Thesis for the Degree of Doctor of Philosophy

Forgery Detection and Counter-Forensics for Image Manipulation

by

Munkhbaatar Doyoddorj

Interdisciplinary Program of Information Security

The Graduate School

Pukyong National University

February 2014

Forgery Detection and Counter-Forensics for Image Manipulation
이미지 변형에 대한 위조 검출 및 안티 포렌식 기술

Advisor: Prof. Kyung-Hyune Rhee

by

Munkhbaatar Doyoddorj

A thesis submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in Interdisciplinary Program of Information Security,
The Graduate School,
Pukyong National University

February 2014

Forgery Detection and Counter-Forensics

for Image Manipulation

A dissertation

by

Munkhbaatar Doyoddorj

Approved by :

(Chairman) Man-Gon Park

(Member) Kyung-Hyune Rhee

(Member) Song Ha-joo

(Member) Sang-Uk Shin

(Member) Weon Shin

February 2014

이미지 변형에 대한 위조 검출 및 안티 포렌식 기술

Munkhbaatar Doyoddorj

부경대학교 대학원 정보보호학협동과정

요약

오늘날 디지털 멀티미디어 콘텐츠 제작 및 분배의 보편화와 더불어 멀티미디어 편집 도구의 발전으로 인하여 디지털 이미지의 위변조와 같은 멀티미디어 콘텐츠에 대한 보안 위협이 증가하고 있으며, 이에 대한 안전성 및 신뢰성을 제공할 수 있는 멀티미디어 보안기술이 각광받고 있다. 특히 이미지 포렌식 기술은 디지털 이미지 상의 불법적인 위변조를 식별 및 검출할 수 있는 기술로써 디지털 이미지 콘텐츠의 무결성 검증을 수행할 수 있을 뿐만 아니라 불법적인 콘텐츠 조작 (e.g., splicing, composition, copy-move)을 검출할 수 있는 기술이다.

본 논문에서는 이미지 포렌식을 위한 핵심 요소기술로써 이미지 변형에 대한 효율적인 위조 검출 기법들을 제안한다. 제안 기법들에서는 불법적인 이미지 스플라이싱 (Splicing)과 합성 (Composition)과 같은 악의적인 공격에 대한 효율적인 검출 방법 제공을 위하여 에지 기반의 호환성 행렬 (Compatibility metrics) 분석을 사용한다. 이와 함께, 본 논문에서는 카피-무브 (Copy-move) 위조 공격에 대응하기 위하여 위조된 지역을 동시에 검출 및 지역화할 수 있는 기술로써 Radon 변형과 Discrete Cosine 변형에 기반한 듀얼 변형 (Dual transform) 기법을 새로이 설계하고, 이를 기반으로 카피-무브 위조 공격에 강건한 위조 검출 기법을 제안한다. 또한 본 논문에서는 불법적인 위조자 관점에서의 멀티미디어 보안기술에 관하여 고려한다. 이를 위하여 핑거프린팅 (Fingerprinting)에 대한 조작을 숨길 수 있을 뿐만 아니라 디지털 이미지로부터 키 포인트의 검출을 제거함으로써 현재까지 제안된 키 포인트 기반의 이미지 포렌식 기술을 무력화시킬 수 있는 안티 포렌식 기법을 제안한다.

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Background

In todays digital age, our daily life is permeated with digital multimedia content as one of the principal means for communication. As a matter of fact, such information can be created, stored, transmitted and processed in digital format in an extremely easy way, thanks to the wide spread of low-cost cameras and computers and user-friendly editing tools. Besides the economic and technical advantages, the digital information revolution has led to problematic issues concerning multimedia security and reliability. Therefore, it is more and more important to be able to automatically provide protection to digital contents in order to guarantee their truthfulness and security. The scientific community is very active in this field, coming up with sophisticated and accurate methods for authentication and protection.

Digital images are everywhere - from our cell phones to the pages of our online news sites. How we choose to use digital image processing raises a surprising host of legal and ethical questions that we must address. Investigators from a diverse set of fields require the best possible tools to tackle the challenges presented by the malicious use of today's digital image processing techniques. Image manipulation have become ubiquitous today in society, especially on the Internet. As the availability of digital cameras and photo-editing software has increased, processes used to manipulate images, which earlier took hours or days, can now be done at the expense of a few clicks, as shown an example

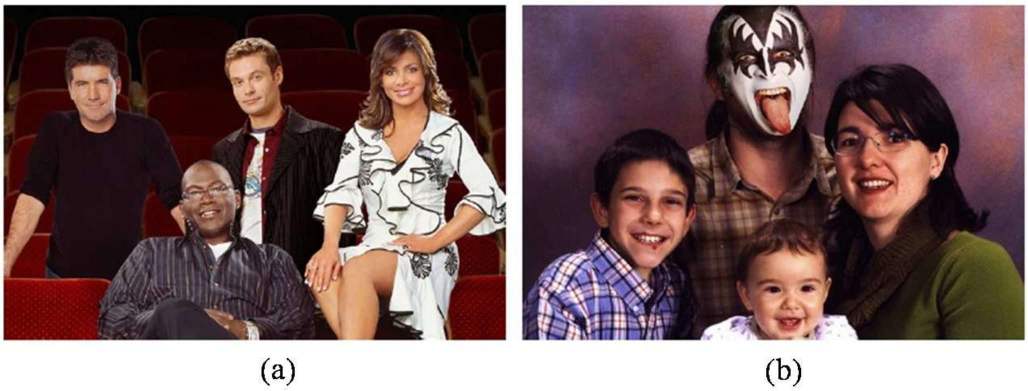


Figure 1.1: Image forgeries. (a) The American Idol forgery and (b) a family portrait with rock star Gene Simmons.

in Figure 1.1. The two images are the American Idol forgery and a family portrait where the fathers face has been replaced with the face of Gene Simmons from the rock band KISS hosted by the website Worth1000 [1]. Not all images are tampered with to mislead. Indeed, most of the forged images on the Internet are created for mere entertainment [2]. However, it is usually quite obvious that these images are tampered with and hence, such images can be easily dismissed as fake. Nevertheless, there are many images which are actually fake, but are advertised as real. It is these very images that need to be analyzed to establish their authenticity. The spread of forged images for illegal or misleading purposes has impacted every aspect of society.

The creation of manipulations can be motivated politically, economically, commercially, socially, or individualistically. Real-world example for a socially motivated manipulation is shown in Figure 1.2. The manipulation was created by just removing unwanted information (in this case, the cigar) from the source image. For an exhibition in London 2010, Churchills cigar was removed

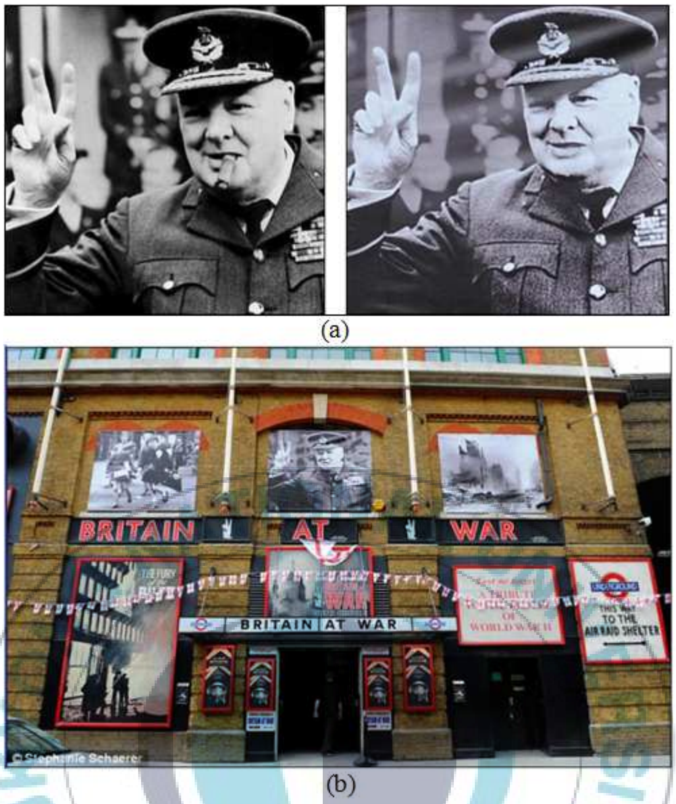


Figure 1.2: Examples of real-world image manipulation. (a) Churchills trademark, the cigar, has been removed from the image, (b) The Britain At War Experience in South-East London with the airbrushed picture of Churchill above the entrance.

from a poster allegedly due to the anti-smoking movement. Thus, an image manipulation is not defined independently of the applied technique. As a side note, not even every motive is considered a manipulation. For instance, photo collages are typically acceptable, because it is not expected that the image shows a real event.

In response to these challenges, the field of digital *image forensics* has been born. Digital image forensics involves the study and development of techniques to determine the authenticity, processing history, and origin of digital image content without relying on any information aside from the digital content itself. This is done by making use of the fact that most signal processing operations leave behind perceptually undetectable traces known as fingerprints in digital content similar to the way that a criminal leaves behind fingerprints at a crime scene. By discovering these fingerprints and developing techniques to detect them, digital forensics researchers can identify digital multimedia forgeries. Because most signal processing operations leave behind unique fingerprints, no universal method of detecting digital forgeries exists. Instead, many forensics tests must be designed to identify the fingerprints of a wide variety of digital content editing operations. It has been posited that if a large set of forensics methods is developed, it will be difficult for a forger to create a digital forgery capable of fooling all forensics authentication techniques [10].

In short, digital image manipulation detection is an interesting and exciting problem that is far from being solved and deservers further research.

1.2 Problem Formulation and Scope

A digital camera performs a series of operations on the incoming lights from the real world scene before it writes the image to the memory card. These operations can be linear or nonlinear, point-wise or spatial, all of which when combined yield visually pleasant, comprehensible images to human eyes.

Even if forensics technologies are usually applied for different purposes, actually it is possible to evidence how a common approach is followed by almost all the forensics algorithms proposed so far, regardless of their application for source identification or tampering detection. Particular, image forensics works

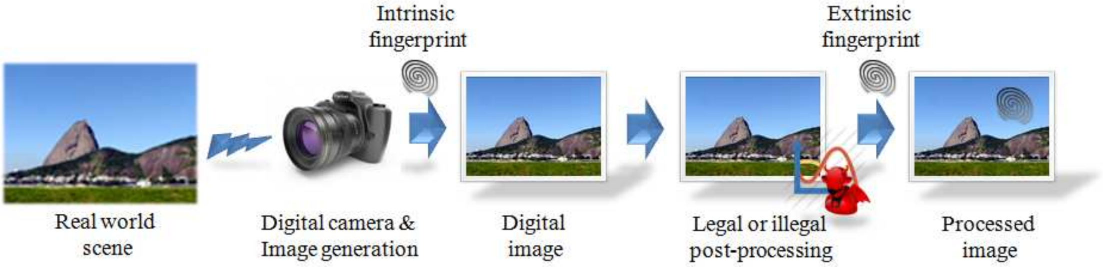


Figure 1.3: System model for fingerprint traces.

by estimating the fingerprint traces that are left behind in a digital image when it goes though various processing blocks in the information processing chain, and uses such traces for estimating component parameters. Generally, the fingerprint traces are grouped into *intrinsic* and *extrinsic* fingerprint traces. A system model for image forensics based on fingerprint traces is shown in Figure 1.3. We classify the intrinsic fingerprint traces into two categories, namely, *in-camera* and *post-camera* fingerprints. Each component in a digital acquisition device modifies the input and leaves intrinsic fingerprints in the final output, due to the specific optical system, color sensor and camera software; furthermore, images and in particular natural images, have general characteristics, regardless of the content, such as inherent noise or behaviour of the luminance or statistical properties that can be seen as in-camera fingerprint. After the image has been produced by the camera, additional processing operations may be done using softwares such as Adobe Photoshop, GIMP, etc. to further improve the picture quality and/or tamper with the image as a legal way. Such processing applied to digital image modifies their properties (e.g., statistical, geometrical, etc.) leaving the post-camera fingerprints (peculiar traces) accordingly to the processing itself.

Extrinsic fingerprints are external signals added to the image by the camera after capture. They can be employed to establish the authenticity of images and determine possible tampering of hidden data with the image as a legal and/or illegal ways. The model for extrinsic fingerprint is shown in Figure 1.3. We only considered the extrinsic fingerprints and its creation mechanisms (namely image manipulation) are concentrated in our research.

In this thesis, we propose several new digital image forensics techniques based on the passive model to detect evidence of editing in digital multimedia content for the content-changing manipulation.

Additionally, we consider the problem of multimedia security from the forgers point of view. Though existing digital forensics techniques are capable of detecting several standard digital image manipulations, they do not account for the possibility that *counterfeiting* operations designed to hide traces of manipulation may be applied to digital content. In reality, it is quite possible that a forger with a digital signal processing background may be able to secretly develop *anti-* or *counter-forensics* operations and use them to create undetectable digital forgeries. As a result, several multimedia forensics techniques may possess vulnerabilities that are unknown to the forensics community at large.

To protect against this scenario, it is crucial for researchers to develop and study counter-forensics operations so that vulnerabilities in existing forensics techniques may be known. This will help researchers to know when digital forensics results can be trusted and may assist researchers in the development of improved digital forensics techniques. Furthermore, the study of counterfeiting operations can also lead to the identification of fingerprints left by counter-forensics operations and the development of techniques capable of detecting when an counter-forensics operation has been used to hide evidence forgery.

