenvalues. When the original image signals in an image block are more correlated, the related eigenvalues from the autocorrelation matrix will have a spreading distribution. The magnitude of each nonzero eigenvalue will be much variant from each other. More specifically, there will be a few eigenvalues have the much larger magnitudes respect to others. On the other hand, the distribution will be closer when the signals are less correlated. From observation of the image blocks, the eigenvalues produced from each of them are in the spreading distribution most of time. There is a largest eigenvalue that its magnitude is far from others for most of the image blocks. The result can be understood because the image blocks with size 8 by 8 in most natural images are often low frequency dominated. The image pixels within a block are much correlated.

As described in this section and the last section, the autocorrelation matrix constructed from the block and subblockk composition of an image block has the information of intensities and correlations of the image pixels within this block. In addition, the magnitudes and distribution of eigenvalues of the autocorrelation matrix reflect the structure of the autocorrelation matrix. The proposed system used the dominated eigenvalues as the feature of an image block.

4.2.4 Feature Extraction

The feature of an image block is extracted from the dominated eigenvalues of the autocorrelation matrix of this block.

For most of the nature images, the pixels within the image blocks with size 8-by-8 are much correlated. The eignenvalues produced from the autocorrelation matrix may have a very large variance between each. Usually the maximum eigenvalue will be far from others. For less correlated blocks that pixels in those block are more activity, the values of the eigenvalues will be a little closer. But in the most of images, the condition of the eigenvalues is still the extra case that

$$l_1 \gg l_{j,\, j\neq 1}$$

Figrue 4.6 illustrates the distribution of the eigenvalues. The curves in this figure are the ratios of each eigenvalue magnitude respect to the maximum eigenvalue.

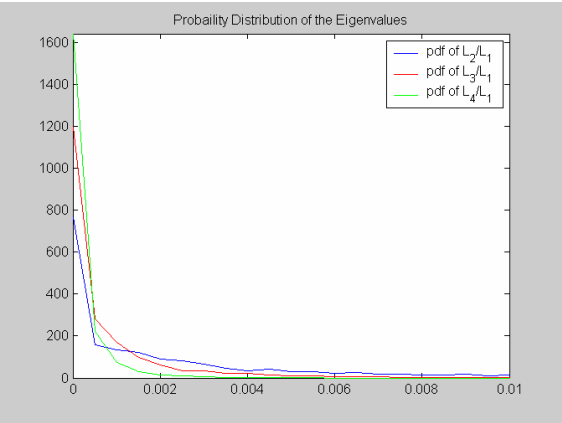


Figure 4.6 The pdf of the distribution of eigenvalues

To extract the feature for an image block, only the dominated eigenvalues that have the larger magnitudes will be considered. The consideration is based on the concept of the principal component analysis in the digital signal processing [15]. The original random signals can be estimated from the largest few eigenvalues and corresponding eigenvectors.

$$\hat{\mathbf{R}}_p = \sum_{n=1}^{k} l_n \mathbf{v}_n \mathbf{v}_n^T , \quad k \le 16$$

Under this situation, the feature extraction process first determines the dominated eigenvalues that have large maginuteds of an image block. Including the consideration that the distribution of the eigenvalues also reflects the activity of the image block, whether the individual eigenvalue is dominated is determined by the ratio of its magnitude respected to the magnitude of the maximum eigenvalue in this block.

Let l_j, j = 1,2,3,4, denote the nonzero eigenvalues of an image block with the descending order that

$$l_1 > l_2 > l_3 > l_4$$

$l_{r,j}$ denotes the ratio of the magnitude of l_j respect to l_1

$$l_{r,j} = l_j / l_1 \ , j \neq 1$$

A predefined threshold is used to determine whether the individual eigenvalue l_j, which $j \neq 1$, is dominated. It is certain that the largest eigenvalue l_1 is always the dominated one under consideration.

Whenever the $l_{r,j}$ is larger than the threshold l_t, this eigenvalue l_j is

regarded as dominated. Otherwise the l_j is not considered in the feature extraction. In the feature extraction process, if the entire ratio $l_{r,j}$ is less than the threshold l_t, l_1 demonstrates the most correlated quantity among all l_j. The image pixels are considered as high correlated in this block. The maximum eigenvalue l_1 can express the main characteristic of the correlation in this block. The maximum eignvalue is taken as the feature of this block.

$$F_b = l_1,$$

On the other hand, the dominated eigenvalues l_j that each ratio $l_{r,j}$ is larger than the threshold l_t are considered to be the feature. Under this situation, the image block is deemed to be less correlated in its image pixels and more active. In these blocks, the feature is extracted by power n of the product of the dominated eigenvalues, l_d, $l_d = l_j$ when $l_{r,j} > l_t$, the feature extraction is illustrated in figure 4.6.

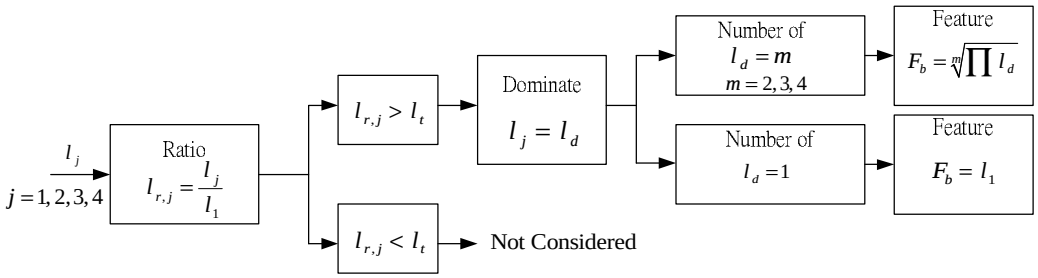


Figure 4.7 The determination of the feature extraction

The maximum eigenvalue is adopted for the reason that it contains the most of the information of the image block. The value indicates the largest correlation quantity of the autocorrelation matrix. The more

correlated the pixels are in an image block, the larger the maximum eigenvalue of the autocorrelation matrix is. Since the sum of the eigenvalues will be equal to the sum of the diagonal elements of the autocorrelation matrix as shown in eq(7) and $l_1 \gg l_{j,j\neq 1}$,

$$l_1 \approx \sum_{j=1}^{4} l_j = tr(\tilde{\mathbf{R}}_{16\times 16}) = \sum_{i=1}^{16} x_{ii} \quad \text{.....(9)}$$

The maximum eigenvalue l_1 approximates the sum of all l_j. Since the sum of the eigenvalues equals to the diagonal elements of the autocorrelation matrix that are the square terms of the average pixel values. The maximum eigenvalue characterizes a much correlated image block associated with the intensities of this image block.

When the blocks those are more active, there exists the eigenvalues that the ratio of the magnitude respect to the maximum one is larger than the threshold l_t. The eq(9) can be rewritten as

$$\sum l_d \approx \sum_{j=1}^{4} l_j = tr(\hat{\mathbf{R}}_{16\times 16}) = \sum_{i=1}^{16} x_{ii} \quad \text{.....(10)}$$

The sum of the dominated eigenvalues approximates the trace of the autocorrelation matrix in this case. In this kind of blocks, the correlations between image pixels are more concerned because the image block is more active. The distribution of the dominated eigenvalues can be used to be the feature characterizing the block. The feature is produced as

$$\mathrm{F_b} = \sqrt[m]{\prod l_d},$$

m is the number of the dominated eigenvalues, l_d.

When the image block is tampered, the change of the intensities, which means the change of pixel values, will directly change the magnitudes of the eigenvalues. Because the pixel values determine the diagonal elements of the autocorrelation matrix and those elements will determine the magnitudes of the eigenvalues. Both the two conditions, only the maximum eigenvalue is considered or more than one eigenvalues are dominated, can demonstrate the intensity information of an image block. The change of the intensities within a image block will influence the feature.

