ACTIVE AND PASSIVE APPROACHES FOR IMAGE AUTHENTICATION

SHUIMING YE

(*M.S., TSINGHUA, CHINA*)

A THESIS SUBMITTED

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

NATIONAL UNIVERSITY OF SINGAPORE

2007

Acknowledgements

I have had the privilege to work with groups of terrific mentors and colleagues over the last four years. They have made my thesis research rewarding and enjoyable. Without them this dissertation would not be possible.

First and foremost, I would like to express my deepest gratitude to my advisors: Qibin Sun and Ee-Chien Chang, for their invaluable guidance and support that direct me towards my research goals. There is no way I could acknowledge enough their help.

I also benefit a lot from the helpful interactions with other members in the media semantics department. Specifically, I would like to thank Dajun He for his kindly help and insightful discussions. I would like to thank Zhi Li for his help of smoothing the writing of every chapters of my thesis. I would also like to thank other current and former department members: Zhishou Zhang, Shen Gao, Xinglei Zhu, Junli Yuan and Yongwei Zhu, for their suggestions and friendships.

I also would like to thank my thesis committee members, Wei Tsang Ooi, Kankanhalli Mohan, and Hwee Hua Pang, for their constructive comments.

I would like to thank Qi Tian, Shih-Fu Chang, Yun-Qing Shi, Min Wu, Ching-Yung Lin, and Tian-Tsong Ng, for their advices.

Last but not least, I would like to thank all members of my family for their perpetual understanding and support of my study. I especially thank my parents for everything. No words can express my gratitude to my wife, Xue Yang, who has provided invaluable and indispensable support of my pursuing such a long term dream and all the future ones.

Table of Contents

Summary

The generation and manipulation of digital images is made simple by widely available digital cameras and image processing software. As a consequence, we can no longer take the authenticity of a digital image for granted. This thesis investigates the problem of protecting the trustworthiness of digital images.

Image authentication aims to verify the authenticity of a digital image. General solution of image authentication is based on digital signature or watermarking. A lot of studies have been conducted for image authentication, but thus far there has been no solution that could be robust enough to transmission errors during images transmission over lossy channels. On the other hand, digital image forensics is an emerging topic for passively assessing image authenticity, which works in the absence of any digital watermark or signature. This thesis focuses on how to assess the authenticity images when there is uncorrectable transmission errors, or when there is no digital signature or watermark available.

We present two error resilient image authentication approaches. The first one is designed for block-coded JPEG images based on digital signature and watermarking. Pre-processing, error correct coding, and block shuffling techniques are adopted to stabilize the features used in this approach. This approach is only suitable for JPEG images. The second approach consists of a more generalized framework, integrated with a new feature distance measure based on image statistical and spatial properties. It is robust to transmission errors for both JPEG and JPEG2000 images. Error concealment techniques for JPEG and JPEG2000 images are also proposed to improve the image quality and authenticity. Many acceptable manipulations, which were incorrectly detected as malicious modifications by the previous schemes, were correctly classified by the proposed schemes in our experiments.

We also present an image forensics technique to detect digital image forgeries, which works in the absence of any embedded watermark or available signature. Although a forged image often leaves no visual clues of having been tampered with, the tampering operations may disturb its intrinsic quality consistency. Under this assumption, we propose an image forensics technique that could quantify and detect image quality inconsistencies found in tampered images by measuring blocking artifacts or sharpness. To measure the quality inconsistencies, we propose to measure the blocking artifacts caused by JPEG compression based on quantization table estimation, and to measure the image sharpness based on the normalized Lipschitz exponent of wavelet modulus local maxima.

List of Figures

List of Tables

Chapter 1

Introduction

We are living in a world where seeing is no longer believing. The increasing popularity of digital cameras, scanners and camera-equipped cellular phones makes it easy to acquire digital images. These images spread widely through various channels, such the Internet and Wireless networks. They can be manipulated and forged quickly and inexpensively with the help of sophisticated photo-editing software packages on powerful computers which have become affordable and widely available. As a result, a digital image no longer holds the unique stature as a definitive recording of scenes, and we can no longer take the integrity or authenticity of it for granted. Therefore, image authentication has become an important issue to ensure the trustworthiness of digital images in sensitive application areas such as government, finance and health care.

Image authentication is the process of verifying the authenticity and integrity of an image. Integrity means the state or quality of being complete, unchanged from its source, and not maliciously modified. This definition of integrity is synonymous with the term of authenticity. Authenticity is defined [1] as "the quality or condition of being authentic, trustworthy, or genuine". Authentic means "having a claimed and verifiable origin or authorship; not counterfeit or copied" [1]. However, when used together with integrity in this thesis, authenticity is restricted in the meaning of quality of being authentic that verified entity is indeed the one claimed to be.

1.1 Motivations

The image trustworthiness is especially important in sensitive applications such as finance and health care, where it is critical and often a requirement for recipients to ensure that the image is authentic without any malicious tampering. Applications of image authentication also include courtroom evidence, insurance claims, journalistic photography, and so on. For instance, in applications of the courtroom evidence, when an image is provided as evidence, it is desirable to be sure that this image has not been tampered with. In electronic commerce, when we purchase multimedia data from the Internet, we need to know whether it comes from the alleged producer and must be assured that no one has tampered with the content. That is to say, the trustworthiness of an image is required for the image to be digital evidence or a certified product.

Image authentication differs from other generic data authentication in its unique requirements of integrity. An image can be represented equivalently in different formats, which may have exactly the same visual information but totally different data representations. Images differ from other generic data in their high information redundancy and strong correlations. Images are often compressed to reduce its redundancy which may not change its visual content. Therefore, robust image authentication is often desired to authenticate the content instead of the specific binary representation, i.e., to pass the image as authentic when the semantic meaning of it remains unchanged. In many applications, image authentication is required to be robust to acceptable manipulations which do not modify the semantic meaning of the image (such as contrast adjustment, histogram equalization, lossy compression and lossy transmission), whereas be sensitive to malicious content modifications (such as object removal or insertion).

The rapid growth of the Internet and Wireless communications has led to an increasing interest towards the authentication of images damaged by transmission errors, where the conventional image authentication would usually fail. During lossy transmission, there is no

guarantee that every bit of the received images is correct. Moreover, compressed images are very sensitive to errors, since compression techniques such as variable length coding lead to error propagations. As a result, image authentication would be required to be robust to transmission errors, but sensitive to malicious modifications at the same time. Previous image authentication approaches may fail in being robust to these errors. Therefore, *error resilient image authentication* is desired, which is the image authentication technique which is robust enough to transmission errors under some levels.

Approaches of image authentication are mainly based on watermarking or digital signatures. This direction is often referred as *active image authentication*, a class of authentication techniques that uses a known authentication code embedded into the image or sent with it for assessing the authenticity and integrity at the receiver. However, this category of approaches requires that a signature or watermark must be generated at precisely the time of recording or sending, which would limit these approaches to specially equipped digital devices. It is a fact that the overwhelming majority of images today do not contain a digital watermark or signature, and this situation is likely to continue for the foreseeable future. Therefore, in the absence of widespread adoption of digital watermark or signature, there is a strong need for developing techniques that can help us make statements about the integrity and authenticity of digital images.

Passive image authentication is a class of authentication techniques that uses the received image itself only for assessing its authenticity or integrity, without any side information (signature or watermark) of the original image from the sender. It is an alternative solution for image authentication in the absence of any active digital watermark or signature. As a passive image authentication approach, digital image forensics is a class of techniques for detecting traces of digital tampering without any watermark or signature. It works on the assumption that although digital forgeries may leave no visual clues of having been tampered with, they may, nevertheless, disturb the underlying statistics property or quality consistency of a natural scene image.

1.2 Research Objectives

The overall purpose of this thesis is to develop new authentication techniques to protect the trustworthiness of digital images. The techniques developed can be put into two research topics: error resilient image authentication and image forensics based on image quality inconsistencies.

1.2.1 Error Resilient Image Authentication

Image transmission over lossy channels is usually affected by transmission errors due to environmental noises, fading, multi-path transmission and Doppler frequency shift in wireless channel [2], or packet loss due to congestion in packet-switched network. Normally errors under a certain level in images would be tolerable and acceptable. Therefore, it is desirable to check image authenticity and integrity even if there are some uncorrectable but acceptable errors. For example, in electronic commerce over mobile devices, it is important for recipients to ensure that the received product photo is not maliciously modified. That is, image authentication should be robust to acceptable transmission errors besides other acceptable image manipulations such as smoothing, brightness adjusting, compressing or noises, and be sensitive to malicious content modifications such as object addition, removal, or position modification.

A straightforward way of image authentication is to treat images as data, so that data authentication techniques can be used for image authentication. Several approaches to authenticate data stream damaged by transmission errors have been proposed. Perrig et al. proposed an approach based on efficient multi-chained stream signature (EMMS) [3]. The basic idea is that the hash of each packet is stored in multiple locations, so that the packet can be verified as long as not all these hashes are lost. However, in this approach there would be large transmission payload due to multiple hashes for one packet. Furthermore, the

computing overhead would be very large if this approach is applied directly to image authentication, since the size of an image is always very large compared with the size of a packet. Golle et al. proposed to use an augmented hash chain of packets [4] instead of Perrig's multiple signatures for one packet. This approach may reduce the communication payload, but very large computing payload can still be expected. In summary, treating images as data stream during authentication does not take advantage of the fact that images are tolerable to certain degree of errors, and the computing payload would be very large. Therefore, it is not suitable for these data approaches to be applied directly to image authentication.

An image can be represented equivalently in different formats, which have exactly the same visual information but totally with different data representation. Image authentication is desirable to authenticate the image content instead of its specific binary representation, which passes the image as authentic when its semantic meaning remains unchanged [5, 6]. Some distortions which do not change the meaning of images are tolerable. It is desirable to be robust to acceptable manipulations which do not modify the semantic meaning of the image (such as contrast adjustment, histogram equalizing, compression, and lossy transmission), while be able to detect malicious content modifications (such as object removed, added or modified). In order to be robust to acceptable manipulations, several robust image authentication algorithms were proposed, such as signature-based approaches [7, 8, 9] and watermarking based approaches [10, 11].

Content-based image authentication, the main robust authentication technique, typically uses a feature vector to represent the content of an image, and the signature of this image is calculated based on this feature vector instead of the whole image. However, content-based authentication typically measures feature distortion in some metrics, so authenticity fuzziness would be introduced in these approaches which may even make the authentication result useless. Furthermore, transmission errors would damage the encrypted

signatures or embedded watermarks. Therefore, previous techniques would fail if the image is damaged by transmission errors.

Although many studies have been done on robust image authentication and error resilient data authentication, no literature is available on error resilient image authentication. Transmission errors affect the image authentication in three ways. Firstly, many of the standard signature techniques at present require that all received bits are correct. As a result, there would be significant overhead due to retransmission and redundancy in applying standard signature techniques to image data, which lead to the unavoidable increase of transmission payload [12]. Secondly, by requiring all bits received correctly, this system cannot verify the received image if there are errors during transmission. In this case, this system cannot take advantage of the fact that multimedia applications are tolerable to some errors in bitstreams, which can be achieved by *error concealment* techniques. Finally, transmission errors can damage embedded watermarks, removing them from the image or reducing the robustness. Therefore, there is an emergent need of authenticating images degraded during lossy transmission. The first problem this thesis focuses on is *how to authenticate images transmitted through lossy channels when there are some uncorrectable transmission errors*.

Accordingly, the first purpose of this thesis is to develop techniques for authenticating images received through lossy transmission when there are some uncorrectable transmission errors. It aims to distinguish the images damaged by causal transmission errors from the images modified by the malicious users. It focuses on the development of error resilient image authentication schemes incorporated with error correcting code, image feature extraction, transmission error statistics, error concealment, and perceptual distance measure for image authentication.

We propose error resilient image authentication techniques which can authenticate images correctly even if there uncorrectable transmission errors. An image feature distance

measure is also proposed to improve image authentication system performance. The proposed perceptual distance measure is quite general that it is able to be used in many content-based authentication schemes which use features containing spatial information, such as edge [7, 13], block DCT coefficients based features [8, 14, 15], highly compressed version of the original image [9], block intensity histogram [16]. The proposed perceptual distance measure, when used as the feature distance function in image authenticity verification stage, will improve the system discrimination ability. Many acceptable manipulations, which were detected as malicious modifications in the previous schemes, can be bypassed in the proposed scheme. The proposed feature distance measure can be incorporated in a generic semi-fragile image authentication framework [15] to make it able to distinguish images distorted by transmission errors from maliciously tampered ones.

Cryptography and digital signature techniques are beyond the scope of this thesis, since they have been well studied in the data security area, and are not the key techniques that make our research different from others. The authentication techniques proposed in this thesis can produce good robustness against transmission errors and some acceptable manipulations, and can be sensitive to malicious modifications. Moreover, the perceptual distance measure proposed for image authentication would improve the system performance of content-based image authentication schemes.

1.2.2 Passive Image Authentication based on Image Quality Inconsistencies

A requirement of active image authentication is that a signature or watermark must be generated and attached to the image. However, at present the overwhelming majority of images do not contain digital watermark or signature. Therefore, in the absence of widespread adoption of digital watermark or signature, there is a strong need for developing

techniques that can help us make statements about the integrity and authenticity of digital images. Passive image authentication is a class of authentication techniques that uses the image itself for assessing the authenticity of the image, without any active authentication code of the original image. Therefore, the second problem this paper focuses on is *how to passively authenticate images without any active side information from signature or watermark*.

Accordingly, the second purpose of this thesis is to develop methods for authenticating images passively by evaluating image quality inconsistencies. The rationale is to use image quality inconsistencies found in a given image to justify whether the image has been maliciously tampered with.

One approach of passive image authentication is to detect specific operations as the traces of image modifications. Several specific operations have been used, such as copy-move forgery [17], color filter array interpolation [18], and so on. Another approach is based on statistical properties of natural image [19, 20], with the assumption that modifications may disrupt these properties. However, these approaches may be effective only in some aspects and may not always be reliable. They may neglect the fact that the quality consistencies introduced during the whole chain of image acquiring and processing would be disrupted by digital forgery creation operations. Few studies have been done based on detection of these image quality inconsistencies.

We propose to use *content independent* image quality inconsistencies in the image to detect the tampering. Images from different imaging systems in different environments would be of different qualities. When creating digital forgery, there are often parts from different sources of images. If the image is a composite from two different sources, there would be quality inconsistencies found in it, which can be as a proof of its having been tampered with. A general framework for digital image forensics is proposed in this thesis to detect digital forgery by detecting inconsistencies of the image using JPEG blocking

artifacts and image sharpness measures. For a given source of digital image, the distortions introduced during image acquisition and manipulation can be served as a "natural authentication code", which are useful to identify the source of image or detect digital tampering. The developed digital image forensics technique would be useful in assisting the human experts for investigation of image authenticity.

The assumption that the digital forgery creation operations will disrupt image quality consistency is adopted in this thesis. Therefore, our work focuses on the discovery of quality consistency introduced in the whole chain of digital image creation and modification, and its use in detecting digital forgeries. The results of this thesis may provide a passive way to protect the trustworthiness of digital images by distinguishing authentic images from digital forgeries. Moreover, the results of our image forensics technique may lead to a better understanding of the role of quality consistencies introduced in digital imaging chain for detecting digital forgeries.

In summary, the objective of our thesis is to develop image authentication techniques to verify the authenticity and integrity of a digital image, when the image is damaged by transmission errors during transmission or there is no side information available from digital signature or watermark. Our approaches make use of techniques from various areas of research, such as computer vision, machine learning, statistics analysis, pattern classification, feature extraction, digital cryptography, digital watermarking, and image analysis.

1.3 Thesis Organization

This thesis is organized as follows. In Chapter 2, a review of state-of-the-art related work is presented, including active image authentication and image forensics techniques. The proposed error resilient image authentication scheme is present in Chapter 3. In Chapter 4,

we describe the feature distance measure for content-based image authentication and its application in error resilient image authentication. Image forensics based on image quality inconsistencies is present in Chapter 5. Chapter 6 concludes this thesis with some comments on future work in image authentication.

Chapter 2

Related Work

Image authentication, an important technique for protecting the trustworthiness of digital images, is mainly based on active approaches using digital signature or watermarking. The rapid growth of Internet and Wireless communications has led to the increasing interest towards authentication of images damaged by transmission errors. On the other hand, today most digital images do not contain any digital watermark or signature, so there is an emerging research interest towards passive image authentication techniques.

This chapter examines previous works on active and passive image authentication that are relevant to this thesis. In Section 2.1, we review active image authentication techniques, including discussions on the differences between image authentication and data authentication, robustness and sensitivity requirements of image authentication, content-based image authentication, error resilient data authentication, and digital signature or watermarking based approaches. In Section 2.2, we review the image forensics techniques, including the analysis of the distortions introduced during the digital image generation and manipulation, image forensics based on the detection of specific manipulation, image forensics based on passive integrity checking, and image quality measures for image forensics. This chapter sets up the context of our research topics of error resilient image authentication and passive image authentication using image quality measures.

2.1 Active Image Authentication

Active image authentication uses a known authentication code during image acquiring or sending, which is embedded into the image or sent along with it for assessing its authenticity or integrity at receiver side. It is different from classic data authentication. Robustness and sensitivity are the two main requirements of active image authentication. The main approaches of active image authentication are based on digital watermarking and digital signatures.

2.1.1 Preliminaries of Active Image Authentication

It is useful to discover the differences between image authentication and data authentication in order to exploit data authentication techniques for image authentication or to develop particular image authentication techniques. Robustness, which is a key requirement of image authentication, makes image authentication different from general data authentication. Based on different level of robustness, image authentication can be classified into complete authentication and soft authentication. Content-based image authentication is a main approach of soft authentication.

Differences between Image Authentication and Data Authentication

The main difference between image authentication and data authentication would be that image authentication is generally required to be robust to some level of manipulation, and data authentication technique would not accept any modification. General data authentication has been well studied in cryptography [21]. A digital signature, which is usually in an encrypted form of the hash of the entire data stream, is generated from the original data or the originating entity. The classic data authentication can generate only a

binary output (tampered or authentic) for the whole data, irrespective of whether the manipulation is minor or severe. Even if one bit changed in the data, the verification will fail due to the properties of the hashing function [22]. On the contrary, image authentication is desirable to be based on the image content so that an authenticator remains valid across different representations of the image as long as the underlying content has not changed.

Authentication methods developed for general digital data could be applied to image authentication. Friedman [23] discussed its application to create a "trustworthy camera" by computing a cryptographic signature that is generated from the bits of an image. However, unlike other digital data, image signals are often in a large volume and contain high redundancy and irrelevancy. Some image processing techniques, such as compression, are usually required to be applied to image signals without affecting the authenticity. Most digital images are now stored or distributed in compressed forms, and would be transcoded during transmission which would change the pixel values but not the content. Due to the characters of image signals, manipulations on the bitstreams without changing the meaning of content are considered as acceptable in some applications, such as compression and transcoding. Classical data authentication algorithms will reject these manipulations because the exact representation of the signal has been changed. In fact, classical data authentication can only authenticate the binary representation of digital image instead of its content. For example, in [23], if the image is subsequently converted to another format or compressed, the image will fail the authentication.

In summary, due to the difference between image authentication and data authentication, it is not suitable to directly apply general data authentication techniques to image authentication. The reason would be that the conventional data authentication techniques are not capable of handling distortions that would change the image representation but not the semantic meaning of the content. In addition, long computation time and heavy computation load are expected since the size of an image could be very large.

Robustness and Sensitivity of Image Authentication

The requirement on a certain level of authentication robustness is the main difference between data authentication and image authentication. An image authentication system would be evaluated based on the following requirements with variable significances in different applications:

- Robustness: The authentication scheme should be robust to acceptable manipulations such as lossy compression, lossy transmission, or other content-preserving manipulations.
- Sensitivity: The authentication scheme should be sensitive to malicious modifications such as object insertion or deletion.
- Security: The image cannot be accepted as authentic if it has been forged or maliciously manipulated. Only authorized users can correctly verify the authenticity of the received image.

In image authentication, these requirements highly depend on the definitions of acceptable manipulations and malicious modifications. Commonly, manipulations on images can be classified into two categories as follows:

- Acceptable manipulations: Acceptable (or incidental) manipulations are the ones which do not change the semantic meaning of content and are acceptable by an authentication system. Common acceptable manipulations include format conversions, lossless and high-quality lossy compression, resampling, etc.
- Malicious manipulations: Malicious manipulations are the ones that change the semantic meaning, and should be rejected. Common malicious manipulations include cropping, inserting, replacing, reordering perceptual objects in images, etc.

Note that different applications may have different criteria of classifying manipulations. The manipulation considered as acceptable in one application could be considered as malicious in another application. For example, JPEG image compression is generally considered as acceptable in most applications, but may be rejected for medical images since loss of details during lossy compression may render a medical image useless.

Complete Image authentication and Soft authentication

Based on the robustness level of authentication and the distortions introduced into the content during image signing, image authentication techniques can be classified into two categories: complete (or hard) authentication and soft authentication. Complete authentication refers to techniques that consider the whole image data, and do not allow any manipulations or transformation. Soft authentication passes certain acceptable manipulations and rejects all the rest malicious manipulations. Soft authentication can be further divided into quality-based authentication, which rejects any manipulations that makes the perceptual quality decrease below an acceptable level, and content-based authentication, which rejects any manipulations that change the semantic meaning of the image.

Early works on image authentication are mostly complete authentication. If images are treated as data bitstreams, many previous data signature techniques can be directly applied to image authentication. Then, manipulations will be detected because the hash values of the altered message bits will not match the information in the digital signature. In practice, fragile watermarks or traditional digital signatures may be used for complete authentication.

On the contrary, normally distortions in images under a certain level would be tolerable and acceptable in many applications. Therefore, it is desirable that image

authentication should be robust to these acceptable image manipulations. These requirements motivate the development of soft authentication techniques.

Content-based Image Authentication

An efficient soft image authentication approach could be content-based authentication, which passes images as authentic if the image content remains unchanged [5]. It typically uses a feature vector to represent image content, and the authentication code of this image is calculated based on this feature vector instead of the whole bit-stream representation. Content-based authentication uses soft decision to judge the authenticity [5], which typically measure authenticity in terms of the distance between a feature vector of the received image and its corresponding vector of the original image, and compares the distance with a preset threshold to make a decision.

Several content-based authentication schemes have been proposed [24, 7, 8, 13, 14, and 10], which could pass certain acceptable manipulations, and reject all the rest. The main difference between these schemes is what kind of feature is used. Moment is used as the feature in [7], edge in [7, 13], DCT coefficients in [8, 14], and Wavelet coefficients in [10].

These content-based authentication schemes have a common problem that there is typically no sharp boundary between authentic images and unauthentic images [14]. This intrinsic fuzziness makes challenges to these authentication schemes. A fuzzy region exists between the surely authentic and unauthentic images in [14], where the authenticity of the images is difficult to ascertain. A solution to do with this problem is to introduce human intervention [25], in which a human is required to distinguish acceptable manipulations from malicious modifications.

Furthermore, it is difficult for these techniques to survive network transmissions and error concealment during transmission over lossy networks. Typically the best-effort networks have no guarantee on the correctness of every received bit of images.

Transmission errors are inevitable in lossy networks such as wireless channel (environmental noises fading, multipath and Doppler frequency shift [2]), or the Internet (packet loss due to congestion when using UDP over IP protocol). In this paper, both the packet loss in Internet and noises in wireless network are referred to as transmission errors.

Error Resilient Authentication for Data Stream over Lossy Channels

Authenticating data stream over lossy channels has been studied in cryptography field, such as signature-based data streaming authentication schemes [3, 4]. In these schemes, a data stream of packets is divided into a number of blocks. Within each block, the hash of each packet is appended to some other packets which in turn generate new hashes appended to other packets. This hash-and-concatenate process continues until it reaches the last packet, which is the only packet in this block signed by the signature algorithm. In these schemes the verification of each packet is not guaranteed in the presence of loss, but instead it is assured that this can be done with a certain probability.

The main difference between these hash-chaining schemes [3, 4] is how to construct the hash chaining topology, that is, in what way the packets should be linked. Perrig et al. proposed an Efficient Multi-chained Stream Signature (EMSS) scheme [4] which is robust against packet losses by storing the hash of each packet in multiple locations and appends multiple hashes in the signature packet. The basic idea this scheme is that when a packet is lost, its hash will be found in other packets unless total packet loss of a segment exceeds a threshold. Golle and Modadugu [3] proposed an Augmented Chain Stream Signature (ACSS) scheme in which a systematic method of inserting hashes in strategic locations so that the chain of packets formed by the hashes will be resistant to a burst loss.

These hash-chaining based schemes would not be suitable to be directly applied to image authentication, because directly applying these schemes to image authentication has

several drawbacks: (1) long computation time and heavy computation load are required. The reason is that the size of an image is still tremendously huge even if it has been compressed; (2) the direct application of digital signatures to an image is vulnerable to image processing such as compression or contrast adjustment which are commonly considered to be acceptable; (3) with the increase of Bit Error Rate (BER) and the need of time synchronization, the transmission overhead will be unavoidably large; (4) in image transmission, the importance and the size of packets vary in different environments. It may not be practical to generate hash functions from pre-defined fixed boundaries; (5) treating an image as data bit stream, it does not taking advantage of the fact that image is tolerable to certain degree of errors.

2.1.2 Approaches of Active Image Authentication

The main approaches of active image authentication are based on digital watermarking or digital signatures, as well as some combinatory methods that use both of them.

Image Authentication based on Digital Signature

A digital signature is an external authentication code generated from the original message, which is usually an encrypted form of some kind of hash values [24]. The signature includes the encrypted authentication code that is to be authenticated, as well as some other information such as the issuer, the owner, and the validity period of the public key. A public key certificate is a digitally signed message consisting of two parts which can be used for authentication using a public key.

Digital signature standard (DSS) is a typical technology for data authentication, which consists of two phases – signature generation and signature verification [21]. Given a

message of arbitrary length, a short fixed-length digest is obtained by a secure hash function. The signature is generated using the sender's private key to sign on the hashed digest. The original message associated with its signature is then sent to the intended recipients. Later on, the recipient can verify whether the received message has been altered, and whether the message were really from the sender, by using the sender's public key to authenticate the validity of the attached signature. The final authentication result is drawn from a bit-bit comparison between two hash codes (one is decrypted from the signature and the other is obtained by re-hashing the received message). Even one bit difference existing in the received message will be deemed unauthentic.

Due to its great success in data authentication, DSS could be also employed in image authentication [7, 26, 27, 28, 29]. In this type of image authentication, the sender's private key is used to sign the feature of the original image to generate a digital signature. During verification, a public key is used to decrypt to get the original feature, and compared with a feature extracted from the received image to determine the image authenticity.

Image Authentication based on Digital Watermarking

Image authentication is classically handled through digital signature by cryptography. However, digital signature can only work when an authentication message is transmitted with the media. In signature-based authentication, the digital signature is stored either in the header of format or in a separate file. Therefore, the risk of losing the signature is always a major concern. It does not protect against unauthorized copying after the message has been successfully received and decrypted. Furthermore, although complex cryptographic techniques generally make the cracking of the system difficult, they are also expensive to implement.

Digital watermarking is an effective way to protect copyright of image data even after transmission and decryption. It is a concept of embedding a special pattern (watermark) into a host signal so that a given piece of information, such as the owner's or authorized consumer's identity, is indissolubly tied to the data. This information can later be used to prove ownership, identify a misappropriating person, trace the marked document's dissemination through the network, or simply inform users about the rights-holder or the permitted use of the data.

Compared with digital signature, digital watermarking takes advantage of the fact that all images contain a small amount of data that does not usually have a discernible effect on their appearances. These data are often treated as "noise" because they are random and usually nonsensical. Digital watermarking creates a message that mimics the noise data and embeds it as a digital watermark. In addition, digital watermarks are very durable. A robust digital watermark can survive many kinds of image manipulations (including blur, rotate, cut, paste, crop, and color separation), data compression, and multiple generations of reproduction across a variety of digital and print media. Watermarking has many applications, such as broadcast monitoring, owner identification, proof of ownership, authentication, transactional watermarks, copy control and covert communication [30].

All digital watermarking techniques consist of two phases: watermark embedding and watermark detection. In watermark embedding, the cover message and the secret key are combined to produce a stego object, which consists of the cover object with a watermark embedded in it. Then, to determine either authenticity or copyright ownership of the stego object, the secret key and the stego object are combined in the process of watermark extraction, which recovers and/or verifies the watermark. Digital watermarking can be divided into various categories in various ways. Generally it can be classified into three types: pixel domain (least significant bit replacement) and frequency domain techniques.

The most straight-forward method of watermark embedding, would be to embed the watermark into the least-significant-bits (LSB) of the cover object, e.g., to insert watermark bits into the least significant bits of an image. LSB substitution is simple, but also brings a host of drawbacks. Although it may survive transformations such as cropping, any addition of noise or lossy compression is likely to alleviate the watermark. In a word, LSB modification proves to be a simple and fairly powerful tool for stenography, but lacks the basic robustness that watermarking applications require. Yeung et al. [31] proposed an fragile scheme that a binary watermark is embedded into the original image in pixel domain, and a key dependent binary look-up-table (LUT) is employed as a watermark extraction function to extract watermark pixel-by-pixel. A similar LUT is used in [32], in which watermarking is performed in the DCT domain. Another improved LUT based scheme was proposed in [33], in which the key dependent LUT for a single pixel is replaced by an encryption map.

There are some more robust watermarking methods which are analogous to spread spectrum communications techniques. Modulators and demodulators of classical spread spectrum communications systems are identical to the watermark embedding and extraction process. The noisy transmission is analogous to the distribution and distortion of watermarked data. The communication channel is viewed as the frequency domain of the data signal to be watermarked. The narrowband signal transmitted over this wideband channel represents the watermark. I. Cox et al proposed a spread spectrum-watermarking method [34]. They place the watermark in a perceptually most significant frequency sequence. The watermark in their system is not a binary identification word but the pseudo-noise itself, i.e., a sequence of small pseudo-random numbers.

In frequency domains, discrete cosine transform (DCT) domain is classic and popular for image processing, which allows an image to be broken up into different frequency bands, making it much easier to embed watermarking information into the middle frequency bands of an image. The middle frequency bands are chosen such that they avoid altering the

most visual important parts of the image (low frequencies) without over-exposing themselves to removal through compression and noise attacks (high frequencies) [35]. Another possible domain for watermark embedding is wavelet transform domain [36, 37]. The Discrete Wavelet Transform (DWT) separates an image into a lower resolution approximation image (LL) as well as horizontal (HL), vertical (LH) and diagonal (HH) detail components. One of the many advantages of wavelet transform is that it is believed to be able to model the Humana Visual System (HVS) more accurately, as compared with the FFT or DCT. This allows us to use higher energy watermarks in regions that the HVS is known to be less sensitive to, such as the high resolution detail bands (LH, HL, and HH). Embedding watermarks in these regions allow us to increase the robustness of our watermark, at a little or no additional impact on image quality.

Image Authentication based on Hybrid Digital Signature and Watermark

Digital signature or watermarking based technologies can be independently used for image authentication; moreover, it is possible to implement both of them in the same authentication application, providing a multiple-layer security. The content may have been watermarked after signature generation. The sending party encrypts the watermarked content to provide the second layer of protection. At the receiving end, the signature is decrypted before watermark detection takes place.

A preferable solution is to embed the signature directly into the image using digital watermarking. It inserts an imperceptible watermark into the image at the time of recording. It eliminates the problem of having to ensure that the signature stays with the image. It also opens up the possibility that we can learn more about what kind of tampering has occurred, since any changes made to the image will also be made to the watermark. With the assumption that tampering will alter a watermark, an image can be authenticated by verifying that the extracted watermark is the same as that which was inserted. Thus, the

authentication system can indicate the rough location of changes that have been made to the image. The major drawback of this approach is that a watermark must be inserted at the time of recording or sending, which would limit this approach to specially equipped digital cameras. This method also relies on the assumption that the watermark cannot be easily removed and reinserted.

In summary, the advantages of hybrid digital signature or watermarking scheme include:

- Additional level of security: The hacker will have to attack both the encryption algorithm and watermarking algorithm.
- Multiple uses: The embedded activating share can be a multi-purpose watermark, representing both the key data and copyright or copy control information.

A robust watermarking protocol for key-based video watermarking are proposed in [38]. This protocol generates keys that are both very secure and content dependent using a cryptographically strong state machine. It is robust against many types of video watermarking attacks and supports many kinds of embedding and detection schemes.

However, some applications demand the same security solution on a semi-fragile level, i.e., some manipulations on the content will be considered acceptable (e.g. lossy compression) while some are not allowable (e.g. content modifications). At the semi-fragile level, watermarking-based approaches only work well in protecting the integrity of the content [39], but are unable to identify the source if without other associated solutions. This is because watermarking makes use of a symmetric key for watermark embedding and extracting. Once the key or watermark is compromised, attackers can use the key or watermark to fake other images as authentic. Signature based approaches can work on both the integrity protection of the content and the repudiation prevention of the owner. However, a shortcoming exists that the generated signature is unavoidably large because its size is usually proportional to the image size.