It is clear that the authentication of multimedia signals poses a great challenge to information security researchers. Not only must new forensics techniques be developed, but counter-forensics techniques must also be uncovered and their effects mitigated. The reactions of forgers to the development of more sophisticated forensics methods must be predicted and the dynamic interplay between a forger and forensics investigator must be understood. Additionally, unintended uses of forensics techniques must be anticipated and protected against. In this thesis, we address these problems and show how information security can be provided through the study of both digital image forensics and counter-forensics.

1.3 Extrinsic Fingerprints for Post-Processing

In this section, we formulate the model of the post-processing for the image space. Let us assume that images are smooth bidimensional functions and consequently that the image space is the space of the smooth functions. We further assume that the considered image transformations are smooth functionals defined on this space. Then, the set containing the manipulations of the part S for the original image I can be defined (subset $S \subseteq I$) as follows

$$\mathcal{M}(S) = \{f_i(S,p),\ p \in \mathcal{C}_i,\ i = 1,2,...,N\}, \tag{1.1}$$

where the $f_i(S,p)$ are the N considered transformation functions, p stands for each functions parameter, and the set C_i give the possible values of the parameters. To give a better intuitive feeling, lets consider that $f_1(S, p)$ corresponds to the rotating transformation function or operation. In this case, p stands for the rotation angle and C_1 corresponds to the range of allowed rotation angles. For instance C_1 is given by the interval $[0, 360^o]$. The subset S is implicitly part of $\mathcal{D}(S)$ because we assume that for any transformation f there exists an invariant parameter p such that $S = f(S, p)$. For example, in the case of the resizing operation it implies that the corresponding parameter, C_1, contains the real number one, which creates a manipulated image. Additionally, we formulate the multiple manipulation of the image. Let us now consider that duplicates resulting from n-level of composition can be expressed by a single function $g_n(S, p)$. The first variable S is the subset of the original image I (subset $S \subseteq I$), and the variable p is a vector of parameters that controls the manipulation aspect, for example p_1 can be the rotation angle and p_2 the scaling factor, respectively. Such a function can be recursively constructed by using the previously introduced transformation functionals $f_i(S, p)$ as follows

$$g_n(S, p) = f_n(g_{n-1}(S, p), p_n), \tag{1.2}$$

$$g_1(S, p) = f_1(S, 1) \tag{1.3}$$

The order of operations can be modified by permuting the indices i of the transformations $f_i(S, p)$. In this simplified case, the set of manipulations for n-level of composition is

$$\mathcal{M}_{set}(S, n) = \{g_n(S, p),\ p \in \mathcal{C}_1 \times \mathcal{C}_2 \times \mathcal{C}_3 \times ... \times \mathcal{C}_n\} \tag{1.4}$$

The duplicate set $\mathcal{M}_{set}(S, n)$ is thus defined by a bounded smooth high-dimensional surface, or smooth manifold, embedded within the image space. The manifold intrinsic dimensionality is upper bounded by n, the number of considered compositions, since the manifold is created by a function controlled by $n+1$ parameters and one of them is the original image.

1.4 Image Manipulation

Image manipulation is the art of altering an image to express what you want, rather than what the original image may have displayed. It is done for deceitfulness and artistic reasons. In digital editing, images are usually taken with a digital camera and input directly into a computer. Transparencies, negatives or printed images may also be digitized using a scanner, or images may be obtained from stock photography databases. With the advent of computers, graphics tablets, and digital cameras, the term image editing encompasses everything that can be done to a photo on a computer. Image manipulation is often much more explicit than subtle alterations to color balance or contrast and may involve overlaying a head onto a different body or changing a sign's text, for example. Image editing software can be used to apply effects and warp an image until the desired result is achieved. The resulting image may have little or no resemblance to the photo (or photos in the case of compositing) from which it originated [90].

In generally, digital image manipulation can be labelled into two main categories as follows:

- *Content-Changing Manipulation (CCM).* Manipulations in this category change the perceptual quality or semantic meaning of image and thus should be rejected.

- *Content-Preserving Manipulation (CPM).* It do not change or alter the perceptual quality and content meaning of image should be accepted by an authentication system.

In the first category, forensics methods concentrate on identifying the CCMs including image splicing [3], and copy-move [4], which reshape the image content visually and semantically. In the second category, CPMs such as resampling [5], compression [6], contrast enhancement [7], sharpening [8] and median filtering [9] are detected or estimated passively. Besides the wide application in the general image processing pipeline, the both categories of techniques are often used to conceal visual tampering trail and destroy the forensically significant statistical fingerprints. From a forensics perspective, several changes in a photograph are widely acceptable. For instance, it is well accepted to improve the image quality (e.g., to enhance the contrast, denoise an image, or highlight important regions, etc). Forensics investigators search for changes in an image that create a different statement of the image. Thus, an image forgery is semantically defined, by considering the information communicated by the original image and the manipulated image.

1.5 Contributions and Outlines

In this thesis, we propose several new image forensics methods to detect the use of content-changing manipulations. From the above discussion, we can clearly see the need for image forensics techniques capable of authenticating digital images.

Furthermore, we consider the image authentication from the point-of-view of the forger. We propose a counterfeiting method and demonstrate that a forger can use them to fool the detection of content-changing manipulation such as copy-move forgery. We show how both a forger and forensics investigator can respond to the actions of each other, and develop the relation to understand the interaction between these two parties.

The major contributions of this thesis can be summarized as follows:

- **Chapter 2** : We first give an overview of image forensics and their characteristics. In particular, the research area of image forensics and the categories of forgery detection techniques. Following the principles of counter-forensics are briefly introduced.

- **Chapter 3** : We develop two efficient and closely related methods for exposing image splicing and composition detection through an edge based analysis of two different terms of compatibility metrics. In the first method, the key insight of this work is that the image edge (feature) blurring with a Gaussian kernel, the location of the image features such as edges are detectible even if the feature strength is weakened. When a content of the tampered scene is unknown, we detect edges and predict the underlying sharp edges that created the blurred observations, under the assumption that detected edge was a step edge before blurring. When the two regions of different images are spliced to create a tampered region into a single image, the edges of two regions should connect that according to the scene. In other words, manipulation operations will destroy the local coherence of the image as the result of consistent imaging conditions of forged image regions. Therefore, inconsistency of blurred edge width can act as the evidence for image splicing.

  In the second method, we considered that knowledge of illuminant directions is necessary both in computer vision for shape reconstruction, and in image based computer graphics, in order to realistically manipulate an existing image. In this case, when the regions of different images are composed, each region will be illuminated by a different set of lighting sources. Differences in lighting can, therefore, be a telltile sign of digital tampering. Moreover, to the extent that the direction of light source can be estimated for different objects or people in an image, inconsistencies in the lighting direction can be used as evidence of image manipulation.

When creating a digital composite of, for example, two people standing side-by-side, it is often difficult to match the lighting conditions from the individual photographs. Lighting inconsistencies can therefore be a useful tool for revealing traces of digital tampering. The illuminant direction tool estimates the direction to the light source from several objects in an image; widely varying estimates are evidence of tampering.

- **Chapter 4** : We develop a robust copy-move forgery detection method to detect and locate the forged regions. The detection process uses a construction of the invariant features from dual-transform, such as Radon and discrete cosine transforms. The key insight of our work is that the copied region concealed with post-processing operations before being pasted in same image, the invariant image features are detectable by using the ability of such transform even if the feature strength is weakened. When the position of the copied part is unknown, we able to detect the exact pasted position that using the extracted invariant features, under the assumption that the pasted regions will yield similar features with the copied regions.

  In this method, Radon transform is utilized to project the image onto directional projection space, and then 1-D DCT is used to extract significant frequency features from the Radon space. Dual-transform largely reduces the influence of geometrical and image processing operations, and the invariant feature of the dual-transform coefficients is found to be stable. Extensive comparative studies show the superiority and robustness of the proposed method.

- **Chapter 5** : We demonstrate a counter-forensics research against image forgery detection, where with the study of methods to counter-forensics techniques by concealing manipulation traces is to be intended. The actual reliability of such methods can only be estimated by considering what an attacker can try to do to invalidate detection techniques.

  The key insight of our work is investigated by analyzing countermeasure method against SIFT algorithm to recreate keypoints in a keypoints removed image while still avoiding keypoint matching for a copy-move forgery detection. The keypoint creation sometimes unavoidably accompanies keypoint removal. In addition to keypoint removal and keypoint insertion is harmful to scale-space image feature extraction. Also, keypoint removal or creation mechanism is not suitable for image counter-forensics, because the forensics analyst can easily identify the manipulation traces. In order to solve the problem, our proposed attack that is successful in deluding a SIFT-based copy-move forgery detection method can simultaneously remove and create keypoints in the image to be conceal traces left with same keypoints removal and creation rate.

  To provide an experimental validation, we need to choose a specific scenario. This consists in selecting a detector for the forensics analist and a processing tool for the adversary. During the whole procedure, the adversary can exploit the knowledge of the detector used by the forensics analist since we are aiming at a *targeted* counter-forensics method.

Basically, our attack aims at identifying the security weakness of the SIFT that employ scale-space keypoint detection mechanism and should not be interpreted as the conventional attacks (e.g., signal processing or geometric attacks) that are blind in destroying the keypoints.

Chapter 2. Preliminaries

The purpose of this chapter is to explain the image forensics techniques, in order to give you the fundamental backgrounds which are needed to understand our proposed methods in this thesis. We firstly give an overview of image forensics and its characteristics in Section 2.1. The research directions for passive approach are presented in Section 2.2, then the categories of image forgery detection techniques are introduced in Section 2.3. A brief review of counter-forensics is provided in Section 2.4. Category of counter-forensics techniques is presented in Section 2.5 Finally, the measures of image forensics and the performance evaluation metrics are explained in Section 2.6 and Section 2.7, respectively.

2.1 Image Forensics

We consider digital images created by using an electronic imaging device to capture a real world scene. We adopt the following a framework for the digital image forensics as shown in Figure 2.1. In respect the image forgeries, techniques for detecting image manipulation can be classified into two main categories - *active* and *passive* approaches. Active approaches mainly uses techniques like signatures and watermarks which are stored within the image at the time of its creation. In most cases such insertion must fall below human perception levels so that human eyes cannot detect the inserted signatures and watermarks. At the receiving end, if the copyright is ever in question,

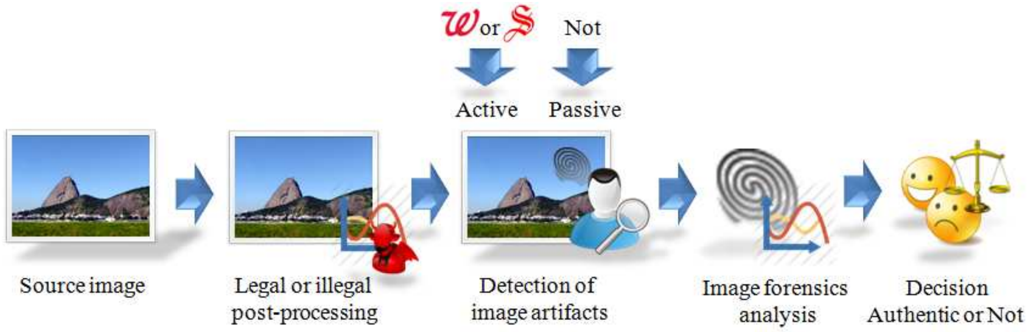

W or S
Not
Active
Passive
Source image
Legal or illegal post-processing
Detection of image artifacts
Image forensics analysis
Decision Authentic or Not

Figure 2.1: A framework for image forensics.

the watermark is extracted and verified to determine the ownership and the authenticity of an image. This methodology calls for specialized hardware and software in order to create and save the watermarks. In active approaches, although proven effective in terms of robustness and accuracy, has its fundamental limitations. With the ease of access to image editing tools nowadays, almost everyone can generate tampered images and it is difficult to ensure every image goes through the standard watermarking process. Even if no watermark is extracted from an image, one still cannot claim this image being tampered. Therefore watermarking has limited use in practice.

But, passive techniques are regarded as the new direction in digital multimedia security as they operate, in contrast of active forensics techniques, in absence of any special equipped device and do not require the knowledge of any prior information about the content. The core assumption for this class of techniques is the assumption that original non-forged content owns some inherent statistical pattern introduced by the generative processing. Such patterns are always consistent in the un-forged content, but they are very likely to be altered after some tampering processes. Although visually imperceptible, such changes can be detecting by statistical analysis of the content itself, without the need of any a-priori information. Thus, passive approaches have a wider range of applications, and hence, it is useful to develop these techniques in order to be able to detect image tampering in general.

2.1.1 Image Generation and Artifacts

Understanding of the image generation (formation) process is necessary to develop passive image forensics. An authentic image is generated from following steps: the light is diffused or reflected from the objects in the scene, then these light rays are recorded by a capturing device (digital cameras), and finally some processing is applied (to generate required compressed formats or meet certain storage constraints) [34]. Digital images are projections of observations of the infinite set of all conceivable scenes $O \in \mathcal{N}$ to vectors $I \in \mathcal{I}, \mathcal{I} \equiv \mathcal{X}^N$, over a finite alphabet $\mathcal{X}$ of discrete symbols. An universal image generation process helps us to conveniently formalize such projections. Therefore an *image generation* is defined as: $\mathcal{G} : \mathcal{N} \times \Theta \rightarrow \mathcal{I}$ maps observations of the infinite set of all conceivable natural real-world phenomena $O \in \mathcal{N}$ to the finite set of digital images $I \in \mathcal{I}$. The mapping is parametrized with a collection of parameters $\theta \in \Theta$.

The parameters include, inter alia, the perspective, the time of the acquisition, the choice of the acquisition device, and its configuration (e.g., settings, lenses). It is convenient to understand the image generation process as a combination of both the *image acquisition* and the *processing* operations.