To consider the other point of the tamper, the frequency change in an image block, the frequency change will change the correlation of the random variables observed from the image pixels within this block. The change of the correlation will influence the distribution of the eigenvalues. The maximum eigenvalue is regarded as the feature when an image block is deemed to be inactive. When these kinds of blocks are tampered with the frequency movement, the correlation between the image pixels will change. The changed correlation will also change the distribution of the eigenvalues. The projection quantity, magnitude of each eigenvalue will be various from the original. From the concept of the digital signal processing, the frequency movement within an inactive image block indicates the spreading of the energy compaction of the image signals. The original maximum eigenvalue of this block will change. As to the active image blocks, the frequency change influences the distribution of

the eigenvalues. The defined feature is associated with the distribution.

The feature of an image block is sensitive to both the intensities and frequency change in the block. The assumptions will be certain confidentially that most of the meaningful tampering can not be designed just to meet the feature by adapting both the intensities and correlations of image pixels suitable.

4.2.5 Quantization of Block Feature

The feature is extracted from each independent block. The features of all the blocks in the image will be over a range. The final signatures will be formed by a quantization step.

The maximum feature value is found to be a normalized range. And all the feature values then are normalized from 0 to 1. The quantization step is dividing 1 into 32 levels and 5 bits express each level. The 5 bits finally form the signature of a block.

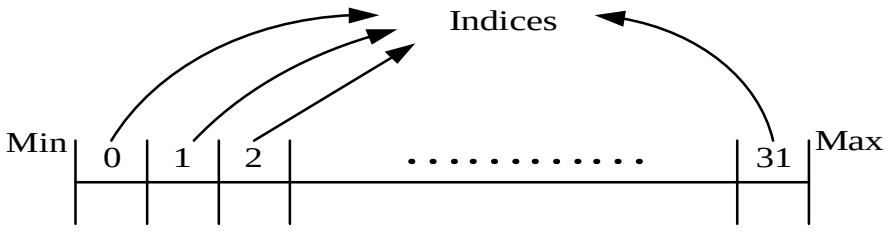


Figure 4.8 Quantization of the feature

4.3 The Authentication Process

The authentication process was shown in figure 4.2 in the section 4.1. The authenticator will have the encrypted signature and the received image. As shown in figure 4.2, the signature will first be decrypted. And the received image will go through the same process in the signature generation as transmission end.

The received image will be first divided into blocks and each block will be further divided into subblocks. The autocorrelation matrix will be formed for each block as described in the section 4.2.2. The eigenvalue decomposition of the autocorrelation matrix will be done. The resulted eigenvalues will be analyzed to be the feature value of the block as described in 4.2.4. The authentication results will be depended on comparing the decrypted signatures and the quantized features that from the received image with a neutral zone consideration.

4.3.1 Neutral Zone

The neutral is designed to tolerant the small change of the feature value. The quantization step is to range the feature into several levels. When the feature extracted by the eigenvalues in an image block just changes slightly because of the incidental distortion, the quantization result should be the same with the original. But the quantization step is considered as a hard decision with each boundary between the

neighboring quantization levels. When the feature is just allocated near to the boundary, the small change of it may result the different quantization result. The neutral zone is considered on the both side of each boundary to tolerant the small change of the feature value that is near the boundary.

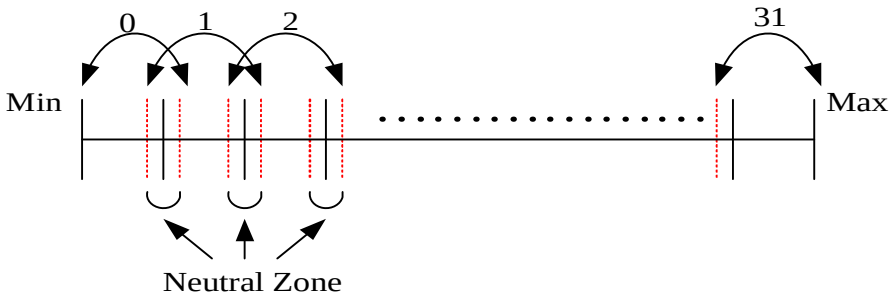


Figure 4.9 The neutral zone consideration

The final authentication result will be dependent on a threshold and distribution judgment as described in 4.1.3.

Chapter 5 Simulations and Analysis

This chapter includes the simulations and analysis of the proposed image authentication system based on eigenvalue decomposition. Section 5.1 describes the simulation environment. Section 5.2 analyzes the ability of incidental tolerance under neutral zone consideration. Section 5.3experiments the detection accuracy of the malicious tampering and the neutral zone influence.

5.1 Simulation Environment

- **Tested Images**
  - The tested images are all in the gray level with size 512 by 512.
  - The tested images :

(a) Lena

(b) Pepper

(c) Boat (d) Goldhill (e) Zelda

(f) Barbara (g) Mandrill

Figure5.1 The tested images
(a) Lena (b) Pepper (c) Boat (d) Goldhill (e) Zelda (f) Barbara (g) Mandrill

- **System Parameters**
  - The independent image blocks are with size 8 by 8.
  - The subblocks that are regarded as the observation data space are with size 2 by 2.
  - The neutral zone is tested from 0.005 to 0.01.
  - The authentication result threshold is set to be 0.01.

- **Incidental Distortions**
  - JPEG compression.
  - JPEG 2000 compression.
  - Low Pass Filter
  - Medium Filter..

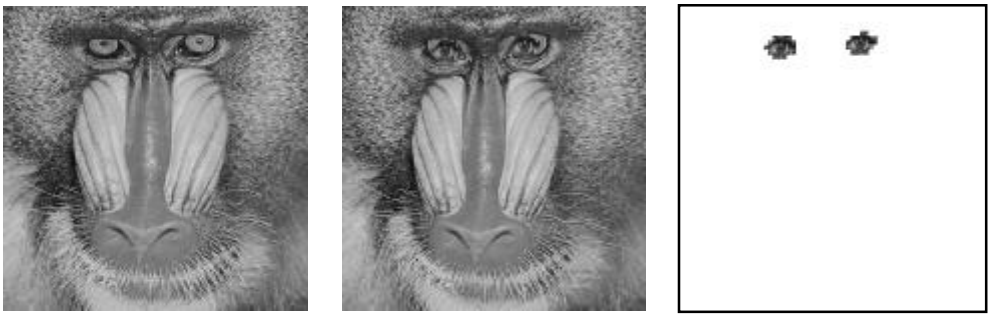
- **Incidental Tolerance Testing**

The incidental tolerance is the main requirement in the proposed image authentication system. The experiment will generate the signature

from the original images and these tested images will go through several image processing. In the authentication end, the signatures will be regenerated from the processed images and compared with the original signatures to produce the authentication result. The final decision that an authenticated image is pass or fail to the authentication will dependent on a mistake percentage threshold and distribution judgment.

- **Detection Accuracy of Malicious Tampering**

The detection accuracy of the malicious tampering is also the principle of the authentication system. The detection accuracy is much related to the neutral zone in the proposed system. On the other hand, in the decision of the final decision, even the mistake percentage is very small; the specific pattern might still indicate the meaningful tampering in the image. For example, the eyes of "mandrill" were replaced by others. The changed regions are small but the authentication result illustrates the specific pattern so that the tampering should be regarded as meaningful.