A hybrid digital signature or watermarking system as present in [15] generates short and robust digital signatures based on the invariant message authentication codes (MACs). These MACs are obtained from the quantized original frequency-domain coefficients and ECC-like embedded watermarks. The invariance of MACs is theoretically guaranteed if the images are under lossy compression or other acceptable minor manipulations such as smoothing, brightness change, etc. The whole MACs generated from the signing end have to be preserved in the receiving end. Thus, the size of digital signature is proportional to the image size. The MACs are generated strictly invariant in the signing end and the receiving end, so the hash function can be applied to significantly reduce the size of digital signature [40]. This scheme is robust to transmission errors by using error correction concepts, and is secure by adopting crypto signature.

2.2 Passive Image Authentication

The major drawback of active image authentication based on digital signature or watermarking is that a signature or watermark must be available for authenticity verification, which would limit this approach to special imaging equipments. Passive image authentication is an alternative solution to active authentication when there is no active side information provided by digital signature or watermark. It is a class of authentication techniques that uses the image itself for assessing the authenticity or integrity of the image, without any side information available from the image or the original reference image.

Digital forensics has been defined by the Digital Forensic Research Workshop (DFRWS) as “the use of scientifically derived and proven methods towards the preservation, collection, validation, identification, analysis, interpretation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal or helping to anticipate the unauthorized

actions shown to be disruptive to planned operations" [41]. We use the phrase of digital image forensics as a passive image authentication technique for the purpose of evaluation of the image authenticity or integrity. Image forensics, in this context, is to examine the characteristics of content or to detect the traces of some underlying forgery creation operation trails in the image for detecting forgery.

For image authentication based on digital signature or watermarking, there is a authenticaiton code (side information) embedded in the image or sent with it. For image forensics, there is no such side information available at the receiver. In order to check of image authenticity, it works in a passive blind way, in a very different way compared with active image authentication. It is often based on some prior knowledge about image acquiring, image statistics, and traces of forgery creation operations.

A typical authentication decision is based on the comparison between a preset threshold and the distance of the pattern vector extracted (P_t) from the test image and the original pattern (P_o) from the original image. The main differences between active and passive authentication schemes are:

- For image authentication based on digital signature, the original vector P_o is from a feature vector extracted from the image or the source entity, followed by an optional data-reduction stage and another optional lossless compression to reduce amount of data in the feature vector. And this pattern vector is stored as side information along with the image.
- For image authentication based on watermarking, the original vector P_o is from a feature vector extracted from the image or a predefined pattern. And this pattern vector is embedded into the image to be extracted from it in the stage of verification.

- For passive authentication, both the vectors P_o and P_t come from pattern learning stage or prior knowledge of some operations during image acquiring, processing and transmission.

Therefore, prior knowledge of digital imaging system is useful for digital image forensics. Knowledge from traditional forensics experts would also be useful or incentive for image forensics. Tampered analog photos can be detected by forensic experts in several levels [42]: (1) At the highest level, one may analyze what are inside the image, the relationship between the objects, and so on. Even very advanced information may be used, such as George Washington cannot take photos with George Bush [43]; (2) At the middle level, one may check the image consistency, such as consistency in object sizes, color temperature, shading, shadow, occlusion, and sharpness; (3) At the low level, local features may be extracted for analysis, such as the quality of edge fusion, noise level, and watermark.

Human is very good at high level and middle level analysis and has some ability in low level analysis. On the contrary, computers now still have difficulties in high level analysis, but can be very helpful in middle level and low level analysis, as complement of human examination. Therefore, general approaches of passive digital image authentication could be based on distortion ballistics (detection of the trace of distortions caused by some specific manipulation), image statistics or pattern classification. Image quality measures would also be useful in image forensics.

2.2.1 Image Forensics based on Detection of the Trace of Specific Operation

Although there may be an uncountable number of ways to tamper with digital images, the most common forgery creation operations are:

- Compositing: Two or more digital images are spliced together to create a composite image. It is one of the most common forms of digital forgery creation;
- Resampling, rotating, or stretching portions of the images;
- Brightness, contrast, or color adjustment, such as white balance and gamma correction;
- Filtering or introducing noise to conceal evidence of tampering;
- Compressing or reformatting the result image.

Recently, some digital image forensics approaches have been proposed to detect the traces of specific manipulation applied to the image using statistical techniques, such as detecting the resampling [44], copy-paste [17], JPEG recompression [18], and color filter array interpolation [45, 46, 47, 48].

Most digital cameras are equipped with a single charge-coupled device (CCD) or complementary metal oxide semiconductor (CMOS) sensor, and capture color images using an array of color filters. At each pixel location, only a single color sample is captured. The missing color samples are then inferred from neighboring values. This process, known as color filter array (CFA) interpolation or demosaicking, introduces specific correlations between the samples of a color image. These correlations are typically destroyed when a CFA interpolated image is tampered with, and can be employed to uncover traces of tampering. Using an approach similar to the resampling detection [44], the authors in [45] employed the expectation/maximization (EM) algorithm to detect if the CFA interpolation correlations are missing in any portion of an image. An advantage approach over EM algorithm was proposed in [49], which first assumes a CFA pattern, thereby discriminates between the interpolated and un-interpolated pixel locations and values, and estimates the interpolation filter coefficients corresponding to that pattern for each of three clusters.

In [20], the authors proposed to detect photomontage by a passive-blind approach using improved bi-coherence features (mean of magnitude and negative phase entropy).

Photomontage refers to a paste-up produced by sticking together photographic images. Creation of photomontages always involves image splicing, which refers to a simple putting together of separate image regions, without further post-processing steps. Among all operations involved in image photomontage, image splicing can be considered the most fundamental and essential operation. The block level detection results can be combined in different ways to make global decision about the authenticity of a whole image or its sub-regions

When tampering with an image, a typical pattern is to load the image into some software (e.g., Adobe Photoshop), do some processing, and resave the tampered image. If JPEG format is used to store the images, the resulting tampered image would be double compressed. Double JPEG compression introduces specific correlations between the discrete cosine transform (DCT) coefficients of image blocks. These correlations can be detected and quantified by examining the histograms of the DCT coefficients. While double JPEG compression of an image does not necessarily prove malicious modifications, it raises suspicions that the image may not be authentic. If these histograms of the DCT coefficients contain periodic patterns, then the image is very likely to have been double compressed [18].

2.2.2 Image Forensics based on Feature Inconsistency

The second approach of image forensics is based on statistic properties of the natural images [20, 50, 51, 52, 53], linear filter estimation by blind de-convolution [54], or inconsistencies based on scene lighting direction [55] and camera response normality [43, 56, 57], with the assumption that image forgery creation perturbs the natural images statistics or introduce inconsistent lighting directions. Pattern noise can be used as the other way to detect the origin of image acquired by digital cameras [18]. The pattern noise of a camera can be

considered as a high-frequency spread spectrum watermark to identify the camera from a given image, whose presence in the image is established using a correlation detector.

In [58], a statistical model based on Benford's law for the probability distribution of the first digits of the JPEG coefficients is used to estimate the JPEG quantization factor. In [19] the authors propose a method which could reliably discriminate between tampered images from the original ones. The basic idea is that a doctored image would have undergone some image manipulations like rescaling, rotation, brightness adjustment, etc. They designed classifiers that can distinguish between images that have and have not been processed using these basic operations. Then equipped with these classifiers they applied them successively to a suspicious sub-image of a target image and classify the target as doctored if a sub-image classifies differently from the rest of the image. Natural scene statistics [59, 60] are also used in this scheme. In [19] the authors present a technique for capturing image features that, under some assumptions, are independent of the original image content and hence better represent the image manipulations. They employed several image quality metrics as the underlying features of the classifier. The features are selected as two first-order moments of the angular correlation and two first-order moments of the Czenakowski measure.

If the light source can be estimated for different objects/people in an image, inconsistencies in the lighting direction can be used as evidence of digital tampering. Lighting inconsistencies are applied for revealing traces of digital tampering in [55]. The authors proposed a technique for estimating the light source direction from a single image. The light direction estimation requires the localization of an occluding boundary. These boundaries are extracted by manually selecting points in the image along an occluding boundary. This rough estimate of the position of the boundary is used to define its spatial extent. The boundary is then partitioned into approximately eight small patches. Three points near the occluding boundary are manually selected for each patch, and fit with a quadratic curve. The surface normalcy along each patch is then estimated analytically from

the resulting quadratic fit. The intensity at the boundary is then determined by evaluating intensity profile function, and repeated for each point along the occluding boundary.

The problems faced in image forensics are extremely difficult. A basic problem is to determine the model of the digital camera that was used to capture the image. An approach based on feature extraction and classification is proposed for the camera source identification problem by identifying a list of candidate features [61]. A vector of numerical features is extracted from the image and then presented to a classifier built from a training set of features obtained from images taken by different cameras. Then a multi-class support vector machine (SVM) was used to classify data from all of the different camera models. The feature vector is constructed from average pixel values, correlation of RGB pairs, center of mass of neighbor distribution, RGB pairs energy ratio, and it also exploits some small scale and large scale dependencies in the image expressed numerically using a wavelet decomposition previously used for image steganalysis [62].

Fridrich et al. proposed to use the sensor's pattern noise for digital camera identification from images [63, 64]. Instead of measuring the noise, they used a wavelet-based denoising filter described in [65] to extract the pattern noise from the images. For each camera under investigation, they first determine its reference pattern, which serves as a unique identification fingerprint. To identify the camera from a given image, they consider the reference pattern noise as a high-frequency spread spectrum watermark, whose presence in the image is established using a correlation detector.

2.2.3 Image Quality Measures

Digital images are subject to a wide variety of distortions during acquisition, processing, compression, and transmission, any of which may result in a degradation of the visual quality. Image quality measures are figures of merit used for the evaluation of imaging

systems, image coding, and processing techniques. The image acquiring and post-processing operations will introduce some pattern distortions which would result a quality consistency in the final image. Malicious modifications of the image would disrupt this quality consistency. Therefore, if the image qualities of regions are inconsistent with each other, then the image may be a forgery.

Digital Image forensics based on Image Quality Measure

Image quality measure is generally based on some specific distortions. Using quality measure for digital image forensic analysis is actually to measure the distortion. There are many sources of distortions introduced in the whole chain of digital imaging and processing, as shown in Figure 2.1. Several stages exist in this chain. In a typical consumer digital camera, the light from the photographed scene passes through the camera lens, and then it is converted to digital signal by the sensors. In the third stage, the signal is then processed by digital imaging processor such as Canon DIGIC chips. Interpolation, color correction, white balance adjustment, and gamma correction are the usual operations by the processor. Finally, the raw data may be compressed, and saved to the camera memory [66]. All these stages will introduce some distortions or correlations into the final image, such as optical distortion by lens, noises introduced by the sensor, and artifacts by compression.

In each stage, the distortions introduce a kind of quality consistency, which can be served as an "authentication code". On the other hand, the image forgery creation operations usually involve decompression, transformation, composition of the image fragments, and retouching of the final image. These manipulations may disturb the intrinsic quality consistency of the image.

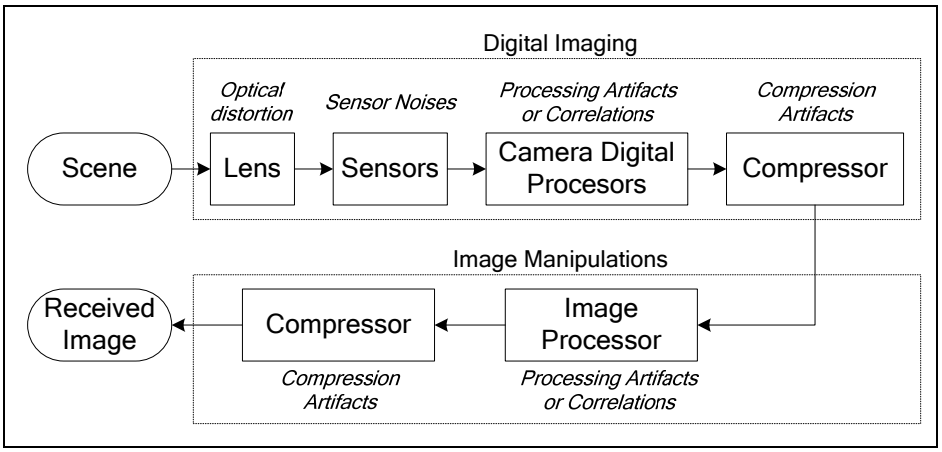


Figure 2.1: Distortions of digital imaging and manipulations

Our idea is to check the quality consistency of different regions of the whole image. For each region, a quality measure is calculated. Then the variance of these measures can be served as the authenticity of the image. The assumption is that if the image is authentic, then different regions of it should be quality consistent. Therefore, if the image qualities of regions are abnormal or inconsistent with each other, then the image may be a forgery.

Image Quality Measures

Many quality measures have been proposed in different research areas such as image coding, image processing, camera design and visual psychology. In practice, subjective evaluation is usually too inconvenient, time-consuming and expensive, therefore objective quality measure becomes important in many applications. A good objective quality measure should reflect the distortion on the image well due to, for example, blurring, noise, compression, and sensor inadequacy. An objective image quality measure could be instrumental in predicting the performance of vision-based algorithms such as feature extraction, image-based measurements, detection, tracking, and segmentation, etc., tasks. It can be used to dynamically monitor and adjust image quality, optimize algorithms and parameter settings of image processing systems, and benchmark image processing systems and algorithms.

Image quality depends on many factors, such as the initial capture system and its image processing, compression, and transmission. There are two key aspects of image quality.

- Factors intrinsic to the imaging system, such as cameras, lenses, or printers: Sharpness, noise, dynamic range, color accuracy, color gamut, etc.
- Factors affected by post-processing, such as contrast adjustment, compression and transmission: Contrast, color balance, color saturation, lossy transmission, etc.

In the image coding and computer vision literature, the raw error measures based on deviations between the original and the coded images are overwhelmingly used [67], with MSE or PSNR varieties being the most common measures. The reason for their widespread choice is their mathematical tractability and that it is often easy to design systems that minimize the MSE. Raw error measures such as MSE may quantify the error in mathematical terms, and they are at their best with additive noise contamination. However, they do not necessarily correspond to all aspects of the observer's visual perception of the errors [7, 8], nor do they correctly reflect structural coding artifacts.

In order to quantify the similarity between the test and the reference images in a perceptually meaningful manner, researchers have explored measuring error strength after processing the test and the reference images with HVS models [68, 69]. The underlying premise is that the sensitivities of the HVS are different for different aspects of the visual signal that it perceives, such as brightness, contrast, frequency content, and the interaction between different signal components, and it makes sense to compute the strength of the error between the test and the reference signals once the different sensitivities of the HVS have been accurately accounted for. Methods of this type are useful at determining whether the distortions are below or beyond the threshold of visual detection.

Different from traditional error-sensitivity based approach, structural similarity based image quality assessment has been recently proposed [70]. This approach is based on the

following philosophy: the main function of the human visual system is to extract structural information from the viewing field, and the human visual system is highly adapted for this purpose. Models of this group, i.e. the structural similarity index [70], are based on a measurement of structural information loss.

This error sensitivity paradigm is a bottom-up approach in which researchers model the low-level features of the HVS to achieve consistent quality predictions. Although such methods have met with good success, there are many questions that arise in their design [70]. Some researchers have therefore explored arbitrary signal fidelity criteria that are not affected by assumptions about HVS models, but are motivated instead by the need to capture the loss of visual structure in the signal that the HVS hypothetically extracts for cognitive understanding. Such top-down methods have also met with good success [70].

For image applications with very low bit rate coding, quality measures based on human perception are being more frequently used [9, 10, 11, 12, 13, 14]. Since a human observer is the end user in image applications, an image quality measure that is based on a human vision model seems to be more appropriate for predicting user acceptance and for system optimization. This class of distortion measures gives in general a numerical value that will quantify the dissatisfaction of the viewer in observing the reproduced image in place of the original (though Daly's VPD map [13] is a counter example to this). The alternative is the subjective tests where the subjects view a series of reproduced images and rate them based on the visibility of artifacts [15, 16]. Subjective tests are tedious and time consuming and the results would depend on various factors such as observer's background, motivation, etc., and furthermore actually only the displayed quality is being assessed. Therefore an objective measure that accurately predicts the subjective rating would be a useful guide when optimizing image compression algorithms.

Quality measures can be classified into the following categories:

- Full-reference (FR) measures perform a direct comparison between the image or video under test and a reference or "original", such as MSE and PSNR belong to this class as well.
- No-reference (NR) measures only look at the image or video under test and have no need of reference information. Our proposed blocking artifacts and sharpness measures belong to this class. It is also called blind quality measure.
- Reduced-reference (RR) measures lie between these two extremes. They extract a number of features from the reference image. The quality measure is then based only on those features.

Only NR measures could be exploited for image forensics, where the reference image is unavailable. NR measures work with the assumption that all images and videos are perfect unless distorted during acquisition, processing or reproduction. Hence, the task of blind quality measurement simplifies into blindly measuring the distortion that has possibly been introduced during the stages of acquisition, processing or reproduction. The reference for measuring this distortion would be the statistics of natural images and videos, measured with respect to a model that best suits a given distortion type or application. For example, natural images do not contain blocking artifacts, and any presence of periodic edge discontinuity at the boundaries of blocks, is probably a distortion introduced by block-DCT based compression techniques.

The following quality measures would be the most important no-reference measures to be exploited in image forensics.

- Blocking artifacts: Blocking artifacts refers to a block pattern (discontinuities at the boundaries of adjacent blocks) in the compressed image or video. It is due to the independent quantization of individual blocks during block-DCT based JPEG compression. Due to the regularity and extent of the resulting pattern, the blocking effect is easily noticeable.

- Sharpness/blurriness: One of the most important quality factors is sharpness, which determines the amount of detail an image can convey. Blurriness, which is measured by the same way as sharpness, manifests itself as a loss of spatial detail and a reduction of edge sharpness due to the attenuation of the high spatial frequencies during filtering or visual data compression.
- Ringing artifacts: Ringing in an image is caused by the quantization or truncation of the high frequency transform coefficients resulting from DCT- or wavelet-based coding. In the spatial domain this causes ripples or oscillations around sharp edges or contours in the image. This is also known as the Gibbs phenomenon.
- Noise: Noise is a random variation of image density, visible as grain in film and pixel level variations in digital images. It is a key image quality factor. Noise can get ugly in compact digital cameras with small pixels, especially at high ISO speeds. There are many sources of noise in images obtained using CCD arrays, such as dark current, shot noise, circuit noise, fixed pattern noise, etc. In most cases noise is perceived as the degradation of quality. The factors that affect noise of a digital image are: pixel size of the sensor, sensor characters, ISO, exposure time, digital processing in camera (noise reduction and sharpening, etc.).
- Color bleeding: It is the smearing of the color between areas of strongly differing chrominance. It results from the suppression of high-frequency coefficients of the chroma components by compression due to chroma subsampling, or by CCD color filter array interpolation. Color bleeding is also considered as a loss of colorfulness.

2.3 Summary

Although many studies have been done on robust image authentication and error resilient data authentication, no literature of others is available on error resilient image authentication. Therefore, there is an emergent need of authenticating images degraded by lossy compression or transmission. The first purpose of this thesis is to authenticate images received through lossy transmission when there are some uncorrectable transmission errors. On the other hand, image forensics is an emerging research topic. Several passive authentication approaches have been proposed, which are effective in some aspects but are by no means always reliable or form a complete solution. Therefore, the second aim of this thesis is to authenticate images passively by evaluating image quality inconsistencies through detecting the traces of forgery creation operations.

Chapter 3

Error Resilient Image Authentication for JPEG Images

With the pervasive use of digital images over the Internet and wireless channel, there is an emergent need of authenticating degraded images despite lossy transmission. When transmission errors exist, the digital signature or watermark used in authentication schemes would be damaged or even made unusable. This situation motivates us to design an image authentication scheme that allows two parties to exchange images while guaranteeing content integrity and source identity, even if there are errors during transmission. The problem this chapter focuses on is how to assess the authenticity of an image when there are uncorrectable transmission errors during transmission over lossy channels.

This chapter presents a content-based error resilient image authentication combining watermark embedding and feature hashing: embedding a good amount of side information for robust feature extraction in presence of unrecoverable errors during image transmission; robust image features are converted to cryptographically secure hash of small size for integrity verification. The proposed scheme integrates feature extraction, quantization-based watermarking, error correction coding, error concealment, cryptographic hashing, and digital signature into a unified framework. We first discuss the role of error concealment in this scheme. The proposed error resilient authentication scheme is then present with details. Experimental results are present at the last of this chapter to support the proposed scheme.

3.1 Introduction

Watermark-based authentication approaches usually work for protecting the integrity of the image but not for preventing sender's repudiation [39]. On the contrary, signature-based approaches can work on both the integrity protection of the image and the repudiation prevention of the sender, but the signature could be easily removed and no protection remains then. Furthermore, previous robust digital signature is unavoidably very large because its size is usually proportional to the image size [71, 29].

In order to solve these problems, this chapter presents a hybrid digital signature or watermarking scheme. It generates short and robust digital signatures based on the invariant message authentication codes (MACs). These MACs are obtained from the quantized original DCT coefficients, Error Correction Coding (ECC) coded, and embedded into the image using quantization based watermarking. Similar approaches based on MACs for robust digital signature generation were proposed in [71, 72]. The invariance of MACs is theoretically guaranteed if images are under lossy compression or other acceptable minor manipulations such as smoothing, and brightness adjustment. However, the MACs in [71] are only weakly invariant, which has some exceptional ambiguous cases when two coefficients are the same after manipulations. Because of these ambiguous cases, the whole MACs generated from the signing end have to be preserved in the receiving end. Furthermore, the size of these MACs is proportional to the image size. In this chapter, we propose a method to generate the MACs that are strictly invariant in the signing end and the receiving end. Thus, hashing function can be applied to significantly reduce the size of digital signature. We use watermarks to store ECC check information and localize attacks. The Public Key Infrastructure (PKI) [21] is incorporated to address the authentication problems over various networks.

3.2 Feature-based Adaptive Error Concealment for JPEG Images

It is efficient and advisable to apply error concealment before image authentication since the feature of the error-concealed image would be much closer to the original one than that of the damaged image [77]. As a result, the content authenticity of the error concealed image is higher than that of the damaged image, which was validated in our experiments in Section 3.4.

Error resilient techniques have been developed for image and video transmission over lossy networks, which can be classified into three categories: source coding, such as error control coding (ECC) and data embedding for error concealment [73]; joint source-channel coding, which aims at lossless recovery, such as Forward Error Correction (FEC), and automatic re-transmission request (ARQ) [74]; and error concealment which strive to obtain a close approximation of the original or attempt to make the output least objectionable to human eyes, such as error concealment using residual coefficient correlations [75, 76]. Error concealment is an important technique, since there are always uncorrectable errors in the final received multimedia signals even after applying the other two kinds of error resilient techniques. Error concealment techniques are usually applied by either using contextual relationship of adjacent blocks [77, 78], or through embedded watermarking information [79, 80].

Various image (spatial) error concealment algorithms have been proposed, which makes use of the smoothness assumption through a minimization approach. For JPEG images, transmission errors can be concealed by exploring the contextual relationship between the damaged image blocks and their non-damaged neighboring blocks, which is a common solution in image transmission [74]. One approach recovers a lost block by minimizing the sum of squared differences between the boundary pixels of the lost block

and its surrounding blocks [78]. This smoothness measure often leads to blurred edges in the recovered image. The other approaches proposed to minimize variations along edge directions or local geometric structures [81, 82]. They require accurate detection of image structures, and mistakes can yield annoying artifacts. A classification was proposed in [83] to takes advantage of various concealment algorithms by adaptively selecting the suitable algorithm for each damaged image area.

This section presents a content-based adaptive error concealment algorithm to improve image quality after lossy transmission. The procedure of our proposed algorithm is illustrated in Figure 3. Firstly, we use some features extracted from the neighboring blocks of the damaged block to classify this damaged block into three types: smooth block, texture block and edge block. Five eigenvalues obtained from statistical measures of its neighboring blocks are selected as features and a Minimum Distance Weighted Linear Classification (MDWLC) algorithm is adopted for block classification. Different error concealment methods are then applied to each type of blocks: Linear Interpolation method is used for smooth blocks, DCT Coefficient Prediction [84] for textural blocks, and Directional Interpolation [85] for edge blocks.

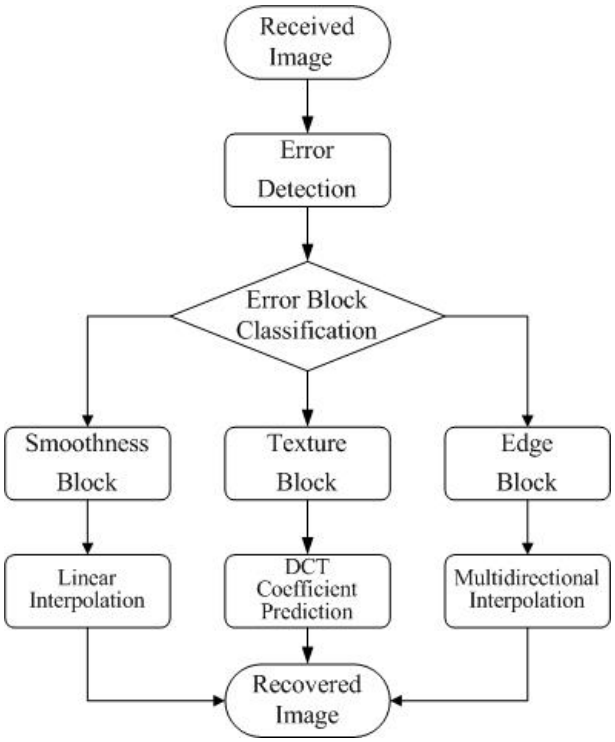


Figure 3.1: Adaptive error concealment

3.2.1 Error Block Classification

Roughly the areas of natural images could be characterized into three types:

- Smooth Area: where the pixel values usually vary slowly and within a small range. Both mean and variance of the gradients are small.
- Texture Area: where the pixel values usually vary in a periodical way. Both mean and variance of the gradients are quite large.
- Edge Area: where the pixel values usually vary significantly and within a large range. The mean value of the gradients is between that of type 1 and 2 while its variance value is the biggest among these three types.

Five measures are calculated for block classification. They are: Pixel Variance (*PV*), Range of Pixel Variance (*RPV*), the Gradient Mean (*GM*), the Gradients Variance (*GV*), and the Number of Pixels whose gradient values are within some range (*PN*). After error

detection, the correctly received neighboring blocks (**N**) of the damaged block (**M**) are used to extract the PV and RPV of M, by:

$$\begin{cases} Mean(\mathbf{M}) = \dfrac{1}{Num(\mathbf{N})} \sum\limits_{(x,y)\in\mathbf{N}} f(x,y) \\ DV(\mathbf{M}) = \dfrac{1}{Num(\mathbf{N})} \sum\limits_{(x,y)\in\mathbf{N}} [f(x,y) - Mean(\mathbf{M})]^2 \\ RPV(\mathbf{M}) = Max(f(x,y)) - Min(f(x,y)), \quad (x,y)\in\mathbf{N} \end{cases} \tag{3.1}$$

where *Mean*(**M**) is the pixel mean of **M** estimated from **N**. *Num*(**M**) is the element number of **N**.

We use the Sobel Operator to calculate the gradient (*G*) of the **N**, and then calculate the *GM*, *GD* and *PN* by:

$$\begin{cases} GM(\mathbf{M}) = \dfrac{1}{Num(\mathbf{N})} \sum\limits_{(x,y)\in\mathbf{N}} G(x,y) \\ GD(\mathbf{M}) = \dfrac{1}{Num(\mathbf{N})} \sum\limits_{(x,y)\in\mathbf{N}} [G(x,y) - GM(\mathbf{M})]^2 \\ PN(\mathbf{M}) = Num\{(x,y) \mid T_1 < G(x,y) < T_2\} \end{cases} \tag{3.2}$$

A feature vector (**F**) is composed of these five features and the MDWLC is used to classify the damaged block, which is shown below.

$$d_k(\mathbf{M}) = \sum_{i=1}^{5} \lambda_i (f_i - u_{ki})^2 \qquad (k = 1,2,3) \tag{3.3}$$

In the above equation, f_i is the *i*-th value of the feature vector **F**(**M**), u_{ki} is the *i*-th value of the cluster center ($\mathbf{U}_k$), and λ_i is the weighted value corresponsive. The cluster center value ($\mathbf{U}_k$) and weight value (λ_i) are both from training with a set of standard test image, and d_k(**M**) is the distance between **M** and $\mathbf{U}_k$.

3.2.2 Error Concealment Methods for Different Block Types

In view of the properties of smooth block, texture block or edge block, different error concealment approaches are selected for different types of blocks: linear interpolation method for smooth blocks, DCT Coefficient Prediction for textural blocks, and directional interpolation for edge blocks.

Linear interpolation is used to conceal every erroneous smooth blocks. It recovers the damaged block through interpolation from pixels in adjacent correctly received blocks, as shown in Figure 3.2. M is the damaged block, and N is its neighboring block set. f(x,y) is the pixel waiting to be recovered. A, B, C and D are the nearest pixel in M's neighboring blocks respectively. d_x, 8- d_x, d_y and 8-d_y are the distance respectively to point f(x,y). Then we get:

$$f(x,y) = \frac{d_x f_A + (8-d_x) f_B + d_y f_C + (8-d_y) f_D}{16} \tag{3.4}$$

This method uses the smoothness property of the image, so it can achieve good result for concealing the smooth blocks.

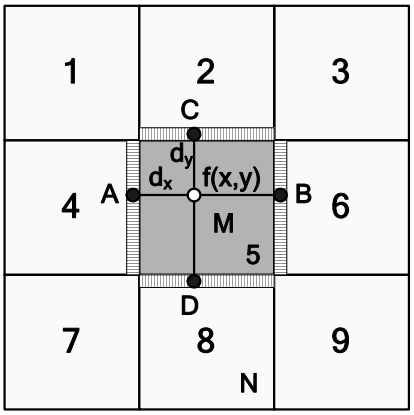


Figure 3.2: Spatial linear interpolation

DCT Coefficient Prediction estimates the DCT coefficients of the lost blocks using the adjacent error-free blocks. The DC coefficients and the first 5 low frequency AC coefficients (*DC* and AC_1-AC_5 from the zigzag like order of the 64 DCT frequencies) of the lost 8×8 blocks can be estimated using the adjacent error-free blocks' DC coefficients [84]:

$$
\begin{cases}
DC_5 = \dfrac{1}{\sum_{j=1,\, j\neq 5}^{9} W_j} \sum_{j=1,\, j\neq 5}^{9} W_j \, DC_j, \text{ wherer } W_j = \begin{cases} 1, & \text{when } j \text{ is even} \\ 1/\sqrt{2}, & \text{when } j \text{ is odd} \end{cases} \\
AC_1 = 1.13885\,(DC_4 + DC_6)/8 \\
AC_2 = 1.13885\,(DC_2 + DC_8)/8 \\
AC_3 = 0.27881\,(DC_2 + DC_8 - 2DC_5)/8 \\
AC_4 = 0.16213\,(DC_1 + DC_9 - DC_3 - DC_7)/8 \\
AC_5 = 0.27881\,(DC_4 + DC_6 - 2DC_5)/8
\end{cases} \tag{3.5}
$$

This method can reproduce the areas of high details with high accuracy, so it is appropriate to recover texture blocks.

Directional interpolation [85] utilizes spatially corrected edge information from a large scale neighborhood of the damaged block and performs multi-directional interpolation to restore the damaged block. The edge direction of the damaged block is determined by convoluting the neighborhood pixels with a set of directional masks and then finding the top three maximum directions. For each selected direction, one-dimensional interpolation is carried out along this direction to obtain one version of restored the damaged block. A block mixing scheme is performed to extract the strong characteristic features of two or more image, and merge them into one image. This procedure [85] can be explained as Figure 3.3. This method can greatly restore the edges of the damaged block, so it is appropriate to recover the edge block.

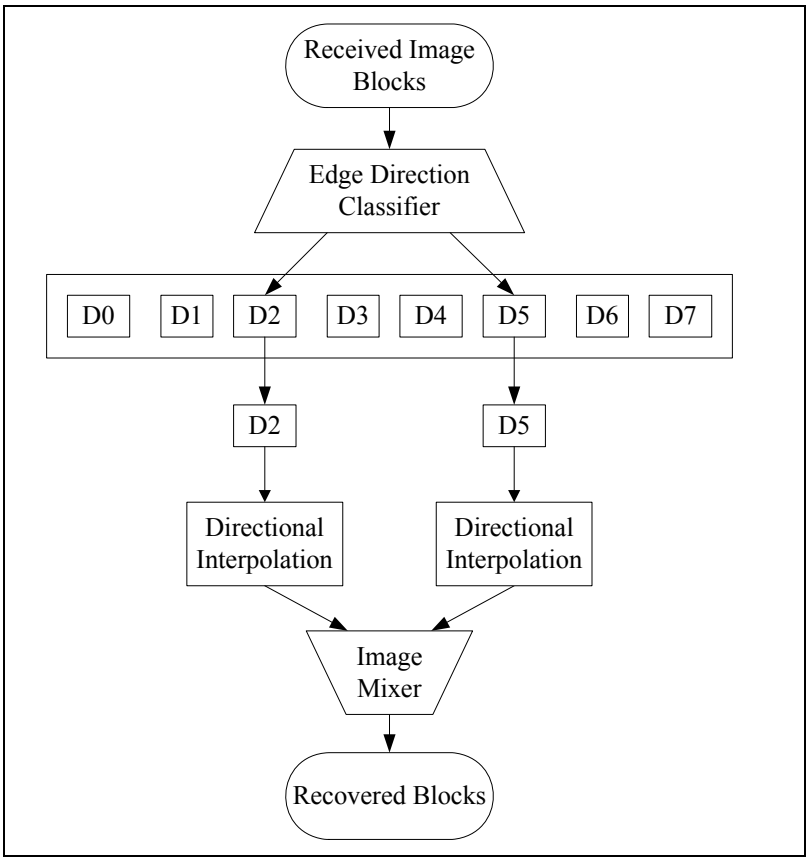


Figure 3.3: Directional interpolation

The aim of image mixing is to extract the strong characteristic features sensitive to human eyes, such as the contrast of the mixed image. According to the histogram of each image, the pixels can be classified into three types: background, bright foreground and dark foreground. Pixels with values within the range of one variance distance to the mean may be considered as the background ones. Any pixels out of this range are considered as foreground, in which the ones greater than the mean are considered as bright foreground and the left are dark foreground.