The image acquisition is the interface between the real and the digital world, where a scene is projected to a discrete representation. Such projections can take place via an image acquisition device - camera, scanner or others, it is useful to study fingerprint traces that the acquisition left traces in the output (e.g., in-camera fingerprint traces). These traces are used for source identification. As far as image forensics is concerned, the acquisition device of the images in question is usually a digital camera.

In order to further improve the image quality or meet the practical constraints in storage or transmission, various processing steps are often employed in the imaging pipeline. Upon observing the artifacts generated in processing domain (e.g., post-camera fingerprint traces). Note these processing cues are independent of device signatures. They are generally used for tampering detection instead of source identification.

The complete image generation process is composed of a concatenation of an image acquisition $\mathcal{A} : \mathcal{N} \rightarrow \mathcal{I}$ and an image processing $\mathcal{P} : \mathcal{I}^+ \rightarrow \mathcal{I}$, where $\mathcal{A}$ and $\mathcal{P}$ are the respective families of functions of all possible image acquisition methods and all possible image processing operations, respectively. Operation $^+$ is a given set of $\mathcal{I}$, which defined as $\mathcal{I}^+ = \bigcup_{n=1}^{\infty} \mathcal{I}^n$. Hence, process $\mathcal{P}$ take an arbitrary positive number of digital images as input. The exact composition is defined by the parameters θ of $\mathcal{G}$.

Each of steps leaves inherent traces in the final output image as mentioned above. Any image that lacks any of the three sets of natural characteristics is subject to the suspicion of being non-authentic. These cues can be categorized as natural scene (e.g., lighting, shadow, geometry, etc.), device characteristics (e.g., sensor noise statistics, color filtering array, Camera Response Function CRF, etc.), or post processing artifacts (e.g., JPEG quantization settings, video de-interlacing settings, etc.).

2.1.2 Semantic Meaning and Authenticity

A further important attribute with regard to the image generation process is the notion of authenticity. Image authenticity is a central idea for addressing the image forensics problems. An image is authentic if it represents a witness to an actual event, place, or time. Image authenticity is used to prove the truth of multimedia based on various security techniques, thus which faces a great challenge. A definition of image authenticity should enable us to distinguish an authentic image from the forged images while maintaining the ability to discriminate malicious (illegitimate) manipulations from incidental (legitimate) manipulations. The digital image preserving the original perceptual quality or semantic meaning is desirable to be considered as authentic. A digital image I is called *authentic* if it is a valid projection of the natural phenomenon O. Instances of process may impair authenticity.

In [34], note that the projection of one particular natural phenomenon O to an authentic image is not necessarily unique. There may exist many different mappings that yield *semantically equivalent* images. This means each element in a set of many different images $I_1 \neq I_2 \neq \cdots \neq I_n$ is a valid representation of the same realization of nature O. For example, in many cases it makes no difference with which digital camera a given event is captured, but each camera will produce a slightly different image. Within certain limits also the change of resolution or lossy compression may retain an images authenticity. In this sense, authenticity is an attribute of the tuple (I, θ, O) where O must be the realization of $\mathcal{N}$ under parameters θ. Intuitively, also the *semantic meaning* of an image refers to the link between a depicted scene and the corresponding natural phenomenon. Yet this is more difficult to formalize, as the association of semantic meaning requires interpretation and it is highly context-dependent in general. The assumption that semantic equivalence is measurable between images.

Two images I_1 and $I_2 \in \mathcal{I}$ are *semantically equivalent* if there exists $O \in \mathcal{N}$ such that $|\mathsf{dist}(I_1, O) - \mathsf{dist}(I_2, O)| < t$, where $\mathsf{dist} : \mathcal{I} \times O \rightarrow \mathbb{R}_+$ is a measure of the *semantic distance* between an image and a real or imaginary natural phenomenon, and t is a given threshold.

The *semantic resolution* is the ability of function dist to differentiate between very similar natural phenomena for a fixed image I_2. This resolution depends on the quality of an image, or, more precisely, on the information conveyed in an image I_2 about O. Threshold t has to be chosen commensurate with the semantic resolution of the image with the lowest quality. Equipped with the notion of semantic equivalence, we can finally define what qualifies an image as authentic.

All original natural images are authentic. Furthermore, for a given authentic image $I_1 = \mathcal{G}(O, \theta)$, a processed version $I_2 = \mathcal{P}(I_1)$ is called *authentic* if I_1 and I_2 are semantically equivalent with respect to O.

2.2 Passive Approaches

Passive forensics techniques can be primarily divided in three categories [12].

- Image source identification

- Discriminating between real and computer generated images

- Image forgery detection

Image source identification aims at establishing a link between an image and the device it was generated from (e.g., camera, scanner or cell phone... etc.). The basic assumption is that digital pictures taken by the same device are overlaid by a specific pattern, that is a unique and intrinsic fingerprint of the acquisition device. A remarkable works has been presented in the literature for source identification, exploiting distortions introduced by demosaicing artifacts [46, 47], while others rely on the sensor imperfections, such as defective pixels [48], fixed pattern noise [49] and photo-response non-uniformity noise ($PRNU$) [50, 51].

The second class of digital forensics techniques aims at *discriminating between real and computer generated images*, based on the assumption that computer and imaging technologies are nowadays so sophisticated and accurate that the distinction between virtual and real images is increasingly difficult to be done at simple visual inspection due their high photorealism [45]. Therefore, techniques based on demosaicing and chromatic aberration have been proposed [52], as long as pattern noise based approaches [53]. In [54], where higher order statistics of wavelet transform coefficients are used to train a classifier and are shown to be effective in discriminating natural and computer generated images. After this, several techniques have been developed, among which, physics-based [55] and features-based methods [56].

The third class of passive forensics aims at *uncovering tampering* or *image forgery detection* that possibly occurred in the content. Generally, when an image is forged, no visual artifacts are introduced in the digital image and it is hard to disclose the manipulation at simple visual inspection. However, the underlying image statistics are heavily affected, thus allowing forgery to be traceable [19, 40]. We review the state-of-the-art of image forgery detection techniques shortly in Section 2.3, following the classification presented in [19].

2.3 Image Forgery Detection

Image forgery detection is a task that aims at detecting the forgeries of an original image. Consequently, it is first necessary to define what a manipulation is. In short, an image forgery is a transformed version of an original artwork that keeps a visual artifacts. In other words, *being a forgery* is a pairwise equivalence relationship that links the original to any of its variations through a transformation operation, for example, compression, brightness changes or cropping, etc. Forged images of the original artwork can subsequently be detected by checking the artifacts presence within images. On the other hand, the passive approach relies, as suggested by its name, on the analysis of the images content in order to extract relevant visual features. Image forgery is then identified when their features are close to those of the original image. In particular, image forensic tools designed for image authentication and forgery detection can be grouped into five categories: (1) Pixel-based techniques detect statistical anomalies introduced at the pixel level, (2) Format-based techniques leverage the statistical correlations introduced by a specific lossy compression scheme, (3) Camera-based techniques exploit artifacts introduced by the camera lens, sensor or on-chip post-processing, (4) Statistical-based techniques explicitly model and detect anomalies in the three dimensional interaction between physical objects, light, and the camera, (5) Geometric-based techniques make measurements of objects in the world and their positions relative to the camera. In the next subsections, a short review of the state-of-the-art of forgery detection techniques [19] are presented in more details.

2.3.1 Pixel-Based Techniques

The basic assumption of this class of techniques is that any form of manipulation, if applied properly, is not visually detectable but alters specific statistics at a pixel level. Depending on the occurred forgery, pixel-level correlations can be analyzed.

Resampling is a process required when resizing, rotating or stretching an image, operations that are likely to happen when creating a fake image. It introduces some specific correlations in neighboring pixels, that can be analyzed as evidence of manipulation. In [57], relying on finding traces of resampling in the image. The idea is based on the observation that tampering may alter the underlying statistics. In fact, when an image is modified, operations like resizing, rotating and stretching must be typically performed, which require to resample the original image. This process introduces correlations that, once detected, can be considered as evidence of a digital tampering. The detection process is based on estimating, through the expectation or maximization algorithm, a set of periodic samples that are correlated to their neighbors.

Copy-move forgery is probably one of the popular form of forgeries, usually performed in order to conceal an object in the scene by covering it with other parts of the image itself [58]. Although visually challenging to be disclosed, such kind of tampering can be detected by looking for statistically similar parts within the content. However, it results computationally unaffordable to perform a brute search all over the image.

When creating a composite, two or more images are spliced together. Such operation has been demonstrated to alter higher order Fourier statistics, alteration that can be used as evidence of tampering [65]. Recently, in [66] the co-occurrence matrix of thresholded edge image of image chroma is analyzed.

2.3.2 Format-Based Techniques

Format-based methods take advantage of the specific format of images. This class of techniques aims at disclosing statistical correlations introduced by compression schemes. The $JPEG$ compression is well known to be a lossy scheme, i.e. some information is lost during the process. Since most images are $JPEG$ compressed, to detect a tampering it is possible to exploit the blocking effect introduced by $JPEG$, which gives rise to the so-called Block Artifact Grid (BAG). In fact, manipulating images in this format causes an alteration of these artifacts, mainly in the case of copy-move processing, since the BAG of the original image and that of the copied region very likely mismatch. In [62], a simple method is proposed to identify this type of forgery. The basic idea is to extract the horizontal and vertical edges due to $JPEG$ artifacts by means of a second order derivative followed by a thresholding operation in order to eliminate edges relative to signal discontinuities. A further enhancement is then realized to obtain the block artifact grid. If the image has been subject to a copy-move processing a BAG mismatching can be detected when lines are present within a 8×8 block. The procedure delineated in [62] tries to determine this presence through summations along rows and columns both inside and at the boundaries of the block. Block artifacts introduced by a $JPEG$ compression at the border of neighboring pixels are studied in [63] for forensics purposes supposing that manipulations are likely to alter such artifacts.

Assuming that, when creating a composite, it is unlikely to match the same level of quantization of the two spliced parts, [64] analyzes the quantization coefficients to prove tampering.

2.3.3 Camera-Based Techniques

A very powerful approach in detecting for forgeries relies on artifacts introduced by the digital camera itself, and in particular the Photo-Response Non Uniformity ($PRNU$) which can be considered as a sort of intrinsic fingerprint of a specific digital camera. The $PRNU$ arises from differences and imperfections in the silicon wafer used to manufacture the imaging sensor: these physical differences provide a unique sensor fingerprint which can be used for forgery detection. The method in [59], requires the preliminary estimation of the camera $PRNU$ from a large number of images taken by the camera itself. Then, the $PRNU$ of the image under investigation is estimated and compared with the reference. This step is quite challenging, since this fingerprint is much weaker than the image, therefore a denoising step is used, which removes much of the image content increasing the signal-to-noise ratio. In [60], the authors are replaced the original denoising algorithm with state-of-the-art nonlocal filtering, obtaining a significant performance improvement. The $PRNU$ comparison is carried out by sliding an analysis window of dimension 128×128 over the image: if the camera $PRNU$ is present, the block (or more correctly its central pixel) is labeled as genuine, otherwise it is considered tampered. The test statistic used for detection is the normalized correlation value with a decision threshold selected so as to obtain the required false acceptance rate. A similar algorithm has been recently proposed in [61]. It makes use of canonical correlation analysis (CCA) to measure the linear correlation between the two $PRNU$ estimates.

2.3.4 Statistical-Based Techniques

Simple forgery detection, based on naive physical constraints on the scene, may involve detection of inconsistencies of lighting direction (shadows) [67], color balancing, intelligent reasoning etc. These inconsistencies, however, are easily avoided even by a novice forger while usually hard to detect automatically using software. Computer Vision based algorithms that detect lighting inconsistencies, for example, must introduce strong assumptions on the scene (e.g. that all surfaces of interest are Lambertian). Intelligent reasoning typically requires segmenting most of the image into meaningful objects and classifying them correctly as a prior to any further analysis. Such a task is highly complex and demanding and is currently feasible only for a limited set of objects (e.g. faces, cars, airplanes... etc.) [68].

2.3.5 Geometrical-Based Techniques

Typically images may undergo a variety of post-processing and re-compression, which may impair the effectiveness of traditional techniques for forgery detection. In contrast to statistical techniques, geometric-based forensic techniques have been proposed, which exploit measurements of objects in the world and their position relative to the camera analyzing the projection geometry. Their major advantage over techniques based on low level image statistics is that the modeling and estimation of geometry is less sensitive to resolution and compression that can easily confound the statistical analysis of images and videos.

The algorithm in [69] is based on the assumption that generally the principal point is located near the center of the image and translations in the image plane correspond to an equivalent shift of the principal point across the image. Exploiting the known geometry of a pair of eyes, inconsistencies in the principal points across persons in an image are used as evidence of tampering. Similarly, in [70] describe a technique for detecting image composites by enforcing two geometrical constraints on the homography. The approach can detect fake regions efficiently on a pair of images at the same scene, but requires two images correlated with H (planar homography) or F (fundamental matrix) constraints. When making composites, also the matching of shadow is a challenging task. The imaged shadow can be modeled by a planar homography. By imposing geometric constraints on it from a single image, it is possible to detect digital forgeries [71].