(a) (b) (c)

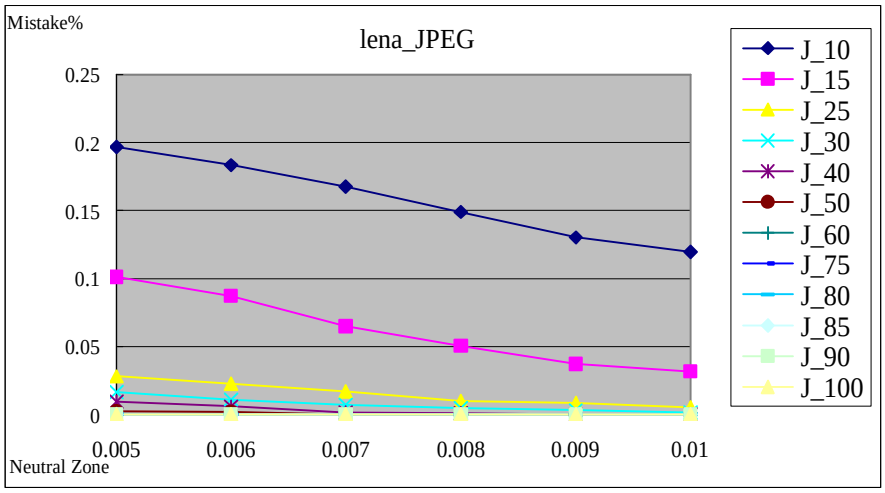
Figure5.2 The authentication result（Mandrill）
(a) original image (b) tampered image (c) authentication result

5.2 Incidental Tolerance

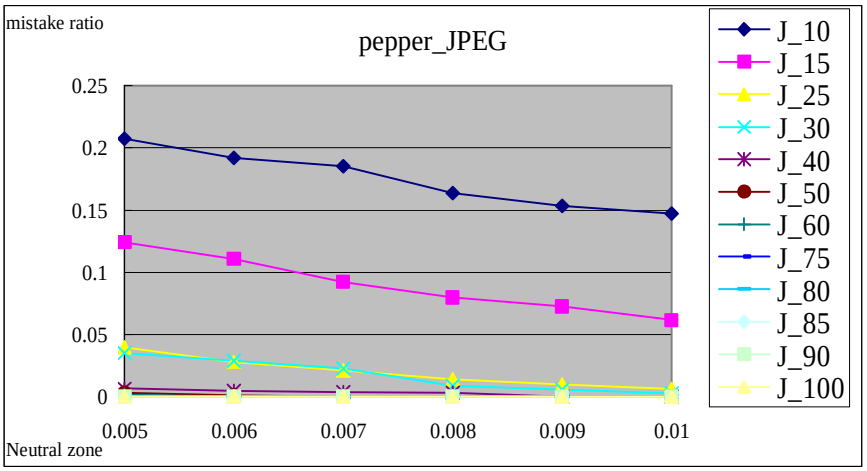
The authentication results will dependent on the percentage of the mistake blocks under distortions. The neutral zone influences the tolerance of the incidental distortions. The experiments are worked on the different range of the neutral zone. Tested images are processed by several image processing with different qualities.

5.2.1 JPEG compression

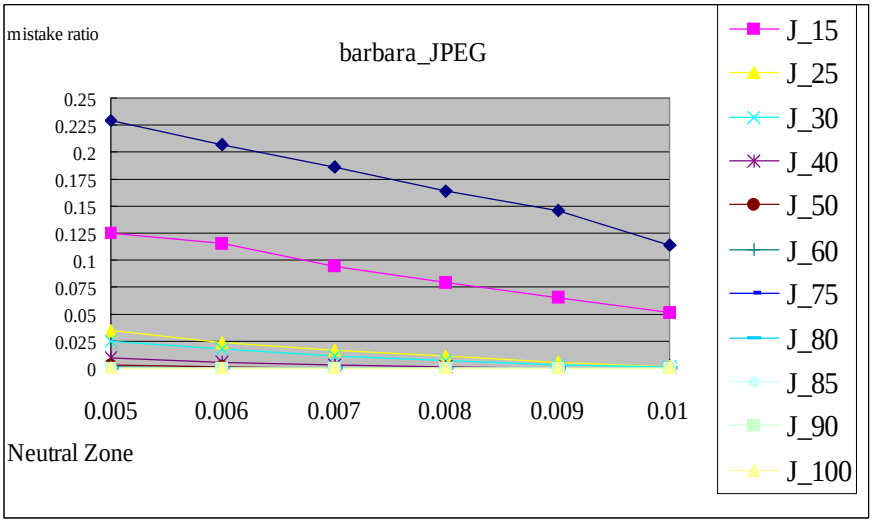
The tested images are compressed by the JPEC standard for different compression ratio. The neutral zone is range from 0.005 to 0.01.



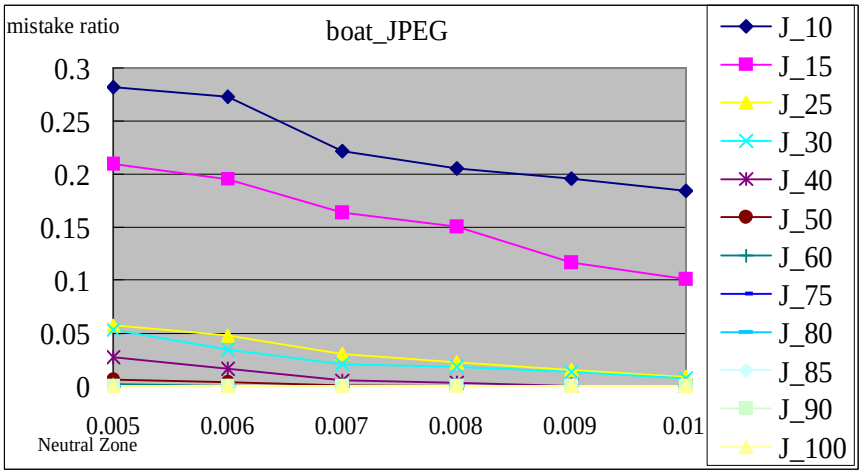
(a)


mistake ratio
pepper_JPEG
0.25
0.2
0.15
0.1
0.05
0
0.005
0.006
0.007
0.008
0.009
0.01
Neutral zone
J_10
J_15
J_25
J_30
J_40
J_50
J_60
J_75
J_80
J_85
J_90
J_100

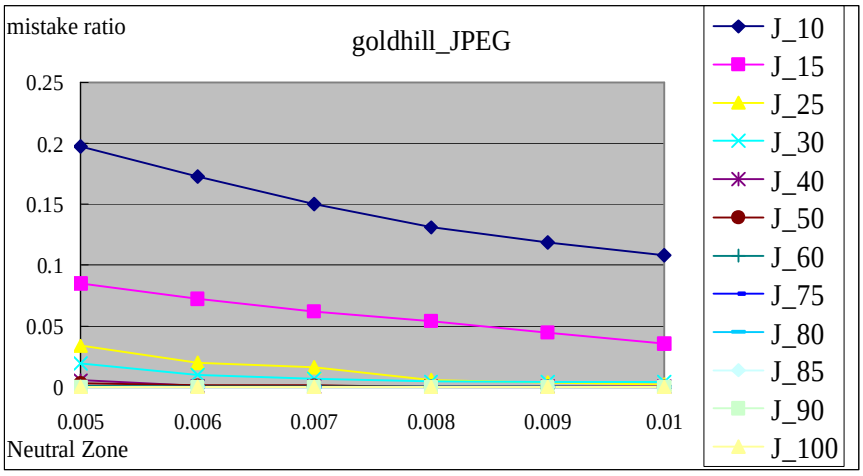
(b)


mistake ratio
barbara_JPEG
0.25
0.225
0.2
0.175
0.15
0.125
0.1
0.075
0.05
0.025
0
0.005
0.006
0.007
0.008
0.009
0.01
Neutral Zone
J_15
J_25
J_30
J_40
J_50
J_60
J_75
J_80
J_85
J_90
J_100