3.3 Error Resilient Image Authentication Scheme for JPEG Images

3.3.1 Feature Generation and Watermark Embedding

Our proposed content-based error resilient image authentication uses combined watermark embedding and feature hashing. An amount of side information is embedded for robust feature extraction in presence of unrecoverable errors during image transmission, and image features are converted to cryptographically secure hash of small size for integrity verification. It is possible to design a robust feature without using side information for an embedding-only approach. Fox example, the adopted feature is first ECC coded to improve its robustness, and then embedded into the image. However, the proposed approach using both hashing and embedding would not only improve the feature robustness, but also provide a multiple-layer security: the hacker will have to attack both the hashing algorithm and watermarking algorithm. Furthermore, the embedded watermarks can also be used to detect the rough location of changes may would be made to the image.

The proposed authentication scheme uses DCT coefficients to generate content-based message authentication codes (MACs), and stores some auxiliary ECC information of MACs in the image using watermarking. A robust digital signature of image is generated as follows. The original image is partitioned into 8×8 blocks. Those blocks are further labelled as either **T** block or **E** block. We choose **T** blocks for extracting content-based features (MACs) and **E** blocks for watermarking.

All **E** blocks are shuffled by a random number seed *RGN*. One example of partitioning blocks into E and T is shown in Figure 3.4. The final bit-stream is assembled in this shuffled block order before transmission. The reasons doing so are as follows. Firstly we want to ensure that most damaged blocks caused by packet loss are isolated. Such techniques have

already been adopted in [74, 86] to achieve a better result of error concealment. Secondly such shuffling makes the watermarks from those smooth blocks remain embeddable.

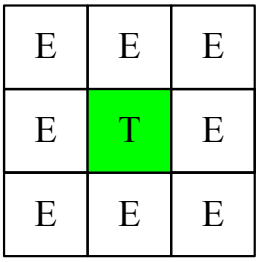


Figure 3.4: Example of partitioning image blocks into **T** and **E**

For each **T** block, we pick up its DC and 3 AC to generate MACs. These 4 coefficients are quantized by the preset authentication strength matrix Q_a. The quantization process is shown as follows. Assume the original value is D, the quantization step size specified in the quantization table is Q, and the output of the quantizer is quotient F (integer rounding) and remainder R, respectively: $D/Q = F$, and $D\%Q = R = D - F*Q$. Suppose the incidental distortion introduced by acceptable manipulations on the original coefficient D can be modeled as noise whose maximum absolute magnitude is denoted as N.

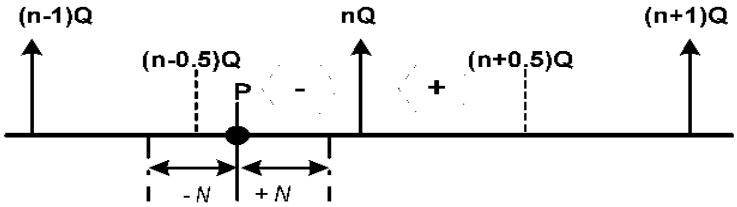


Figure 3.5: Illustration on the concept of error correction

Refer to Figure 3.5, assuming a pre-determined $Q>4N$ is known at both the signing end and the receiving end. If the original value D is located at the point nQ, then no matter how this value is corrupted, if the distortion is in the acceptable bounds, the distorted value will still be in the range $((n-0.5)Q, (n+0.5)Q)$, and the quantized value with step Q will remain unchanged as nQ before and after noise addition [87]. However, if the original value D drops into the range of $((n-0.5)Q, nQ)$ (Point P in Figure 3.5), its quantized value (with step Q) is still nQ before adding noise, but there is also a possibility that the noisy value could

drop at the range((*n*-1)*Q*, (*n*-0.5)*Q*) and will be quantized as (*n*-1)*Q*, not *nQ,* after adding noise. Thus the noise causes a different quantization result.

To avoid such a case, we propose an ECC-like procedure to record the sign of *R*. ECC codes are stored as watermarks (in other blocks) and can be retrieved by the authenticator. We record an ECC bit '0' if the original value *D* drops between ((*n*-0.5)*Q*, *nQ*) (i.e., $R < 0$). In the authentication procedure, if this value *D* was corrupted, the following steps will be adopted. If we retrieve a 0 bit (i.e. R<0), we add the value 0.25*Q* from the corrupted value. Then, using the quantization step *Q*, we can obtain the same quantized value as *nQ*, which is the same as the original quantized value. Similar process is applied to the case when the original value *D* is in (*nQ*, (*n+0.5)Q*). Based on such an error correction procedure, all quantized values can be used to form MACs that will stay unaltered before and after distortion. These MACs can then be hashed and encrypted to form crypto signature, which is short, fix-length and robust to signal distortion with acceptable manipulation bounds. Here the original value *D* could be in the DCT domain, wavelet domain or pixel domain, as long as the acceptable manipulation constraint is predictable. As discussed in [8], several HVS models can be used to determine the setting of such constraints.

For each **T** block, the 4 bits of each ECC-like codewords are then watermarked into its corresponding **E** blocks. Assuming Q_a is used for generating features and watermarking while Q_c is for actual JPEG compression. In [88], the authors have proved that as long as Q_c is less than or equal to Q_a, the robustness of generated features as well as embedded watermarks is guaranteed. Based on this principle, we embed the watermark of **T** block by directly modifying some AC coefficients in **E**. A typical ratio of **T** and **E** blocks is 1:8. Among 8 **E** blocks of a **T** block, we only embed the watermark into those 3 blocks with highest AC energy (i.e., the most 3 textual blocks).

3.3.2 Signature Generation and Watermark Embedding

The image signing procedure is shown in Figure 3.6. Given an image, the user generates a crypto signature by performing the following signing process on the image sequentially: (1) perform block-based pre-processing; (2) extract the DCT features and generate the watermarks; (3) shuffle image blocks and select the blocks for watermarking; (4) embed the watermarks and obtain the watermarked image; (5) cryptographically hash the extracted features, generate the crypto signature by the image sender's private key; (6) send the watermarked image and its associated crypto signature to the recipients.

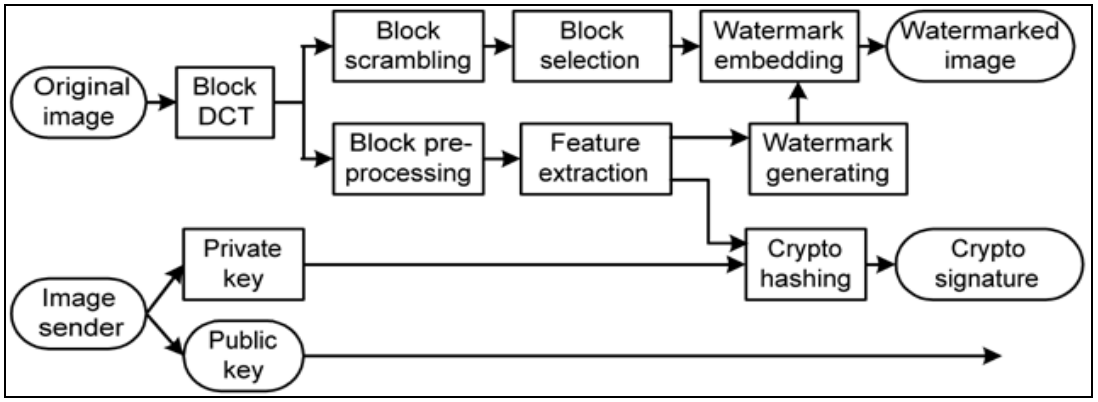


Figure 3.6: Diagram of image signing

During the image signing procedure, it is impossible to know which blocks will be damaged in advance (i.e., which packets will be lost during the transmission is unknown). However, only two cases exist: either **T** is damaged or **E** is damaged. If it is an **E** block, it will affect the correctness of watermark extraction. If it is a **T** block, it will affect the stability of MAC extraction because **T** has to be reconstructed at the receiver end by the error concealment methods. Usually such reconstruction is just a roughly approximated version of original **T** and eventually affects either system robustness or system security because a large Q has to be set for feature extraction in order to tolerate a large N. Therefore some preprocessing is required. Assuming **T** is lost during transmission and is reconstructed as **T'** by our error concealment algorithm [77]. We check the distortion between **T** and **T'**. If it is greater than our preset N, we then recursively modify **T** with decreasing difference

values on randomly selected coefficients until the modified coefficients can generates the same MACs as in **T'**. In the worst situation this recursive method results in worse visible quality than that of **T'**, but the system can choose to make **T** equal to **T'** at the signing end.

A one-way crypto hash function such as SHA-1 is applied to the MACs concatenated from all **T** blocks. In addition to these hash values, other auxiliary information includes the size of image, and the authentication strength matrix ($\boldsymbol{Q}_a$) is combined together and is encrypted using the image sender's private key to obtain the crypto signature.

3.3.3 Image Authenticity Verification

The image authentication procedure is shown in Figure 3.7. Given the degraded image and its associated digital signature, the proposed solution authenticates both the integrity and the source of the received image by performing the following process on the image sequentially: (1) perform content-adaptive error concealment, if some blocks are damaged; (2) extract message authentication codes and watermark respectively; (3) correct the perturbations in the extracted feature set by the extracted watermark based on the ECC concept; (4) cryptographically hash the corrected feature set, obtain a short and fixed-length bit stream *A*; (5) decrypt the signature by using the sender's public key and obtain another bit string *B*; (6) bit-by-bit compare *A* and *B*; Deem the image authentic if they are the same; Otherwise (7) locate the possible attacks by correlating the extracted feature and the watermark.

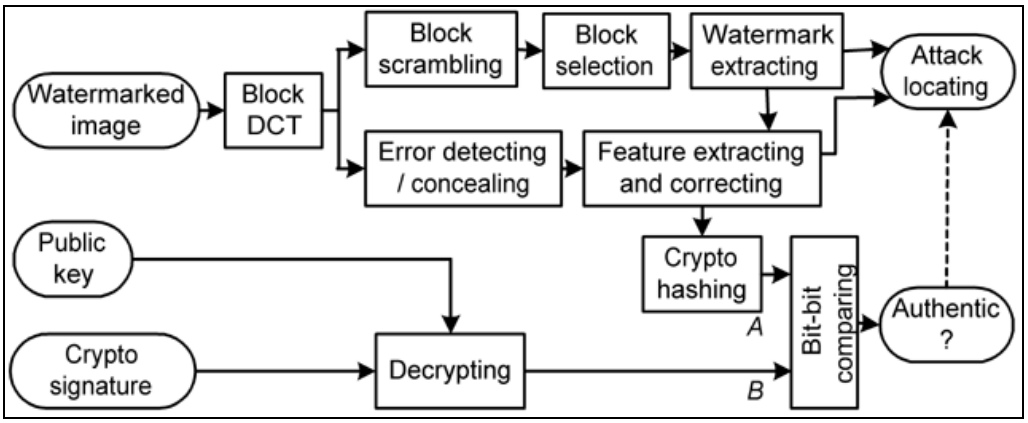


Figure 3.7: Diagram of image authentication

Error concealment technique proposed in [77] is used, with an additional block shuffling method in order to evenly distribute the corrupted blocks. The preprocessing process guarantees the invariance of the reconstructed images message authentication codes.

If the image is verified as unauthentic, the attacked locations may be detected by correlating between the extracted watermarks and the remainders of DCT features quantized by $\boldsymbol{Q}_a$. This advantage could help in further convincing the authentication results. Note that some false alarms may exist because of other incidental distortions. This may be acceptable because the major system performances are system robustness and system security. Such false alarms can be further reduced by removing isolated detected blocks.

3.4 Experimental Results and Discussions

The proposed error concealment algorithm has been evaluated on a number of standard test images. We implemented our error detection and adaptive error concealment to the damaged images received from the simulated wireless fading channel. The Bit Error Rate (BER) is 3×10^{-4}. The PSNRs of test results are shown in Figure 3.8. We can see that for different images, the error concealment achieved different improvements, depending on the image content. For example, image *Barbara* contains much richer texture and details than *Lena*, so it can achieve better improvement than *Lena*.

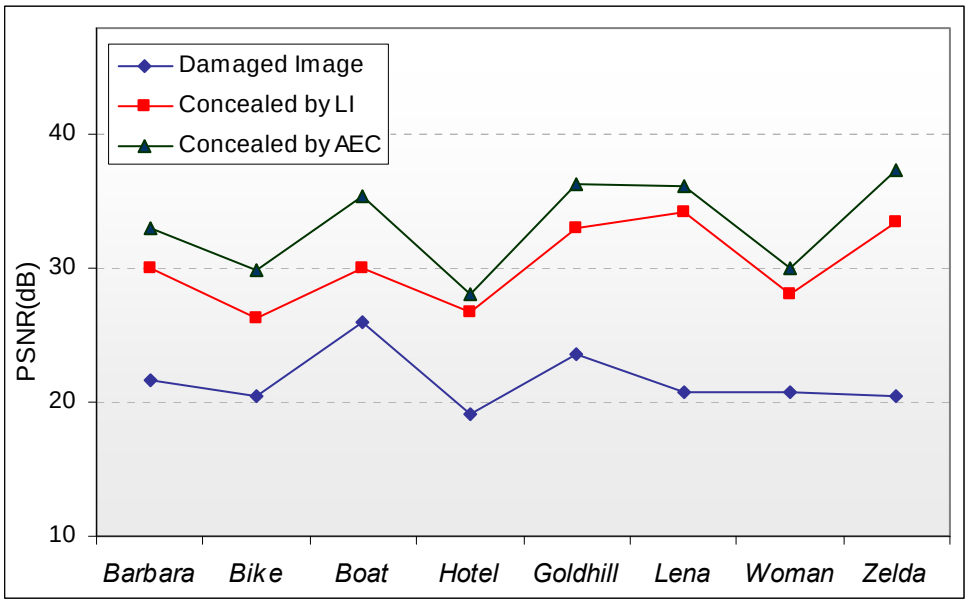


Figure 3.8: PSNR (dB) results of images restored by proposed algorithm (AEC) and linear interpolation (LI)

Figure 3.9 shows the error concealment results of JPEG test image *Barbara*. We can observe that the proposed algorithm can achieve better results than the linear interpolation on the texture areas and edge areas, such as the back of chair, trousers and tablecloth. The PSNR gain of this proposed algorithm is better than the conventional algorithms mentioned in [78, 85]. The algorithms in paper [78, 85] can achieve about 1 dB better than linear interpolation. The proposed algorithm can achieve about 3 dB better than the linear interpolation.

(a) error-free Image (b) damaged Image (BER=3×10^{-4})

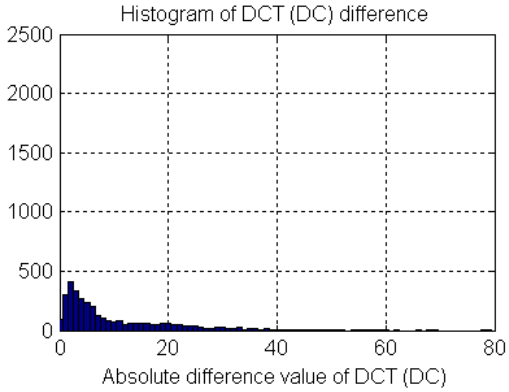
(c) restored Image by LI (d) restored Image by AEC

Figure 3.9: Error concealment results of the image *Barbara*

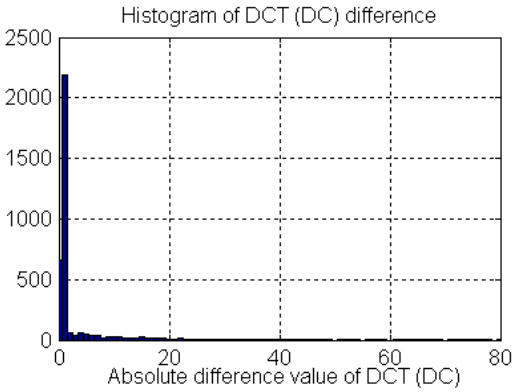
Figure 3.10 shows the merits of using block shuffling before image transmission on the stability of extracted features (MACs), by comparing the DCT value difference between the original and the concealed. Figure 3.10(c) is the histogram without block shuffling (the corrupted image is shown in Figure 3.10(a) and Figure 3.10(d) is with block shuffling (the corrupted image is shown in Figure 3.11(c). The number of DCT coefficients having small difference in Figure 3.10(d) is much smaller than that in Figure 3.10(c). Such improvement allows us to choose smaller $\boldsymbol{Q}_a$ given the same $\boldsymbol{Q}_{c,}$ which consequently improves system security with fewer false negative on missing manipulations. Furthermore, block shuffling also make the burst packet loss distributed evenly in the image (Figure 3.11), which improve the error concealment performance and authentication feature robustness.

(a) corrupted image without block shuffling

(b) image with block shuffling



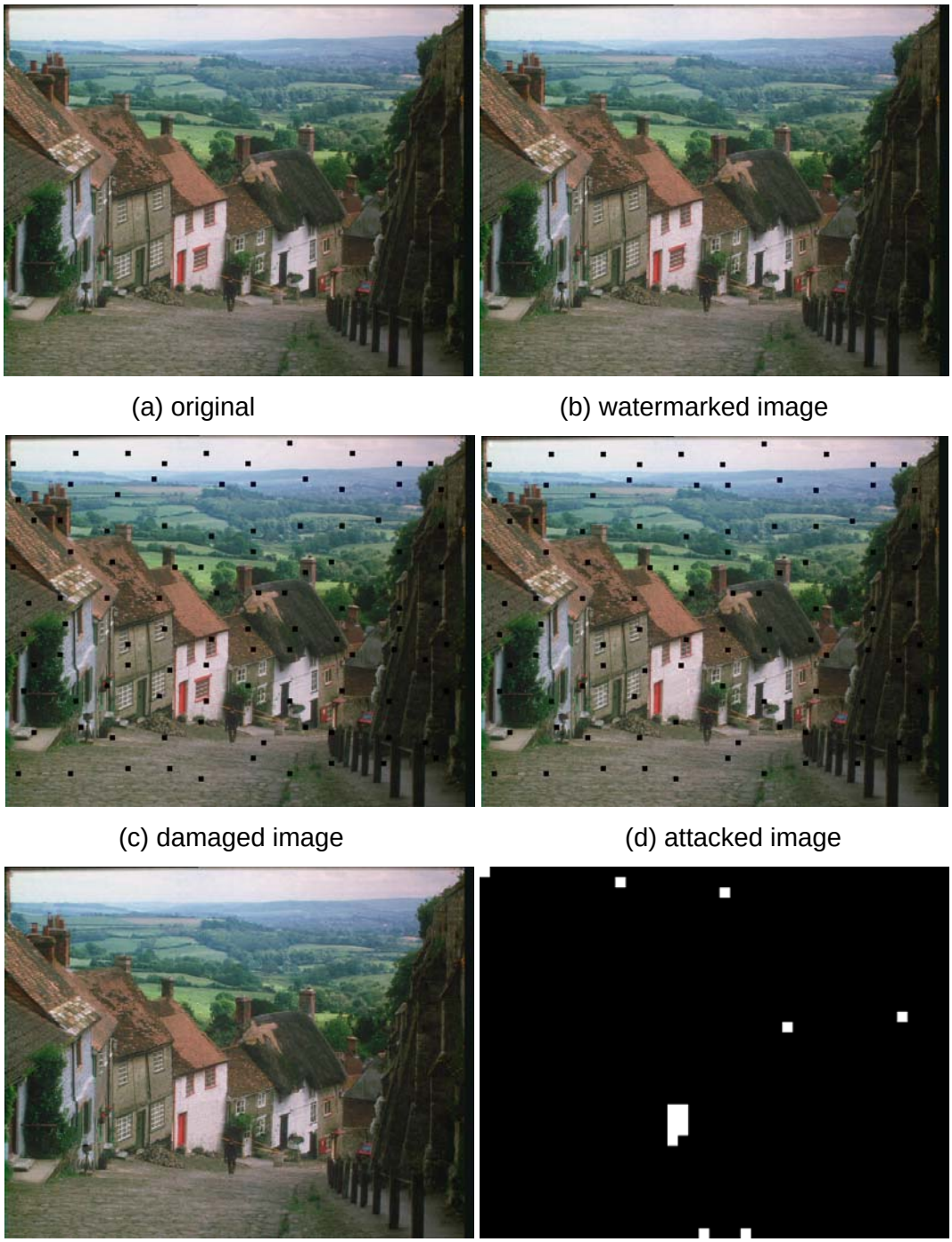
(c) MAC differences without shuffling



(d) MAC differences with shuffling

Figure 3.10: MAC differences between reconstruction without and with shuffling

Figure 3.11(a) shows the original image. Figure 3.11(b) is the watermarked image compressed with JPEG quality factor 8 in Adobe Photoshop, the robustness of authentication and watermarking is set to JPEG quality factor 7. Figure 3.11(c) is the damaged image due to packet loss (The BER is $3x10^{-4}$). We have tested that the corrupted image below the BER of $3x10^{-3}$ can still pass the authentication after error concealment. Note that some damaged blocks may not be detected and therefore can escape from being concealed. However, such misses did not affect the authentication. Fox example, Figure 3.11(d) is the attacked image on the damaged image (window removed). Figure 3.11(e) is the recovered image and Figure 3.11(f) shows the detected attacked location.



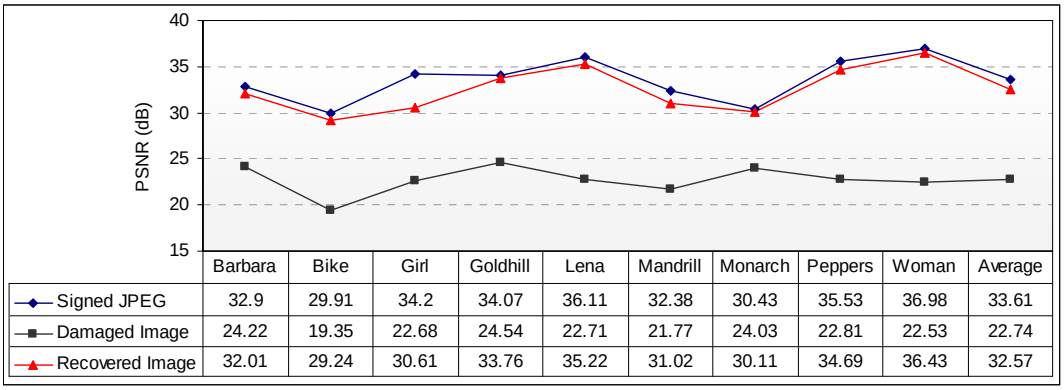
(a) original (b) watermarked image

(c) damaged image (d) attacked image

(e) recovered image (f) detected attack areas

Figure 3.11: Image authentication results

The image quality is also measured in terms of objective quality measure PSNR, as shown in Figure 3.12. We see that the quality of the damaged images recovered by our error concealment method is very close to the original watermarked image.



	Barbara	Bike	Girl	Goldhill	Lena	Mandrill	Monarch	Peppers	Woman	Average
Signed JPEG	32.9	29.91	34.2	34.07	36.11	32.38	30.43	35.53	36.98	33.61
Damaged Image	24.22	19.35	22.68	24.54	22.71	21.77	24.03	22.81	22.53	22.74
Recovered Image	32.01	29.24	30.61	33.76	35.22	31.02	30.11	34.69	36.43	32.57

Figure 3.12: Image quality evaluation in terms of PSNR

3.5 Summary

An error resilient image authentication scheme is present in this chapter, which is robust to transmission errors in JPEG images. Pre-processing and block shuffling techniques are adopted to stabilize the features for signature generation. The experimental results indicate that the proposed scheme successfully creates a signature redundancy in the selected block and its neighbourhoods, which is exploited for verification when uncorrectable errors exist in the received image. Furthermore, the proposed error resilient scheme can improve the trustworthiness of digital images damaged by transmission errors by providing a way to distinguish them from digital forgeries. Limitations of the proposed scheme are that it is suitable only for JPEG images and that it may not be robust to some acceptable image manipulations. For example, the authentication failed when auto contrast adjustment was done on the watermarked image Figure 3.11(b). In the next chapter we will present a feature distance measure to improve the performance of error resilient image authentication to make it also suitable for wavelet-based images (JPEG2000 format), and robust to some other acceptable manipulations.

Chapter 4

Feature Distance Measure for Content-based Image Authentication

Content-based image authentication typically assesses authenticity based on a feature distance measure between the test image and the original image. Commonly employed distance measures such as the Minkowski measures (including Hamming and Euclidean distances) may not be adequate for content-based image authentication since they do not exploit statistical and spatial properties of features.

This chapter presents a feature distance measure for content-based image authentication, which is based on statistical and spatial properties of the feature differences. This statistics- and spatiality-based measure (SSM) is motivated by the observation that most malicious manipulations are localized whereas acceptable manipulations result in global distortions. Based on SSM, an error resilient image authentication scheme is then presented, which is an improvement of the scheme present in the previous chapter. The experimental results have confirmed that our proposed measure is better than the previous measures in distinguishing malicious manipulations from acceptable ones, and can improve the performance of content-based image authentication.

4.1 Introduction

Content-based image authentication is a main robust authentication technique, which accepts an image as authentic if its semantic meaning remains unchanged [5, 6]. The main

requirement for content-based image authentication is that minor modifications which do not alter the content preserve the authenticity of the image, whereas modifications which do modify the content render the image not authentic. In order to be robust to acceptable manipulations, several content-based image authentication schemes have been proposed [7, 9, 10]. These schemes may be robust to one or several specific manipulations, however, they would classify the image damaged by transmission errors as unauthentic [89].

General image authentication may evaluate authenticity and integrity of images via a hypothesis test:

- Authentic (H_0): the image is authentic, not maliciously modified;
- Unauthentic (H_1): the image is not authentic with some malicious modifications.

A typical authentication decision is based on the comparison between a preset threshold (T) and the distance of the pattern vector extracted (P_t) from the test image and the original pattern (P_o) from the original image:

$$d(P_t, P_o) \underset{H_0}{\overset{H_1}{\gtrless}} T \tag{4.1}$$

Content-based image authentication typically measures the authenticity in terms of the distance between a feature vector from the received image and its corresponding vector from the original image, and compares the distance with a preset threshold to make a decision [14, 16, 40]. Commonly employed distance measures, such as the Minkowski metrics [90] (including Hamming and Euclidean distances), may not be suitable for robust image authentication. The reason is that even if these measures are the same (e.g., we cannot tell whether the question image is authentic or not), the feature differences under typical acceptable modifications or malicious ones may still be distinguishable (feature differences are differences between the feature extracted from the original image and the feature extracted from the test image). That is to say, these measures do not properly exploit statistical or spatial properties of image features. For example, the Hamming distance

measures of Figure 4.1(b) and Figure 4.1(d) are almost the same, but yet, one could argue that Figure 4.1(b) is probably distorted by malicious tampering since the feature differences concentrate on the eyes.

4.2 Statistics- and Spatiality-based Feature Distance Measure

Content-based image authentication generally verifies authenticity by comparing the distance between the feature vector extracted from the test image and the original with some preset thresholds. Various feature distance functions, such as Minkowski metrics [90] and Figure of Merit (FoM) ([91]), have been used to measure similarity between the feature vectors representing images. Minkowski metric $d(X, Y)$ [90] is defined as:

$$d(X,Y) = (\sum_{i=1}^{N} | x_i - y_i |^r)^{1/r} \tag{4.2}$$

where X, Y are two N dimensional feature vectors, and r is a Minkowski factor. Note that when r is set as 2, it is actually Euclidean distance; when r is 1, Manhattan distance (or Hamming distance for binary vectors).

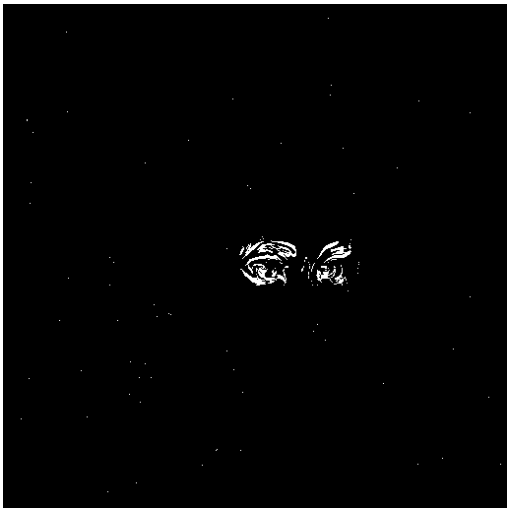
FoM is commonly used at measuring image similarity based on edge feature, which is defined [91] by:

$$\text{FoM} = \frac{1}{\max(N_O,\ N_C)} \sum_{i=1}^{N_C} \frac{1}{1 + \lambda \times d_i^2} \tag{4.3}$$

where N_C and N_O are the number of detected and original edge pixels, respectively. The d_i is the *Euclidean* distance between the detected edge pixel and the nearest original edge pixel, and λ is a constant typically set to 0.1.

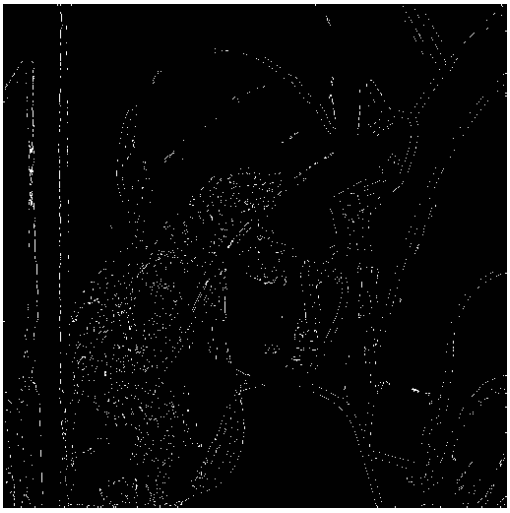
(a) original image

(b) tampered image



(c) feature difference of (b)

(d) blurred image (by Gaussian 3×3 filter)



(e) feature difference of (d)

Figure 4.1: Discernable patterns of edge feature differences caused by acceptable image manipulation and malicious modification

Unfortunately, these measures do not exploit spatial information of feature or the statistics property of the distortion patterns, so they are not adequate for content-based image authentication scheme. Therefore, the image authentication scheme based on Minkowski metric or FoM may not be suitable to distinguish the tampered images (e.g., small local objects removed or modified) from the images by acceptable manipulations such as lossy compression. On the other hand, we found that even if the Minkowski metric distances are the same, the feature difference under typical acceptable manipulations and malicious ones are still distinguishable especially in the case that the feature contains spatial information such as edges or block DCT coefficients. Therefore, the Minkowski metric is not a proper measure for content-based image authentication.

A new feature distance measure based on the distinguishable difference patterns is proposed to differentiate distortions caused by acceptable and malicious image manipulations. Essentially, under this distance measure, spatially clustered differences are less likely to be authentic compared to scattered differences. This measure is quite general and can be incorporated into many existing content-based image authentication schemes.

4.2.1 Main Observations of Image Feature Differences

Many features used in content-based image authentication are composed of localized information about the image such as edge [7, 13], block DCT coefficients based features [8, 14, 15], highly compressed version of the original image [9], and block intensity histogram [16]. To facilitate discussions, we let x_i be the feature value at spatial location i, and X be an N-dimension feature vector, for example, $N=W\times H$ when using edge feature (W and H are the width and height of the image). We define the feature difference vector δ as the difference between feature vector X of the test image and feature vector Y of the original image:

$$\delta_i = | x_i - y_i | \tag{4.4}$$

where δ_i is the difference of features at spatial location i.

After examining many discernable feature difference patterns from various image manipulations, we could draw three observations on feature differences:

(1) The feature differences by most acceptable operations are evenly distributed spatially, whereas the differences by malicious operations are locally concentrated.

(2) The maximum connected component size of the feature differences caused by acceptable manipulations is usually small, whereas the one by malicious operations is large.

(3) Even if the maximum connected component size is fairly small, the image could have also been tampered with if those small components are spatially concentrated.

These observations are supported by our intensive experiments and other literature mentioned previously [7, 89]. Image contents are typically represented by objects and each object is usually represented by spatially clustered image pixels. Therefore, the feature to represent the content of the image would inherit some spatial relations.