2.4 Image Counter-Forensics

Image forensics has promised to reestablish trust in digital images, which otherwise were deemed too easy to manipulate. But what stops perpetrators, spies and swindlers, who make efforts to manipulate images for their own profit anyway, from finding out forensics investigators latest tricks and techniques? Image forensics methods have benefit from research on countermeasures in a similar way as reasoning about attacks in multimedia security in general is useful to improve the security. In this sense attacks on image forensics algorithms can be understood as methods to sistematically mislead the detection methods. A framework for counter-forensics are described in Figure 2.2. In general, such

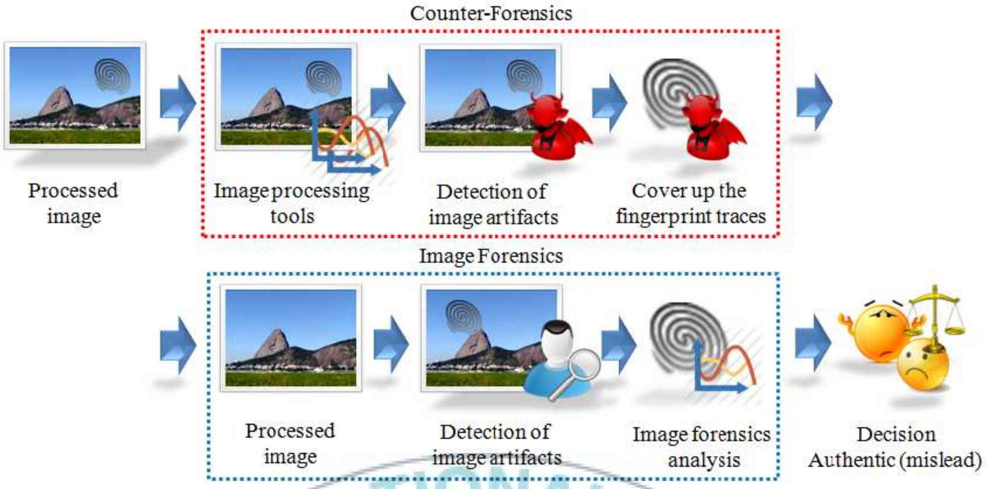

Counter-Forensics
Processed image
Image processing tools
Detection of image artifacts
Cover up the fingerprint traces
Image Forensics
Processed image
Detection of image artifacts
Image forensics analysis
Decision Authentic (mislead)

Figure 2.2: A framework for counter-forensics.

attacks can be assigned to one of the following three objectives: *the camouflage of malicious post-processing* or tampering of an image, *the suppression of correct image origin identification* and furthermore *the forgery of image origin.* Attacker can use this knowledge to cover up fingerprint traces or misleading the detection of fingerprint traces. Generally, digital image counter-forensics, *the art and science of impeding and misleading forensics analyses of digital images*. Counter-forensics stands in a equally productive relation to forensics like cryptanalysis to cryptography. Therefore we borrow from the cryptanalysis terminology and call a counter-forensics attack (against forensics). This reflects the strategic intention of the counterfeiters action.

Only few papers are presented on this topic, because research on this theme is only in its infancy. In [72], where a methods is developed to reveal counter-forensics activities in which an attacker estimates the camera fingerprint from a set of images and pastes it onto an image from a different camera with the intent to introduce a false alarm and frame an innocent victim. Another interesting work is presented in [73], authors proposed a targeted method to hide traces of contrast enhancement, a common enhancement operator that leaves traces in the histogram of the image, so to deceive the detector developed by [74]. A method [73] is based on the introduction of local random dithering in the enhancement step, so it can be classified as integrated attack. Nevertheless, the authors also mention the possibility of turning this attack into a post-processing one. In [75], several works for hiding traces of $JPEG$ compression, that also allow to hide some kinds of tampering that are revealed thanks to $JPEG$ compression side effects [76]. The basic idea underlying these works is to remove an important trace left by $JPEG$ compression into the image, namely the quantization of DCT coefficients. Since the goal is pursued by introducing additive noise to remove discontinuities in DCT coefficients values, these methods can be thought of as post-processing counter-forensics attacks. We will introduce a category of specific techniques in Chapter 5.

2.5 Category of Counter-Forensics Techniques

The research on attacks against image forensics techniques is important to evaluate and ultimately improve detectors, as is steganography for steganalysis and vice versa. In general, the counterfeiters can exploit robustness or security weaknesses to mislead forensics analysis. Kirchner *et al.* [37] introduced the concept of fighting against image forensics. The distinction of this concept is between *post-processing* and *integrated* techniques, and between *targeted* and *universal* ones.

A counter-forensics technique belongs the post-processing class if it consists of two steps: first the attacker performs the tampering, thus obtaining a desired modified content, then she processes the content so to conceal or erase the detectable traces left during the first step. On the contrary, an integrated counter-forensics technique modifies the image so that by construction it does not expose the detectable traces. It is easy to guess that, developing integrated methods is much harder in most cases.

The second distinction regards the target of the counter-forensics method: if it aims at removing the trace searched for by a specific detector or exploits particulars and weaknesses of one specific forensics algorithm, which the counterfeiter usually knows, then it belongs to the targeted family. A universal method, instead, attempts to maintain or correct as many statistical properties of the image as possible, in order to conceal manipulations even when presented to unknown forensics tools, so to make the processed image hard to detect also with tools unknown to the attacker. This is by far the more difficult task, and it is open research question whether image models can be found good enough to sustain analysis with combinations of forensics algorithms. A weaker, more practical form of universal attacks exploits poor robustness properties and uses lossy but legitimate processing whenever plausible.

2.6 Measures for Security and Robustness of Image Forensics

The design of image forensics techniques is a strong academic perspective, as it allows to study the security and the robustness of forensics algorithms. Before we define what exactly we mean by security and robustness, it is useful to come back to distinction between legitimate and illegitimate post-processing and thereby to introduce the notion of efficiency first.

- *Efficiency.* The efficiency of a digital image forensics technique is the detection capacity of the technique in case no legitimate or illegitimate attack has been applied to forged images.

- *Robustness.* The robustness of a digital image forensics technique is its reliability even if legitimate image post-processing is performed [34]. Forensics investigators generally wish to operate highly robust forensics algorithms, which are barely sensitive to any form of legitimate post-processing. Most forensics techniques in the literature are tested with some common post-processing operations such as $JPEG$ compression, downscaling, and Gaussian noise addition in order to measure their reliability. If quality reduction is considered plausible and thus inconspicuous, a counterfeiter can always try to eliminate subtle identifying traces of the original image by reducing the semantic resolution of image. The authors of [37] show that the common manipulations allow to judge the reliability of the forensics techniques only on an average. In fact, based on the knowledge of a forensics technique, which are mostly published, adversaries can design deliberated attacks in order to defeat the technique.

- *Security.* The security of a digital image forensics technique is defined by its reliability to detect forgeries even in case intentionally concealed illegitimate post-processing has been applied to forged images [34]. In other words, security is the ability to withstand counter-forensics. Counterfeiters attacking security properties exploit specific knowledge about and shortcomings of the image model used by forensics investigators. Thus, the security of an image forensics technique can be evaluated by examining its resistance against targeted attacks.

2.7 Performance Evaluation Metrics

A performance metric is a meaningful and computable measure used for quantitatively evaluating the performance of any forgery detection technique [77]. The image dataset is suitable for evaluations at two levels of detail.

- *Image level.* The evaluation focuses on the number of images in dataset that were correctly detected as original or manipulated.
- *Pixel level.* The second possibility is to evaluate the detection performance within an individual image. In this case, we count the number pixels that were correctly detected. This can be done using the ground truth map together with the pixelwise output of the respective benchmark image.

At both levels, it is possible to count the number of correct detection (true positives - TP), where a forged image is correctly identified as forged, false detections (false positives - FP), where an original image is incorrectly identified as forged, correctly omitted images or pixels (true negatives - TN), where an original image is correctly identified as original, and (false negatives - FN), where a forged image is incorrectly identified as original. These metrics have been used in several papers in [78, 79]. From these measures, related performance metrics can be computed. The true positive rate (TPR) and the false positive rate (FPR) are defined as

$$TPR = \frac{TP}{TP + FN}, \tag{2.1}$$

$$FPR = \frac{FP}{TN + FP}. \tag{2.2}$$

The detection rate is the fraction of the number of images detected as forged and the total number of testing images. In a test with a dataset of all forged images, the true positive rate is equal to the detection rate. Similarly, in the test with a dataset containing only original images, the false positive rate is computed as the fraction of the number of original images which have been detected as forged and the total number of testing images. The detection rate and false positive rate are general metrics, some other evaluation metrics are only suitable for a certain forgery type.

Copy-move forgery is a very popular problem in image forensics, where an image is judged as forged if there are two similar regions in the image. Typically, these algorithms are able to identify the copied regions pretty accurately. However, the detection algorithm may produce false positives when the detected results are, for example, parts of homogeneous image regions or produce errors when estimating the forged regions. In order to evaluate the accuracy of detection techniques, we use a specific form of the metrics, which described in Section 4.5.

Chapter 3. Image Splicing and Composition Detection Using Compatibility Metrics

3.1 Introduction

Sophisticated digital cameras and photo editing software packages are becoming ubiquitous. As a results, it has become relatively easy to manipulate digital images and create forgeries which are difficult to distinguish from authentic photographs. The goal of the forgery creators is to create image forgery as a fabrication of the truth, while the forgery detectors try to uncover any possible act of the fabrication by assessing the authenticity of a given image. A common manipulations in tampering with an image are to *splice* or *composite* portions of the image to conceal a person or object in the scene as shown in Figure 3.1. One form of photo manipulation is the digital splicing of two or more images into a single image. When performed carefully, the border between the spliced regions can be visually imperceptible that generated by pre-processing, such as blurring, adding a noise, etc. But in case of digital composition, the border between two regions is noticeable by the human visual system. Basically, the main difference between these two manipulations is defined by the pre-processing operation.

Recently, a passive techniques are regarded as the new direction in multimedia forensics as they operate, in contrast of active techniques, in absence

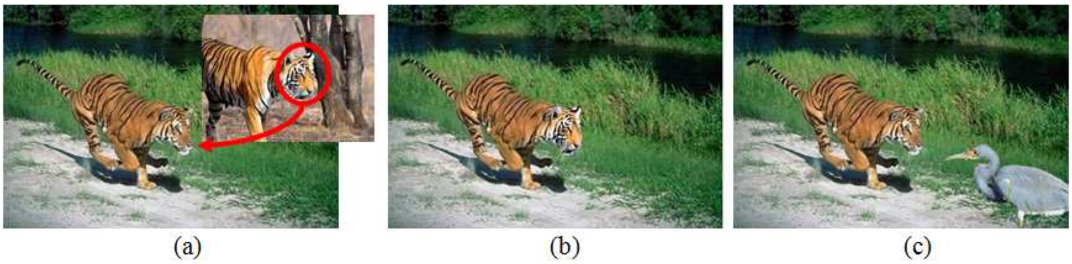


Figure 3.1: Image tampering. (a) Original image, (b) Splicing (a head of tiger), (c) Composition.

of any special equipped device and do not require the knowledge of any prior information about the content. The core assumption for this class of techniques is the assumption that original non-forged content owns some inherent statistical pattern introduced by the generative processing. Such patterns are always consistent in the un-forged content, but they are very likely to be altered after some tampering processes. Although visually imperceptible, such changes can be detecting by statistical analysis of the content itself, without the need of any prior information [11]. Image manipulations very often involve local sharpness or blurring adjustments. Hence, the blurriness characteristics in the tampered parts are expected to differ in non-tampered parts, which is measured by compatibility metric. The *compatibility metric* is a mechanism to detect image tampering based on consistent variations of selected features across the image. These variations may be in the form of abrupt deviations from the image norm or unexpected similarities over the image [12].

A representation of image information in terms of edges is also compact in the sense that the two-dimensional image pattern is represented by a set of one-dimensional curves. For these reasons, edges have been used as main features in a large number of computer vision algorithms. A non-trivial aspect of edge based analysis of image data, however, concerns what should be meant by a compatibility metric in image. Real-world image data are inherently discrete, and for a function defined on a discrete domain, there is no natural notion of ”discontinuity”. This means that there is no inherent way to judge what are the edges in a given discrete image. Therefore, the concept of an image edge is only what we define it to be [13]. In this respect, edges in the image domain constitute a strong link to physical properties of the world. From this viewpoint, it is easy to understand that a large number of approaches have been developed for detecting edges. However, the earliest schemes focused on the detection of points at which the gradient magnitude is high.

The reminder of this chapter is organized as follows: Background about the image edge model and our proposed tampering detection mechanisms are presented in Section 3.2. The experimental results are provided in Section 3.3. Discussions are drawn in Section 3.4.

3.2 Image Edge Formation

Edge detection is a fundamental tool in image processing, machine vision and computer vision, particularly in the areas of feature detection and feature extraction, which aim at identifying points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities. The same problem of finding discontinuities in 1-D signals is known as step detection. The purpose of detecting sharp changes in image brightness is to capture important events and changes in properties of the world. It can be shown that under rather general assumptions for an image formation model, discontinuities in image brightness are likely to correspond to:

- Discontinuities in depth
- Discontinuities in surface orientation
- Changes in material properties
- Variations in scene illumination

In the ideal case, the result of applying an edge detector to an image may lead to a set of connected curves that indicate the boundaries of objects, the boundaries of surface markings as well as curves that correspond to discontinuities in surface orientation. Thus, applying an edge detection algorithm to an image may significantly reduce the amount of data to be processed and may therefore filter out information that may be regarded as less relevant, while preserving the important structural properties of an image. If the edge detection step is successful, the subsequent task of interpreting the information contents in the original image may therefore be substantially simplified. However, it is not always possible to obtain such ideal edges from real life images of moderate complexity. Edges extracted from non-trivial images are often tampered by fragmentation, meaning that the edge curves are not connected, missing edge segments as well as false edges not corresponding to interesting phenomena in the image - thus complicating the subsequent task of interpreting the image data.