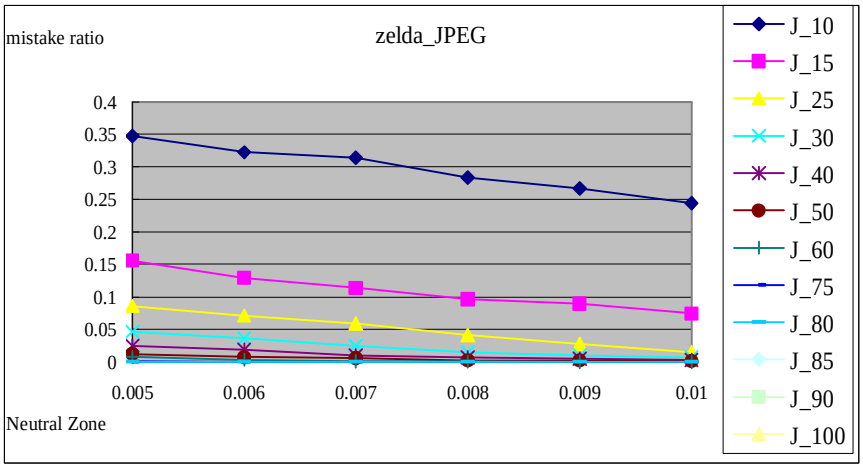
(c)


mistake ratio
boat_JPEG
0.3
0.25
0.2
0.15
0.1
0.05
0
0.005
0.006
0.007
0.008
0.009
0.01
Neutral Zone
J_10
J_15
J_25
J_30
J_40
J_50
J_60
J_75
J_80
J_85
J_90
J_100

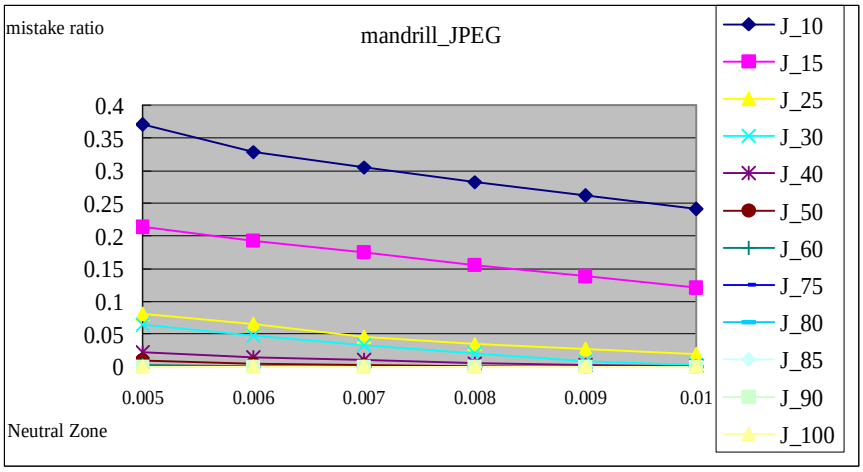
(d)



(e)



(f)

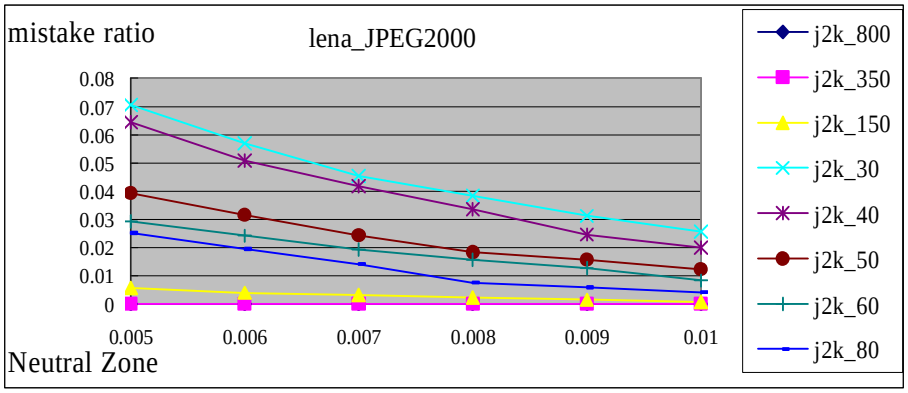


(g)

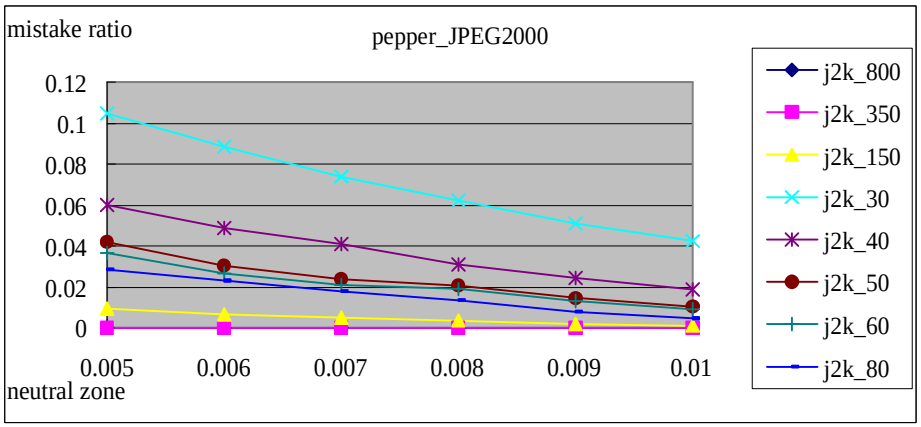
Figure5.3 The neutral zone via mistake ratio in JPEG compression
(a) Lena (b) Pepper (c) Boat (d) Goldhill (e) Zelda (f) Barbara (g) Mandrill

5.2.2 JPEG 2000 compression

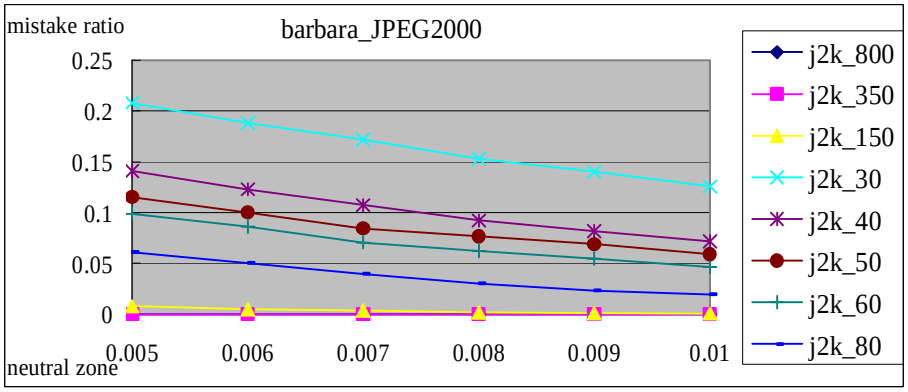
The JPEG2000 compression is also be experimented in the test images.