A malicious manipulation of an image is usually concentrated on modifying objects in image, changing the image to a new one which carries different visual meaning to the observers. If the contents of an image are modified, the features around the objects may also have been changed, and the affected feature points tend to be connected with each other. Therefore, the feature differences introduced by a meaningful tampering would typically be spatially concentrated.

On the contrary, acceptable image manipulations such as image compression, contrast adjustment, and histogram equalization introduce distortions globally into the image. The feature differences may likely to cluster around all objects in the image, therefore they are not as concentrated locally as those by malicious manipulations. In addition, many objects may spread out spatially in the image, thus the feature differences are likely to be evenly

distributed with little connectedness. The distortion introduced by transmission errors would also be evenly distributed since the transmission errors are randomly introduced into the image [99].

The above observations not only prove the unsuitability of Minkowski metric to be used in image authentication, but also provide some hints on how a good distance function would work: it should exploit both the statistical and spatial properties of feature differences. These observations further lead us to design a new feature distance measure for content-based image authentication.

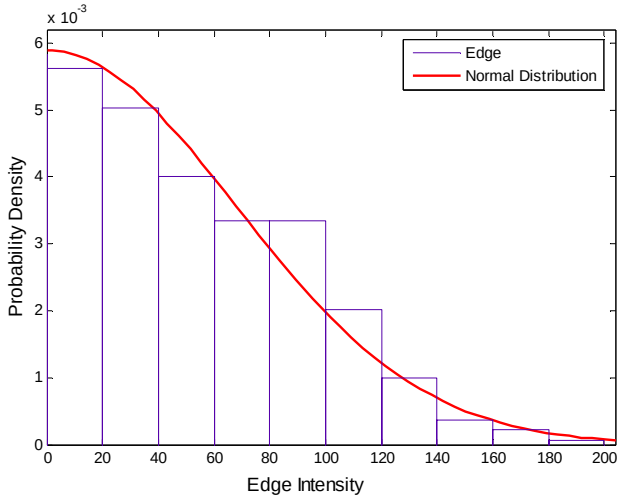
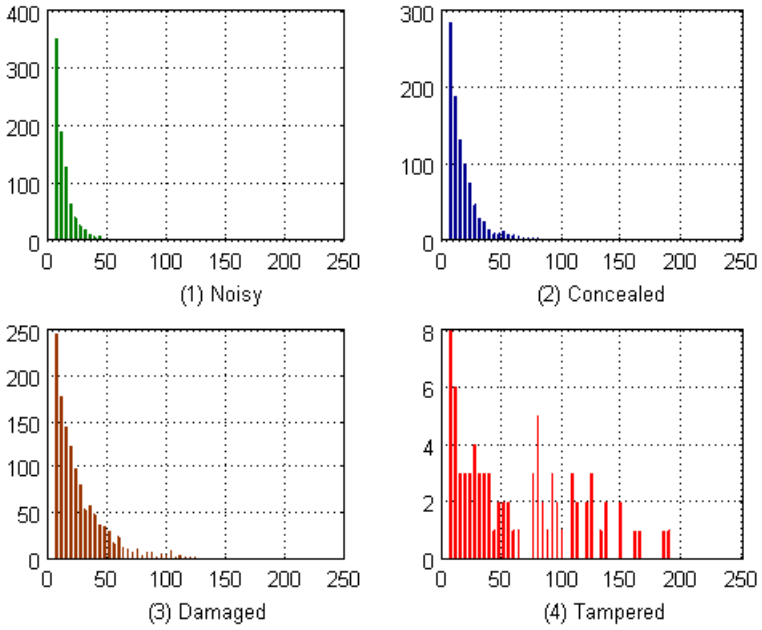
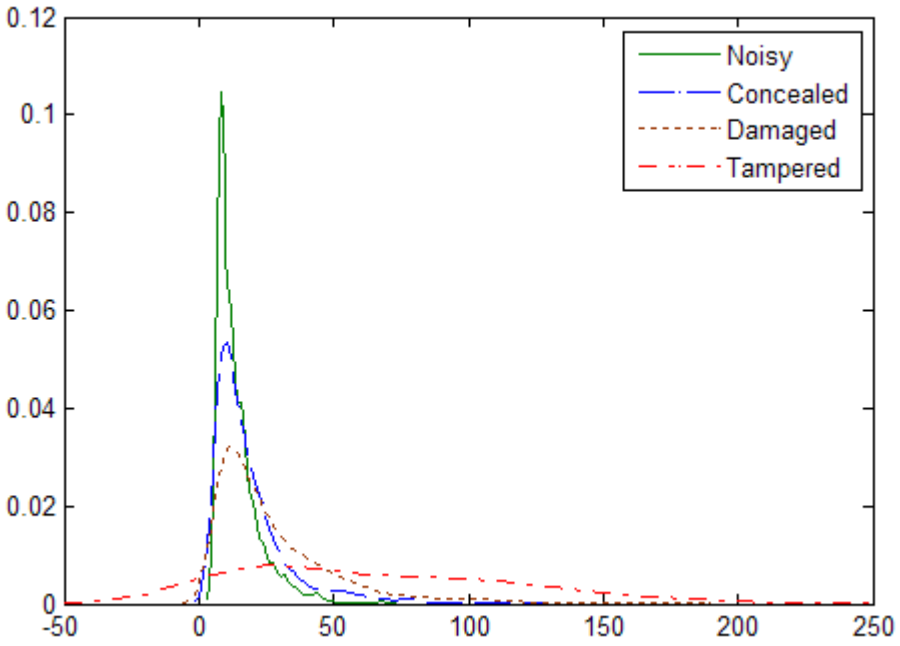


Figure 4.2: Edge distribution probability density estimation

As shown in Figure 4.1, the distortion of the attacked image is concentrated on some objects (eyes and eyebrows in this example), and the distortion from transmission errors are much more randomly and evenly distributed. The reason is that the distribution of edge prorogation in one block is somewhat Gaussian distributed (Figure 4.2). The edge distortion of the acceptable manipulations can be considered as the subtraction of two independent Gaussian variables, thus it is a new Gaussian distribution. On the contrary, the malicious attacked image has strong localized relations with the original image, so the distortion distribution cannot be treated as subtraction of two independent distributions.



(a) Histograms of edge differences



(b) Probability density estimation

Figure 4.3: Edge distortion patterns comparisons

Figure 4.3 shows the histogram of edge difference and their respective probability density estimates of noisy, error concealed, damaged and maliciously tampered images. Binary edge detected by [100] is selected as feature in our evaluations. We can find that the

distribution of feature differences between maliciously tampered image and the original image have a much longer tail than that of the error-concealed image. The damaged, error-concealed and noisy images all have smaller right tails. The maliciously tampered image has a different distortion pattern from those of the acceptable manipulations. This difference can be exploited to distinguish malicious modifications from acceptable operations. These results support our observations that the maliciously tampered image has a different pattern of feature differences from that of the acceptable manipulations.

4.2.2 Feature Distance Measure for Content-based Image Authentication

Natural images exhibit strong dependencies, especially when they are spatially proximate, and these dependencies carry important information about the image content. Therefore, the feature used in image authentication will inherit some spatial dependencies or statistics properties. The motivation of our proposed measure is to find a way to exploit the spatial and statistical information in the feature differences between the test and the original image.

Based on the observations and rules discussed so far, a feature distance measure is proposed in this section for image authentication. The distance measure is based on the differences of the two feature vectors from the test image and from the original image. Two measures are used to exploit statistical and spatial properties of feature differences, including the kurtosis (*kurt*) of feature difference distribution and the maximum connected component size (*mccs*) in the feature difference map. Observation (1) motivates the use of the kurtosis measure, and observation (2) motivates the use of the *mccs* measure. They are combined together since any one of the above alone is still insufficient, as stated in observation (3).

The proposed Statistics- and Spatiality-based Measure (SSM) is calculated by sigmoid membership function based on both *mccs* and *kurt*. Given two feature vectors *X* and *Y*, the proposed feature distance measure SSM(*X*, *Y*) is defined as follows:

$$\mathrm{SSM}(X,Y) = \frac{1}{1+e^{\alpha(mccs\cdot kurt\cdot\theta^{-2}-\beta)}} \tag{4.5}$$

The measure SSM(*X*,*Y*) is derived from the feature difference vector δ defined in Equation(4.4). The *mccs* and *kurt* are obtained from δ, and their details are given in the next few paragraphs. θ is a normalizing factor.

The parameter α controls the changing speed especially at the point $mccs\cdot kurt\cdot\theta^{-2}=\beta$. β is the average $mccs\cdot kurt\cdot\theta^{-2}$ value obtained by calculating from a set of malicious attacked images and acceptable manipulated images. In this thesis, the acceptable manipulations are defined as the contrast adjustment, noise addition, blurring, sharpening, compression and lossy transmission (with error concealment); the malicious tampering operations are object replacement, addition or removal. During authentication, if the measure SSM(*X*, *Y*) of an image is smaller than 0.5 (that is, $mccs\cdot kurt\cdot\theta^{-2}<\beta$), the image is identified as authentic, otherwise it is unauthentic.

Kurtosis

Kurtosis describes the shape of a random variable's probability distribution based on the size of the distribution's tails. It is a statistical measure used to describe the concentration of data around the mean. A high kurtosis portrays a distribution with fat tails and a low even distribution, whereas a low kurtosis portrays a distribution with skinny tails and a distribution concentrated towards the mean.

Two distributions may have the same variance, but differ markedly in kurtosis. For example, the variance of feature differences of Figure 4.1(b) and Figure 4.1(d) are almost

the same, but the kurtosis of Figure 4.1(b) is much larger than that of Figure 4.1(d). Therefore, kurtosis is particularly helpful to distinguish feature difference distribution of the malicious manipulations from that of the acceptable manipulations.

Let us partition the spatial locations of the image into neighborhoods, and let N_i be the i-th neighborhood. That is, N_i is a set of locations that are in the same neighborhood. For example, by dividing the image into blocks of 8×8, we have a total of $WH/64$ neighborhoods, and each neighborhood contains 64 locations. Let D_i be the total feature distortion in the i-th neighborhood N_i:

$$D_i = \sum_{j \in N_i} \delta_j \tag{4.6}$$

We can view D_i as a sample of a distribution D. The *kurt* in the Equation (4.5) is the kurtosis of the distribution D. It can be estimated by:

$$kurt(D) = \frac{\sum_{i=1}^{N} (D_i - \mu)^4}{Num\ \sigma^4} - 3 \tag{4.7}$$

where *Num* is the total number of all samples used for estimation. μ and σ is the estimated mean and standard deviation of D, respectively.

Maximum Connected Component Size

Connected component is a set of points in which every point is connected to all others. Its size is defined as the total number of points in this set. The maximum connected component size (*mccs*) is usually calculated by morphological operators. The isolated points in the feature difference map are first removed and then broken segments are joined by morphological dilation. The maximum connected component size (*mccs*) is then calculated by using connected components labelling on the feature map based on 8-connected neighbourhood. Details can be found in [92, 93].

Normalizing Factor

Since natural scene images may contain different number of objects, details as well as dimensions, normalization is needed. Instead of using traditional normalization (i.e., the ratios of the number of extracted feature points to image dimension), we employ a new normalizing factor θ to make the proposed measure more suitable for natural scene images, which is defined as:

$$\theta = \frac{\mu}{W H} \tag{4.8}$$

where W and H are the width and height of the image, respectively. μ is the estimated mean of D, same as that in Equation(4.7).

It is worth noting that the two measures *mccs* and *kurt* should be combined together to handle different malicious tampering. Usually tampering results in three cases in terms of the values of *mccs* and *kurt*: (1) the most general case is that tampered areas are with large maximum connected size and distributed locally (Figure 4.1b). In this case, both *kurt* and *mccs* are large; (2) small local object is modified such as a small spot added in face (Figure 4.4a). In this case, the *mccs* is usually very small, but *kurt* is large; (3) tampered areas are with large maximum connected size but these areas are evenly distributed within the whole image (Figure 4.4c). In this case, the *mccs* is usually large, but *kurt* is small. Therefore, it is necessary for SSM to combine these two measures so that SSM could detect all these cases of malicious modifications.

(a) small object tampered (*kurt*: large; *mccs*: small); (b) feature differences of (a)

(c) object tampered with global distortions (*kurt*: small; *mccs*: large); (d) feature differences of (c)

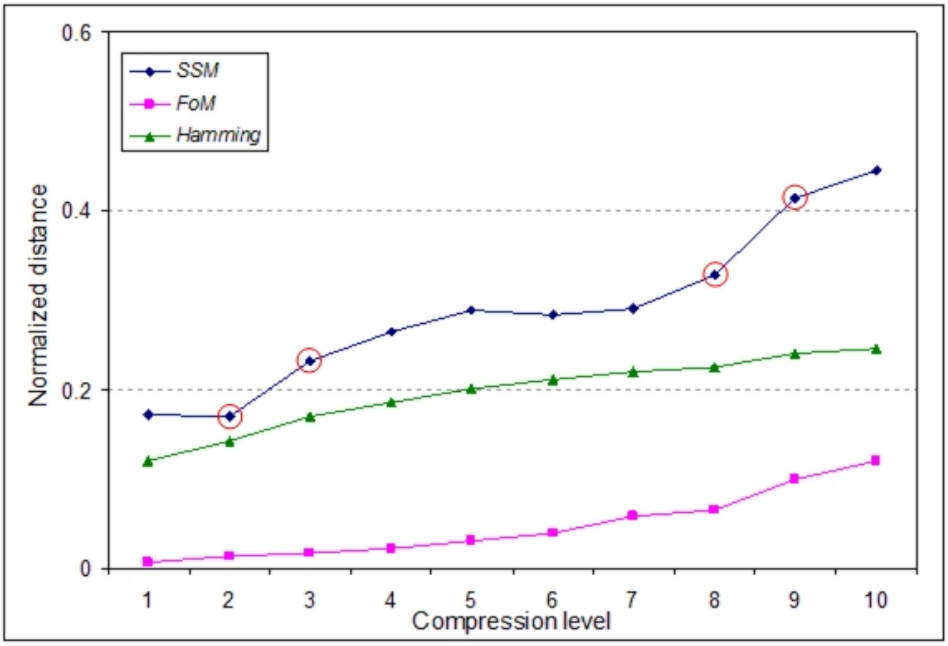
Figure 4.4: Cases that required both *mccs* and *kurt* to work together to successfully detect malicious modifications

4.2.3 Feature Distance Measure Evaluation

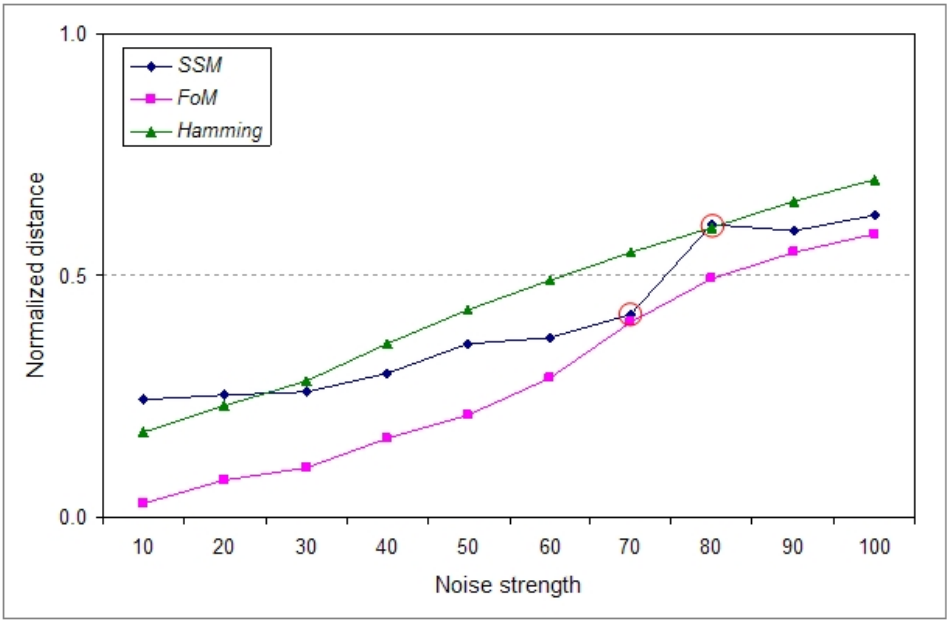
The proposed SSM has been evaluated by experiments, compared with Minkowski metrics and FoM. Edge detected by [100] was selected as the feature in our evaluations. Figure 4.3 shows the histogram of edge difference and their respective probability density estimates of noisy, error concealed, damaged and maliciously tampered images. We can find that the

distribution of the feature differences between malicious tampered image and the original image have a much longer tail than that of the error-concealed image. The damaged, error-concealed and noisy images all have smaller right tails. These results support our observations that a maliciously tampered image has a different pattern of feature differences compared to that of the acceptable manipulations.

Some acceptable distortions and malicious attacks were introduced into the original images for robustness evaluation. The proposed SSM was compared with Hamming (Minkowski Metric with r=1 for binary feature) as shown in Figure 4.5. Pratt's Figure of Merit (FoM) [91] was also used for comparison since it is commonly used at measuring image similarity based on edges. Figure 4.5(a) shows the experimental results of the proposed SSM for image *Lena* after JPEG compression, and Figure 4.5(b) shows the experimental results for Gaussian noisy images. These figures show that the Hamming and FoM distances are almost linear to the compression level or Gaussian noise strength. On the contrary, there were some sharper changes (such as the circled points in Figure 4.5) in SSM curves which may be good choices for authenticity threshold. As an image can be considered as points in a continuous space, it is typically difficult to set up a sharp boundary between authentic and unauthentic images [14]. This intrinsic fuzziness makes the content-based authentication design challenging and, likely, ad hoc in most cases [14]. Therefore, the sharper change of authenticity based on the proposed measure around threshold may lead to a sharper boundary between the surely authentic and unauthentic images, which is desirable for image authentication.

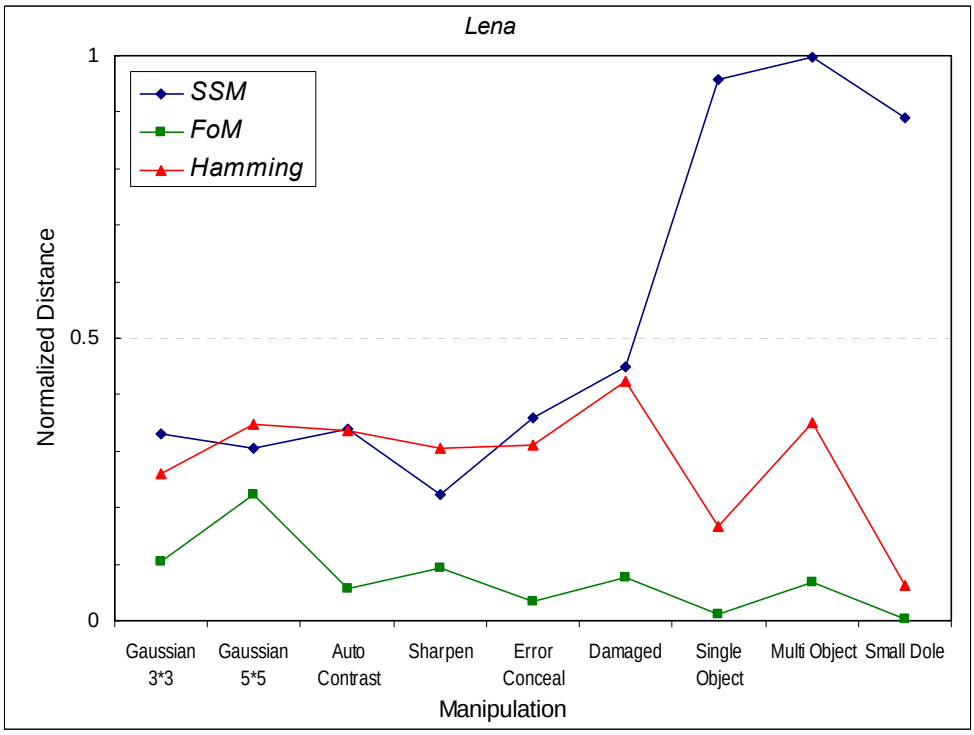


(a) JPEG compressions

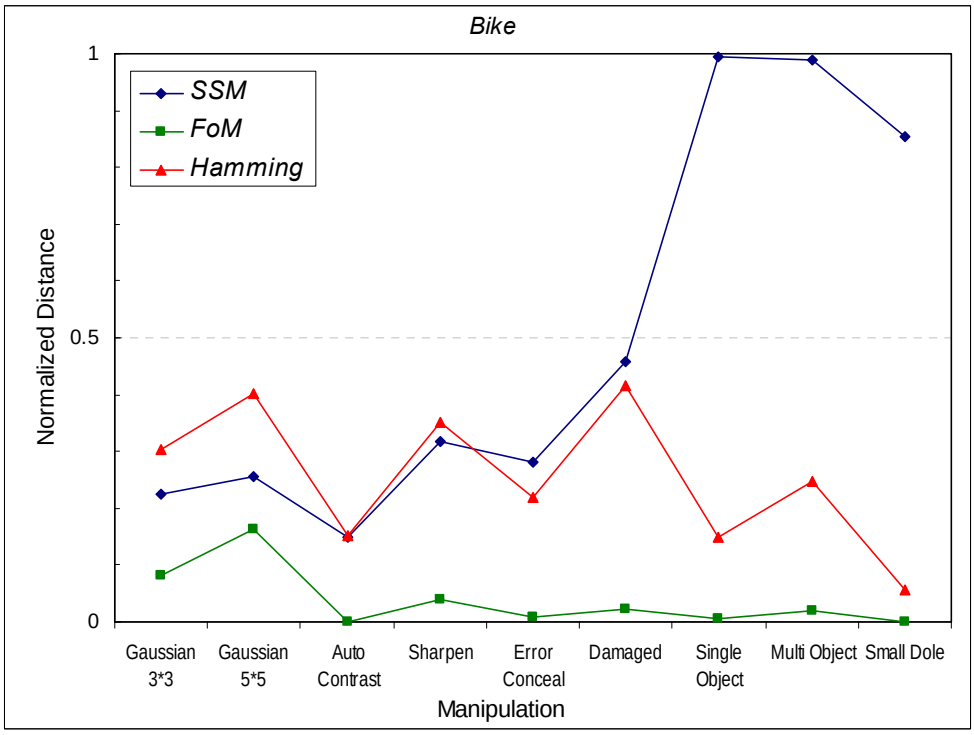


(b) *Gaussian* noises

Figure 4.5: Distance measures comparison



(a) results of image *Lena*;



(b) results of image *Bike*

Figure 4.6: Comparison of distinguishing ability of different distance measures: only the proposed measure can successfully distinguish malicious manipulations from acceptable ones

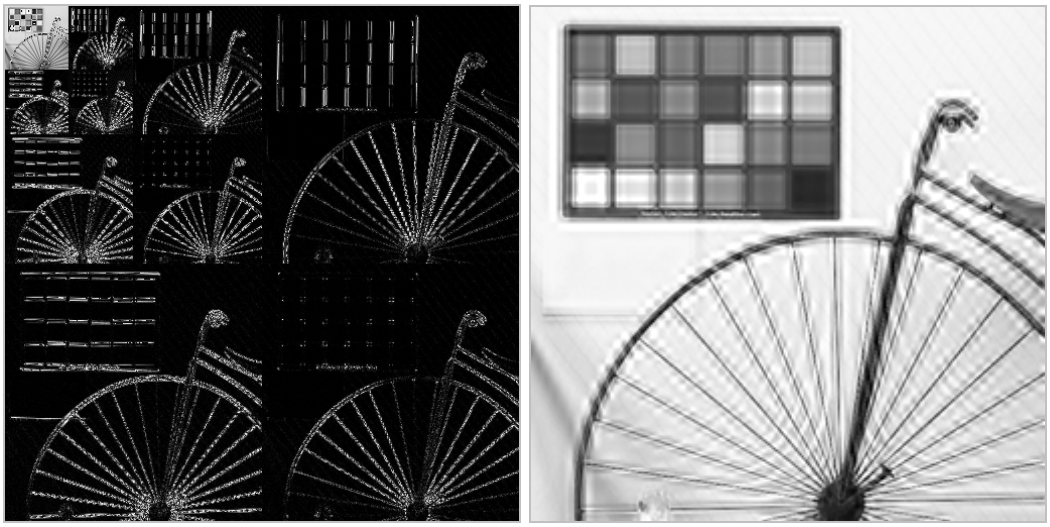
Figure 4.6 shows the comparison results of different distance measures in terms of their discernable abilities. In Figure 4.6(a), the last three columns are images maliciously tampered from the original portrait image *Lena*, by enlarging the eyes, modifying multiple objects in the image, and adding a small spot on the face. The others are images from acceptable manipulations including Gaussian noise addition, auto contrast adjustment, sharpening, and lossy transmission (with error concealment). Figure 4.6(b) shows results of image *Bike* with much stronger edges than image *Lena*. The last three columns of Figure 4.6(b) are images tampered by deleting the saddle, modifying multiple objects (changing logo at the left top, modifying the display of the clock at right top, and deleting the saddle), and adding a small spot in the center of the right circle. Note that the SSMs were all below 0.5 for acceptable manipulations and all above 0.5 for maliciously attacked images. On the contrary, the Hamming and Figure of Merit (FoM) measures of maliciously attacked images were among the range of acceptable manipulations especially the measures of the attacked image in which there was a small local object changed (last column). The results show that the proposed SSM was able to distinguish the malicious manipulations from acceptable ones, i.e., identify lossy transmission as acceptable, and was sensitive to malicious manipulations. On the contrary, the Hamming and FoM measures were not sensitive to small localized object modification. The results indicate that the proposed SSM is more suitable for content-based image authentication than the Hamming and FoM measures.

4.3 Error Concealment using Edge Directed Filter for Wavelet-based Images

As discussion in the previous chapter, error concealment would be applied before image authentication if the image has been damaged by transmission errors. Error concealment technique for JPEG images has been discussed in the previous chapter. Given an image to be verified, the first step is to conceal the errors if some transmission errors are detected. As

a result, the content authenticity of the error concealed image is higher than that of the damaged image, which is validated in our experiments of the error resilient image authentication. This section presents an error concealment technique for wavelet-based images.

The effects of errors in wavelet-based images depend on which parts of wavelet coefficients are corrupted. Errors in JPEG2000 bitstreams would result in the loss of a bitplane of the wavelet coefficients of the affected subband. The LL subband can be considered as a subsampled version of the original image, so errors in the LL coefficients are similar to the lost blocks in JPEG images. Therefore, recovery of LL coefficients can be achieved by some error concealment techniques developed for block-based image [74]. There, we only focus on concealing the errors of the high frequency coefficients.



(a) Wavelet decomposition; (b) High frequency errors

Figure 4.7: Wavelet-based image (*Bike*) error pattern

In the wavelet domain, the energies of high frequency coefficients are mainly concentrated around edges in image (Figure 4.7a). When errors occur in these subbands, errors have effects like ring or ripple artifact around edges (Figure 4.7b) in the damaged image. However, edges in a natural image have important effects on the subjective visual quality, since edges are always associated with the boundary of an object, or with marks on

the object. An image with blurred edges is always annoying to the spectator. Our proposed algorithm aims to remove the noises around edges and then to improve the image quality.

4.3.1 Edge Directed Filter based Error Concealment

The proposed algorithm is inspired by how experts repair damaged images, which involves determining the areas to be corrected, examining the boundary of these regions, continuing lines into these regions, gradually filling in, and painting small details [94, 95]. However, it cannot be applied to conceal errors in JPEG2000 images directly, because the errors in JPEG2000 images do not result in lost blocks. Based on the idea of how experts repair damaged images, we propose an error concealment scheme based on a new edge directed filter for wavelet-based images. It makes use of the redundancy residuals in spatial domain combined with those in wavelet domain. The process of proposed error concealment scheme can be summarized as:

$$I^{n+1} = \mathrm{W}^{-1} \circ \mathrm{C} \circ \mathrm{W} \circ \mathrm{F}\left(I^{n}\right) \tag{4.9}$$

where I^{n+1} is the recovered image after the n+1 iterations, and I^{0} is the received image. "$\circ$" is concatenation operation of two functions. Function F is the edge directed filtering in the spatial domain to remove artifacts around edges. Function W is the wavelet transform, and function W^{-1} is the inverse wavelet transform. C is a function that rectifies the recovered results, taking in information regarding which bit-planes are lost, as well as I^{0} as the input.

In other words, the damaged image is firstly filtered using edge directed filter F, and then transformed into WT domain (W). The recovered WT coefficients are then constrained to their statistical characteristics in the WT domain by using function C. These recovered wavelet coefficients are then transformed into the image domain again (W^{-1}) to get a valid image I^{n+1}. The constraint function C comprises the known WT coefficient values constraint function C_1 and the WT statistical characteristics constraint function C_2. That is, $\mathrm{C} = \mathrm{C}_1 \circ \mathrm{C}_2$.

Let ψ be the set of images comprised by those which satisfy the two WT constraints discussed in Subsection 2.4. The above algorithm can be viewed as an attempt to find a recovered image I in ψ that minimizes the distortion between I and the image $F(I_0)$, that is, find *I* which:

$$\begin{aligned} &\min_{I_c} \left(\mathrm{F}(I_0) - I\right) \\ &s.t. \quad I \in \boldsymbol{\psi} \end{aligned} \tag{4.10}$$

4.3.2 Edge Directed Filter

Based on the error pattern of the wavelet-based images, we can construct an edge directed filter to remove the noises around edges caused by errors in the damaged images. Anisotropic diffusion techniques have been widely used in image processing for its efficiency of smoothing the noisy images while preserving the sharp edges [96]. When some proper function is constructed in anisotropic diffusion, it can form direction diffusion or edge directed filter to remove the ring or ripple artifacts around edges of damaged images caused by errors in high frequency subbands. An edge directed filter using a new diffusion function is proposed in this section.

The original anisotropic diffusion equation is presented by Perona and Malik [97], which can be written as a Partial Differential Equation (PDE):

$$\begin{cases} \dfrac{\partial I}{\partial t} = div\left(f\left(\left|\nabla I\right|\right)\nabla I\right) \\ I\left(t\ =0\right) = I_0 \end{cases} \tag{4.11}$$

where I_0 is the initial image, and ∇ is the gradient operator:

$$\nabla I = \frac{\partial I}{\partial x}\vec{x} + \frac{\partial I}{\partial y}\vec{y} \tag{4.12}$$

The gradient magnitude is used to detect an image edge or boundary as a step discontinuity in intensity. In our scheme, *Sobel* operator is adopted to generate the gradient

for the damaged image. The *div* is the divergence operator and $f(x)$ is a decreasing positive diffusion function. Perona and Malik suggested two diffusion functions:

$$f(x)=\frac{1}{1+(x/K)^2} \tag{4.13}$$

$$f(x)=\exp\left(-(x/K)^2\right) \tag{4.14}$$

where K is a constant with fixed value. The scale-spaces generated by these two functions are different: the function (4.13) privileges wider regions over smaller ones, while the function (4.14) privileges high-contrast edges over low-contrast ones [97].

We adopt the anisotropic diffusion as a direction diffusion operation, and design a new diffusion function for error concealment. Since we only aim to construct edge directed filter to remove the ring or ripple artifacts caused by errors, in our solution the diffusion function *f(x)* is:

$$\begin{cases} f(\nabla I)=\dfrac{k\exp(-|\nabla I|/M)}{\max(\exp(\Delta I),\ 1+|\nabla I|)} \\ M=\max\limits_{P\in\mathbf{\Gamma}}(|\nabla I_P|) \end{cases} \tag{4.15}$$

where $\mathbf{\Gamma}$ is the $N\times N$ pixels blocks where the damaged pixel belongs to (N=16 and k=1 in this chapter), and $|\nabla I|$ is the magnitude of ∇I . ΔI is the *Laplacian* of image I, a second order derivative of I, defined as:

$$\Delta I=\nabla(\nabla I)=\frac{\partial^2 I}{\partial x^2}+\frac{\partial^2 I}{\partial y^2} \tag{4.16}$$

If $|\nabla I|$ is close to M, $f(\nabla I)$ is approximate to the minimum, and we have the direction filter along with the direction of edges. If $|\nabla I|$ is very small, $f(\nabla I)$ is approximate to the maximum, and now we could achieve isotropic diffusion (like Gaussian filter).

M and ΔI are used to make that the conduction coefficients of the ring or ripple artifacts are small, because the errors are concentrated around edges and the gradient values

are always smaller than those of edges. Thus the artifacts are made to be filtered smoothly, but edges are kept sharp.

A discrete form of Equation (4.16) is given by:

$$I^{n+1} = I^n + \frac{\Delta t}{N}\sum_{i=1}^{N} f\left(\nabla I_i^n\right)\cdot\nabla I_i^n \tag{4.17}$$

This process constructs edge directed filter served as edge directed filter F in Equation(4.9).

4.3.3 Wavelet Domain Constraint Functions

Two WT domain constraint functions are applied in wavelet domain: known-value constraint function C_1, and WT statistical constraint function C_2 to rectify the recovered coefficients.