3.2.1 Edge Blurriness Model

An edge is modeled as step function $\mathcal{F}(x) = A(x) + H$ with amplitude A and height H on the position x. For each pixel on the chosen edge segment, the edge $\mathcal{F}(x)$ has not been retouched. It is a widely used effect in graphics software,

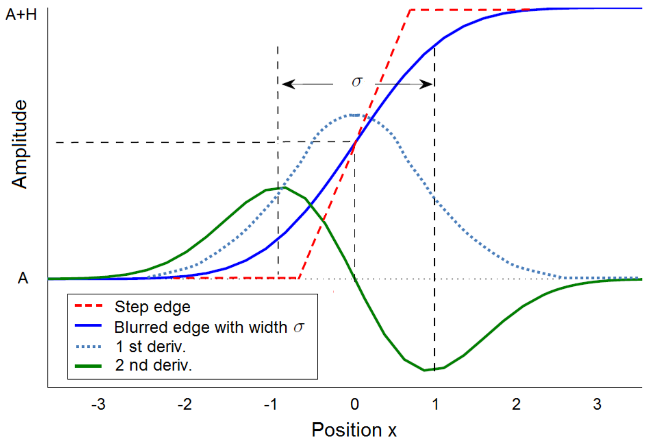


Figure 3.2: Step edge, blurred edge width and its derivatives.

typically to reduce image noise and reduce detail. The visual effect of this blurring technique is a smooth blur resembling that of viewing the image through a translucent screen, distinctly different from the bokeh effect produced by an out-of-focus lens or the shadow of an object under usual illumination. The blur kernel of the edge is modeled by a 1-D Gaussian function:

$$\mathcal{G}(x,\sigma) = \frac{1}{(2\pi)^{1/2}\sigma}\, e^{-\frac{x^2}{2\sigma^2}}, \qquad x \in Z \tag{3.1}$$

where σ is the blur width. As a normalized Gaussian function:

$$\sum_{n\in I} \mathcal{G}(x,\sigma) = \sum_{n\in I} \frac{1}{(2\pi)^{1/2}\sigma}\, e^{-\frac{x^2}{2\sigma^2}} = 1 \tag{3.2}$$

Blurring can be generated by the convolution between the edge segment and a Gaussian blur kernel. Convolution is similar to cross-correlation that two functions $\mathcal{F}$ and $\mathcal{G}$, producing a third function that is typically viewed as a modified version of the original function, giving the area overlap between the two functions (as a function of the amount) that one of the original functions is translated [14], as follows:

$$\begin{aligned}\mathcal{C}(x) &= \mathcal{F}(x-m) * \mathcal{G}(x,\sigma)\\ &= \begin{cases} \frac{H}{2}(1+\sum\limits_{n=-x}^{x}\mathcal{G}(x,\sigma))+A, & x \geq 0,\\ \frac{H}{2}(1-\sum\limits_{n=x+1}^{-x-1}\mathcal{G}(x,\sigma))+A, & x < 0.\end{cases}\end{aligned} \tag{3.3}$$

As a result of convolution, the blurred edge segment is exposed with different width. It can be found that such two widths exhibit different slope at the edge center, as shown in Figure 3.2. In order to obtain a size of the blurred width, we calculate a derivative of blurred edge $\widehat{\mathcal{C}(x)}$:

$$\begin{aligned}|\widehat{\mathcal{C}(x)}| &= |\widehat{\mathcal{F}(x)} * \mathcal{G}(x,\sigma)|\\ &= |H\delta(x) * \mathcal{G}(x,\sigma)|\\ &= \frac{|H|}{(2\pi)^{1/2}\sigma}\, e^{-\frac{x^2}{2\sigma^2}},\end{aligned} \tag{3.4}$$

where $|\cdot|$ is absolute value. We find absolute values that achieves edge center $x = 0$ and the blur width is obtained as follows:

$$\sigma = \frac{1}{(2\pi)^{1/2}} \cdot |\frac{H}{\widehat{\mathcal{C}(x)}}|, \qquad x = 0. \tag{3.5}$$

According to the edge line, the blur kernel is calculated with all edge pixels and obtained the blur widthes,

$$\sigma = \{\sigma_1 ... \sigma_i\}, \qquad i \in 1...n. \tag{3.6}$$

where n denotes the total number of pixels in the edge line.

3.2.2 Discontinuity in Edge Line for Image Splicing Detection

Preserving edge smooth filtering can be formulated by linear fitting. In statistics, a linear fitting (regression) is an approach to modeling the relationship between a dependent variable σ_c and one or more regression variables denoted σ_i, is linear. This relationship is modelled through a disturbance term or error variable ε_i an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors. The case of one regression variable is called simple fitting. Thus the model takes the form of consistency metric, is formulated as,

$$\begin{aligned}\sigma_c &= \beta_1\sigma_1 + \beta_2\sigma_2 + \cdots + \beta_i\sigma_i + \varepsilon_i \\ &= \beta\sigma_i + \varepsilon_i, \qquad\qquad i \in 1...n. \end{aligned} \tag{3.7}$$

where β denotes a dimensional parameter vector. Its elements are also called effects, or regression coefficients σ_i, and ε_i is called the error term, disturbance term, or noise. This variable captures all other factors which influence the dependent variable σ_c other than the regressors σ_i. After preserving edge smooth filtering, the spliced edge can be shown clearer that discriminant strategy is used to identify and discern the spliced fraud edge pixels in image.

$$\sigma_i - \varepsilon_i < \varepsilon < \sigma_i + \varepsilon_i \tag{3.8}$$

Ideally, ε should be a negligible. According to equation (3.8), deviation of consistency metric ε are to be considered as abnormal.

3.2.3 Dependence Between the Edge Blurriness and Illumination for Image Composition Detection

Detection of composed region is based on term of inconsistency metric for lighting direction. In this sense, the method is established mathematical formulation of dependence between defined edge width and the intensity of image.

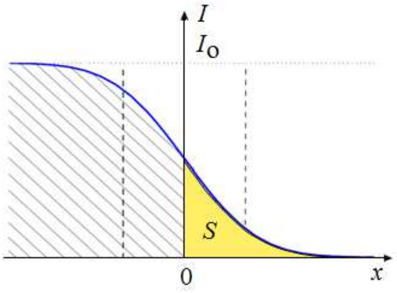


Figure 3.3: Blurred edge segment and its intensity model.

The blur width σ in a specific image point has to be computed by numerical integration of image intensity curve along the orthogonal direction of the stripe edge. A step light projected on the object surface, the blur width is proportional to time rate flow of irradiant light energy in the blurred area. We consider the illumination and its intensity profile. Intensity function on the blurred edge segment is illustrated in Figure 3.3.

The brightness on illuminated scene is convolution of Gaussian kernel and source intensity curve:

$$\begin{aligned}
\mathcal{I}(x) &= \mathcal{I}_0(m) * \mathcal{G}(x-m,\sigma) \\
&= \mathcal{I}_0 \int_{-\infty}^{0} \frac{1}{(2\pi)^{1/2}\sigma}\, e^{-\frac{(x-m)^2}{2\sigma^2}}\, dm \\
&= \mathcal{I}_0 \int_{-\infty}^{0} \frac{1}{(\pi)^{1/2}}\, e^{-(\frac{x-m}{(2)^{1/2}\sigma})^2}\, d(\frac{x-m}{(2)^{1/2}\sigma}) \\
&= \frac{\mathcal{I}_0}{(\pi)^{1/2}} \int_{\frac{x}{(2)^{1/2}\sigma}}^{+\infty} e^{-v^2}\, dv, \qquad (v = \frac{x-m}{(2)^{1/2}\sigma}).
\end{aligned} \tag{3.9}$$

where $\mathcal{I}$ denotes an intensity on edge segment. The area size under blurring curve is the integration of $\mathcal{I}(x)$ from 0 to $+\infty$,

$$\mathcal{S} = \int_0^{+\infty} \mathcal{I}(x)\, dx = \frac{\mathcal{I}_0}{(\pi)^{1/2}} \int_0^{+\infty} \int_{\frac{x}{(2)^{1/2}\sigma}}^{+\infty} e^{-v^2}\, dvdx$$

$$
\begin{aligned}
&= \frac{\mathcal{I}_0}{(\pi)^{1/2}} \int_0^{+\infty} \int_0^{(2)^{1/2}\sigma v} e^{-x^2}\, dx dv \\
&= \frac{\mathcal{I}_0}{(\pi)^{1/2}} \int_0^{+\infty} (2)^{1/2}\sigma v e^{-v^2}\, dv \\
&= \frac{\mathcal{I}_0 \sigma}{(2\pi)^{1/2}} \int_0^{+\infty} e^{-v^2}\, d(v^2) = \frac{\mathcal{I}_0 \sigma}{(2\pi)^{1/2}} \qquad (3.10)
\end{aligned}
$$

where $\mathcal{S}$ denotes an illuminated area. Therefore, a step light projected on the object surface, the blur width is proportional to time rate flow of irradiant light energy in the blurred area.

$$
\mathcal{I}_0 = \frac{(2\pi)^{1/2}\mathcal{S}}{\sigma}. \qquad (3.11)
$$

Here, we defined a relation between the image intensity and the edge blur width. The direction of lighting is determined by the highest intensity point of edge segment. Different lighting directions on the objects are produced that the image has faked by the composition.

3.3 Experimental Results

3.3.1 Image Database

In this section, we describe our experiments and discuss the results. We simulated our scheme under a PC with 1.8G Hz Dual CPU, 6G RAM, and Windows Vista platform. The simulation was carried out using Matlab version R2008a. We test our proposed algorithm on the (CASIA TIDE v1.0) image splicing detection dataset [15] including 800 authentic and 926 spliced color images of size 384×256 pixels with JPEG format of different indoor/outdoor scenes, as shown in Figure 3.4. The authentic images were mostly collected from the Corel image dataset and others are taken by digital cameras. The

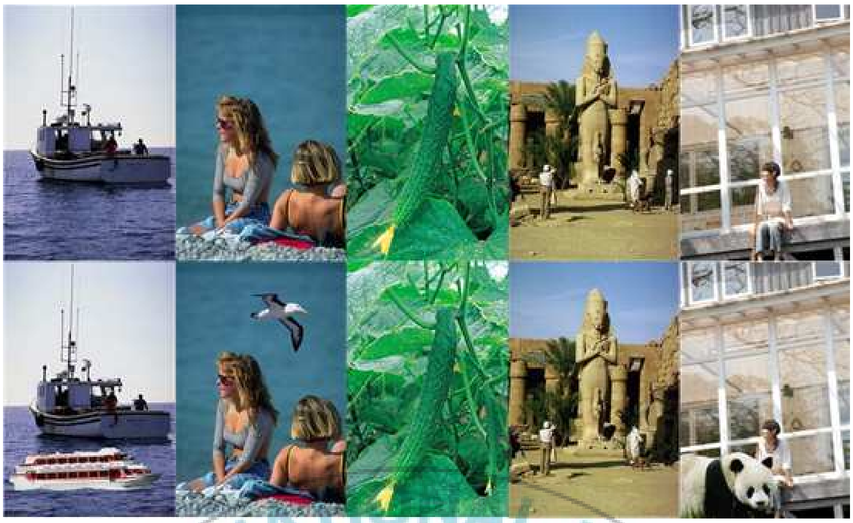


Figure 3.4: Example images for authentic (top row) and forged version (bottom row) from CASIA TIDE v1.0 dataset.

authentic images are divided into several categories (scene, animal, architecture, character, plant, article, nature and texture) according to image content and also consider some criteria based on the information of categories when making spliced images. All tampered images in this database are made only by splicing operation. Spliced images are generated from authentic images by crop-and-paste operation using Adobe Photoshop CS3 version 10.0.1 on Windows XP. Spliced region(s) are either from the same authentic image or from another image.

The spliced set of images [15] are generated by following structures:

- Randomly crop-and-paste image region(s) of different shapes (circle, triangle, rectangle and arbitrary boundaries).
- Cropped image region(s) can be processed with resizing, rotation or other distortion then be pasted to generate a spliced image.

- Difference sizes (small, medium and large) of spliced regions are concerned.

- Most generated spliced images are considered to be realistic images judged by human eyes.

3.3.2 Estimating the Edge Blurring

In order to estimate the edge blurring, we calculated the edge blur width σ_i, $(i \in 1...n)$ according to stripe edge segment, and compared with Hu *et al.* [14] method. We applied same test synthetic images with such method, as shown in Figure 3.5(a). Distance between adjacent step edges is size of 50 pixels. The multiple step edges are blurred by a 1-D Gaussian kernel, with the blur width increasing along the edge segment from 0.1 to 5, as shown in Figure 3.5(b). Figure 3.6 shows our blurring method can perform well on the whole range of blur widthes. The actual curve is defined as the linear increased range. However, our method can provide better the edge blurring than Hu *et al.* [14] method.

3.3.3 Splicing Detection Using Edge Blurriness

In detecting spliced portions of image, the blurred edge segment is applied to estimate the reliable detection of image. Detection process of our scheme is illustrated in Figure 3.7. The blurriness of edge is applied according to detected edge line. In order to show our detection process, we selected two points of blurred edge in forgery image as shown in Figure 3.7(d). The widthes of blurred edge clearly presented the inconsistency of edge. The exposed widthes for selected point 1 and point 2 are illustrated by different edge widthes in

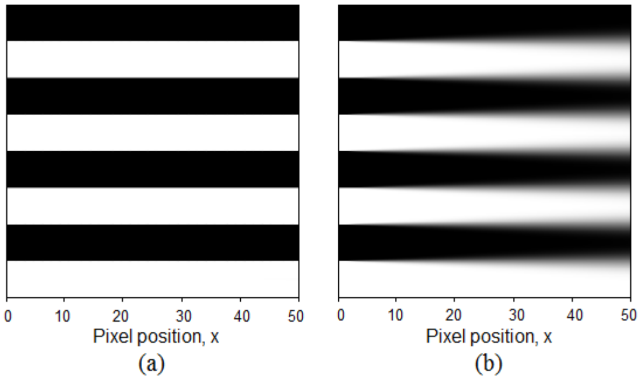


Figure 3.5: (a) Multiple step edges and (b) its blurred version with the blur width 0.1 to 5.