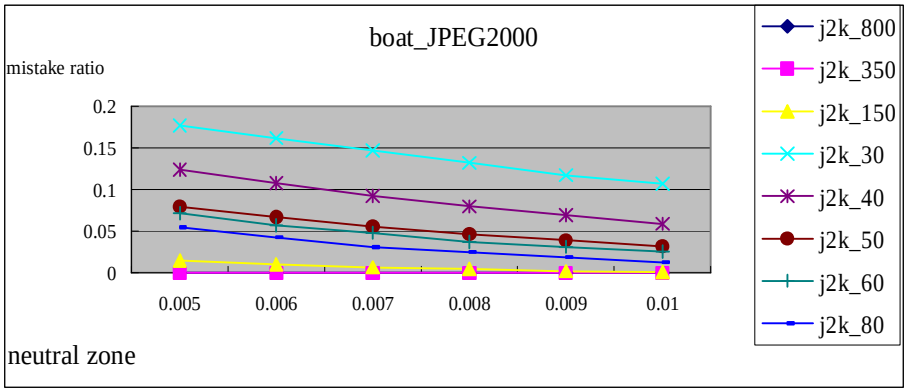
(a)



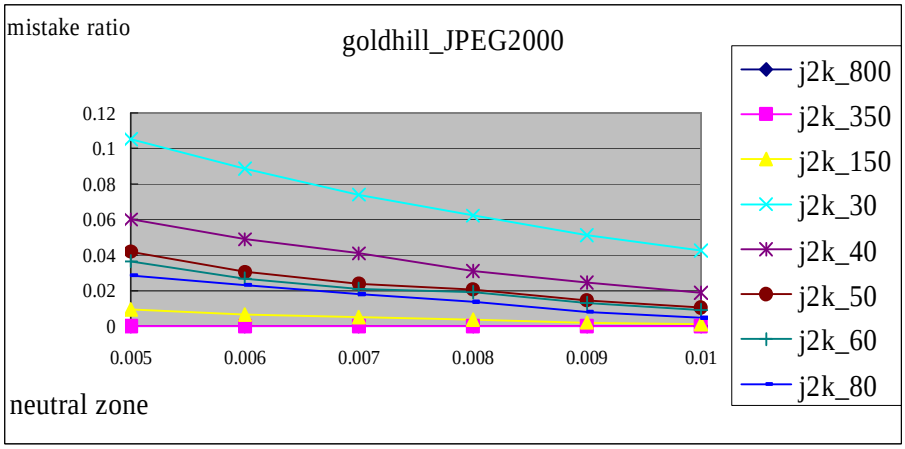
(b)



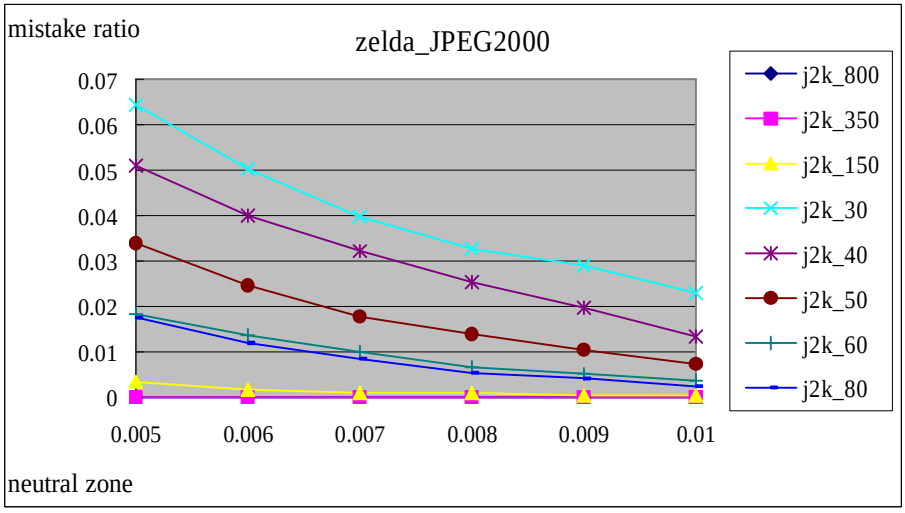
(c)


boat_JPEG2000
mistake ratio
0.2
0.15
0.1
0.05
0
0.005
0.006
0.007
0.008
0.009
0.01
neutral zone
j2k_800
j2k_350
j2k_150
j2k_30
j2k_40
j2k_50
j2k_60
j2k_80

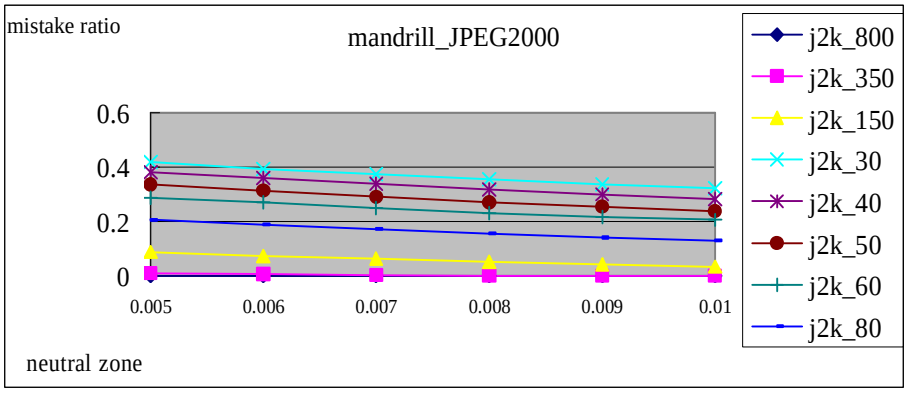
(d)


mistake ratio
goldhill_JPEG2000
0.12
0.1
0.08
0.06
0.04
0.02
0
0.005
0.006
0.007
0.008
0.009
0.01
neutral zone
j2k_800
j2k_350
j2k_150
j2k_30
j2k_40
j2k_50
j2k_60
j2k_80

(e)


mistake ratio
zelda_JPEG2000
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0.005
0.006
0.007
0.008
0.009
0.01
neutral zone
j2k_800
j2k_350
j2k_150
j2k_30
j2k_40
j2k_50
j2k_60
j2k_80

(f)



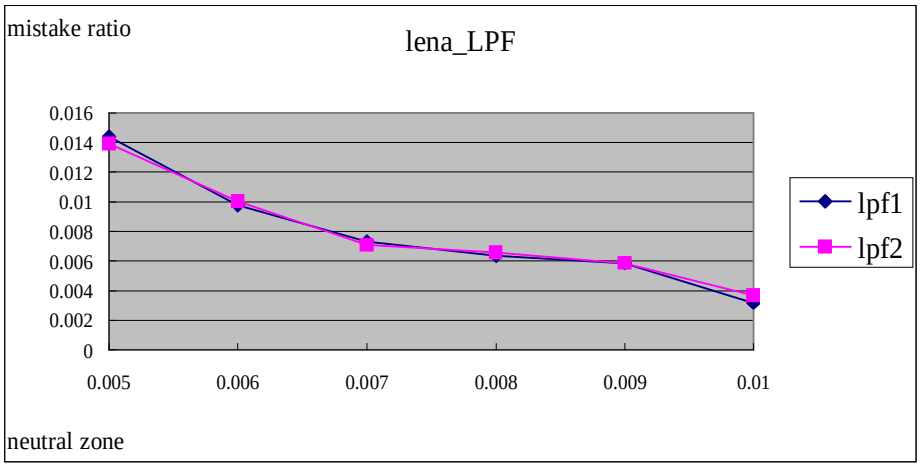
(g)

Figure5.4 The neutral zone via mistake ratio in JPEG 2000 compression.
(a) Lena (b) Pepper (c) Boat (d) Goldhill (e) Zelda (f) Barbara (g) Mandrill