After the damaged image is filtered by edge directed filter, the lost WT coefficients are recovered. However, the correctly received WT coefficients (denoted as $\mathbf{\Phi}$) may also be altered at the same time. We should discard these changes, with known-value constraint function:

$$\mathrm{C}_1(x) = \begin{cases} x_0, \text{ if } x \in \mathbf{\Phi} \\ x, \text{ else} \end{cases} \tag{4.18}$$

where x_0 is the original wavelet coefficients of x before edge directed filtering.

Furthermore, although WT almost decorrelates WT coefficients, the distribution of one coefficient conditioned on its parent P usually is a linear function of P [8]. It means that the coefficients are still statistically dependent. For high-amplitude coefficients, if the parent is less than some threshold (e.g., one standard deviation) then the child is also most likely to be less than the threshold [8]. Moreover, wavelet coefficients also show their statistical dependency across their neighborhoods in spatial domain. After using the function C_1, such

characteristics may not be kept anymore. Thus, the image set **Ω** within statistical characteristics is used to construct a function to discard the recovered wavelet coefficients which violate these statistical characteristics constraint function:

$$C_2(x) = \begin{cases} x, \text{ if } x \in \mathbf{\Omega} \\ 0, \text{ else} \end{cases} \tag{4.19}$$

Then we get **ψ** (**ψ** = **Ω** ∩ **Φ)**, and C (C = $C_1 \circ C_2$).

4.3.4 Error Concealment Evaluation

In our experimental evaluation, the error detection can be done by the error resilience tools issued in JPEG2000 [98]. We use five-level wavelet decomposition.

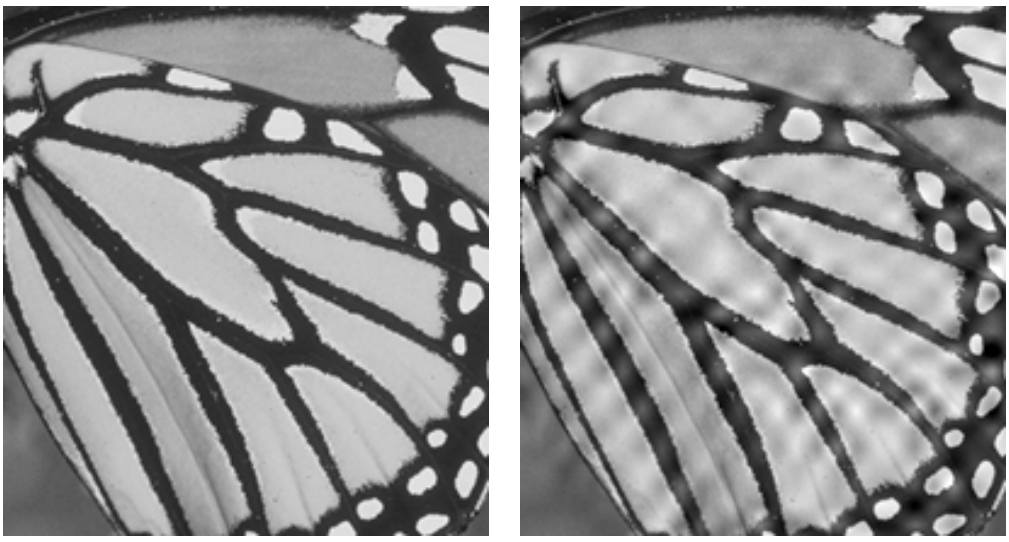
The details of improvement of the damaged image *Monarch* by our proposed algorithm is shown in Figure 4.8. We can see that the annoying noises around edges have been almost removed. And the recovered areas have high continuity, which visually makes the spectators much comfortable. The cost is the decrease in the contrast. But such change is not easily caught by human eyes.

The PSNR results are listed in Table 4.1. The bit error rate (BER) is set to 10^{-4}. In terms of PSNR, the improvement on the quality of the damaged images is significant, though the improvement varies from image to image. For example, image *Monarch* contains much more clear edges and in stronger contrast than *Lake*, so it can achieve better result than Lake.

Considering that the criterion of PSNR does not always provide an accurate measure of the visual quality for natural images. To evaluate the performances on edge preservation of the proposed algorithm, we further use Figure of Merit (FoM). The FoM values of the results are also listed in Table 4.1. We used *Canny* edge detector, and the standard deviation of the Gaussian kernel in the Canny detector is set to 0.4. We can see that the FoM values of

the recovered images are close to 1, which shows high edge preserving ratios are achieved by the proposed error concealment algorithm.

The efficiency of our proposed diffusion function (Equation (4.15)) is illustrated in Figure 4.9, compared with the classic Laplacian edge enhancement filter (detailed in [10]), and the two anisotropic diffusion functions proposed by Perona and Malik (Equation (4.13) and (4.14), $K = 25$). These filters are all used in the same error concealment scheme defined by Equation (4.11).



(a) original image; (b) damaged image;

(c) recovered image

Figure 4.8: Edges enhanced by the proposed error concealment (*Monarch*)

Table 4.1: Image quality evaluation of error concealment

Image Quality		*Actor*	*Bike*	*Chart*	*Fruits*	*Hotel*	*Lake*	*Lena*	*Monarch*	*Peppers*	Average
Damaged Images	PSNR (dB)	30.76	30.34	33.85	33.63	33.83	31.37	33.31	29.49	33.05	32.18
	FoM (%)	88.5	88.0	91.3	88.7	88.0	89.1	89.0	89.2	87.3	88.8
Recovered Images	PSNR (dB)	38.24	38.55	40.13	39.21	41.79	38.37	40.02	39.57	40.77	39.63
	FoM (%)	94.5	95.2	96.3	94.2	95.5	93.9	93.7	94.2	91.2	94.3

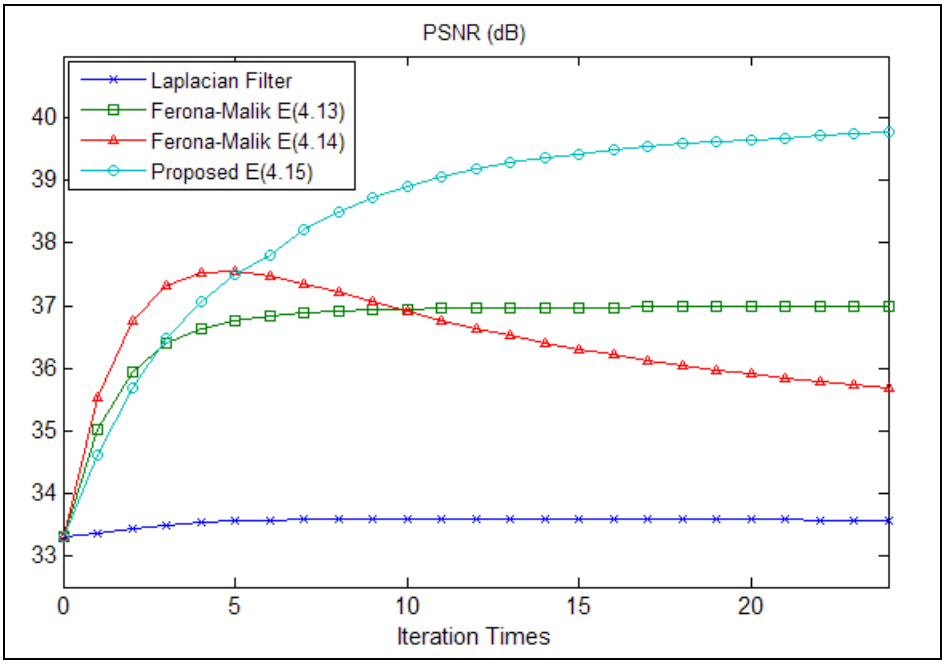


Figure 4.9: Comparison of diffusion functions (*Lena*)

Figure 4.9 shows the progressive visual results of image *Lena*, done by our proposed error concealment algorithm. We can see that after a few iterations (5 to 10 times), the noises around edges are almost removed.

4.4 Application of SSM in Error Resilient Wavelet-based Image Authentication

The proposed feature distance measure SSM is quite general that it is able to be used in many content-based authentication schemes which use features containing spatial information. When used as the feature distance function in image authenticity verification stage, it would improve the system discrimination ability between acceptable and malicious

manipulations. Many acceptable manipulations, which were detected as malicious modifications in the previous schemes, can be bypassed in the scheme using SSM.

JPEG2000, an emerging wavelet-based image compression standard, can operate at higher compression ratios without generating the typical blocking artifacts of the previous DCT-based JPEG standard. It also allows more sophisticated progressive downloads. The error resilient image authentication scheme present in the previous chapter is only suitable for JPEG images. It is not suitable for wavelet-based images. This section presents an improved error resilient image authentication scheme which is applicable to both JPEG and JPEG2000 images.

The improved error resilient scheme exploits the proposed feature distance measure SSM in a generic semi-fragile image authentication framework [15] to distinguish images distorted by transmission errors from maliciously modified ones. The experimental results support that the proposed feature distance measure can improve the performance of the previous scheme in terms of robustness and sensitivity. The results of this error resilient authentication scheme validate that the proposed SSM can improve the authentication performance.

4.4.1 Feature Extraction

One basic requirement for selecting feature for content-based image authentication is that the feature should be sensitive to malicious attacks on the image content. Edge-based features would be a good choice because usually malicious tampering will incur the changes on edges. Furthermore, edge may also be robust to some distortions. For instances, the results in [99] show that high edge preserving ratios can be achieved even if there are uncorrectable transmission errors. Therefore, the remaining issue is to make the edge more robust to the defined acceptable manipulations. Note that this is main reason why we

employ the normalization in Equation (4.8) to suppress those "acceptable" distortions around edges.

In [100], a method based on fuzzy reasoning is proposed to classify each pixel of a gray-value image into a shaped, textured, or smooth feature point. We adopt their fuzzy reasoning based detector because of its good robustness.

4.4.2 Signature Generation and Watermark Embedding

The image signing procedure is outlined in Figure 4.10. Binary edge of the original image is extracted using the fuzzy reasoning based edge detection method [100]. Then, the edge feature is divided into 8×8 blocks, and edge point number in each block is encoded by error correcting code (ECC) [10]. BCH(7,4,1) is used to generate one parity check bit (PCB) for ECC codeword (edge point number) for every 8×8 block. The signature is generated by hashing and encrypting the concatenated ECC codewords using a private key. Finally, the PCB bits are embedded into the DCT coefficients of the image. In our implementation, the PCB bits are embedded into the middle-low frequency DCT coefficients using the same quantization based watermarking as present in [15].

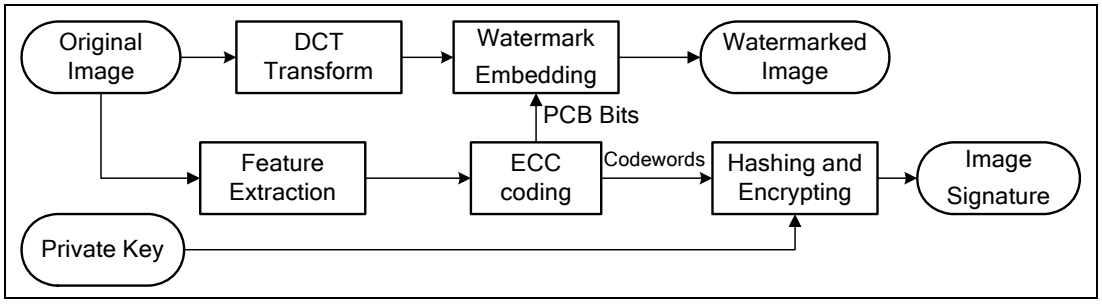


Figure 4.10: Signing process of the proposed error resilient image authentication scheme

Let the total selected DCT coefficients form a set **P**. For each coefficient c in **P**, it is replaced with c_w which is calculated by:

$$c_w = \begin{cases} Q\,\mathrm{round}(c/Q), \text{ if } \mathrm{LSB}(\mathrm{round}(c/Q)) = w \\ Q\left(\mathrm{round}(c/Q) + \mathrm{sgn}\left(c - Q\,\mathrm{round}(c/Q)\right)\right), \text{ else} \end{cases} \tag{4.20}$$

where w (0 or 1) is the bit to be embedded. Function round(x) returns the nearest integrate of x, *sgn(x)* returns the sign of x, and *LSB(x)* returns the least significant bit of x. Equation (4.20) makes sure that the LSB of the coefficient is the same as the watermark bit.

The watermarking procedure would introduce some distortions into the image, which makes the re-extracted features different from those of the original image. Therefore, the embedding procedure should not affect the feature extracted. In order to exclude the effect of watermarking from feature extraction, a compensation operator C_w is adopted before feature extraction and watermarking:

$$\begin{cases} I_c = C_w(I) \\ I_w = f_e(I_c) \end{cases} \tag{4.21}$$

$$C_w(I) = \mathrm{IDCT}\left\{\mathrm{IntQuan}\left(d_i, 2Q, \mathbf{P}\right)\right\} \tag{4.22}$$

where d_i is the i-th DCT coefficient of I, and IDCT is inverse DCT transform. $f_e(I)$ is the watermarking function, and I_w is the final watermarked image. The IntQuan(c, **P**, Q) function is defined as:

$$\mathrm{IntQuan}\left(c, Q, \mathbf{P}\right) = \begin{cases} c, \text{ if } c \notin \mathbf{P} \\ Q\,\mathrm{round}(c/Q), \text{ else} \end{cases} \tag{4.23}$$

C_w is designed according to the watermarking algorithm, which uses 2Q to pre-quantize the DCT coefficients before feature extraction and watermarking. That is, from Equation (4.20), (4.22) and (4.23), we can get $C_w(I_w) = C_w(I)$, thus $f_e(I_w) = f_e(I)$, i.e., the feature extracted from the original image I is the same as the one from the watermarked image I_w. This compensation operator ensures that watermarking does not affect the extracted feature.

4.4.3 Image Authenticity Verification

The image verification procedure, shown in Figure 4.11, can be viewed as an inverse procedure of image signing. Firstly, error concealment is carried out if transmission errors are detected. The feature of the image is extracted using the same method as used in image signing procedure. Watermarks are then extracted. If there are no uncorrectable errors in the ECC codewords, the authentication is based on bit-wise comparison between the decrypted hashed feature and the hashed feature extracted from the image [10]. Otherwise, image authenticity is calculated by the SSM based on differences between the PCB bits of the re-extracted feature and the extracted watermark. Finally, if the image is identified as unauthentic, the attacked areas are then detected.

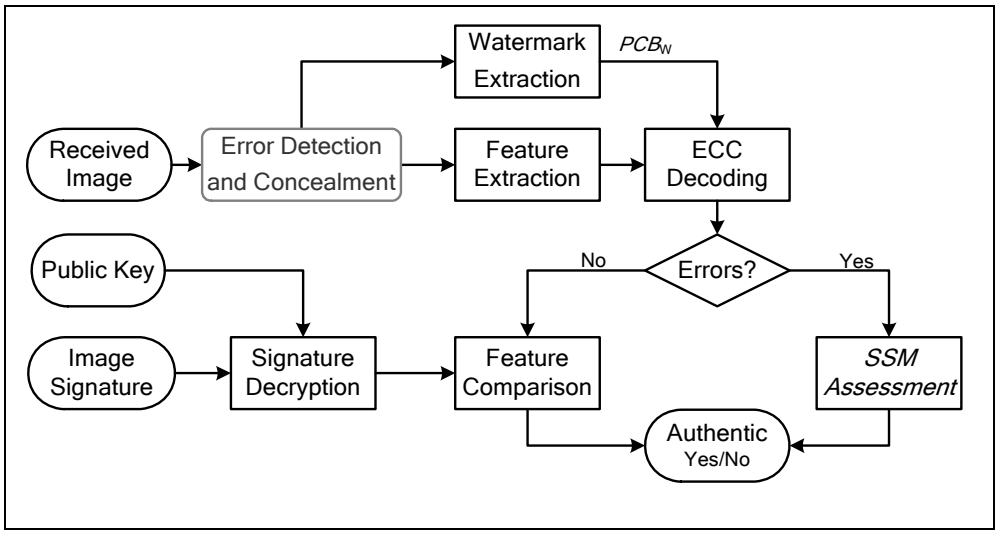


Figure 4.11: Image authentication process of the proposed error resilient image authentication scheme

Image Authenticity Verification

Image authenticity is calculated based on the binary difference map which is created by comparing the PCB bits decoded from extracted watermark and the recalculated PCB bits from feature extracted from the image to be evaluated. The PCB bits of the image feature

are recalculated using the same method as used in image signing procedure. The difference map is obtained by differentiating these two PCB vectors.

Given an image to be verified, we repeat feature extraction described in image signing procedure. The corresponding PCB bits (PCB_W) of all 8×8 blocks (one bit/block) of the image are extracted from the embedded watermarks. Then the feature set extracted from the image is combined with the corresponding PCB bits to form ECC codewords. If all codewords are correctable, we concatenate all codewords and cryptographically hash the result sequence. The final authentication result is then concluded by bit-by-bit comparison between these two hashed sets. If there are uncorrectable errors in ECC codewords, image authenticity is calculated based on the proposed distance measure. The two feature vectors in the proposed measure are PCB_W from watermarks and the recalculated PCB bits (PCB_F) from ECC coding of the re-extracted image feature set. If the distance measure between PCB_W and PCB_F is smaller than 0.5 (SSM(PCB_W, PCB_F)<0.5), the image is deemed authentic. Otherwise, the image is deemed unauthentic.

Feature Aided Attack Location

If the image is verified as unauthentic, the tampered areas could be detected. Attack location is an important part of the authentication result since the detected attacked areas give the users a clear figure where the image has possibly been tampered with. The diagram of our feature aided attack location algorithm is shown in Figure 4.12. The attack areas are detected using information from watermarks and image feature. The difference map between PCB_W and PCB_F is calculated, and then morphological operations are used to compute the connected areas, with isolated pixels and small connected areas removed. After these operations, the difference map is masked with the union of the watermark and the feature. The masking operation can refine the detected areas by concentrating them on the objects in the tampered image or in the original image. The areas in the difference map

which do not belong to any object (defined by edge feature) are removed, which may be a false alarm of some noises.

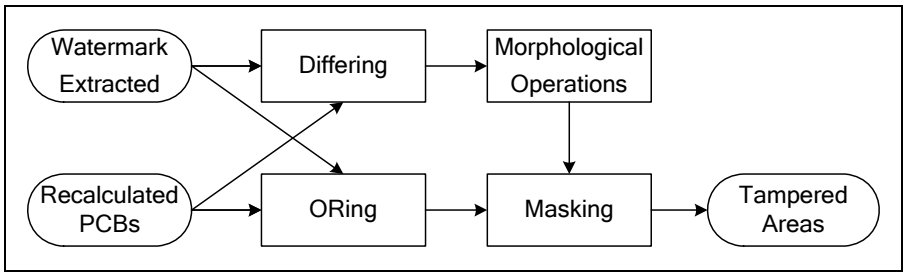


Figure 4.12: The diagram of feature aided attack localization

It is worth noting that the authentication result of our scheme is friendly to users. Since human perceptivity treats image as a combination of objects, some objects may be the region of interest (ROI) to users. If the image fails to pass the authentication, our scheme provides possible attacked areas which concentrate on objects. If these detected areas are not the user's ROI, further decision can be made by the user on a case by case basis. Finally, this scheme can also provide a degree of authenticity (by SSM measure) to the user which gives the user a confidence on the trustiness of the image.

4.5 Experimental Results and Discussions

To support our solutions, experimental results have been collected on the proposed error concealment, feature distance measure, and error resilient image authentication. In our experiments, JPEG and JPEG2000 images were used. Test images include *Actor, Barbara, Bike, Airplane, Fruits, Girl, GoldenHill, Lena, Mandrill, Monarch, Pepper, Woman*, and so on. The dimensions of these images include 512×512, 640×512, 640×800, and 720×576. *Daubechies* 9/7 wavelet filter is used for the wavelet transform (which is also applied in the JPEG2000 standard [101]). The parameters α and β in Equation(4.5) were set to 0.5 and 48.0, respectively.

4.5.1 SSM-based Error Resilient Image Authentication Scheme Evaluation

Robustness to Transmission Errors and other Acceptable Manipulations

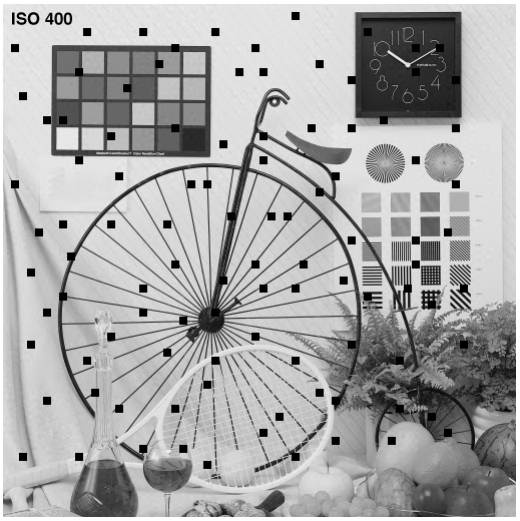
The transmission errors in wireless networks were simulated based on the Rayleigh model [102] which is commonly used for wireless networks. Figure 4.13(b) is an example of wavelet-based images damaged by transmission errors, and Figure 4.13(c) is its error-concealed result. Figure 4.13(d) is a DCT-based image damaged by transmission errors, and Figure 4.13(e) is its error concealed result. The SSM values of image Figure 4.13(c) and Figure 4.13(e) are 0.134 and 0.250, i.e., the error-concealed images are both authentic.

With the set of images produced, the average peak signal-to-noise ratio (defined by PSNR) of our watermarked images is 44.46 dB (Table 4.2), which is above the usually tolerated degradation level of 40 dB [103] and much higher than the average 33.45dB in [15]. It is also better than the 42.47 dB obtained by the paper [103]. The quantization table used in these experiments is JPEG recommended quantization table of Q50. These results indicate the embedding procedure did not introduce visual artifacts in the images.

(a) original image

(b) damaged image (wavelet based)

(c) error concealed image of (b)



(d) damaged image (DCT based)

(e) error concealed image of (c)

Figure 4.13: Robustness against transmission errors

Table 4.2 shows the evaluation results of the system robustness of the proposed error resilient image authentication scheme based on the proposed SSM. PSNR and SSM measures of the images damaged by transmission errors with different bit error rate of the transmitted images (BER) 10^{-4} and 2×10^{-4}. The corresponding PSNR and SSM of the error-concealed images are also listed in this table. 60% of the damaged images at BER 10^{-4} and 100% at BER 2×10^{-4} in our experiments were verified as unauthentic. On the contrary, all error-concealed images were verified as authentic. These results indicate that our proposed scheme could obtain a good robustness to transmission errors. Note that on the contrary, the authentication scheme [103] was not robust to transmission errors. These results further confirm that it is effective and advisable for error concealment to be applied before image authentication. The reason why the authenticities of the recovered images were better than those of the damaged images may be the image quality improvement by using error concealment on the damaged images [99, 77]. For example, the recovered image had much better objective qualities than the damaged images (evaluated by PSNR). This quality improvement made features of the error-concealed images closer to those of the original images than damaged images, so that the image authenticities (evaluated by SSM) of the error-concealed images were much larger than the damaged images.

Table 4.2: Comparison of objective quality reduction introduced by watermarking: PSNR(dB) of watermarked images

PSNR	*Barbara*	*Bike*	*Airplane*	*Girl*	*Goldhill*	*Lena*	*Mandrill*	*Monarch*	*Pepper*	*Woman*
Proposed	44.17	44.40	44.56	44.39	44.32	44.60	44.14	44.75	44.46	44.79
Ref. [15]	32.90	29.91	32.01	34.20	34.07	36.11	32.38	30.43	35.53	36.98
Ref. [103]	42.72	/	43.15	/	/	/	/	/	/	/

Table 4.3: Authentication performance improved by error concealment: PSNR (dB) and SSM of damaged images and error-concealed images (BER1:10^{-4}; BER2:2×10^{-4})

Images		*Actor*	*Bike*	*Chart*	*Flight*	*Fruits*	*Hotel*	*Lake*	*Lena*	*Pepper*	*Woman*
Damaged PSNR	BER1	30.78	31.26	33.95	32.41	33.68	33.87	31.39	33.31	33.07	35.50
	BER2	25.87	25.76	28.51	26.05	27.81	26.71	25.68	30.34	27.74	30.72
Damaged SSM	BER1	0.948	0.939	0.707	0.297	0.794	0.365	0.143	0.391	0.729	0.989
	BER2	0.812	0.999	0.987	0.951	0.942	0.568	0.883	0.638	0.865	0.955
Recovered PSNR	BER1	38.03	41.76	41.11	41.03	39.90	42.40	38.54	40.21	41.25	42.96
	BER2	32.06	34.99	34.74	34.06	31.68	33.26	31.64	36.03	33.85	36.84
Recovered SSM	BER1	0.158	0.134	0.141	0.035	0.204	0.067	0.057	0.345	0.089	0.329
	BER2	0.220	0.099	0.446	0.072	0.406	0.045	0.280	0.059	0.182	0.015

Our scheme was also tested on other acceptable manipulations such as image contrast adjustment, histogram equalization, compression and noises addition. The results are shown in Table 4.4, with the parameter for each manipulation. The SSM values of these images were all less than 0.5, i.e., all these images can pass the authentication. These results validate that the proposed scheme is not only designed to be robust to transmission errors, but also robust to general acceptable manipulations.

Table 4.4: Robustness against acceptable image manipulations

Manipulations	*Histogram Normalizing*	*Brightness Adjustment*	*Contrast Adjustment*	*JPEG Compression*	*JPEG2000 Compression*
Parameter	Auto	-40	Auto	10:1	1bpp
SSM	0.159	0.159	0.262	0.017	0.057

Sensitivity to Malicious Content Tampering

An important aspect of our SSM-based authentication scheme is that it is sensitive to the malicious content tampering. For that reason, we tampered the previous watermarked *Bike* and *Lena* images and tested the ability of our system to detect and highlight the attacked areas. All the attacked images were detected and the possibly attacked areas were located. The attack location results are shown in Figure 4.14.

These results indicate that the ability of our system to detect tampering is good even in the presence of multiple tampered areas (Figure 4.14e), or noises (Figure 4.14a), or very small area modified (Figure 4.14c). Furthermore, the attack detection result of our scheme is friendly to the users. If the image fails to pass the authentication, our scheme provides detected attacked areas which concentrate on the objects. Further authentication decision can be made by the user with the aid of the attack location results.

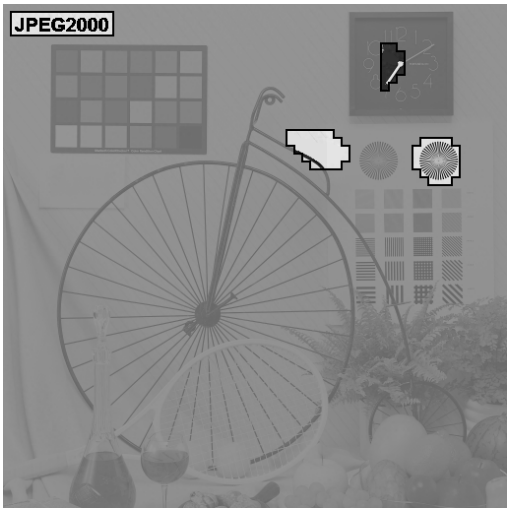
(a) Noisy tampered image *Lena* (0.995)

(b) Attacked areas detected of (a)

(c) Lena with mole added (0.569)

(d) Attacked areas detected of (c)

(e) Attacked image Bike (0.995)



(f) Attacked areas detected of (e)

Figure 4.14: Detected possible attacked locations

4.5.2 System Security Analysis

The security of our scheme can be justified by the false acceptance rate (*FAR*) and false rejection rate (*FRR*). For an image authentication system, *FAR* represents the rate that an image is actually modified by a malicious modifications but some tampered areas are not detected. On the other hand, *FRR* is the rate that an image is detected to be maliciously tampered but in fact it is not. A practical signature system should ensure that both *FAR* and *FRR* are reasonably small.

The image authenticity verification of our scheme is based on the SSM, which is calculated using two statistical parameters (*mccs* and *kurt*) of the edge difference map. Referring to Figure 4.3, the longer the tail of the distribution of the difference map is, the larger the *kurt* is. Since malicious attack concentrates on image objects, the difference pixels are more likely to be connected (thus larger *mccs*) than acceptable manipulations. These two observations are combined to achieve good performance in our scheme.

Let λ be the probability of connection with neighboring pixels in the feature difference map. From the region size model stated above, the *FAR* and *FRR* of our scheme can be approximated as:

$$\begin{aligned} FAR &= p(kurt_m\ mccs_m < \alpha) \\ &= \sum_{s=1}^{\alpha / kurt_m} p(kurt_m\ s < \alpha \mid mccs_m = s)\, p(mccs_m = s) \\ &= \sum_{s=1}^{\alpha / kurt_m} p(kurt_m\ s < \alpha \mid mccs_m = s)\, p(\prod_{i=1}^{s} \lambda_m Cs^k) \end{aligned} \tag{4.24}$$

$$\begin{aligned} FRR &= p(kurt_a\ mccs_a > \alpha) \\ &= \sum_{s=\alpha / kurt_a}^{\infty} p(kurt_a\ s > \alpha)\, p(mccs_a = s) \\ &= \sum_{s=\alpha / kurt_a}^{\infty} p(kurt_a\ s > \alpha \mid mccs_a = s)\, p(\prod_{i=1}^{s} \lambda_a Cs^k) \end{aligned} \tag{4.25}$$

For acceptable manipulations, the $kurt_a$ is small because its difference map is evenly distributed and concentrated on its average. Therefore, *FRR* is accumulated from large

initial ($\alpha/kurt_a$), and $p(mccs_a = s)$ is also *accumulated* from large initial far away from the high density center of the distribution, i.e. Cs^k is very small. Furthermore, the connection factor λ_a is also small because the difference pixels are not likely to be connected, thus (λ_a) is close to zero.

On the contrary, for malicious modifications, the $kurt_m$ is large due to the object-concentrated modification. Therefore, *FAR* is accumulated from 0 to small end ($\alpha/kurt_m$). The probability for $mccs_m$ equal to s is also accumulated from large initial far away from the high density center of the distribution, i.e. Cs^k is very small, and λ_m is also large.

The tradeoff between *FAR* and *FRR* can be adjusted by the parameter α. Since nonlinear perceptual distance is used, the probabilities in the equations above are difficult to calculate. However, it is a reasonable conclusion from discussions above that both *FAR* and *FRR* will be very low.

4.6 Summary

An error resilient image authentication scheme using statistics and spatiality based measure (SSM) is presented in this chapter, which is robust to transmission errors in JPEG and JPEG2000 images. Many acceptable manipulations, which were incorrectly detected as malicious modifications by the previous schemes, were correctly classified by the proposed scheme in our experiments. These results support the observation that the feature difference patterns under typical acceptable image modifications or malicious ones is distinguishable. The results may indicate that the statistical and spatial properties of the image feature are helpful and useful in distinguishing acceptable image manipulations from malicious ones. The proposed SSM would improve system performance for content-based authentication schemes which use features containing spatial information, such as edge [7, 13], block DCT coefficients based features [8, 14, 15], highly compressed version of the original image [9],

or block intensity histogram [16]. Furthermore, the proposed error resilient scheme based on SSM can improve the trustworthiness of digital images damaged by transmission errors by providing a way to distinguish them from digital forgeries.

A limitation of the proposed measure is that it is suitable only for schemes using features containing spatial information since it is based on the statistical and spatial properties of feature differences. Further work would be needed to expand the use of the proposed measure by exploiting new discernable patterns in feature differences when the features contain no spatial information.

Chapter 5

Image Forensics based on Image Quality Inconsistency Measure

Digital watermarking and signature are main tools for active image authentication. However, most images captured today do not contain any digital watermark or signature, which motivate us to design techniques for passively checking the integrity of digital images. During digital forgery creation processing, there are always different sources of images spliced to create the final forgery. If the forgery is a composite image, it is hard to make various quality measures of the different parts consistent. Therefore, quality inconsistencies found in an image can serve as a proof for the existence of tampering.

This chapter presents a digital image forensics technique based on image quality inconsistencies to detect the traces of image tampering. It is a passive image authentication approach, which checks image integrity in the absence of any active authentication code. For a given digital image, the distortions introduced during image acquisition are used as a "natural authentication code" to detect image integrity. A general framework of digital image forensics is proposed which is based on measuring the image quality inconsistencies of JPEG blocking artifacts and sharpness. To measure the quality inconsistencies, we propose to estimate blocking artifacts caused by JPEG compression based on quantization table estimation, and to measure the image sharpness based on the normalized Lipschitz exponent of wavelet modulus local maxima. Discovery of more quality measures related to distortions by image acquiring and operations may also be useful for image forensics. Experimental results have shown the effectiveness of the proposed measures in detecting quality inconsistencies for exposing digital forgeries.

5.1 Detecting Digital Forgeries by Measuring Image Quality Inconsistency

Digital image forensics is an alternative solution to the active image authentication based on signature or watermark. Several statistical techniques have been proposed to detect the traces of specific manipulation applied to the image, such as detecting the resampling [44], copy-paste [17], JPEG recompression [18], and color filter array interpolation [45, 47]. Image forensics can also be based on the natural scene statistics [52], lighting direction inconsistencies [55], or camera response normality [43]. All these approaches are effective in some aspects, but new approaches are still desirable for practical applications.