Figure 3.7(e),(f), respectively. Based on this discontinuity of edge, we can detect the boundary of splicing.

3.3.4 Composition Detection Under Illuminant Lighting Direction

A composite image is created by using a cutting and joining process that two or more objects copied from a different image and pasted into single image. Figure 3.8(d) and 3.8(e) show a composed image and detected edges are indicated. The image edge is detected and the blurring is calculated by edge blur model. The intensity or irradiant light energy is exposed using equation (3.11). However, the inconsistency of lighting direction has a close relationship with image illumination, which is then used as an evidence of image compositions. As shown in Figure 3.8(f), the tampered regions are detected and marked as blue line (edge). The results of splicing and composition detection are described in Figure 3.8, respectively.

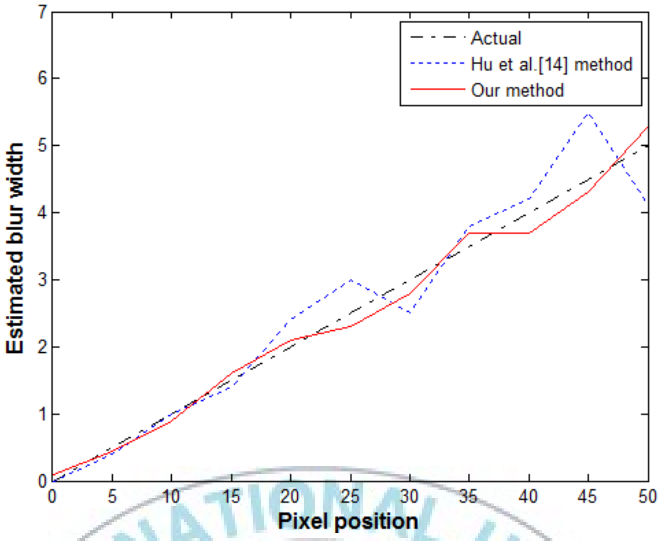


Figure 3.6: Results of the edge blur estimation.

3.3.5 Performance Analysis with ROC Curves of PDA/PFP Rates

We use two quantitative measures to evaluate the performance of our method. Denote Ω as pixels in the true duplicated regions (both the source and its duplicates), and $\tilde{\Omega}$ as pixels in the detected duplicated regions, we define the pixel detection accuracy (PDA) rate as the fraction of pixels in duplicated regions that are correctly identified,

$$PDA = \frac{\mid \tilde{\Omega} \cap \Omega \mid}{\mid \Omega \mid}. \tag{3.12}$$

and the pixel false positive (PFP) rate as the fraction of pixels in untampered regions that are detected as from duplicated regions,

$$PFP = \frac{\mid \tilde{\Omega} - \Omega \mid}{\mid \tilde{\Omega} \mid}. \tag{3.13}$$

Combining the PDA and PFP rates in a Receiver Operator Characteristics (ROC) curve [16] provides a comprehensive evaluation of the detection per-

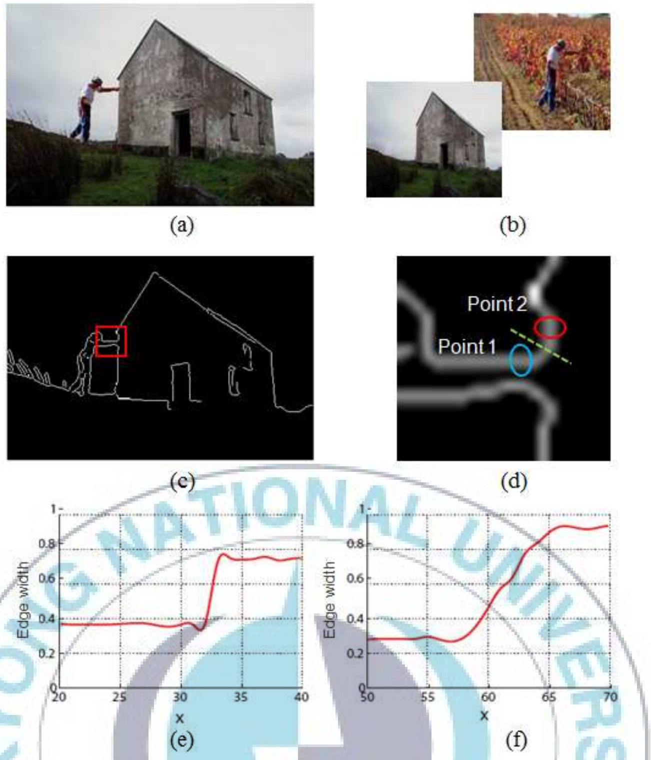


Figure 3.7: Splicing detection test. (a) Spliced image and (b) its authentic images, (c) Detected the edge segment, (d) Blurred edge segment, and Exposed edge widthes for on (e) Point 1 and (f) Point 2.

formance. We evaluated the images, either spliced or authentic images, into three categories from the database: animal, character, and nature. The total number of samples used in experiment is 300 images.

a) Detection Performance for Image Splicing. To evaluate the detection performance for splicing, we conduct number of tests on 300 spliced images. Figure 3.9 confirms the effectiveness for detecting spliced boundary of our scheme with *ROC* curve. As shown in Table 3.1, we compared this result with other previous works for relating splicing detection. In Zhongwei

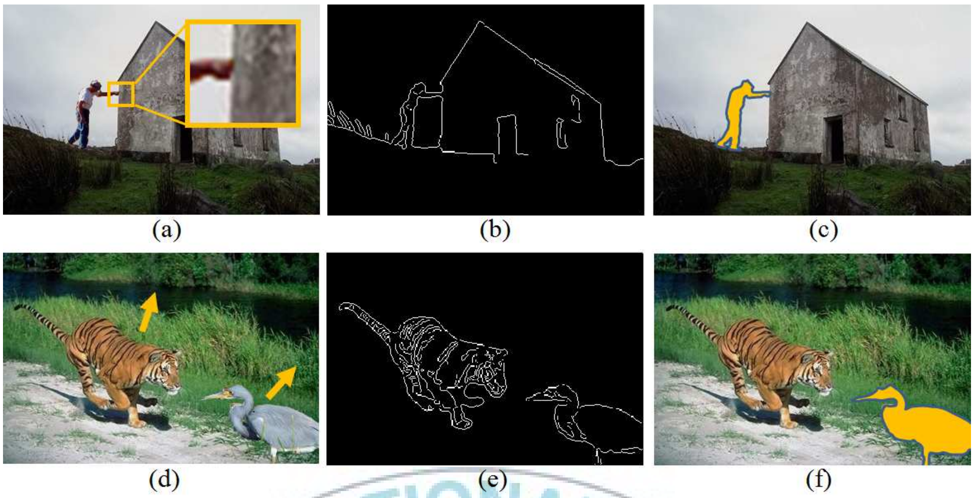


Figure 3.8: Tampering localization. (a),(d) Spliced and composed images, (b),(c) Detected edges, (c),(f) Localization of tampered portions.

Table 3.1: The performance comparison.

Feature Set	*Accuracy*(%)	*Extration Time*(*s*)
SP [18]	74.27	0.2005
ARL + *SP* [17]	80.02	0.2346
Our Method	89.30	0.3520

et al. [17] method, they analyzed the discontinuity of image pixel correlation and coherency caused by splicing in terms of image run-length representation and sharp image characteristics. features are extracted by exploiting both magnitude and phase information of a given image. This image model has been shown to be able to catch changes caused by image splicing. Our performance comparison is selected the edge based feature vector [18] (denoted as SP feature set) and its improved feature vector [17] (denoted as $ARL + SP$). The feature set SP and $ARL + SP$ can reach a detection accuracy 74.27% and 80.02%, respectively. But, our scheme can provide the higher detection

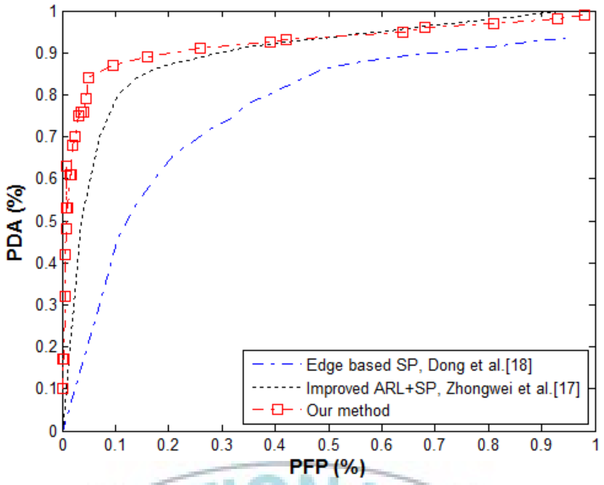


Figure 3.9: ROC curves of PDA and PFP rates for splicing detection.

accuracy 89.30% than both of methods.

a) Detection Performance for Image Composition. We classified the spliced image dataset into three categories as tampering rates, such as 5-20% of whole image is tampered and 21-40%, 41-60%, respectively. Our scheme can detect the image composition by applying irradiant light energy of the edge segment. Detection performance is higher, when tampering rate is lower, as shown in Figure 3.10.

3.4 Discussion

In this work, we described a passive method for detecting digital image splicing and composition based on image edge blurring with a Gaussian kernel. These manipulations are detected by compatibility metrics, such as inconsistency of edge widthes and irradiant lighting direction. The detection of image splicing is assessed based on linear fitting metric. The edge blurriness is efficient and

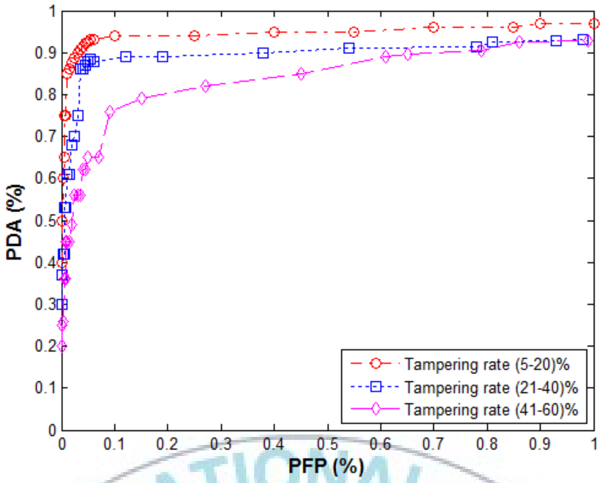


Figure 3.10: ROC curves of PDA and PFP rates for composition detection.

reliable for detecting image splicing and composition both on theory and practical test. Experimental results show promising performance in detecting and locating trials of image splicing and composition manipulations, respectively.

Chapter 4. Robust Copy-Move Forgery Detection Based on Dual-Transform

4.1 Introduction

With the ever increasing diffusion of simple and powerful software tools for digital source editing, image tampering is becoming more common, stimulating an intense quest for algorithms, to be used in the forensics field, which help deciding about the integrity of digital images. Furthermore, it is necessary for us to develop automatic methods to authenticate the images and indicate potential forgeries. Due to the variety of manipulations and the diversity of individual characteristics of media, passive approach usually faces difficulties at a larger scope, and suffers from complicated and time consuming problems [19].

One of the most common type of image forgeries is the copy-move forgery [20], where a region from one part of an image is copied and pasted onto another part in same image, thereby concealing the image content in the latter region. Such concealment can be used to hide an undesired object or increase the number of objects apparently present in the image. Two real-world examples have been as shown in Figure 4.1. The underlying assumption is that a region is copied and pasted within the same image. Although a simple translation may be sufficient in many cases, additional operations are often performed

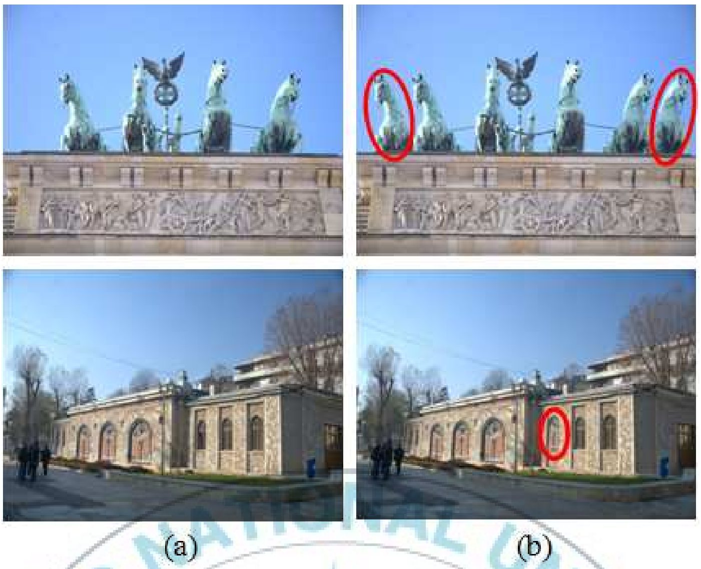


Figure 4.1: Real-world examples of the copy-move forgery. (a) Authentic images and its (b) manipulated versions.

in order to better hide the tampering. These include rotation, scaling, lossy compression, noise contamination, blurring, and among others. In general, if a duplicated region within the same image is found, the manipulation is directly proven. Also, the copied part comes from the same image, all of its properties and statistic information are the same as the rest of the image. Thus, it is difficult to detect forgeries by techniques that compare statistics of different part of an image to each other. Hence, in order to be able to reliably detect such forgeries, a several techniques have been recently proposed which try to be robust to some of these transformations in following Section 4.2.