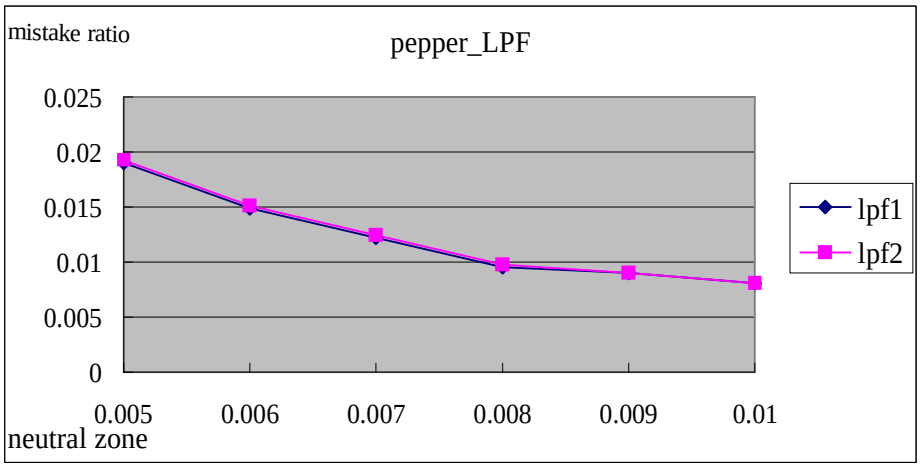
Both the JPEG and JPEG2000 experiment results show that the neutral zone influence the mistake ratio of blocks in an image. The image that contains more active blocks such as mandrill will induced the authentication system performs worse. Since the decision of the authentication result will depend on the threshold of mistake ratio, the suitable neutral zone in the experiment may be 0.01.

5.2.3 Low Pass Filter

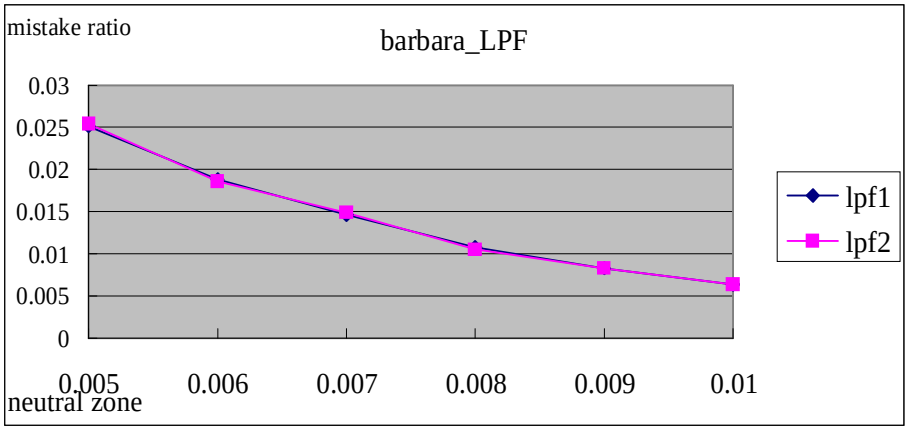
The low pass filter is the Gaussian filter with size 3x3 and 5x5. The image goes through the low pass filtering changes little. The performance of the authentication system is well expect of the mandrill.


mistake ratio
lena_LPF
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
0.005
0.006
0.007
0.008
0.009
0.01
lpf1
lpf2
neutral zone

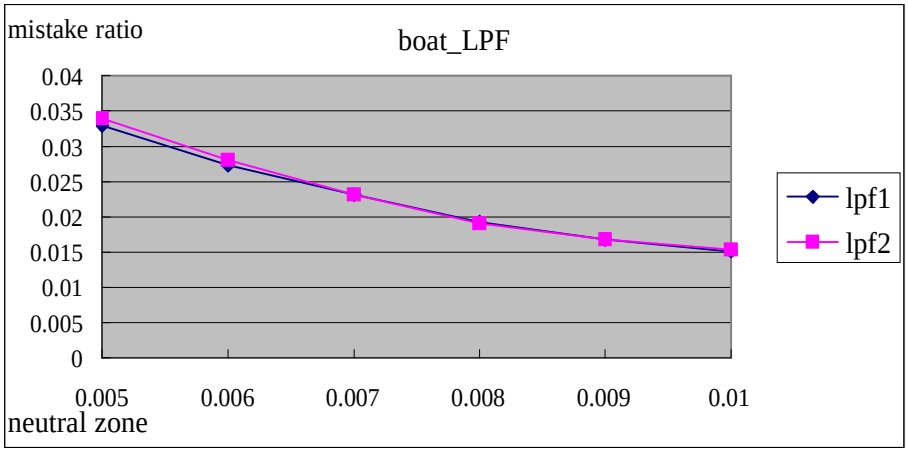
(a)


mistake ratio
pepper_LPF
0.025
0.02
0.015
0.01
0.005
0
0.005
0.006
0.007
0.008
0.009
0.01
lpf1
lpf2
neutral zone

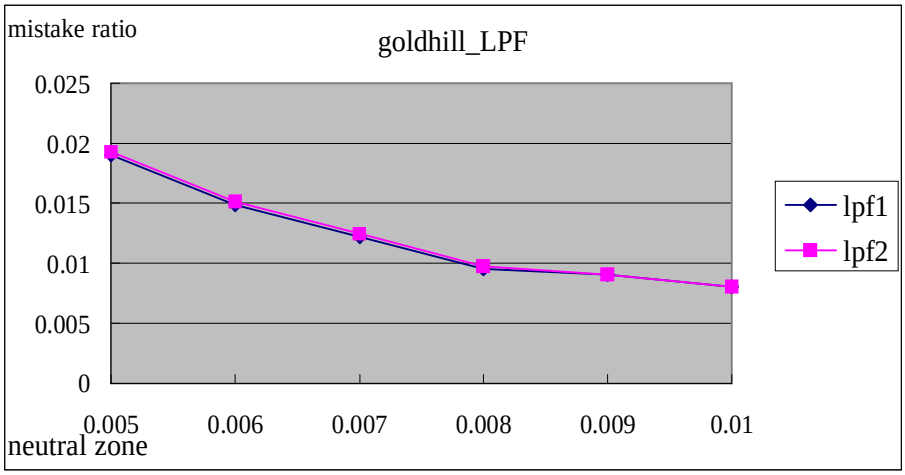
(b)


mistake ratio
barbara_LPF
0.03
0.025
0.02
0.015
0.01
0.005
0
0.005
0.006
0.007
0.008
0.009
0.01
lpf1
lpf2
neutral zone

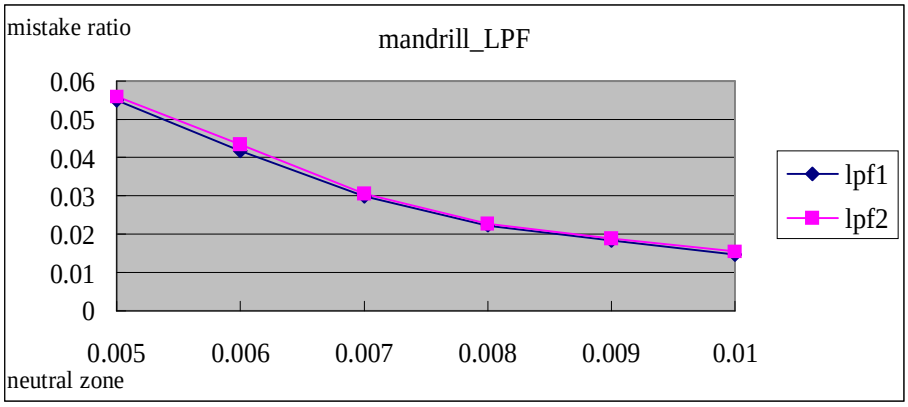
(c)



(d)



(e)