We propose to detect forgeries by measuring the image quality inconsistencies based on distortions introduced during image acquisition and processing. Digital images would bear the characteristics of their acquisition devices. The image acquisition and processing would introduce some pattern distortions into the images, which can be used as an intrinsic authentication code for detecting image integrity. Therefore, measures of these distortions are potentially useful for image forensics.

There are many sources of distortions in the whole chain of images acquisition and manipulation, such as the distortion introduced by the lens, sensor noise, postprocessing distortions, and compression artifacts. General pipeline of digital imaging has been discussed in Chapter 2.2.3. The whole distortion introduced to an image can be summarized as:

$$I = S + \sum_k D_k \tag{5.1}$$

where D_k is the distortion introduced by the k-th operation during image acquiring and processing. I and S are the test image and the scene radiance, respectively.

Digital forgeries may be created in various ways, but in general, the image forgery creation process involves selection, transformation, composition of the image fragments, and retouching of the final image. Although these manipulations are often imperceptible to the human eye, they may disturb the intrinsic image quality consistencies, which can be used as the evidence of existence of digital tampering. A digital forgery composed using different sources of images will inherit different image qualities, which deduces quality inconsistencies within different parts of the image. The quality inconsistencies in the image would be a good hint that it is a forgery.

Our assumption is that if the image is authentic, then different regions of it should be consistent. Therefore, if the image qualities of some regions are abnormal or inconsistent with the others, then the image may be a forgery. Given two regions R_1 and R_2 in the image, if they come from two different sources, they will have different type of distortions $D_1(k)$ and $D_2(k)$:

$$\begin{cases} R_1 = S_1 + \sum_k D_{1k} \\ R_2 = S_2 + \sum_k D_{2k} \end{cases} \tag{5.2}$$

The quality differences could be detected by eliminating the image content (S_1 and S_2, respectively). Then the difference between these two regions can be calculated by the distance of the distortions:

$$diff(R_1, R_2) = d(\sum_k D_{1k}, \sum_k D_{2k}) \tag{5.3}$$

Image quality measures are figures of merit used for the evaluation of imaging systems or coding/processing techniques. We consider several image quality metrics and study their statistical behavior when measuring various distortions. A good objective quality measure should well reflect the distortions on the image.

Detecting image quality inconsistencies requires appropriate models of distortions introduced in the whole chain starting from acquisition to the final representation of the

image. However, the total effect of the distortions during image acquisition and processing includes various sources, and would interfere with each other. For example, if Gaussian noises are added into a JPEG compressed image, the compression artifacts of the image would change a lot and cannot be detected correctly. On the other hand, perfect detection of every distortion is very difficult, if it is not impossible, without the original scene *S*. However, it is still possible and feasible to extract some features from each distortion based on the prior knowledge of the processing characteristics.

After the distortions from image regions are measured, the distance of them can be used to determine the image authenticity. A typical authenticity verification decision rule is made by:

$$diff(R_1, R_2) \underset{H_0}{\overset{H_1}{\gtrless}} T \tag{5.4}$$

To check the integrity of an image, we segment it into areas and then check the quality consistency of these segments. Suspicious area is selected as R_1 for evaluation, and other areas are grouped as R_2. If quality inconsistencies are detected, the image is deemed suspicious. By detecting the inconsistency of the segments, we could possibly tell whether or not the image is from one simple shot of one camera. For example, if the image contains a segment with high level of blocking artifacts, but others contains no or low level of blocking artifacts, then the image has high probability of being a forgery. Tools for detecting specific manipulations can be applied to each segments for detect quality measure. We have used blocking artifacts and sharpness as the measure to check the image quality consistency for image authentication. Other distortions can also be exploited, such as CCD noise, ringing artifacts caused by JPEG2000 compression, non-linear distortion caused by the lens, and color distortions caused by color array interpolation.

5.2 Detecting Image Quality Inconsistencies based on Blocking Artifacts

JPEG image format is popularly used in most digital cameras and image processing software. Usually JPEG compression would introduce blocking artifacts. Manufacturers of digital cameras and image processing software typically use different JPEG quantization table to balance compression ratio and image quality. Such differences will also cause different blocking artifacts in the images acquired. When creating a digital forgery, the resulted tampered image may inherit different kind of blocking artifacts from different sources. These inconsistencies, when detected and measured, are used to check image integrity. Besides, forgeries creation process would also change the blocking artifacts, because the blocking artifacts of the affected blocks will change a lot by tampering operations such as image splicing, resampling, and local object operation such as skin optimization. Therefore, the blocking artifact inconsistencies found in a given image may tell the history of the processing the image has undergone.

We present a passive way of detecting digital image forgery by measuring its quality inconsistency based on JPEG blocking artifacts. A new quantization table estimation based on power spectrum of the histogram of the DCT coefficients is firstly introduced, and the blocking artifact measure is calculated based on the estimated table. The inconsistencies of the JPEG blocking artifacts are then checked as a trace of image forgery. Our proposed approach is able to detect spliced image forgeries using different quantization table, or forgeries which would result in the blocking artifact inconsistencies in the whole images, such as block mismatching and object retouching. In addition, our proposed quantization table estimation algorithm is much faster than maximum likelihood based methods [104, 105].

5.2.1 Blocking Artifacts Caused by Lossy JPEG Compression

JPEG compression is a block DCT-based image compression technology. A simplified pipeline of JPEG compression is illustrated in Figure 5.1. The image is firstly converted from *RGB* into a different colour space *YCbCr*. The *Y* component represents the brightness of a pixel. The *Cb* and *Cr* components together represent the chrominance. Then chroma subsampling (or called downsampling) is carried out to reduce the *Cb* and *Cr* components. Each component (*Y*, *Cb*, *Cr*) of the image is partitioned into blocks of eight by eight pixels each, then each block is converted to frequency space using discrete cosine transform. Quantization is done by simply dividing each component in the frequency domain by a constant for that component, and then rounding to the nearest integer. Finally, entropy coding is used to compress the DCT coefficients.

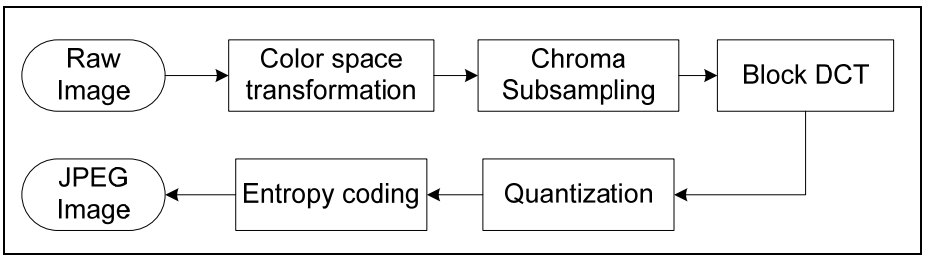


Figure 5.1: Diagram of JPEG compression

Digital cameras from different brands may use different JPEG quantization tables. Table 5.1 lists the default table of the finest quality setting of three brands cameras. The different tables used in JPGE compression brings with different quantization artifacts in the output images taken by the cameras. This intrinsic difference in the digital cameras provides us a "natural" authentication code for forensic analysis.

From the JPEG compression pipeline, we can find that the main lossy operation in JPEG compression is the quantization process. In order to achieve low bit rates, quantization is normally used to compress the transform coefficients. Since the quantization process is lossy, the compressed image exhibit blocking artifacts, which are the visual artifacts found

at block boundaries of DCT-based compressed images. The blocking artifacts are caused by amplified differences between the boundary pixel values of neighboring blocks, or by undesirable high frequency components. The degradation is a result of a coarse quantization of the DCT coefficients of each image block without taking the inter-block correlations into account.

Table 5.1: Quantization table of the finest settings for different cameras

(a) Nikon Coolpix5400

1	1	1	1	1	2	2	2
1	1	1	1	1	2	2	2
1	1	1	1	2	2	3	2
1	1	1	1	2	3	3	2
1	1	1	2	3	4	4	3
1	1	2	3	3	4	5	4
2	3	3	3	4	5	5	4
3	4	4	4	4	4	4	4

(b) Canon Ixus500

1	1	1	2	3	6	8	10
1	1	2	3	4	8	9	8
2	2	2	3	6	8	10	8
2	2	3	4	7	12	11	9
3	3	8	11	10	16	15	11
3	5	8	10	12	15	16	13
7	10	11	12	15	17	17	14
14	13	13	15	15	14	14	14

(c) Sony P10

1	1	1	1	1	2	3	3
1	1	1	1	1	3	3	3
1	1	1	1	2	3	3	3
1	1	1	1	3	4	4	3
1	1	2	3	3	5	5	4
1	2	3	3	4	5	6	5
2	3	4	4	5	6	6	5
4	5	5	5	6	5	5	5

JPEG compression introduces specific correlations in the form of blocking artifacts. When creating a digital forgery, a typical pattern is to load images into an image processing software such as Adobe Photoshop, splice them together, and then save the composite image. If the forgery comes from different JPEG compressed images, it would inherit different kind of blocking artifacts.

Digital tampering will also change the blocking artifacts. For example, the blocking artifacts of the affected blocks would be changed a lot by image splicing, resampling, and some other special manipulations such as skin optimization for portrait images. The inconsistencies of the JPEG compression artifacts would also be caused by block misaligning, which may be usual during the creation of digital forgery.

5.2.2 Blocking Artifact Measure based on Quantization Table Estimation

Blocking artifact measure plays an important role in the areas of image and video processing such as optimal bit allocation and post-processing [106]. Here we explore it to detect the image forgeries. As discussed above, the blocking artifacts are caused mainly by quantization. Therefore, the quantization table used during JPEG compression is useful for precise estimation of the blocking artifacts.

Blocking artifact for each block is estimated via:

$$B(i) = \sum_{k=1}^{64} | D(k) - Q(k) round(\frac{D(k)}{Q(k)}) | \tag{5.5}$$

where $B(i)$ is the estimated blocking artifact measure for the testing block i, and $D(k)$ is the DCT coefficient at position k. $Q(1{:}64)$ is the estimated DCT quantization table. The blocking artifact measure (BAM) for the whole image is then calculated based on the blocking artifacts of all blocks:

$$BAM = \frac{1}{N} \sum_{i} B(i) \tag{5.6}$$

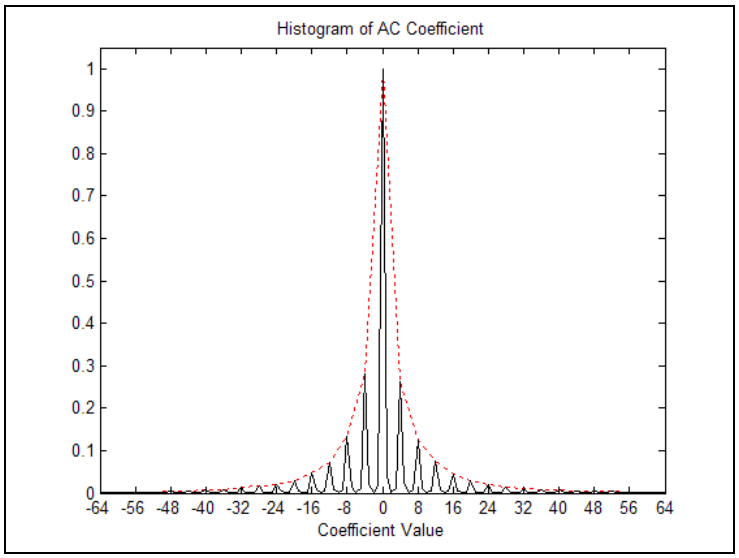
where N is the total number of image blocks.

JPEG quantization table estimation has been useful for JPEG artifact removal [107], image enhancement or JPEG re-compression. To estimate the JPEG quantization table, a method called maximum likelihood estimation (MLE) is proposed in [104, 105] based on the total estimated blocking artifact at DCT frequency i given an estimated step $Q(i)$. In [58], a statistical model based on Benford's law for the probability distribution of the first digits of the JPEG coefficients is used to estimate the JPEG quantization factor. For a given candidate $Q(i)$, a complicated maximum likelihood estimation based on quantization artifacts of the whole coefficients must be computed. Therefore, these methods are very time consuming since they are all based on exhaustive searching for estimation. In the next

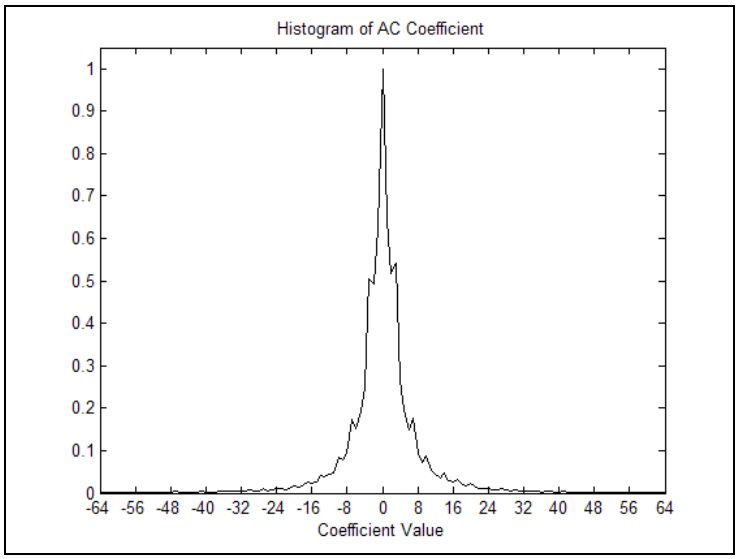
sub-section, we present a faster quantization table estimation algorithm based on histogram power spectrum of DCT coefficients.

We propose to use the power spectrum of the DCT coefficient histogram of each portion to estimate the quantization step of it. It is observed that if the histogram of DCT coefficients contain periodic patterns, then the coefficients are very likely to have been quantized with a step of this period [104]. These periodic artifacts are particularly visible in the Fourier domain as strong peaks in the mid and high frequencies. Therefore, the derivatives of the histogram power spectrum of DCT coefficients could be used to estimate the quantization table.

We firstly calculate the histogram (H) of the DCT coefficients at position i, and then calculate the histogram power spectrum (P) using the Fast Fourier Transform (FFT). The second order derivative (S) of P is then low-pass filtered. Finally, the number of negative local minimum of S is found to be ($Q(i)$-1). Suppose $f(x)$ is a function of x that is twice differentiable at a stationary point x. A stationary point may be a local minimum, maximum, or inflection point. If $f'(x)=0$ and $f''(x)<0$, then has a relative maximum at x. For this aim, we filter S to get a more clear pattern for calculating $Q(i)$, by eliminating the positive values and low pass filtering.



(a) Histogram of AC Coefficient (q=4)

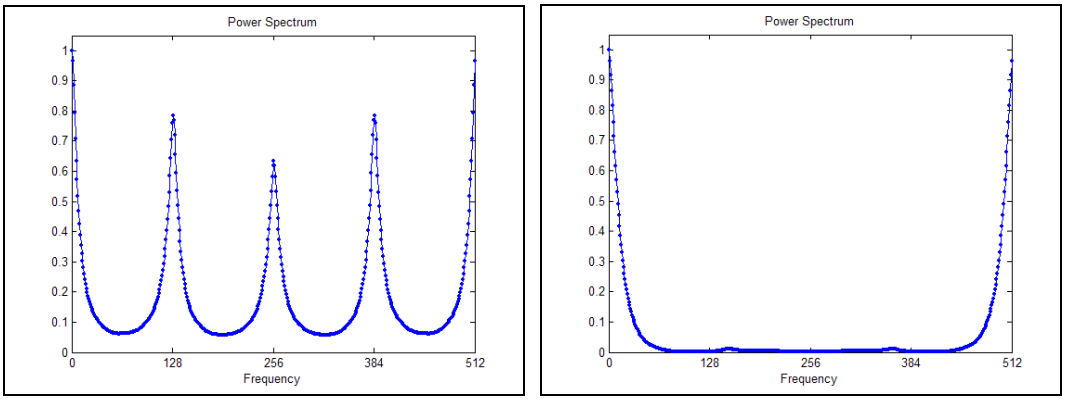


(b) Histogram of AC Coefficient (q= 1)

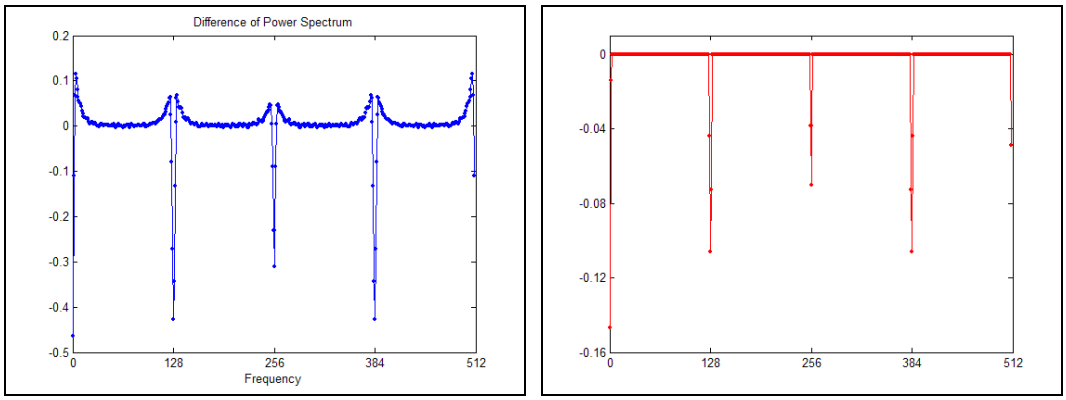
Figure 5.2: Histogram of DCT coefficients

Figure 5.2 shows the DCT coefficient histograms for a JPEG compressed image and the original uncompressed image. The histogram was calculated for the coefficients at DCT frequency $Q(6)$ (the fifth AC component). We can see that these two histograms are very similar, except that for JPEG compressed image (Figure 5.2a) there are some peaks at the positions of multiple of q (here q=4). Note also that this type of period is not present in the uncompressed image (Figure 5.2b).

The periodic peaks in the histogram are particularly visible in its power spectrum as strong peaks (Figure 5.3). In Figure 5.3, we show the power spectrum of the histograms of an uncompressed image and the JPEG compressed image. The histogram (H) of the DCT coefficients at position i is firstly calculated, and then the histogram power spectrum (P) is achieved using the Fast Fourier Transform.



(a) Power Spectrum of Histogram (q=4) (b) Power Spectrum of Histogram (q= 1)



(c) 2nd order Difference of Power Spectrum (q=4) (d) Filtered 2nd order Difference (q=4)

Figure 5.3: Power spectrum of DCT coefficient histogram

The estimated power spectra of the two images are shown in Figure 5.3(a) and Figure 5.3(b), respectively. A combined view of those peaks provides us with a clear view of the quantization step used. The second order derivative (S) of P is then low-pass filtered, and the positive values are eliminated. The reason is that if there is a local maximum $f(x)$ at x, then the derivatives $f'(x)=0$, $f''(x)<0$, and $f''(x)$ is a local minimum. The number of negative

local minimum of S is found to be equal to ($Q(i)$-1). The second differentiable S and its filtered version are shown in Figure 5.3(c) and Figure 5.3(d), respectively.

We use the whole image to estimate quantization table. The procedure of the quantization table estimation is: (1) Calculate DCT coefficients of each 8×8 image block; (2) Calculate the power spectrum (P) of the histogram of DCT coefficients for each of the 64 frequencies; (3) Calculate the second derivative of P, and then low-pass filtering it; (4) Calculate the local minimum number (Num) of the filtered second derivative of P; (4) the estimated quantization step of the DCT frequency is estimated as Num+1.

Our proposed blocking artifact estimation algorithm is faster than that of [104, 105, 58]. The reason is that for a DCT coefficient at a given position, the algorithm proposed in [104, 105] need to calculate the blocking artifact for every possible Q based on the DCT coefficients of the whole image. On the contrary, our algorithm only computes once. Furthermore, our algorithm can estimate arbitrary quantization table which is often adopted in different brand of digital camera, whereas the algorithm proposed in [58] can only detect a standard compression factor, since it re-compress the image by a sequence of preset Q-factors. This step also makes the algorithm in [58] slower than our proposed one.

5.2.3 Detection of Quality Inconsistencies based on Blocking Artifact Measure

Given a digital image, the blocking artifacts introduced during image compression could be used as a "natural authentication code" to check its integrity. We observed that when an adversary forges an image, JPEG images from different sources such as different digital cameras or by different manipulations are usually used. Such forgery usually makes the blocking artifact measurements inconsistent. Therefore, image forgeries could be done by finding the inconsistencies of blocking artifacts.

To check the integrity of an image, we first segment it into areas and then check the blocking artifact consistency of these segments. Suspicious areas are selected for evaluation, the other areas are used to estimate the quantization table, and the *BAM* of the image is calculated based on the estimated table. If the blocking artifact inconsistencies are detected, the image is deemed suspicious. By detecting the inconsistency of the segments, we could possibly tell whether or not the image is from one simple shot of one camera.

5.2.4 Experimental Results and Discussions

Our test images are photos taken by digital cameras including Nikon Coolpix5400, Canon Ixus500, Sony P10, and Canon A85, which are all commercially available cameras, with 4 or 5 mega-pixels CCD, and 3× or 4× optical lens. All photos are saved in JPEG format. The test images were taken in different time under various environments.

First of all we evaluate the DCT statistics of the digital images and also the differences within images by different cameras, as shown in Table 5.1. The quantization tables used by cameras of different brands are different.

Our proposed quantization table estimation algorithm is much faster than that of [104, 105, and 58]. Table 5.2 shows the results of quantization table estimation of our proposed method, compared with the optimization based method used in [104, 105]. The test image is *Lena* with dimension 512×512, quantized with JPEG factor from 100 to 50. These results were generated with program compiled by Visual C++ 6.0 on Dell Dimension 8250 PC (3060 MHz CPU, 512MB memory, and Windows XP operation system). From the results we can see that our method is much faster. The reason would be that for each given DCT coefficient, the algorithms in [104, 105] need to calculate blocking artifact for every possible $Q(i)$ based on the DCT coefficients of the whole image. On the contrary, our algorithm only computes CG_{NLE} once**.** Furthermore, our algorithm can estimate arbitrary

quantization table which is often adopted in different brand of digital cameras, whereas the algorithm proposed in [58] can only detect a standard compression factor, since it re-compress the image by a sequence of preset Q-factors. This step also makes the algorithm in [58] slower than our proposed one. On the other hand, the estimation errors of 64 quantization steps grow when quantization factor decreases. The reason would be that the high frequency DCT coefficients would be all zero when quantized by large step size. Therefore, we only use the first 32 DCT frequencies in blocking artifact estimation.

Table 5.2: Quantization table estimation time (ms)

Quality factor	100	90	80	70	60	50
MLE based method	15091	14957	14893	14950	14828	14737
Proposed method	241	227	228	225	222	228

We evaluated the inconsistencies of the JPEG blocking artifacts to detect the digital tampering in the image. A portrait taken from JPEG2000 test image *Woman* is extracted using Adobe Photoshop and spliced into a landscape taken by Canon Ixus500. There is typical inconsistencies in this composed forgery (Figure 5.4a). We detected blocking artifact measures for all blocks of the image (Figure 5.4b). Figure 5.4(b) shows there are actually two different types of artifacts in different areas of the image, which may denote the inconsistent areas.

We also evaluated the inconsistencies of the JPEG compression blocking artifacts due to resampling and misaligned blocks. Figure 5.5 shows a forgery photo, which was composed by two photos taken by the same cameras under the same conditions. Due to resampling and misaligning, the DCT coefficients of the spliced regions are shuffled, and no period bumps occur in the histogram power spectrum any more.

Some other tampering operations during composing forgery will also render some inconsistencies. Figure 5.6 shows a typical face skin optimization operation widely used by photographers. The face is polished using Photoshop *blur* tool. The forgery shown in Figure

5.6(b) can be detected by our technique, with a blocking artifact measure of 45.419, but only 5.854 for the original untouched photo.

(a) tampered image

(b) blocking artifact detected

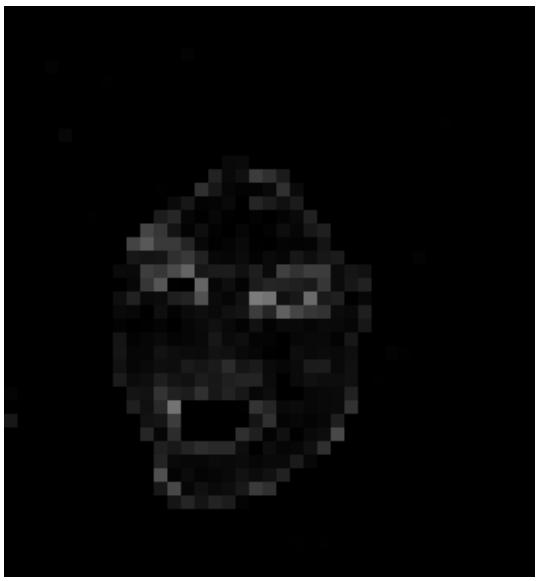
Figure 5.4: Forgery from two images by different sources (spliced from JPEG2000 image and photo by Canon Ixus500)

(a) tampered Image (a pile of stones added)

(b) blocking Artifact Detected

Figure 5.5: Forgery from two images by the same camera (Nikon Coolpix5400)

(a) Original image Cropped; (b) Face skin optimized;



(c) Blocking artifact detected

Figure 5.6: Face skin optimized detection

We also test our algorithm with various photos, 400 taken from Nikon Coolpix 5400, and 100 from Sony P10. We generate a tampered photo by randomly selecting another photo and splicing it into to the original one. The detected artifact measures are shown in Figure 5.7. For untouched photos, the artifact measures are all smaller than those of the spliced images.

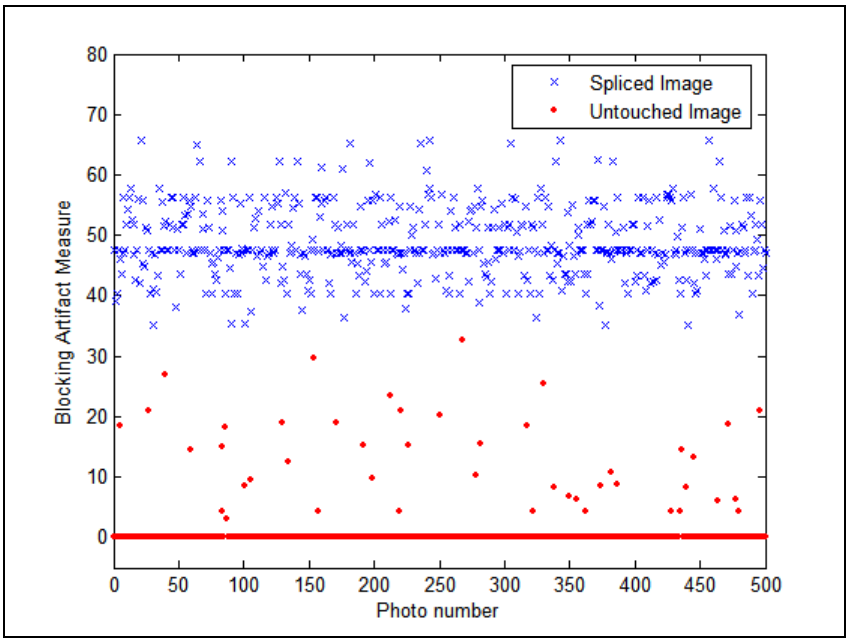
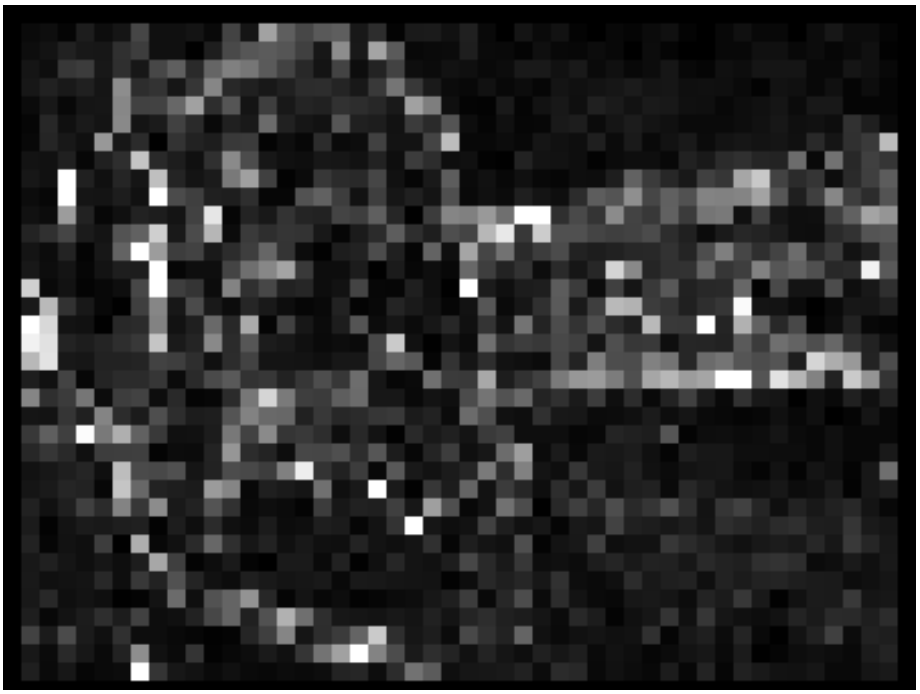


Figure 5.7: Measures for tampered or authentic images

As a passive way of image authentication, in some cases, quality based image forensics would fail to detect the forgeries. When the quality of the tampered image is very low (for example, the image has been highly compressed or reduced to small dimension), it is difficult for image forensics to detect the trace of the tampering based on image quality. For example, in Figure 5.8(a), we reduced the dimension of the tampered image shown in Figure 5.4(a) from 2592×1944 to 400×300, and compressed it with JPEG compression factor 20. The blocking artifacts in Figure 5.8(a) can be found easily, especially on the high contrast areas. Figure 5.8(b) shows the detected blocking artifacts, in which there are no clear areas identified with different class of blocking artifact measures. That is to say, in this case, blocking artifact based image forensics would fail. The reason may be that the low image quality has concealed the traces, the inconsistencies of blocking artifacts, of the tampering. For low quality images, new forensics techniques may be required, such as image forensics based on high level content consistency analysis.

(a) tampered image with low quality (downsampled and highly compressed)



(b) detected blocking artifact measure

Figure 5.8: Failure example: tampered image with low quality

5.3 Sharpness Measure for Detecting Image Quality Inconsistencies

Sharpness is a very important photographic image quality factor. It is defined by the boundaries between zones of different tones or colors. Sharpness is affected by:

- Lens: design and manufacturing quality, focal length, aperture, and distance from the image center;
- Sensor: pixel count and anti-aliasing filter;
- Shooting variances: camera shaking, focus accuracy, lighting, and atmosphere disturbances (thermal effects and aerosols).

One way to measure sharpness is in pixel domain, including analysis of statistical properties and correlation between pixels. In addition, techniques based on image gradient and Laplacian, and which detect the slope of the edges in an image, for example, the perceptual blur metric [108]. It detects edge first, and then scans each row of the image to locate edge pixels. The start and end positions of the edge are defined as the locations of local maxima closest to edge. The edge width is calculated as the distance between the end and start position. The overall metric is calculated as the average of the edge widths or the local blur values over all edges found. The relative contrast at a given spatial frequency (output contrast/input contrast) is called the Modulation Transfer Function (MTF) or Spatial Frequency Response (SFR). MTF is widely used by photographers to evaluate the camera sharpness. However, the calculation of MTF requires an image of variously sized bar patterns of the camera, or manually select the similar pattern from the image. It is not suitable for no-reference image sharpness assessment. In addition, these pixel based approaches is sensitive to noise, and the required edge selection is complicated, and usually done in a manual way.

To get around these problems, the measurements in the frequency domain can be used [109]. An approach based on the occurrence histogram of non-zero DCT coefficients through all 8×8 blocks of the image is proposed in [109]. The assumption is that sharper edges increase the high frequency components. The blurriness metric is estimated by examining the number of coefficients that are almost always zero by counting the number of zeroes in the histogram. Note that the metric is higher for sharper images. However, sharpness metrics that use the whole frequency spectrum of the image cannot separate the sharpness information from the scene content. The sharpness metrics that use spatial gradients of the edges work only for comparisons among images of the same scene. Therefore, a content independent, no-reference sharpness metric is desirable. A sharpness measure based on Gaussian lines and edges is proposed in [110]. It locates these lines and edges in the image, and then the sharpness of these lines and edges is determined by fitting a Gaussian line or edge profile to the Gaussian derivative signature. Another feasible measure is the sharpness metric based on local kurtosis, edge and energy information [111], which is based on averaged edge profile kurtosis. This algorithm is a combination of the spatial domain edge profile acutance, and the kurtosis of the frequency spectrum algorithms.

A combined pixel domain and frequency domain approach based on edge and kurtosis of DCT has been proposed [111]. An edge profile is detected by detecting edge pixels and enclosing them with 8×8 pixel blocks. For each block, sharpness using Kurtosis of the DCT is computed. The final metric is the average sharpness of the blocks in the edge profile.

In this section, we propose to use wavelet transform based sharpness measure for sharpness inconsistency detection. The proposed measure is based on normalized Lipschitz exponent of wavelet as well as relation with module maxima, which is robust to noise without need of manual edge selection.