The remainder of this chapter is organized as follows: The related works for copy-move forgery detection are presented in Section 4.2. General framework for copy-move forgery detection in Section 4.3. Section 4.4 introduces the concept of dual-transform, which includes Radon and *DCT* transforms. The proposed method is presented in Section 4.5. The experimental results are provided in Section 4.6. Discussion is drawn in Section 4.7.

4.2 Review on Copy-Move Forgery Detection

Image copy-move forgery detection methods follow existing two processing alternatives. The methods are either classified into *keypoint-based* and *block-based* categories as shown in Figure 4.2. The block-based methods can be grouped in four categories: moment-based [23, 24], frequency-based [28, 80], intensity-based [81, 82], and dimensionality reduction-based [83, 84] features. But, keypoint-based [43, 85, 86] methods rely on the identification and selection of high-entropy image regions. Whereas in block-based techniques a feature vector was computed per block, keypoint-based approaches extract a feature vector per keypoint. Consequently, fewer feature vectors are estimated, resulting in reduced (by an order of magnitude) computational complexity of feature matching and post-processing. We examined two different versions of keypoint-based feature vectors. One uses the SIFT [87] features while the other uses the SURF [86] features.

In the literature, researchers have developed various techniques. According to review in previous works, the methods are vulnerable to post-processing operations. Huang *et al.* [21] proposed improved robustness using a discrete cosine transform (DCT) to noise addition, global blurring and lossy compression, but does not deal with geometrical transformations of the tampered region. The method of Khan *et al.* [22] reduces the time complexity of the PCA-based approach by using a discrete wavelet transform (DWT), but also does not address geometrical transformations. In [23], Mahdian *et al.* took ad-

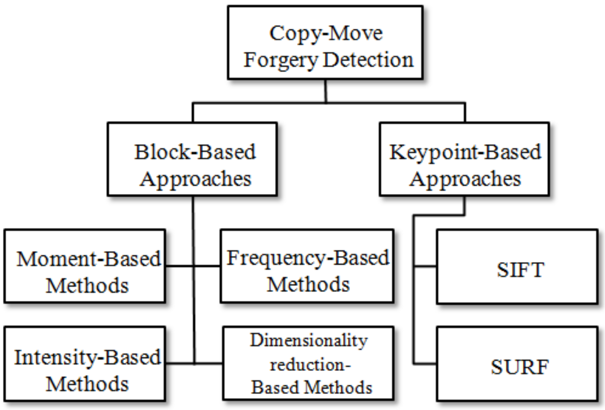


Figure 4.2: Categories for copy-move forgery detection methods.

vantage of the blur invariant moments to extract the block features. Though these methods can detect the copy-move forgery in most cases, they may fail if the copied regions are rotated or flipped. Ryu *et al.* [24] employed Zernike moments to extract the features for block matching. This method achieved an average detection precision rate of 83.59% in the case of region rotation. In [25], Liu *et al.* proposed a method using Hu moments to extract the features of the blocks. This method is robust not only to noise contamination, *JPEG* compression and blurring, but also to moderate rotation.

4.3 General Framework

In this section, we describe a feature vector computation and detection mechanism in detail. A number of methods have been proposed, the general workflow is typically similar [20] as following:

- **Feature extraction**. Block-based methods subdivide the image in rectangular regions or blocks. For every block, a feature vector is computed, and then similar feature vectors are matched. By contrast, keypoint-based methods do not perform explicit image subdivision. Instead, their feature vector is computed at image regions containing high entropy. Subsequently, similar features within an image are matched. A forgery shall be reported if regions of such matches cluster to larger areas. Both, keypoint and block-based methods include further filtering for removing spurious matches. An optional post-processing step of the detected regions may also be performed, to group matches that jointly follow a transformation pattern.

- **Similarity matching**. High similarity between two feature descriptors is interpreted as a cue for a manipulated region. Most authors propose the use of lexicographic sorting in identifying similar feature vectors [81]. In lexicographic sorting, a matrix of feature vectors is built so that every feature vector becomes a row in the matrix. This matrix is then row-wise sorted. Thus, the most similar features are in consecutive rows to each other. Other methods use the Best-Bin-First search method [88] derived from the kd-tree algorithm to get approximate nearest neighbors with lower computational cost than the original kd-tree [89].

- **Filtering**. Filtering schemes have been proposed in order to reduce the probability of false matches. For instance, a common noise suppression measure is the removal of matches between spatially close regions. Neighboring pixels often have similar intensities, which can lead to false forgery detection. Additionally, different distance criteria have been proposed to filter out weak matches. For example, several authors proposed the Euclidean distance between matched feature vectors such as in paper

[24]. In contrast, the method [81] use the correlation coefficient between two feature vectors as a similarity criterion.

4.4 The Concept of Dual-Transform

4.4.1 Radon Transform (RT)

Applying Radon transform on an image $f(x, y)$ for a given set of angles can be thought of as computing the projection of the image along the given angles [26]. The resulting projection is the sum of the intensities of the pixels in each direction, *i.e.* a line integral. For an image $f : \mathbb{R} \times \mathbb{R} \to [0, 255]$ containing an object, the result g of Radon transform is a function $\mathcal{R} : \mathbb{R} \times [0, 2\pi) \to \mathbb{R}_+$ defined as:

$$g(s, \vartheta) = \mathcal{R}(f(x, y)) = \int_{-\infty}^{\infty} f(s\cos\vartheta - t\sin\vartheta, s\sin\vartheta + t\cos\vartheta)dt \tag{4.1}$$

$$\begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{4.2}$$

Radon transform of the translated, rotated and scaled images exhibits interesting properties, which can be employed to construct a method for invariant object recognition. Therefore, the behavior of the transform for these three variations in the input image should be defined. Any translation in spatial domain leads in the Radon domain to translation in the s direction. The amount of the translation varies with the ϑ dimension. The scaling of the original image along both axes results in the scaling along the s axis in the Radon domain. The value of the transform is also scaled. The rotation in spatial domain leads to circular translation along the ϑ axis in the Radon domain.

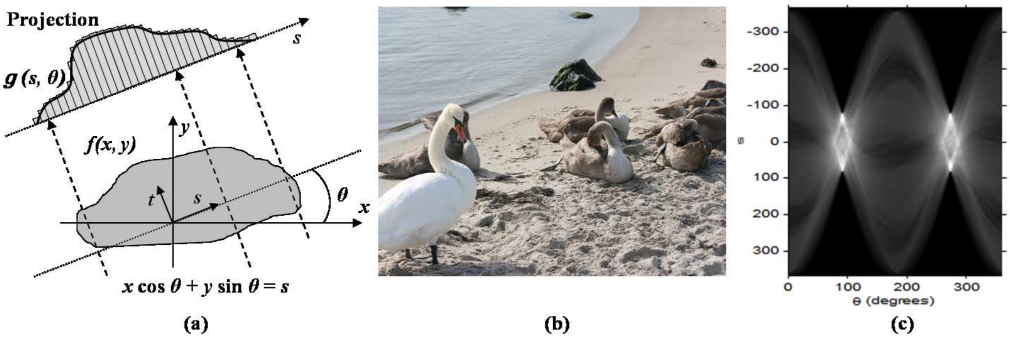


Figure 4.3: Radon transform. (a) Image projection, (b) Test image, and (c) Its projection on Radon space.

The behaviour of Radon transform is summarized in Table 4.1, and depicted in Figure 4.3.

Table 4.1: Behavior of Radon transform for rotated, scaled and translated images.

Behavior	Image function, f	Radon transform, $g = \mathcal{R}(f)$.
Original	$f(x, y)$	$g(s, \vartheta)$
Rotated	$f_{polar}(r, \vartheta_0 + \varphi)$	$g(s, (\vartheta + \vartheta_0)\ mod\ 2\pi)$
Scaled	$f(\alpha x, \alpha y)$	$\frac{1}{\lvert\alpha\rvert}\ g(\alpha s, \vartheta)$
Translated	$f(x - x_0, y - y_0)$	$g(s - x_0 \cos\vartheta - y_0 \sin\vartheta, \vartheta)$

4.4.2 Discrete Cosine Transform (DCT)

Discrete cosine transform is used to know frequency components present in a image [27]. DCT mainly reduces the redundant information present in the image by omitting the undesired parts of the image. Orthogonality, symmetry, separability, and decorrelation are important properties of DCT. The most common *DCT* definition of a 1D sequence of length N is

$$C(u) = \alpha(u) \sum_{x=0}^{N-1} f(x) \cos\left[\frac{\pi(2x+1)u}{2N}\right], \tag{4.3}$$

for $u = 0, 1, ..., N-1$. In equation (4.3), $\alpha(u)$ is defined as

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & \text{for} \quad u = 0 \\ \sqrt{\frac{2}{N}} & \text{for} \quad u \neq 0. \end{cases} \tag{4.4}$$

The *DCT* coefficients for the transformed output image $C(u)$ with an input image $f(x)$ can be calculated by using the equation (4.3). N is the pixel dimensions of the input image $f(x)$. The intensity value of the pixel N of the image is given by $f(x)$ and $C(u)$ is the *DCT* coefficients in u of the *DCT* matrix.

4.5 Robust Copy-Move Forgery Detection

In this section, we present the proposed robust copy-move forgery detection method based on dual transform. At first, we describe a model for copy-move forgery in digital images, and then introduce our proposed method to detect such specific artifact.

4.5.1 Model for Copy-Move Forgery

The task of finding the copy-move forgery is that of finding at least two large similar regions. Given an image $I(x, y)$, the tampered image $I'(x, y)$, must subject to: $\exists$ regions D_1 and D_2 are subsets of D and a shift vector $d = (dx, dy)$, (we assume that $|D_1| = |D_2| > |D| * 0.85\%$ and $|d| > L$), $I'(x, y) = I(x, y)$ if $(x, y) \notin D_2$ and $I'(x, y) = I(x - dx, y - dy)$ if $(x, y) \in D_2$, where

D_1 is the source and D_2 is the target region, $D_2 = D_1 + d$. We consider that the similarity of the target region is larger than 0.85% of the image size. It would be easy to detect above forgery via exact match. However, to make the tampered image harder to detect, the attacker may perform various processing on $I'(x, y)$. Then the tampered image becomes $I''(x, y) = \xi(I'(x, y))$, where ξ is the post-processing operator, which includes geometrical and image processing operations. The post-processing attack makes the task of detecting forgery significantly harder. In the next section, we present an efficient method for detecting copy-move forgery which is also robust against various forms of post-processing operations.

4.5.2 Proposed Method

Our proposed method is based on dual-transform, which includes Radon and discrete cosine transformations. This set of transformations were designed for an efficient and robust approach. The main issue in directly applying these tools to image forgery detection is that these tools were designed to find duplicate but separate, images, whereas we are trying to find identical regions in same image. We perform modifications in the feature extraction and matching processes to efficiently detect such forgery. Firstly, we apply Radon transform on each divided blocks to project the image into a directional projection space, then perform 1-D *DCT* to derive the frequency features from the Radon space. Following we select the *DCT* coefficients with low frequency by using a dimension reduction. Finally, an invariant robust features are extracted. The details of the proposed method is given as the following:

1. **Pre-processing.** Given image I is tiled by overlapping blocks of $b \times b$

pixels. Blocks are horizontally slid by one pixel rightwards starting with upper left corner and ending with the bottom right corner. The total number of overlapping blocks for an image of $M \times N$ pixels is $S_{blocks} = (M-b+1) \times (N-b+1)$, for each block $B_l(l = 1, ..., S_{block})$. For instance, an image with the size of 640×480 with blocks of size 8×8 yields $299,409$ overlapping blocks.

2. **Feature extraction.** Each block is applied Radon transform, the space is projected on the Radon space. The results of Radon transform are contained in the columns of a matrix with the number of projections generated being equal to the number of the defined angles, $(\vartheta_1, \vartheta_2, ..., \vartheta_n)$. Then, delete the rows in projection matrix, which are composed of 0. This will remove the redundancy data generated by Radon transform.

   On each projection (represented by column of the projection matrix) according to projection angles, we apply 1-D DCT to derive the frequency features from the Radon space. We quantize the coefficients according to the $JPEG$ quantization table using a predetermined quality factor Q. The quantized coefficients can be denoted as $\mathrm{c}_k = \{c_1, c_2, ..., c_k\}$. The dimension reduction can make the sorting and matching faster. The frequency features are the nature of 1-D DCT that the energy of transformed DCT coefficients will be focused on the first several values (lower frequency values). Thus, those higher frequency coefficients can be truncated. The truncation can be done by saving only a part of vector components. Here, we define a factor $p, (0 < p \leq 1)$, that only first $\lceil p \times k \rceil$ DCT coefficients are saved for further processing. $\mathrm{c}_r = \{c_1, c_2, ..., c_r\}$, $(r = \lceil p \times k \rceil,\ r < k)$, where p denotes a saved the percentage of DCT coefficients and k denotes the number of coefficients on the projections according to angles ϑ_n. For example, we select the projection angle $\vartheta = 8$, and derived the 1-D DCT coefficients (column matrix 15×1) from the projection space. Five coefficients are deleted, which are composed of 0. The concentration of energy in 80% is calculated as, $\lceil p * k \rceil = \lceil 0.8 * 10 \rceil = 8$ coefficients.