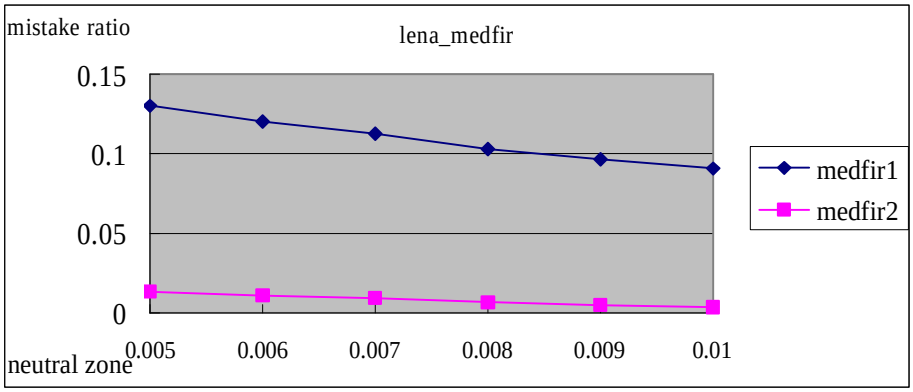


(f)

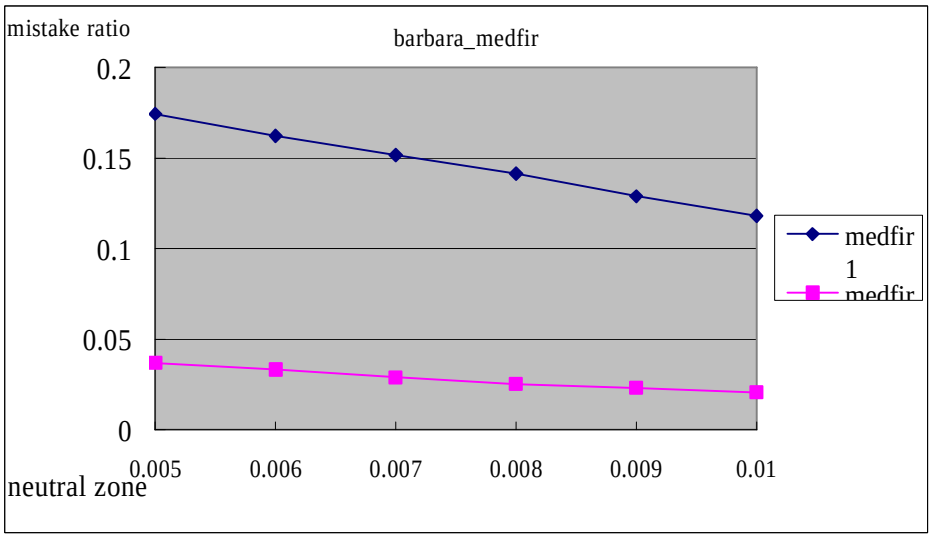
Figure5.5 The neutral zone via mistake ratio in Low Pass Filter
(a) Lena (b) Pepper (c) Boat (d) Goldhill (e) Zelda (f) Barbara (g) Mandrill

5.2.4 Medium Filter

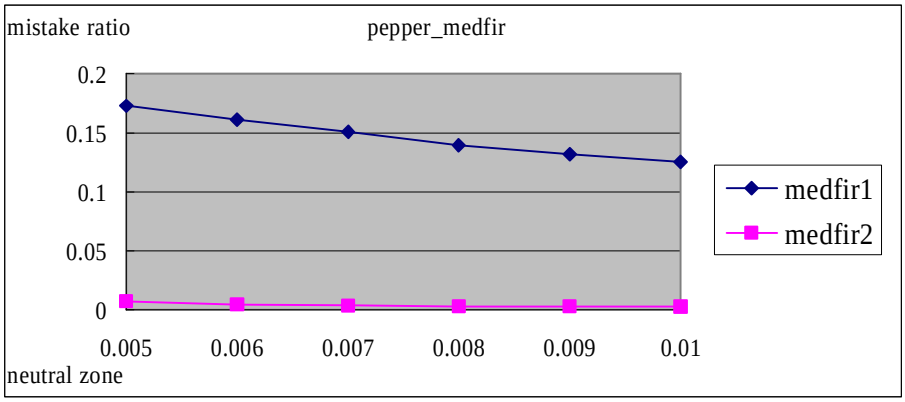
The medium filter is with size 2x2 and 3x3. The performance difference may due to the size and the dominated point.



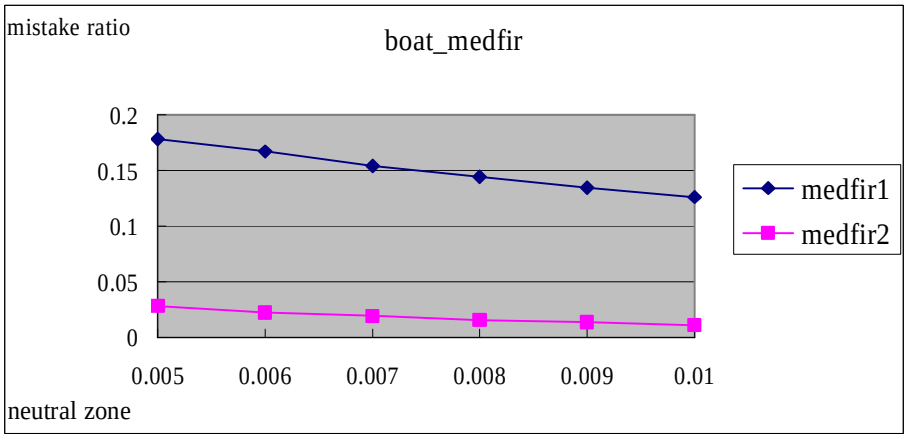
(a)



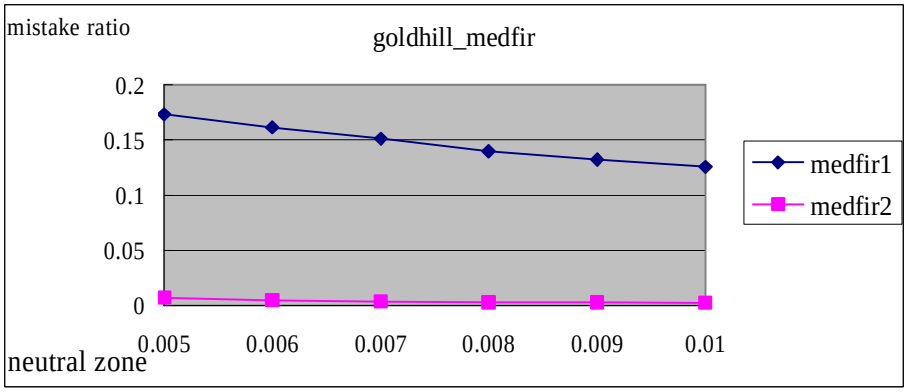
(b)


mistake ratio
pepper_medfir
0.2
0.15
0.1
0.05
0
0.005
0.006
0.007
0.008
0.009
0.01
neutral zone
medfir1
medfir2

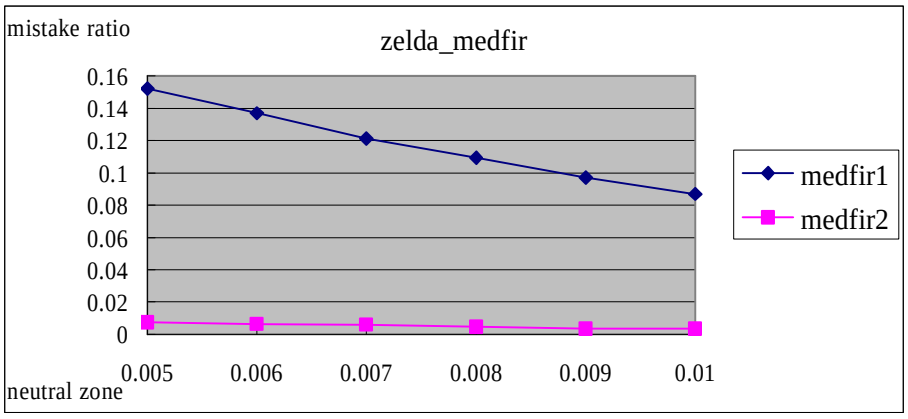
(c)


mistake ratio
boat_medfir
0.2
0.15
0.1
0.05
0
0.005
0.006
0.007
0.008
0.009
0.01
neutral zone
medfir1
medfir2