5.3.1 Lipschitz Exponents of Wavelet

Wavelet domain is a better choice than DCT as the frequency domain for the purpose of sharpness assessment. The reason is that the wavelet transform has the characteristic of the multi-resolution and can describe the local features of the signal both in the time and frequency domain, thus it can get the detail of the signal at the different scales [112]. Wavelet can detect local signal singularity, which is good for edge detection. Multiscale statistics of wavelet can give theoretic explanation why wavelet can measure sharpness well. For example, the effectiveness of edge detection based on multiscale wavelet modulus maxima is validated in [100].

One signal sharp variation produces wavelet modulus maxima at different scales. The value of a wavelet modulus maximum at a scale s measures the derivative of the signal smoothed at the scale, but it is not clear how to combine these different values to characterize the signal variation. The wavelet theory gives an answer to this question by showing that the modulus maximum of each scale of the wavelet transform depends on the local Lipschitz regularity of the signal.

Edge is a significant feature for image on human vision. Image edges are counterpart to image gray singularity, and different types of edge have different singularities. A parameter that depicts singularity is Lipschitz exponent. In mathematics theory, the singularity of the signal as the sharp variation of the signal can be expressed precisely by the Lipschitz exponent.

Definition: A function $f(t)$ is said to be Lipschitz α at t_0, if and only if there exists two constants K and $h_0>0$, and a polynomial of order n (n is a positive integer), $P_n(t)$, such that for $h<h_0$:

$$\left| f(t_0 + h) - P_n(t_0) \right| \le K \mid h \mid^{\alpha} \tag{5.7}$$

The value K gives the amplitude of the sharp variation. Mallat proved [112] that if $f(x,y)$ is Lipschitz α at (x_0, y_0), then there exists a constant K such that for all point (x, y) in a neighborhood of (x_0, y_0) and any scale s:

$$\left|W_s f(x,y)\right| \le Ks^{\alpha} \tag{5.8}$$

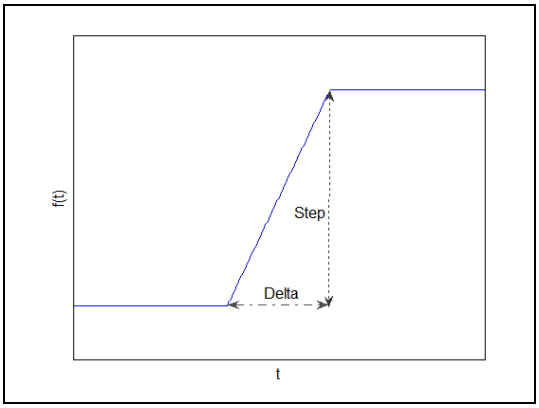
which is equivalent to:

$$\log\left|W_s f(x,y)\right| \le \log K + \alpha \log s \tag{5.9}$$

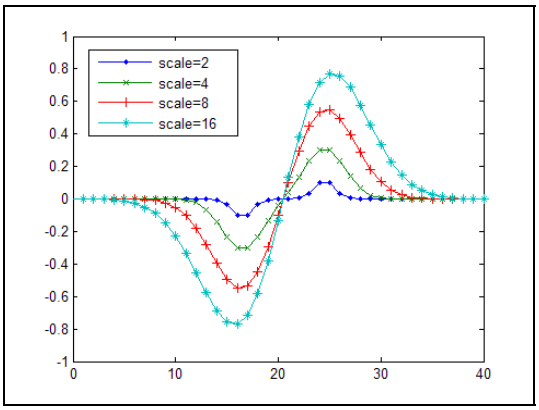
where $W_s\ f(x,y)$ is the wavelet transform of $f(x,y)$ at scale s. $\left|W_s f(x,y)\right|$ represents the modulus of $W_s\ f(x,y)$ at scale s. The Lipschitz regularity is given by the maximum slope of $\log\left|W_s f(x,y)\right|$ as a function of $\log s$ along the lines of modulus maxima that converge towards point (x,y). In the wavelet domain, it is possible to calculate the Lipschitz exponent in a certain point in the image from the evolution of the modulus maxima of the wavelet coefficients corresponding to that point through successive scales.

5.3.2 Normalized Lipschitz Exponent (*NLE*)

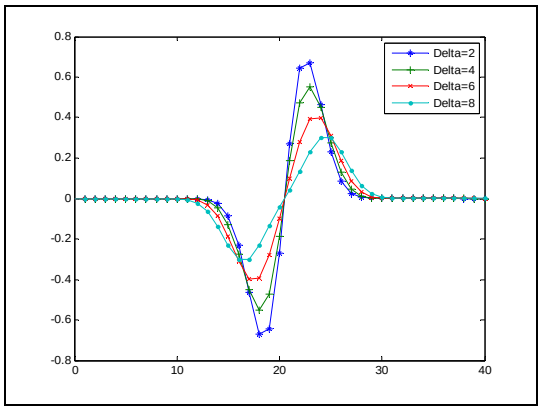
Lipschitz exponent has been used to estimate the blurness in an image [113], but Lipschitz exponent alone is not adequate. The reason may be that Lipschitz exponent only describes the singularity of the signal. For sharpness evaluation, the amplitude of the sharp variation should also be considered. For example, the signal in Figure 5.9(a) with different *Delta* and *Step* settings will have a unique Lipschitz exponent 1. However, for sharpness evaluation, different setting of *Delta* and *Step* would cause different sharpness effect. From the wavelet values of different *Delta* or *Step* (Figure 5.9c and Figure 5.9d), we can find that the local maxima change linearly with *Delta* or *Step*. Therefore, for sharpness evaluation, the amplitude of the wavelet coefficients should be used together with Lipschitz exponents.



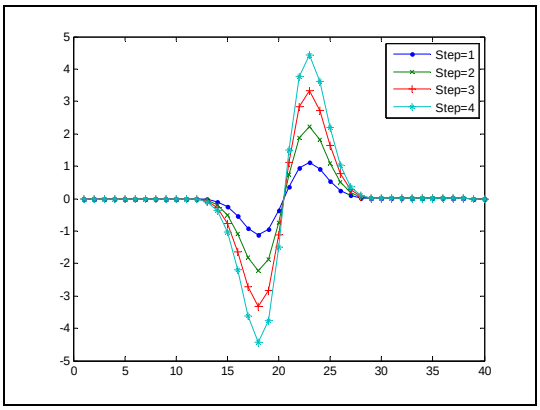
(a) ideal signal that is differentiable once



(b) different scale wavelet transform



(c) wavelet transform of different *Delta*



(d) wavelet transform of different *Step*

Figure 5.9: Multiscale wavelet modulus maxima for different sharp edges

Therefore, we propose to use normalized Lipschitz exponent (*NLE*) as a measure of how sharp the image is at a certain point. The *NLE* is defined as $\alpha / \log K$, and then Equation (5.9) becomes:

$$\frac{\log\left|W_s f(x,y)\right|}{\log K} \le 1 + \frac{\alpha}{\log K}\log s = 1 + NLE \log s \qquad (5.10)$$

Our method for sharpness estimation is based on estimating the *NLE*s of the sharpest edges in the image. To analyze the edges in the image, we calculate the *NLE* in all points where a change in intensity is found either in the horizontal or vertical direction. In the wavelet domain, it is possible to calculate the *NLE* in a certain point in the image from the evolution of the modulus maxima of the wavelet coefficients corresponding to that point through successive scales.

5.3.3 Wavelet *NLE* based Sharpness Measure

The procedure of the proposed sharpness measure can be described as follows: (1) The wavelet decomposition of the image is calculated; (2) the modulus maxima of the wavelet coefficients corresponding to a certain point in the image through different resolution scales are detected; (3) the normalized Lipschitz exponent in that point is calculated by nonlinear fitting an exponential curve to the modulus maxima versus the scale; (4) from the Lipschitz exponents found along the significant edges in the image, a histogram is achieved; (5) the center of gravity (CG) of the histogram is related to the sharpness of the image. The sharpness measure is calculated based on the estimated *CG* of the *NLE*s with sigmoid function.

Continuous Wavelet Transform

We use 2-dimensional Gaussian function $\theta(x, y)$ for wavelet function generation:

$$\theta(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \tag{5.11}$$

$\theta_s(x, y)$, the $\theta(x, y)$ at scale s, is defined as:

$$\theta_s(x,y) = \theta(\frac{x}{s}, \frac{y}{s}) \tag{5.12}$$

The first differentiation of $\theta(x, y)$ in the x, y direction is used to be two wavelet functions:

$$\varphi_s^{(1)}(x,y) = \frac{\partial \theta_s(x,y)}{\partial x} = -\frac{x}{2\pi s} e^{-\frac{x^2+y^2}{2s^2}} \tag{5.13}$$

$$\varphi_s^{(2)}(x,y) = \frac{\partial \theta_s(x,y)}{\partial y} = -\frac{y}{2\pi s} e^{-\frac{x^2+y^2}{2s^2}} \tag{5.14}$$

Therefore, for a 2-dimensional function, in the scale *s*, its two fractions of the wavelet transform are:

$$W_s^{(1)} f(x,y) = f(x,y) * \varphi_s^{(1)}(x,y) \tag{5.15}$$

$$W_s^{(2)} f(x,y) = f(x,y) * \varphi_s^{(2)}(x,y) \tag{5.16}$$

where * represents convolution function. The modulus of the wavelet transform is:

$$W_s f(x,y) = \sqrt{|W_s^{(1)} f(x,y)|^2 + |W_s^{(2)} f(x,y)|^2} \tag{5.17}$$

Local Maxima Detection

The former computation is provided for the edge point of the image signal. The first step is to find the edge of the image. From Equation(5.15) and Equation(5.16), we know that the *x* and *y* fractions of the wavelet transform are the image's gradients, and that the modulus maximum of the wavelet transform is the image's edge point. Considering the very smoothing area of the image will also produce the modulus maximum point with small value which will result in computational error easily. Therefore, a template of edge is used to remove those false edges. Those points are selected as edge points, which are larger than half of the local maxima and larger than the average wavelet modulus at every scale. Because we restrict the Lipschitz exponents to those corresponding to transitions with large amplitude, we already selected the sharpest transitions with large amplitudes in the image.

Normalized Lipschitz Exponent Estimation

In actual computation, the edge points' smoothing factors are different, thus the statistical histogram method is used. Four scales (1, 2, 3 and 4) of wavelet transform are used, since the first 4 scales of wavelet transform carry enough information about the character of local maxima which can be validated by experiments. The normalized Lipschitz exponents

(*NLE*s) of the selected local maxima are estimated by linear least-squares data fitting. The number of the histogram bins is 100, and the Centre of Gravity (*CG*) of the histogram is related to the sharpness of the image.

Sharpness Measure

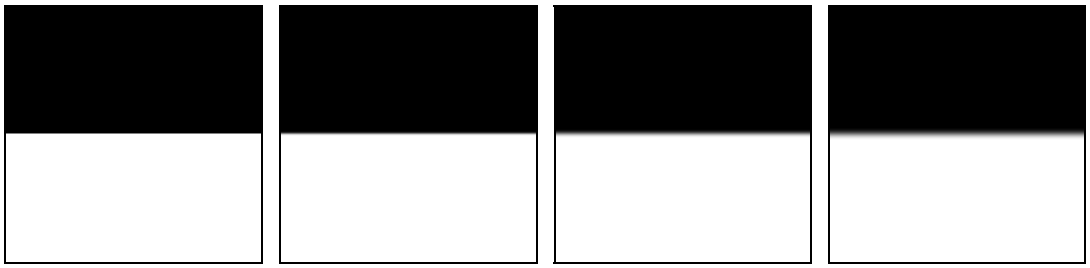
The final sharpness measure is then calculated with the sigmoid function based on the estimated *CG* of the *NLE*s (CG_{NLE}):

$$spn = a\ \text{sigmf}\,(CG_{NLE};[b\ c]) = \frac{a}{1+e^{-b(CG_{NLE}-c)}} \tag{5.18}$$

In our experiments, the parameters [a b c] are set to [2 -2 0.26] empirically.

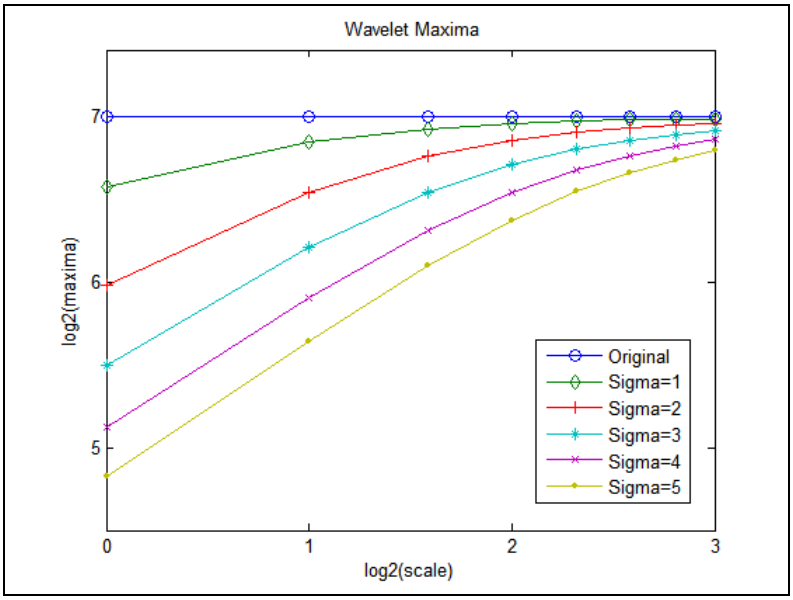
5.3.4 Experimental Results and Discussions

In order to evaluate the proposed sharpness measure, an image of step signal (Figure 5.10a) is used for testing. Its Lipschitz exponent is 0. The modulus maximum of the wavelet transform of the original and its blurred image are in the four apexes of the rectangle. We did the continuous wavelet transform with the scale from 1 to 8, obtaining the corresponding modulus maxima. The results are shown in Figure 5.11(a). The normalized wavelet modulus maxima corresponding *NLE* are shown in Figure 5.11(b).

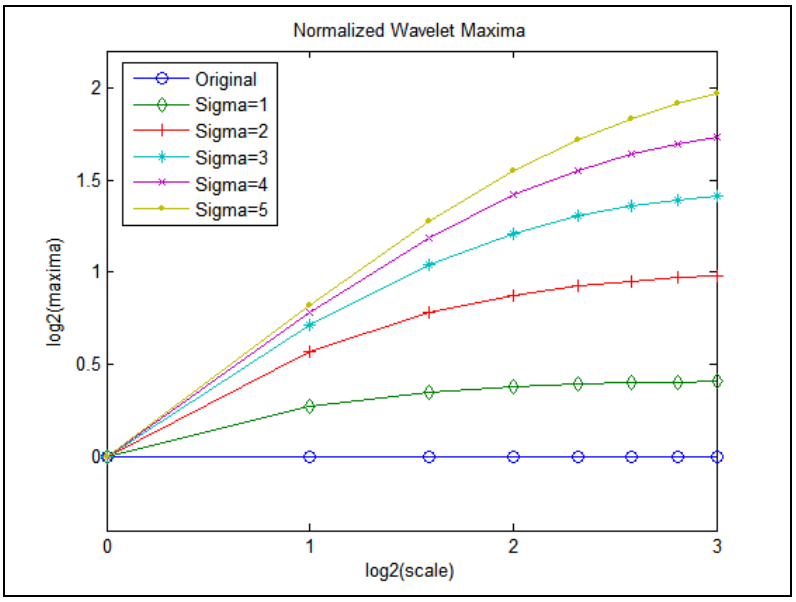


(a) image *Step* (b) blurred image with σ=1 (c) blurred image with σ=2 (d) blurred image with σ=3

Figure 5.10: Test image and its blurred versions



(a) wavelet modulus maxima of different scale



(b) normalized wavelet transform modulus maxima

Figure 5.11: Wavelet transform modulus maxima and its normalized versions

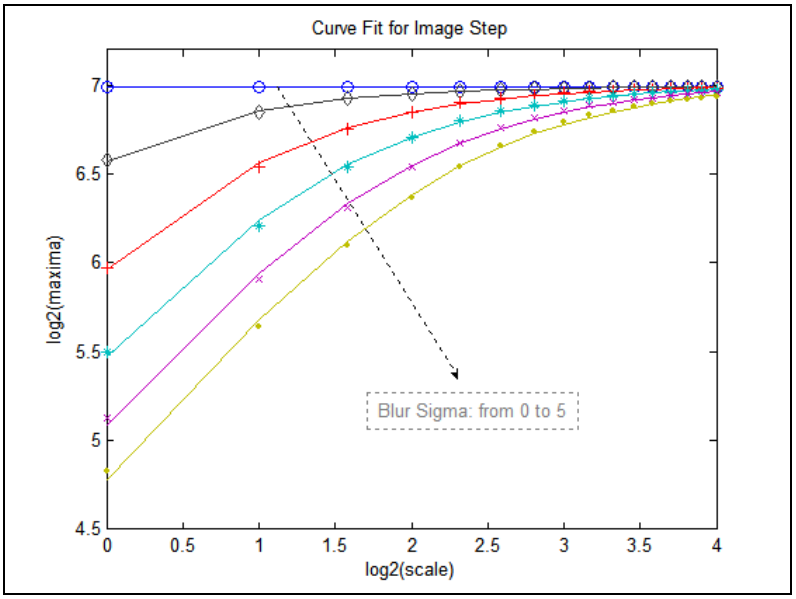
The variance of the Gaussian variance can be estimated by the following equations [112]:

$$\left|W_s g(x,y)\right| \le K \frac{s}{s_0} s_0{}^{\alpha} \tag{5.19}$$

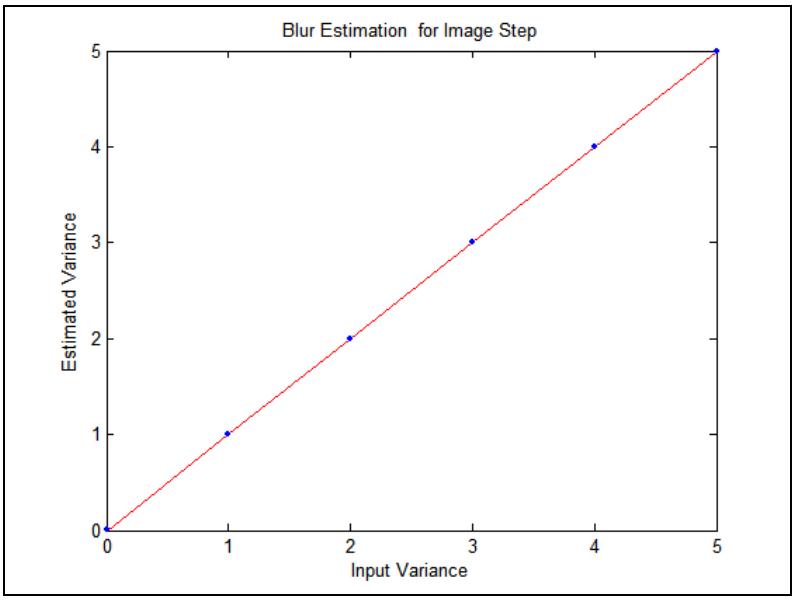
$$s_0 = \sqrt{s^2 + \sigma^2} \tag{5.20}$$

The normalized wavelet transform modulus is useful in sharpness estimation. The Gaussian variances are estimated with non-linear least-squares data fitting by the Gauss-Newton method. The results for image *Step* are shown in Figure 5.12(a) and (b). The results for real image *Lena* are shown in Figure 5.13 (c) and (d). From Figure 5.13 we can find that the model defined by Equation (5.19) and Equation (5.20) is perfect for ideal step signal, but not so well for real image *Lena*. The reason may be the blurness inherent in the real images which could not be modeled as Gaussian.

In Figure 5.14 we show the results of the performance of the proposed *NLE*, compared with the original approach using Lipschitz exponent only. The histograms of the Lipschitz exponent α, K and *NLE* of image *Lena* and its blurred versions (with variance σ equal to 1, 2 and 3) are shown in Figure 5.14 (a), (b) and (c), respectively. The histograms of α of different blurred images have different shapes. The blurred image with larger σ has a smaller K on average than that with smaller σ. After normalization, the histograms of *NLE* are much more regular than those of α. Therefore, *NLE* is better in distinguishing different blurred images than Lipschitz exponent α.

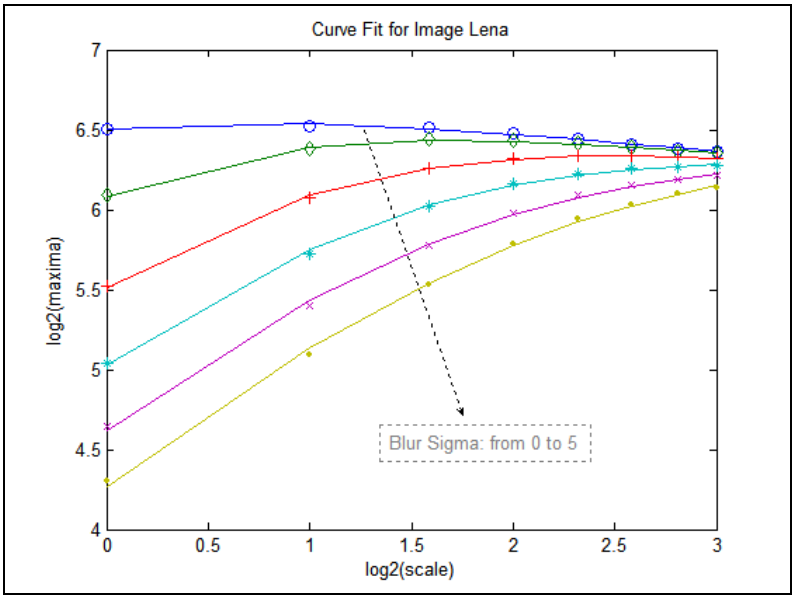


(a) curve fitting for image *Step*

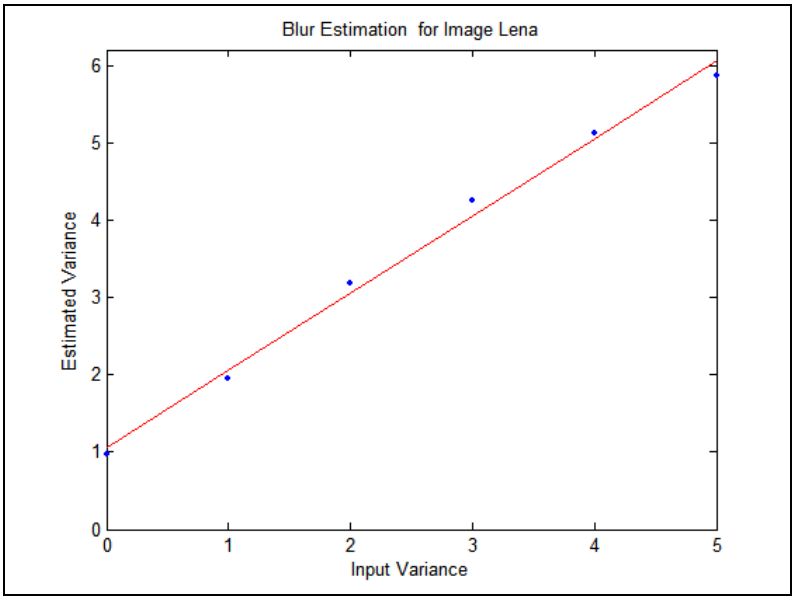


(b) blur estimation of image *Step*

Figure 5.12: Results of Gaussian blur estimation for ideal step signal

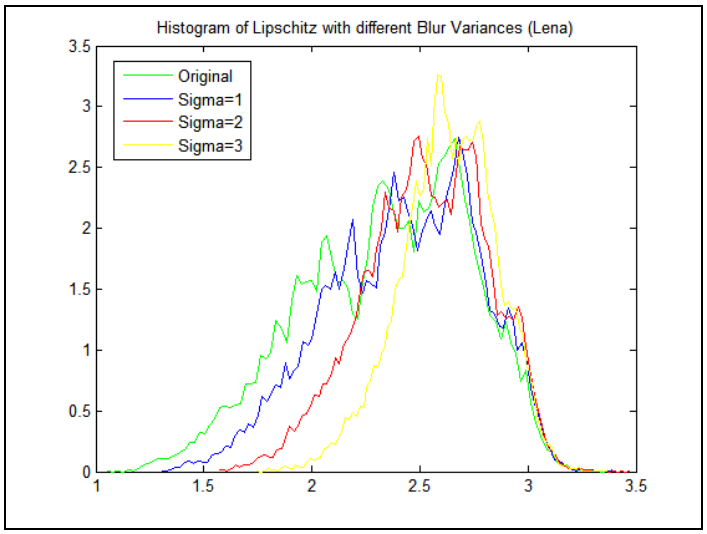


(a) curve fitting for image *Lena*

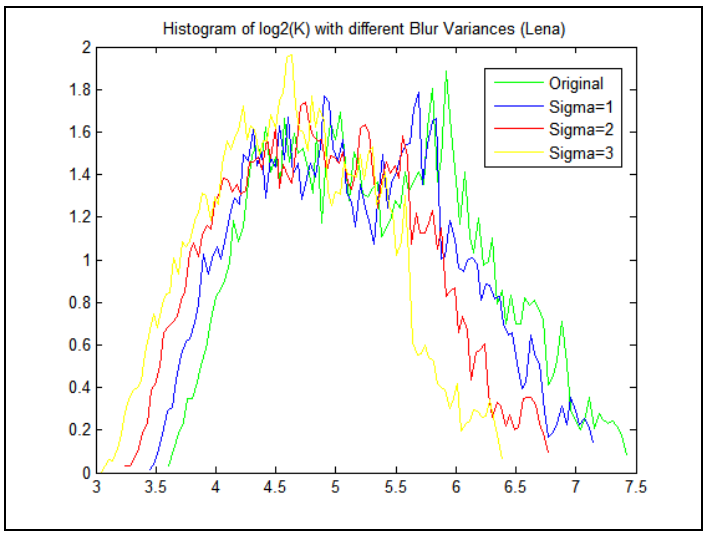


(b) blur estimation of image *Lena*

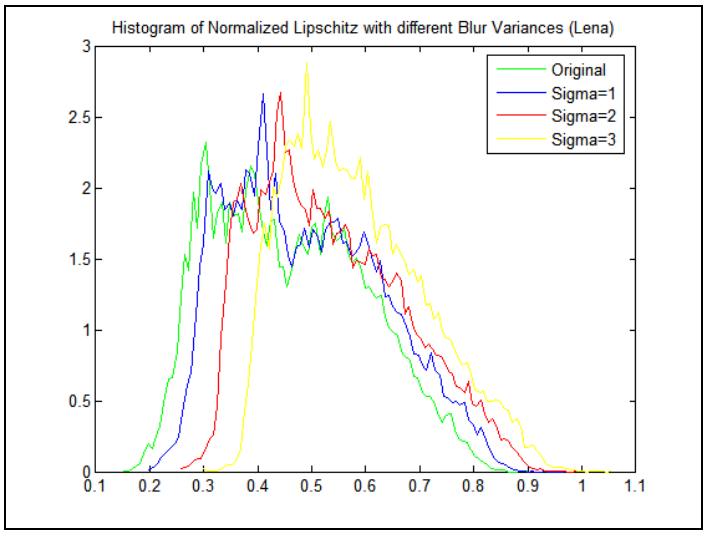
Figure 5.13: Results of Gaussian blur estimation for real image *Lena*



(a) histograms of estimated Lipschitz α

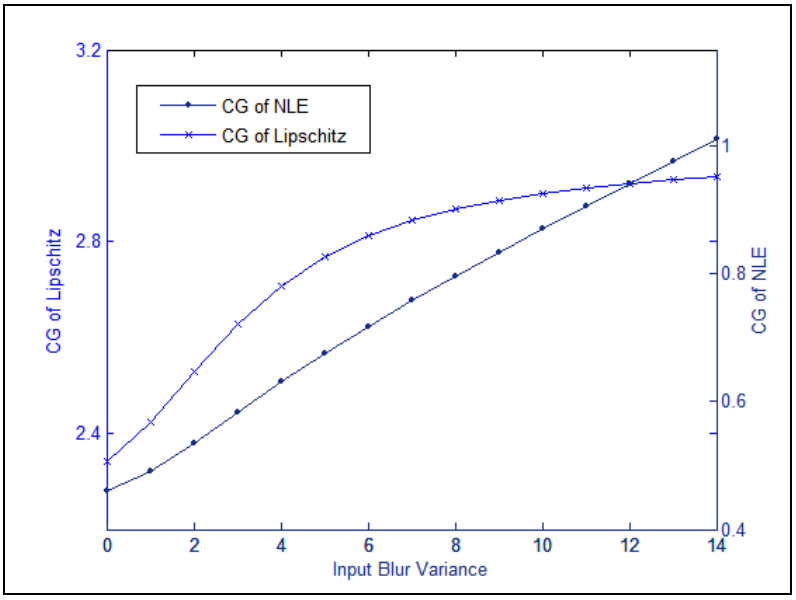


(b) histograms of estimated K

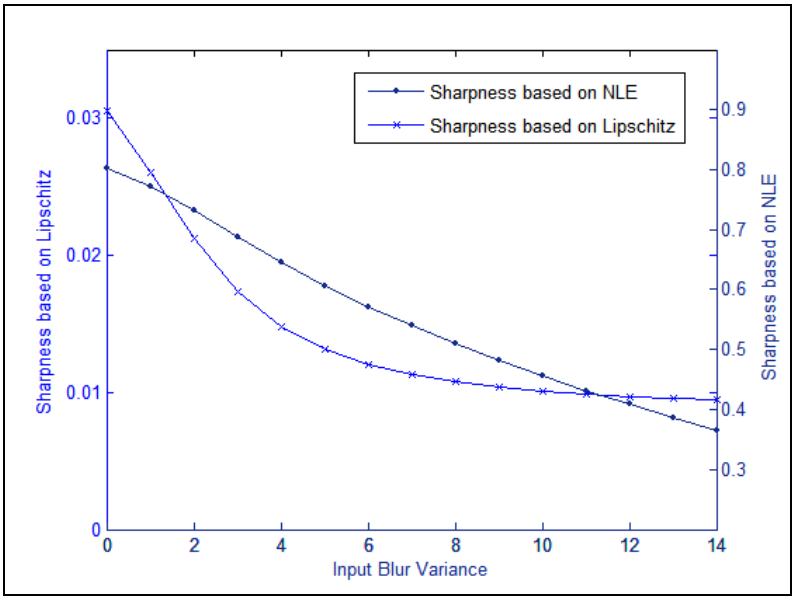


(c) histograms of estimated *NLE*

Figure 5.14: Histogram of Lipschitz α and K for image *Bike* with different blurs



(a) CG of Lipschitz exponent and *NLE* of blurred images



(b) Sharpness estimation for blurred images

Figure 5.15: Comparisons of α and *NLE*

We further tested our algorithm with larger σ of Gaussian blurring, ranging from 0 to 14. The detected CG of Lipschitz exponent and that of *NLE* are shown in Figure 5.15(a). When input σ increases, the CG of α increase nearly linearly when is σ small. But when σ gets larger and larger, the increasing rate of α becomes smaller. Therefore, the whole curve of estimated α is not linear. On the contrary, the curve of the estimated *NLE* is almost linear.

The estimated sharpness measures by Equation (5.18) are shown in Figure 5.15(b). Similarly, the curve of sharpness based on *NLE* is much more linear than that based on the Lipschitz exponent. These results validate that *NLE* is more suitable in evaluation of image sharpness than Lipschitz exponent.

5.4 Summary

The proposed passive approach to detect digital forgery by checking image quality inconsistencies is able to distinguish digital forgeries from authentic images in the absence of any digital watermark or signature. Image quality inconsistencies based on JPEG blocking artifacts and image sharpness are successfully detected as possible evidences that the image had been tampered with. The results support the hypothesis that image quality inconsistencies could serve as a useful intrinsic signature for revealing traces of digital tampering. This is attributed to the observation that a digital forgery composed of different sources of images usually contains quality inconsistencies introduced by forgery creation operations. The proposed image quality inconsistency based image forensics technique provides a passive approach for image authentication, and its development may help provide a better understanding of the role of image quality for detecting digital forgeries. A limitation of the proposed approach is that it is not suitable for low quality images. Further work would be needed to develop forensics techniques to check image content consistency of low quality images, or to discover more quality measures for detecting digital forgeries.

Chapter 6

Conclusions and Further Work

6.1 Conclusions

The main purpose of the work presented in this thesis is to protect the trustworthiness of digital images, either actively when the received image is damaged by transmission errors or passively when there is no side information available. The purpose was achieved by exploiting the statistical and spatial properties of features to authenticate damaged images, and by measuring image quality inconsistencies to detect digital image forgeries passively. This chapter concludes the results of the research work present in the previous chapters, and some areas of future work are suggested.