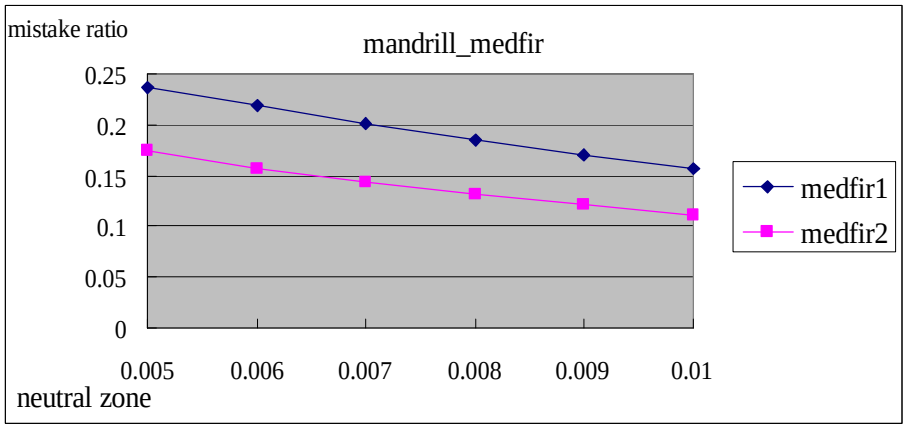
(d)


mistake ratio
goldhill_medfir
0.2
0.15
0.1
0.05
0
0.005
0.006
0.007
0.008
0.009
0.01
neutral zone
medfir1
medfir2

(e)



(f)



(g)

Figure5.6 The neutral zone via mistake ratio in Medium Filter
(a) Lena (b) Pepper (c) Boat (d) Goldhill (e) Zelda (f) Barbara (g) Mandrill

5.3 Detection Accuracy of Malicious Tampering

This section shows the tamper detection with the neutral zone is set to be 0.01.

(a) (b) (c)

Figure5.7 The authentication result（Lena）
(a) original image (b) tampered image (c) authentication result

(a) (b) (c)

Figure5.8 The authentication result（Lena）
(a) original image (b) tampered image (c) authentication result

(a) (b) (c)

Figure5.9 The detection result（Lena）
(a) original image (b) tampered image (c) authentication result

(a) (b) (c)

Figure5.10 The detection result（Pepper）
(a) original image (b) tampered image (c) authentication result

(a) (b) (c)

Figure5.11 The detection result（Barbara）
(a) original image (b) tampered image (c) authentication result

(a) (b) (c)

Figure5.12 The detection result（Boat）
(a) original image (b) tampered image (c) authentication result

(a) (b) (c)

Figure5.13 The detection result（Boat）
(a) original image (b) tampered image (c) authentication result

(a) (b) (c)

Figure5.14 The detection result（Goldholl）
(a) original image (b) tampered image (c) authentication result

(a) (b) (c)

Figure5.15 The detection result（Mandrill）
(a) original image (b) tampered image (c) authentication result

(a) (b) (c)

Figure5.16 The detection result（Zelda）
(a) original image (b) tampered image (c) authentication result

Chapter 6 Conclusions

This thesis proposes an image authentication system that has the incidental distortions tolerance and with the localization ability. The proposed system uses the eigenvalue decomposition to analyze the image blocks. The dominated eigenvalues will be regarded as the feature of one image block. Signature corresponds to an image block will depend on the quantization result of the block feature.

The experiment results show the authentication system takes the image pixels as random variables can dominate the characteristics of an image block. The statistical characteristics of an image block will not change a lot after the incidental distortions since the system can achieve the incidental tolerance.

For the malicious tampering, the proposed system can detect accurately whether to the intensities or frequency modifications.

The eigen-analysis characterizes the images well because the natural images are often much correlated in their pixels. The performance might be improved by designing other classification method to the features.

Bibliography

[1] P. W. Wong, "A public key watermark for image verification and authentication," *in Proc. International Conference*, vol. 1, no. 4-7, pp. 455 - 459, Oct. 1998.

[2] M. M. Yeung, and F. Mintzer, "An invisible watermarking technique for image verification," *in Proc. International Conference*, vol. 2 , no. 26-29, pp. 680 - 683, Oct. 1997.

[3] H. Matthew. and Nasir Memon, "Counterfeiting attacks on oblivious block-wise independent invisible watermarking scheme," *IEEE Trans. Image Processing*, vol. 9, no. 3, pp. 432-441, Mar. 2000.

[4] M. Utku Celik, G. Sharma, E. Saber, and A. Murat Tekalp, "Hierarchical watermarking for secure image authentication with localization," *in Proc., IEEE Trans.*, vol. 11, issue: 6, pp. 585 – 595, June 2002.

[5] D. Kundur, and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *in Proc. IEEE*, vol. 87, issue: 7, pp. 1167 – 1180, July 1999.

[6] C. Y. Lin and S. F. Chang, "A robust image authentication method distinguishing JPEG compression from malicious manipulation," *IEEE Trans. Circuits Syst. Video Technol*, vol. 11, no. 2, pp. 153 – 168, Feb. 2001.

[7] C. S. Lu and H.-Y.M. Liao, "Structural digital signature for image authentication: an incidental distortion resistant scheme," *IEEE Trans. Multimedia*, vol. 5, no. 2, pp. 161 – 173, June 2003.

[8] 游國忠, "應用浮水印技術於影像之智慧財產權保護與認證," *中央大學資訊工程研究所博士論文*, 2001.

[9] Bhattacharjee and M. Kutter, "Compression tolerant image authentication," *in Proc. International Conference*, vol. 1, no. 4-7, pp. 435 - 439, Oct. 1998.

[10] R. B. Wolfgang and E. J. Delp, "Fragile watermarking using the VW2D watermark.," *in Proc. SPIE/IS&T International Conference on Security and Watermarking of Multimedia Contents*, vol. 3657, no. 25-27, pp. 204 -213, 1999, San Jose, CA.

[11] M. U. Celik and G. S, A. M. Tekalp, "Localized lossless authentication watermark (LAW)," *IEEE Trans. Multimedia*, vol. 5, no. 2, pp. 161 – 173, June 2003.

[12] S.C. Byun, I.L. Lee, T.H. Shin, and B.H. Ahn, "A public-key based watermarking for color image authentication," *in Proc. Multimedia and Expo, ICME '02.*, vol. 1, no. 26-29, pp. 593 - 596, Aug. 2002.

[13] I. Kostopoulos, S. Gilani, A.M. and A.N. Skodras, "Colour image authentication based on a self-embedding technique," *in Proc. Digital Signal Processing,14th International Conference on,* vol. 2, no. 1-3, pp. 733 - 736, July 2002.

[14] S. Qibin, S. F. Chang, M. Kurato, and M. Suto, "A quantitative semi-fragile JPEG2000 image authentication system," *in Proc. International Conference on,* vol. 2, no. 22-25, pp. II-921 - II-924, Sep. 2002.

[15] S. Bannour, Azimi-Sadjadi and M.R., "Principal component extraction using recursive least squares learning," *IEEE Trans. Neural Networks,* vol. 6, no. 2, pp. 457 - 469, Mar. 1995.