6.1.1 Error Resilient Image Authentication

An error resilient image authentication scheme has been developed for JPEG images, which incorporates watermarking, ECC, and error concealment into traditional crypto signature scheme to enhance the system robustness. Pre-processing and block shuffling techniques are adopted to stabilize the features for signature generation and verification. This scheme correctly could distinguish JPEG images damaged by lossy transmission from malicious forgeries, which has been validated by our experimental results. It is only designed for images using small block-based coding (8×8 DCT transform in JPEG images). Therefore, an improved scheme using a feature distance measure named statistics and spatiality based measure (SSM) has also then developed. This improved scheme is robust to transmission

errors in images received by lossy transmission. It is not constrained to block-based coded images, and then it is suitable for both JPEG and JPEG2000 images.

The proposed error resilient schemes improve the trustworthiness of the images damaged by transmission errors, by providing solutions to verify their authenticity even if there are uncorrectable errors. Many acceptable manipulations, which were incorrectly detected as malicious modifications by other schemes, were correctly verified by our scheme in our experiments. These results support the observation that the feature difference patterns under typical acceptable image modifications or malicious ones is distinguishable. The results may indicate that the statistical and spatial properties of the image feature are useful in distinguishing acceptable image manipulations from malicious content modifications.

The proposed SSM would improve system performance for content-based authentication schemes which use features containing spatial information, such as edge [7, 13], block DCT coefficients based features [8, 14, 15], highly compressed version of the original image [9], or block intensity histogram [16]. Furthermore, the proposed error resilient scheme based on SSM can improve the trustworthiness of digital images damaged by transmission errors by providing a way to distinguish them from digital forgeries. The images damaged by transmission error can be well error-concealed by the proposed error concealment algorithms, and can be verified by the proposed schemes. Therefore, the damaged images can now be with good quality, and those images that pass the verification are believable. Moreover, the results would lead to a better understanding of the role of image feature statistics and spatial properties for detecting digital forgeries.

6.1.2 Image Forensics based on Image Quality Inconsistencies

Detection of digital forgery without assistance of signature or watermarking is an emerging research task. In this thesis, an image forensics technique has been proposed to detect digital forgeries by checking image quality inconsistencies. It aims to distinguish digital forgeries from authentic images in the absence of any digital watermark or signature. This scheme is based on image inconsistencies using blocking artifact measure and sharpness measure. It can detect digital forgeries if the forgery image is a composite from different sources, or there is resampling, sharpness related operations during forgery construction. In our experiments, image quality inconsistencies based on JPEG compression blocking artifacts and sharpness measures were successfully detected as possible evidences when the image had been tampered with.

The proposed image forensics technique provides an approach for passive image authentication, which makes digital images more trustworthy. Its development may help provide a better understanding of the role of image quality in digital image forensics. The experimental results support the hypothesis that image quality inconsistencies could serve as a useful signature for revealing traces of digital tampering. This may be attributed to the observation that a digital forgery composed of different sources of images usually contains quality inconsistencies introduced by forgery creation operations.

6.2 Summary of Contributions

We describe active and passive approaches for image authentication to protect digital image trustworthiness. These approaches work in active way based on hybrid digital watermark and signature, or in the complete absence of any digital watermark or signature. They

provide a solution of error resilient image authentication and image forensics by exploring the role of image properties or quality measures in detecting digital forgeries. All these techniques have been validated by our experimental results. In summary, the work described in this thesis made the following contributions:

- Unique error resilient image authentication schemes for images transmission over lossy channels. These schemes can authenticate images correctly even if there uncorrectable transmission errors. That is, these schemes can distinguish those images damaged by transmission errors or distorted by some acceptable manipulations from forged images. (Chapter 3 and 4)
- Feature distance measure for content-based image authentication that can improve the performance of image authentication by improve it robustness. This measure is based on statistical and spatial properties of the image feature. (Chapter 4)
- Error concealment techniques for JPEG and JPEG2000 images. They can improve qualities of those images damaged by acceptable errors, which can improve their authenticities and make them more distinguishable from forgeries. (Chapter 3 and 4)
- Image forensics scheme based on measuring quality inconsistencies. It provides a passive way to check the integrity of digital images, and can be extended by using more no-reference quality measures. (Chapter 5)
- Blind measure of blocking artifacts caused by JPEG compression and image sharpness measure based on wavelet Lipschitz. These no-reference measures are useful in detecting quality inconsistencies for image forensics. (Chapter 5)

6.3 Future Work

Seeing may be believing again in the future with the well-developed image authentication techniques. In order to achieve this vision, a lot of works are still required to be done in the future. Possible directions would include robust image authentication that can distinguish acceptable manipulations from malicious contend modification, and passive image forensics tools based on other image quality measure or natural scene statistics.

A limitation of the proposed feature distance measure for content based image authentication is that it is suitable only for schemes using features containing spatial information since it is based on statistical and spatial properties of the feature differences. Further work would be needed to expand the use of the proposed measure by exploiting new discernable patterns of feature differences when the features contain no spatial information. Furthermore, many active image authentication schemes reject manipulations that may preserve better perceptual quality or semantic meaning than acceptable manipulations. Lack of a clear-cut distinction between acceptable and malicious modifications make it difficult to accurately distinguish acceptable manipulations from malicious ones. To be robust to acceptable modifications yet sensitive to malicious content modifications, additional work could be done to extract features that adequately describe the perceptual content of the image signal, or to design feature distance measure that exploits statistics or perceptual properties of image signals.

On the other hand, the proposed image quality based passive image authentication supports our idea of assessing image authenticity by checking quality inconsistencies. This thesis has proposed blocking artifact and sharpness measures to detect image quality inconsistencies for forensic analysis. Discovery of more quality measures related to distortions by image acquiring and operations is then a promising direction of further work of the image forensics. The consistencies related to pattern noise of digital imaging devices or natural scene statistics will be useful for detection of any tampering. The reason is that

the process of creating a forgery is complicated, which would damage the intrinsic quality consistencies of digital images. Further work on careful evaluation of how the image is acquired or tampered with would be required to discover more reliable quality inconsistency measure for image forensics.

The possible image quality measures for image forensics to be explored in future could be based on pattern noise of imaging system. There are many sources of noise in images obtained by imaging sensor, such as dark current noise, shot noise, circuit noise, and fixed pattern noise [114]. Digital images contain an inherent amount of noise that is largely uniformly distributed across an entire image. Statistical properties of the pattern noise, such as variance and kurtosis of noise distribution, may serve as an intrinsic watermark to verify image authenticity. The reason may be that the detected inconsistencies of the pattern noise would indicate that the image may be a faked image. On the other hand, when creating digital forgeries, it is common to add small amounts of localized noise to tampered regions in order to conceal traces of tampering (e.g., at a splice boundary). As a result, local noise levels across the image may become inconsistent.

Noise estimation is useful to detect forgery image regions from different ISO setting or light environment. An image is split into a number of blocks and select smooth blocks that are classified by the standard deviation of intensity of a block, where the standard deviation (σ) is computed from the difference of the selected block images between the noisy input image and its filtered image:

$$\sigma = std(\hat{N}) = std(I - F(I)) \tag{6.1}$$

where $\hat{N}$ is the estimated noise, std($\hat{N}$) is the standard deviation of $\hat{N}$, and $F(I)$ is a filtering function of image I. Several denoising filters [115, 116, 117, 118] can be used for feature extraction. Further works can be done on the selection denoising filter.

The image quality used in this thesis can be called as natural-imaging quality, which captures that the characteristics of images due to the imaging acquisition process, which for

the case of CCD camera consists of low-pass filtering, lens-distortion, color filter array interpolation, white-balancing, quantization, and non-linear transformation [66]. On the other hand, Natural Scene Statistics (NSS) studies aims to observe, discover and explain the statistical regularities in natural images [119]. NSS, being a form of natural image model, has found application in texture synthesis, image compression, image classification and image denoising. Researchers have developed sophisticated models to characterize NSS [120, 121, 122].

Image manipulations would perturb the natural images statistical properties. Images of the visual environment captured using high quality capture devices operating in the visual spectrum are broadly classified as natural scenes. Images of the three dimensional visual environment come from a common class: the class of natural scenes. Natural scenes form a tiny subspace in the space of all possible signals, and researchers have developed sophisticated models to characterize these statistics [120]. The malicious modifications will disturb these natural scene statistics, and introduce some inconsistencies into images. Discovery of how the malicious modifications disturb natural scene statistics may be useful to detect maliciously modifications. To discovery how the malicious modifications disturb the natural scene statistics is another possible solution for detect digital forgeries.

With the rapid development of digital technologies in video application, deliberate attack on valuable video is becoming easier. It is also possible to extend some techniques developed in this thesis to video authentication. In fact, some image authentication solutions can be directly employed in the frame-based video authentication if a video sequence is considered as a series of image frames [123]. Fox example, the hybrid signature and watermark authentication scheme may be useful in video authentication. The feature distance function proposed in this thesis would be helpful to improve video authentication performance. The idea of detecting digital forgeries by quality inconsistencies may also deduce possible passive video authentication techniques.

References

[1] J. P. Pickett, *The American Heritage Dictionary*, Boston, Massachusetts: Houghton Mifflin Company, Fourth edition, 2000.

[2] V. Erceg, K. V. S. Hari, M.S. Smith, and D. S. Baum, "Channel Models for Fixed Wireless Applications", *Contribution to IEEE 802.16.3*, Jul. 2001.

[3] P. Golle and N. Modadugu, "Authenticating Streamed Data in the Presence of Random Packet Loss", in *Proceedings of the Symposium on Network and Distributed Systems Security*, 2001, 2001, pp. 13-22.

[4] A. Perrig, R. Canetti, D. Song, and J. D. Tygar, "Efficient and Secure Source Authentication for Multicast", in *Proceedings of Network and Distributed System Security Symposium*, 2001, pp. 35-46.

[5] B.B. Zhu, M.D. Swanson, and A.H. Tewfik, "When Seeing isn't Believing: Multimedia Authentication Technologies", *IEEE Signal Processing Magazine*, Vol. 21, No. 2, pp. 40-49, Mar. 2004.

[6] M.L. Miller, G.J. Doerr, I.J. Cox, "Applying Informed Coding and Embedding to Design a Robust High-capacity Watermark", *IEEE Transactions on Image Processing*, Vol. 13, No. 6, pp. 792- 807, June 2004.

[7] M.P. Queluz, "Authentication of Digital Images and Video: Generic Models and a New Contribution", *Signal Processing: Image Communication,* Vol. 16, pp. 461-475, Jan. 2001.

[8] C. Y. Lin and S.F. Chang, "A Robust Image Authentication Method Distinguishing JPEG Compression from Malicious Manipulation", *IEEE Transactions on Circuits and Systems of Video Technology*, Vol. 11, pp. 153-168, 2001.

[9] E.C. Chang, M.S. Kankanhalli, X. Guan, Z.Y. Huang, and Y.H. Wu, "Robust Image Authentication Using Content-based Compression", *ACM Multimedia Systems Journal*, Vol. 9, No. 2, pp. 121-130, 2003.

[10] Q. Sun and S.F. Chang, "Semi-fragile Image Authentication using Generic Wavelet Domain Features and ECC", *IEEE International Conference on Image Processing (ICIP)*, Rochester, USA, Sep. 2002.

[11] C.W. Tang and H.M. Hang, "A Feature Based Robust Digital Image Watermarking Scheme", *IEEE Transactions on Signal Processing*, Vol. 51, No. 4, pp. 950-959, Apr. 2003.

[12] Y. Wang, J. Ostermann, and Y.Q. Zhang, *Video Processing and Communications*, New Jersey: Prentice Hall, 2002.

[13] J. Dittmann, A. Steinmetz, and R. Steinmetz, “Content-based Digital Signature for Motion Pictures Authentication and Content-fragile Watermarking”, *IEEE International Conference on Multimedia Computing and Systems*, Vol. 2, pp. 209-213, 1999.

[14] C. W. Wu, “On the Design of Content-based Multimedia Authentication Systems”, *IEEE Transactions on Multimedia*, Vol. 4, No. 3, pp. 385-393, Sep. 2002.

[15] Q. Sun, S. Ye, L.Q. Lin, and S.F. Chang, “A Crypto Signature Scheme for Image Authentication over Wireless Channel”, *International Journal of Image and Graphics*, Vol. 5, No. 1, pp. 1-14, 2005.

[16] M. Schneider and S.F. Chang, “A Robust Content-based Digital Signature for Image Authentication”, in *Proceedings of International Conference on Image Processing*, 1996, Vol. 3, pp. 227 - 230.

[17] F. Fridrich, D. Soukal and J. Lukas, “Detection of Copy-Move Forgery in Digital Images”, *Digital Forensic and Research Workshop*, Cleveland, USA, Aug. 2003.

[18] A.C. Popescu and H. Farid, “Statistical Tools for Digital Forensics”, *International Workshop on Information Hiding*, Toronto, Canada, 2004

[19] I. Avcibas, S. Bayram, N. Memon, M. Ramkumar, and B. Sankur, “A Classifier Design for Detecting Image Manipulations”, *IEEE International Conference on Image Processing*, Singapore, Oct. 2004.

[20] T. Ng, S.F. Chang, and Q. Sun, “Blind Detection of Photomontage using Higher Order Statistics”, *IEEE International Symposium on Circuits and Systems*, Canada, May 2004.

[21] B. Schneier, “Applied Cryptography”, *New York: Wiley*, 1996.

[22] E. Martinian, G. W. Wornell, and B Chen, “Authentication With Distortion Criteria”, *IEEE Transactions on Information Theory*, Vol. 51, No. 7, pp. 2523-2542, July 2005

[23] G. L. Friedman, “The Trustworthy Camera: Restoring Credibility to the Photographic Image”, *IEEE Transactions on Consumer Electronics*, Vol. 39, No. 4, pp. 905-910, 1993.

[24] A. M. Eskicioglu and E. J. Delp, “An Overview of Multimedia Content Protection in Consumer Electronics Devices”, *Signal Processing: Image Communication*, Vol. 16, No.7, pp. 681-699, 2001.

[25] C.S. Lu and H.Y. M. Liao, "Structural Digital Signature for Image Authentication", *IEEE Transactions on Multimedia*, pp. 161-173, June 2003.

[26] D. Lou and J. Liu, “Fault Resilient and Compression Tolerant Digital Signature for Image Authentication”, *IEEE Transactions on Consumer Electronics*, Vol.46, No.1, pp.31-39, 2000.

[27] S. Bhattacharjee and M. Kutter, “Compression Tolerant Image Authentication”, *IEEE International Conference on Image Processing (ICIP),* Chicago, Oct. 1998.

[28] J. Fridrich and M. Goljan, “Robust Hash Functions for Digital Watermarking”, *IEEE International Conference on Information Technology - Coding and Computing*, Las Vegas, 2000.

[29] L. Xie, G. R. Arce, and R. F. Gravemen, “Approximate Image Message Authentication Codes”, *IEEE Transactions on Multimedia*, Vol. 3, No. 2, pp242-253, 2001.

[30] I. J. Cox and M. Miller, “The First 50 Years of Electronic Watermarking”, *EURASIP Journal of Applied Signal Processing*, Vol. 2, pp.126-132, 2002.

[31] M. Yeung and F. Mintzer, “An Invisible Watermarking Technique for Image Verification”, in *Proceeding of International Conference on Image Processing*, Vol. 2, pp. 680-683, 1997.

[32] M. Wu and B. Liu, “Watermarking for Image Authentication”, in *Proceeding of International Conference on Image Processing*, Vol. 2, pp.437-441, 1998.

[33] J. Fridrich, M. Goljan, and A. C. Baldoza, “ New Fragile Authentication Watermark for Images”, in *Proceeding of International Conference on Image Processing,* Vol. 1, pp. 446-449, 2000.

[34] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, “Secure Spread Spectrum Watermarking for Multimedia Information Hiding”, *IEEE Transactions on Image Processing*, Vol. 6, No. 12, pp.1673-1687, Dec. 1997.

[35] G. Langelaar, I. Setyawan, R.L. Lagendijk, “Watermarking Digital Image and Video Data”, *IEEE Signal Processing Magazine*, Vol.17, pp 20-43, Sep. 2000.

[36] P. Meerwald, “Quantization Watermarking in the JPEG2000 Coding Pipeline”, *Conference on Communications and Multimedia Security*, Germany, May 2001.

[37] P. Meerwald and A. Uhl, “A Survey of Wavelet-Domain Watermarking Algorithms”, *SPIE Symposium, Electronic Imaging*, *Conference on Security and Watermarking of Multimedia Contents*, San Jose, USA, Jan. 2001.

[38] E. T. Lin and E. J. Delp, “Temporal Synchronization in Video Watermarking: Further Studies”, *SPIE/IS&T Conference on Security and Watermarking of Multimedia Contents V*, Santa Clara, CA, Jan. 2003.

[39] D. Kundur and D. Hatzinakos, “Digital Watermarking for Telltale Tamper-Proofing and Authentication”, *Proceedings of the IEEE*, Vol. 87, No. 7, pp.1167-1180, Jul. 1999.

[40] J. L. Cannons and P. Moulin, “Design and Statistical Analysis of a Hash-Aided Image Watermarking System”, *IEEE Transactions on Image Processing*, Vol. 13, No. 10, pp. 1393-1408, Oct. 2004.

[41] Digital Forensic Research Workshop, “A Road Map for Digital Forensic Research”, Report from the First *Digital Forensic Research Workshop*, 2001.

[42] J. C. Russ, "Forensic Uses of Digital Imaging", CRC Press, 2001, pp. 121-128.

[43] Z. Lin, R. Wang, X. Tang, and H.-Y. Shum, “Detecting Doctored Images using Camera Response Normality and Consistency”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 1087-1092, Jun. 2005.

[44] A.C. Popescu and H. Farid, “Exposing Digital Forgeries by Detecting Traces of Re-sampling”, *IEEE Transactions on Signal Processing*, Vol. 53, No. 2, pp.758-767, 2005.

[45] A.C. Popescu and H. Farid, “Exposing Digital Forgeries in Color Filter Array Interpolated Images”, *IEEE Transactions on Signal Processing*, Vol. 53, No. 10, pp. 3948-3959, 2005.

[46] S. Bayram, H. T. Sencar, and N. Memon, “Source Camera Identification based on CFA Interpolation”, *IEEE International Conference on Image Processing*, Genova, Sep. 2005.

[47] A. Swaminathan, M. Wu, and K.J.R. Liu, “Non-intrusive Forensic Analysis of Visual Sensors Using Output Images”, *IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), Toulouse, France, May 2006.

[48] A. Swaminathan, M. Wu, and K.J.R. Liu, “Component Forensics for Digital Camera: A Non-intrusive Approach”, *Conference on Information Sciences and Systems*, Princeton, USA, Mar. 2006.

[49] A. Swaminathan, M. Wu, and K. J. R. Liu, “Non Intrusive Component Forensics of Visual Sensors Using Output Images”, *IEEE Transactions of Information Forensics and Security*, Vol. 2, No. 1, pp. 91-106, Mar. 2007.

[50] T. Ng, S.-F. Chang, C.-Y. Lin, and Q. Sun, “Passive-blind Image Forensics”, *Multimedia Security Technologies for Digital Rights*, Elsvier, 2006.

[51] S. Lyu, “Natural Image Statistics for Digital Image Forensics”, PhD Thesis, Dartmouth College, NH, USA, Aug. 2005.

[52] S. Lyu and H. Farid, “How realistic is photorealistic?” *IEEE Transactions on Signal Processing*, Vol. 53, No. 2, pp.845-850, Feb. 2005.

[53] G. N. Ali, P.-J. Chiang, A. K. Mikkilineni, J. P. Allebach, G. T.-C. Chiu, and E. J. Delp, “Intrinsic and Extrinsic Signatures for Information Hiding and Secure Printing

with Electrophotographic Devices", *IS&T International Conference on Digital Printing Technologies*, 2003, pp. 511-515.

[54] A. Swaminathan, M. Wu, and K.J.R. Liu, "Image Tampering Identification Using Blind Deconvolution", *IEEE International Conference on Image Processing* (ICIP), Atlanta, Georgia, Oct. 2006.

[55] M.K. Johnson and H. Farid, "Exposing Digital Forgeries by Detecting Inconsistencies in Lighting", *ACM Multimedia and Security Workshop*, New York, NY, 2005.

[56] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum, "Radiometric Calibration from a Single Image", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, June 2004, pp. 938-945.

[57] S. Lin and L. Zhang, "Determining the Radiometric Response Function from a Single Grayscale Image", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, June 2005, pp. 66-73.

[58] D. Fu, Y.Q. Shi, and W. Su, "A Generalized Benford's Law for JPEG Coefficients and Its Applications in Image Forensics", *SPIE Electronic Imaging*, San Jose, CA, 2007.

[59] H.R. Sheikh and A.C. Bovik, "Image Information and Visual Quality", *IEEE Transactions on Image Processing*, Vol. 15, No. 2, pp. 430-444, 2006.

[60] I. Avcibas, "Image Quality Statistics and their Use in Steganalysis and Compression", PhD Thesis, Bogazici University, Turkey, 2001.

[61] M. Kharrazi, H. Sencar, and N. Memon, "Blind Source Camera Identification", *IEEE International Conference on Image Processing*, Singapore, Oct. 2004.

[62] H. Farid, and S. Lyu, "Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines", *International Workshop on Information Hiding*, New York, 2002.

[63] J. Lukas, J. Fridrich, and M. Goljan, "Determining Digital Image Origin using Sensor Imperfections", in *Proceedings of* the *SPIE Electronic Imaging*, San Jose, CA, Jan. 2005, Vol. 5685, pp. 249-260.

[64] J. Lukas, J. Fridrich, and M. Goljan, "Detecting Digital Image Forgeries using Sensor Pattern Noise", *SPIE Electronic Imaging*, *Photonics West*, Jan. 2006.

[65] T.S. Holotyak, J. Fridrich, and D. Soukal, "Stochastic Approach to Secret Message Length Estimation in ±k Embedding Steganography", in *Proceedings of SPIE Electronic Imaging*, Jan. 2005, pp. 673-684.

[66] Y. Tsin, V. Ramesh, and T. Kanade, "Statistical Calibration of CCD Imaging Process", *IEEE International Conference on Computer Vision*, Vancouver, Canada, Jul. 2001.

[67] H. Ridder, “Minkowsky Metrics as a Combination Rule for Digital Image Coding Impairments”, in *Proceedings of SPIE*: *Human Vision, Visual Processing, and Digital Display III*, 1992, Vol. 1666, pp. 17-27.

[68] S. Daly, “The Visible Differences Predictor: An algorithm for the Assessment of Image Fidelity”, *Digital Images and Human Vision*, A. B. Watson, editor, MIT Press, Cambridge, Massachusetts, pp. 179-206, 1993.

[69] A. Bradley, “A Wavelet Visible Difference Predictor”, *IEEE Transactions on Image Processing*, Vol. 8, pp. 717-730, May 1999.

[70] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Measurement to Structural Similarity”, *IEEE Transactions on Image Processing*, Vol. 13, pp. 1-14, Jan. 2004.

[71] C.Y. Lin and S.F. Chang, “Semi-Fragile Watermarking for Authenticating JPEG Visual Content”, *SPIE Security and Watermarking of Multimedia Contents*, San Jose, USA, Jan. 2000.

[72] Q. Sun and S.F. Chang, “A Robust and Secure Media Signature Scheme for JPEG Images”, *International Workshop on Multimedia Signal Processing*, Virgin Islands, USA, Dec. 2002.

[73] L.W. Kang and J.J. Leou, “Two Error Resilient Coding Schemes for Wavelet-based Image Transmission Based on Data Embedding and Genetic Algorithms”, *International Conference on Image Processing* (*ICIP)*, Barcelona, Spain, Sep. 2003.

[74] Y. Wang and Q. Zhu, “Error Control and Concealment for Video Communication: A Review”, in *Proceedings of the IEEE*, May 1998, Vol. 86, No. 5, p 974-997.

[75] A. C. Ashwin, K.R. Ramskrishnan, and S.H. Srinivasan, “Wavelet Domain Residual Redundancy based Descriptions”, *Elsevier Science Signal Processing: Image Communication*, Vol. 18, No. 7, Aug. 2003, pp.549-560.

[76] P.J. Lee, and L.G. Chen, “Bit-plane Error Recovery via Cross Subband for Image Transmission in JPEG2000”, *IEEE International Conference on Multimedia and Expo* (*ICME*), Lausanne, Switzerland, Aug. 2002.

[77] S. Ye, X. Lin and Q. Sun, “Content based Error Detection and Concealment for Image Transmission over Wireless Channel”, *IEEE International Symposium on Circuits and Systems*, Bangkok, Thailand, May 2003.

[78] W. Zhu, Y. Wang, and Q. Zhu. “Second-order Derivative based Smoothness Measure for Error Concealment in DCT based Codecs”, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 8, No. 6, pp.713-718, Oct. 1998.

[79] C.Y. Lin, D. Sow, and S.F. Chang, "Using Self-Authentication-and-Recovery Images for Error Concealment in Wireless Environments", in *SPIE ITCom/OptiComm*, Vol. 4518, Denver, USA, Aug. 2001.

[80] P. Yin, H. Yu, and B. Liu, "Error Concealment Using Data Hiding", *International Conference on Acoustic, Speech and Signal Processing*, Salt Lake City, USA, 2001.

[81] W. Zeng and B. Liu: "Geometric-structure-based Error Concealment with Novel Applications in Block-based Low-Bit-Rate Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No. 4, pp. 648-665, 1999.

[82] X. Li and M.T. Orchard, "Novel Sequential Error-concealment Techniques using Orientation Adaptive Interpolation", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 12, No. 10, pp. 857- 864, 2002.

[83] M. Chen, Y-F. Zheng, and M. Wu, "Classification-based Spatial Error Concealment for Visual Communications", *EURASIP Journal on Applied Signal Processing*, *Special Issue on Video Analysis and Coding for Robust Transmission*, Vol. 2006, 2006.

[84] M. Ancis, D.D. Giusto, and C. Perra, "Error Concealment in the Transformed Domain for DCT-Coded Picture Transmission over Noisy Channels", *European Transactions on Telecommunications*, Vol. 12, No. 3, pp. 197-204, 2001.

[85] J. Suh and Y. Ho, "Error Concealment based on Directional Interpolation", *IEEE Transactions on Consumer Electronics*, Vol. 43, No. 3, pp. 295-302, Aug. 1997.

[86] W. Zeng and S. Lei, "Efficient Frequency Domain Selective Scrambling of Digital Video", *IEEE Transactions on Multimedia*, Vol. 5, No. 1, pp.118-129, Mar. 2003.

[87] B. Chen and G. W. Wornell, "Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding", *IEEE Transactions on Information Theory*, Vol. 47, No. 4, pp.1423-1443, May 2001.

[88] C.Y. Lin and S.F. Chang, "SARI: Self-Authentication-and-Recovery Image Watermarking System", *ACM Multimedia*, 2001.

[89] S. Ye, Q. Sun, and E.C. Chang, "Error Resilient Content-based Image Authentication Over Wireless Channel", *IEEE International Symposium on Circuits and Systems*, Japan, 2005.

[90] B. Li, E. Chang, and Y. Wu, "Discovery of a Perceptual Distance Function for Measuring Image Similarity," *ACM Multimedia Journal*, *Special Issue on Content-based Image Retrieval*, Vol. 8, No. 6, pp.512-522, 2003.

[91] Y. Yu and S.T. Acton, "Speckle Reducing Anisotropic Diffusion", *IEEE Transactions on Image Processing*, Vol. 11, No. 11, pp.1260-1270, Nov. 2002.

[92] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, New Jersey: Prentice Hall, Second edition, 2002, pp. 128-131.

[93] R. Jain, R. Kasturi and B. G. Schunck, *Machine Vision*, New York: McGraw Hill, 1995.

[94] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image Inpainting", *Computer Graphics* (*SIGGRAPH*), pp.417-424, July 2000.

[95] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous Structure and Texture Image Inpainting", *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Madison, US, June 2003.

[96] B. Tang, G. Sapiro, and V. Caselies, "Direction Diffusion", *International Conference on Computer Vision* (ICCV), Corfu, Greece, Sep. 1999.

[97] P. Perona and J. Malik, "Scale-space and Edge Detection using Anisotropic Diffusion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No.7, pp. 629-639, 1990.

[98] JPWL Editors, "Wireless JPEG 2000 Working Draft Version 1.1", ISO/IEC JTC 1/SC 29/WG 1 N 3039, July 2003.

[99] S. Ye, Q. Sun, and E.C. Chang, "Edge Directed Filter based Error Concealment for Wavelet-based Images", *IEEE International Conference on Image Processing*, Singapore, 2004.

[100] W. Chou, "Classifying Image Pixels into Shaped, Smooth and Textured Points", *Pattern Recognition*, Vol. 32, No. 10, pp.1697-1706, 1999.

[101] M. Boliek (ed.), "JPEG 2000 Final Committee Draft", *ISO/IEC FCD1.5444-1,* Mar. 2000.

[102] S. Ye, G. Wang, and X. Lin, "Feature Based Adaptive Error Concealment for Image Transmission over Wireless Channel", in *Proceedings of SPIE Electronic Imaging*, 2003, Vol. 5022, pp.820-830.

[103] A.H. Paquet, R.K. Ward, and I. Pitas, "Wavelet Packets-based Digital Watermarking for Image Verification and Authentication", *Signal Processing*, *Special Issue on Security of Data Hiding Technologies*, Vol. 83, No. 10, pp. 2117‑2132, Oct. 2003.

[104] J. Fridrich, M. Goljan, and R. Du, "Steganalysis based on JPEG Compatibility", *SPIE Multimedia Systems and Applications*, Vol. 4518, Denver, CO, pp. 275-280, Aug. 2001.

[105] Z. Fan and R. de Queiroz, "Identification of Bitmap Compression History: JPEG Detection and Quantizer Estimation", *IEEE Transactions of Image Processing*, Vol. 12, pp. 230-235, Feb. 2003.

[106] Z. Wang, A. C. Bovik, B. L. Evans, "Blind Measurement of Blocking Artifacts in Images", *IEEE International Conference on Image Processing*, Canada, 2000.

[107] Z. Fan and R. de Queiroz, "Maximum Likelihood Estimation of JPEG Quantization Table in the Identification of Bitmap Compression History", *IEEE International Conference on Image Processing*, Vancouver, Canada, 2000.

[108] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahini, "Perceptual Blur and Ringing Metrics: Applications to JPEG2000", *Signal Proceeding on Image Communication*, Vol. 19, pp. 163-172, 2004.

[109] X. Marichal, W.Y. Ma, and H. J. Zhang, "Blur Determination in the Compressed Domain using DCT Information", *IEEE International Conference on Image Processing*, Kobe, Japan, Oct. 1999.

[110] J. Dijk, M. van Ginkel, R.J. van Asselt, L.J. van Vliet, and P.W. Verbeek, "A New Sharpness Measure based on Gaussian Lines and Edges", *International Conference on Computer Analysis of Images and Patterns*, the Netherlands, Aug. 2003.

[111] J. Caviedes and F. Oberti, "A New Sharpness Metric based on Local Kurtosis, Edge and Energy Information*", Signal Processing: Image Communication*, Vol. 19, No. 2, pp. 147-161, Feb. 2004.

[112] S. Mallat and W. Hwang, "Singularity detection and processing with wavelets", *IEEE Transactions on Information Theory*, Vol. 38, No. 8, pp. 617-643, 1992.

[113] F. Rooms, A. Pizurica, and W. Philips, "Estimating Image Blur in the Wavelet Domain", *Asian Conference on Computer Vision* (ACCV), Melbourne, Australia, Jan. 2002.

[114] G. C. Holst, "CCD Arrays, Cameras, and Displays", 2nd edition, *JCD Publishing & SPIE Pres*, USA, 1998.

[115] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image Denoising using Scale Mixtures of Gaussians in the Wavelet Domain", *IEEE Transactions Image Processing*, Vol. 12, No. 11, pp. 1338-1351, 2003.

[116] H. Faraji and J. MacLean, "Adaptive Suppression of CCD Signal-dependent Noise in Light Space", *IEEE International Conference on Acoustics, Speech and Signal Processing*, Quebec, Canada, May 2004.

[117] D.-H. Shin, R.-H. Park, S. Yang and J.-H. Jung, "Block-based Noise Estimation using Adaptive Gaussian Filtering", *IEEE Transactions on Consumer Electronics*, Vol. 51, No. 1, pp. 218-226, Feb. 2005.

[118] K. Hirakawa and T.W. Parks, "Image Denoising for Signal-Dependent Noise", *IEEE International Conference on Acoustics, Speech and Signal Processing*, Philadelphia, USA, Mar. 2005.

[119] A. Schaaf, “Natural Image Statistics and Visual Processing”, PhD Thesis, Rijksuniversiteit Groningen University, the Netherlands, 1998.

[120] D. J. Field, "Relations between the Statistics of Natural Images and the Response Properties of Cortical Cells”, *Journal of the Optical Society of America A*, Vol. 4, pp. 2379-2394, 1987.

[121] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, “On Advances in Statistical Modeling of Natural Images”, *Journal of Mathematical Imaging and Vision*, Vol. 18, pp. 17-33, 2003.

[122] H. R. Sheikh, A. C. Bovik, and G. de Veciana, “An Information Fidelity Criterion for Image Quality Assessment using Natural Scene Statistics”, *IEEE Transactions on Image Processing*, Vol. 14, No. 12, pp. 2117- 2128, Dec. 2005.

[123] D. He, “Robust and Scalable Video Authentication: Issues and Solutions”, PhD Thesis, National University of Singapore, Singapore, 2005.