The truncated DCT coefficients in projection matrix are sorted by a lexicographically order. Let the matrix C denote the sorted vectors, the size of the matrix will be C_r^m.

$$C = \begin{bmatrix} \mathrm{C}_1^1 & \mathrm{C}_2^1 & \dots & \mathrm{C}_r^1 \\ \mathrm{C}_1^2 & \mathrm{C}_2^2 & \dots & \mathrm{C}_r^2 \\ . & . & \dots & . \\ \mathrm{C}_1^m & \mathrm{C}_2^m & \dots & \mathrm{C}_r^m \end{bmatrix}_{(M-b+1)(N-b+1)} \qquad (4.5)$$

By using a lexicographic sorting, similar features will locate at the neighboring rows and the feature matching can be achieved in a small range.

3. **Similarity matching.** The feature matching is to find out the corresponding similar rows from between m rows of the C matrix. In order to detect the forged region correctly, the similarity threshold τ_s and the distance threshold τ_d should be predetermined, respectively. In our method, we search for the corresponding rows by estimating the Euclidean distance of feature vectors, as follows:

$$D(\mathrm{C}_r^m, \mathrm{C}_r^{m+v}) = \sqrt{\sum_{r=1}^{u} \mathrm{C}_r^m - \mathrm{C}_r^{m+v})^2} < \tau_s \qquad (4.6)$$

If $D(\mathrm{C}_r^m, \mathrm{C}_r^{m+v})$ is smaller than a threshold τ_s, the corresponding features will be regard as correctly matched. Then the locations of two features are stored. The matching will be repeated for all rows of C. Since the feature vectors of the rows are quite similar with each other which have the overlapping pixels, only the rows with the actual distance between two similar features are compared as follows:

$$L(\mathrm{C}_r^m, \mathrm{C}_r^{m+v}) = \sqrt{(x_i - x_{i+j})^2 + (y_i - y_{i+j})^2} > \tau_d \tag{4.7}$$

where x and y are the coordinates of the corresponding features.

4. **Detection.** When all the matched feature pairs are saved, which is achieved by marking the copied and forged regions, respectively. Generally speaking, the regions are stamped on a binary image. That is to say, all the detected features including the forged and un-forged features are marked to generate a detection map. Figure 4.4 shows an example of the proposed method for marking. In general, there are some falsely detected features marked on the initial detection map in Figure 4.4(c), and these falsely detected features should be removed by filtering in Figure 4.4(d). For the filtering, we generate a sliding window with the size of 8 × 8 pixels, and move it from left to right and up to bottom. Each time, the window moves forward by 8 pixels to make sure all the pixels of the image will be filtered and each pixel will be filtered only once. If the number of white pixels are less than 60 in the window, all pixels of the window are marked as black. Otherwise, keep the number of the white pixels and do nothing. After filtering, some small isolated false matches can be removed. Figure 4.4(d) shows the detection result after the filtering operation.

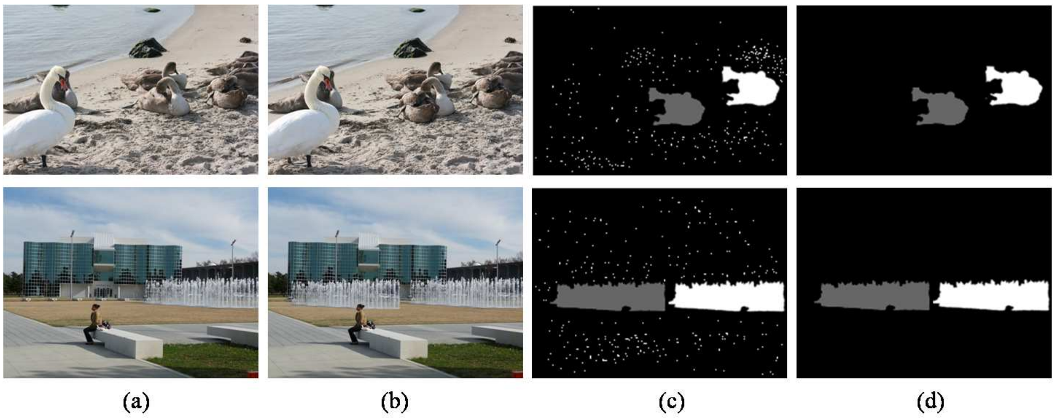


Figure 4.4: Image forgery detection. (a) Original image, (b) Forged image, (c) Detected forgery with similar features, and (d) Results after filtering.

4.6 Experimental Results

In this section, we present the experimental results of our proposed method. We simulated our method under a PC with 3.2G Hz Core i5 CPU, 8G RAM, and Windows 8 platform. The simulation was carried out using Matlab version R2008a. We test our method on Benchmark data for image copy-move detection dataset including 120 authentic and 124 forged color images of size 3888×2592 pixels with different outdoor scenes, as shown in Figure 4.5. The authentic images were taken by different digital cameras. All tampered images in this dataset are generated from the authentic images by crop-and-paste operation using Adobe Photoshop CS3 version 10.0.1 on Windows XP. The tampered regions are from the same authentic image.

Figure 4.5: Samples of test images.

4.6.1 Robustness Test for Feature Vectors

We extracted the features, which expressed by DCT coefficients of 1-D DCT based on the Radon space. These features will not change a lot after some post-processing operations. We have defined the model for copy-move forgery in Section 4.5.1. If an image is contaminated by additive Gaussian noise operation ($AWGN$), then the pixel value will be changed, for each pixel, we define $I(x, y) = \lfloor I(x, y) \rfloor + \xi_{noise}, (0 < \xi < 1)$, where $I(x, y)$ is the corresponding pixel value that contaminated by signal noise, $\lfloor I(x, y) \rfloor$ is the nearest value less than or equal to the original pixel value, ξ_{noise} is the random noise which is independent identically distributed. For instance, each noisy block $B'_i = B_i + \xi_{noise}$, and the extracted features $c'_r = c_r + \xi'_{noise}$, since $E(\xi'_{noise}) = 0$, $D(\xi'_{noise}) = \sum_{i=1}^{b^2} \xi'_{noise}/b^2$, generally $\sum_{i=1}^{b^2} (\xi')^2_{noise} \ll b^2$. Since we get $c'_r \approx c_r$. For the Gaussian blurring only affects in some high frequency components of each blocks, but changes in the low frequency components are a little. The robustness against the geometrical operations are provided by the property of

Table 4.2: The correlation coefficients for the feature vectors, $\vartheta = 8$, (8×8).

Vectors	Extracted, c_r	Post-processed, c_ξ					
		AWGN	AWGN	Blurring	Blurring	JPEG	JPEG
		SNR	SNR	w, σ	w, σ	Q	Q
		$25dB$	$50dB$	3, 1	5, 0.5	5	10
c_1	958.75	959.26	962.31	957.45	959.07	958.26	962.12
c_2	886.37	893.63	896.25	884.16	886.36	884.69	887.02
c_3	875.12	885.02	894.89	873.52	874.85	873.81	878.29
c_4	801.50	820.75	828.20	799.21	802.80	798.68	796.93
c_5	745.25	753.39	761.62	744.03	746.68	748.52	736.84
Correlation coefficients		0.9980	0.9804	1.0000	1.0000	1.0000	1.0000

Radon transform. In order to show the robustness of the feature vectors, we chose a size of block 8×8, 16×16, and 32×32, respectively, from the natural images. Then we applied some post-processing operations with different parameters. The results of robustness test are presented in Table 4.2. c_r and c_ξ are feature vectors that the extracted and post-processed vectors, respectively. After some post-processing, we calculate the correlation coefficients between them, if the result is close to 1, which implies the feature vector is robust and the invariance is more stable. The correlation coefficient is used as a measure of correlation, as it is invariant to intensity change. (Here we note that the extracted feature vectors are reduced by dimension reduction.)

4.6.2 The Evaluation of the Detection Performance.

In order to quantify the accuracy of detection, the true positive ratio (TPR) and the false positive ratio (FPR) are employed, as follows:

$$TPR = \frac{|\Omega_1 \bigcap \Omega_2| + |\overline{\Omega_1} \bigcap \overline{\Omega_2}|}{|\Omega_1| + |\overline{\Omega_2}|}, \qquad FPR = \frac{|\Omega_1 \bigcup \Omega_2| + |\overline{\Omega_1} \bigcup \overline{\Omega_2}|}{|\Omega_1| + |\overline{\Omega_1}|} - 1 \quad (4.8)$$

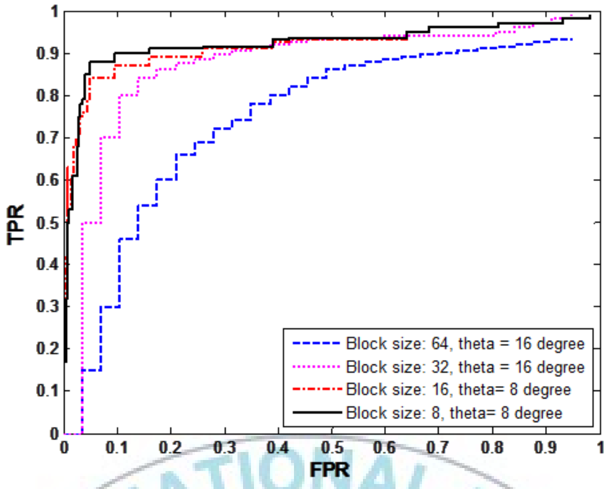


Figure 4.6: Detection results for varying block sizes.

where Ω_1 and Ω_2 are the original copied region and the detected copied region, while $\overline{\Omega_1}$ and $\overline{\Omega_2}$ are the forged region and the detected forged region, respectively. In order to set the threshold parameters, we randomly chose 50 images from the dataset and then make a series of forgeries. After that, we use different the projection angles ranging from 8 to 64 degree with 8 increment, then a set of values for $\tau_s = 0.005$ and $\tau_d = 4$, respectively, from the number of testing results. The threshold parameters are chosen by highest true positive ratio with corresponding lowest false positive ratio. In order to decide the block size, we tested the TPR and FPR curves for various block sizes with a selection of different directional projection angles.

As shown in Figure 4.6, we notice that smaller block size is resulted higher detectability property. But, large block size is indicated lowest detection performance. Therefore, we set the block size of 8×8 pixels in all our following

Table 4.3: The feature matching accuracies with various post-processing operations.

Operations		*Compression*			*Additive Gaussian noise*		
		JPEG 30	JPEG 60	JPEG 90	SNR 10	SNR 20	SNR 30
Rotation	10^o	0.979	0.982	0.987	0.969	0.971	0.975
	30^o	0.971	0.974	0.985	0.950	0.956	0.969
	45^o	0.963	0.966	0.976	0.936	0.938	0.948
Scaling	5	0.984	0.984	0.987	0.974	0.975	0.978
	10	0.982	0.983	0.988	0.968	0.971	0.979
	15	0.965	0.976	0.978	0.956	0.964	0.966
Blurring	3×3	0.970	0.972	0.976	0.931	0.948	0.951
	5×5	0.962	0.968	0.971	0.920	0.927	0.939
	7×7	0.927	0.931	0.935	0.901	0.917	0.919
Contrast changing	10	0.975	0.976	0.976	0.970	0.973	0.976
	30	0.973	0.970	0.974	0.960	0.966	0.968
	45	0.967	0.966	0.966	0.947	0.956	0.957
Rot. + Flip	10^o, Hor.	0.889	0.898	0.897	0.836	0.847	0.848
Sc. + Flip	10, Ver.	0.885	0.890	0.893	0.825	0.826	0.825
Rot. + Sc.	10^o, 10	0.738	0.768	0.787	0.704	0.731	0.747

experiments.

a) The performance of the feature matching. We evaluated the feature matching process that the copied regions have been subjected to various geometrical operations (rotation, scaling and flipping) and image processing operations (blurring and contrast changing). Additionally varying the levels of lossy compression ($JPEG$) and the additive Gaussian noise ($AWGN$) were performed with mixture operations. The purpose of this testing is to highlight the performance of features that we have employed. The accuracies of the feature matching are determined by proportion of true positives in the matching feature pairs. The obtained results are reported in Table 4.3.

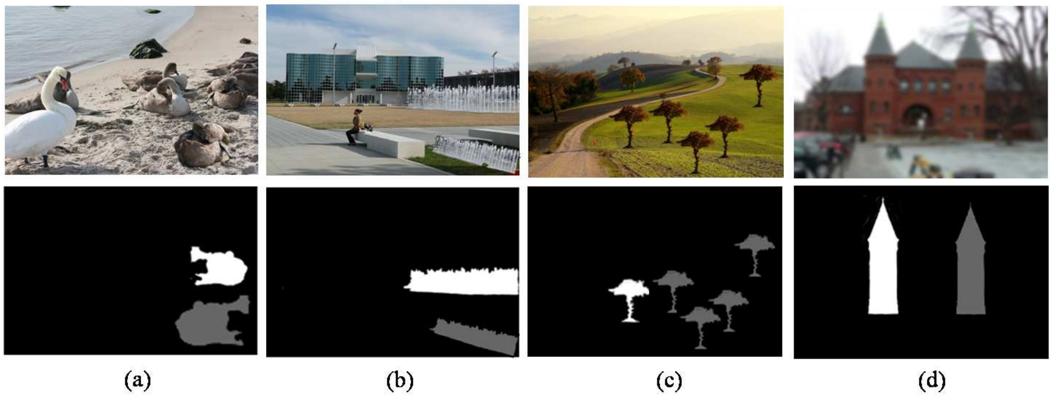


Figure 4.7: Detection results with various mixture operations. (a) Object scaling with horizontally flipping, (b) Object scaling with rotation, (c) Multi-copy with JPEG, and (d) Blurring with scaling.

In Table 4.3, the mixture operations tend to have somewhat lower accuracy than other operations, which is shown at low quality factors and signal noise ratio (SNR). Especially, the accuracies for blurring and contrast changing indicate lower layer among of individual operations, respectively. Nevertheless, TPR and FPR are quite acceptible even with low quality factors and signal noise ratio.

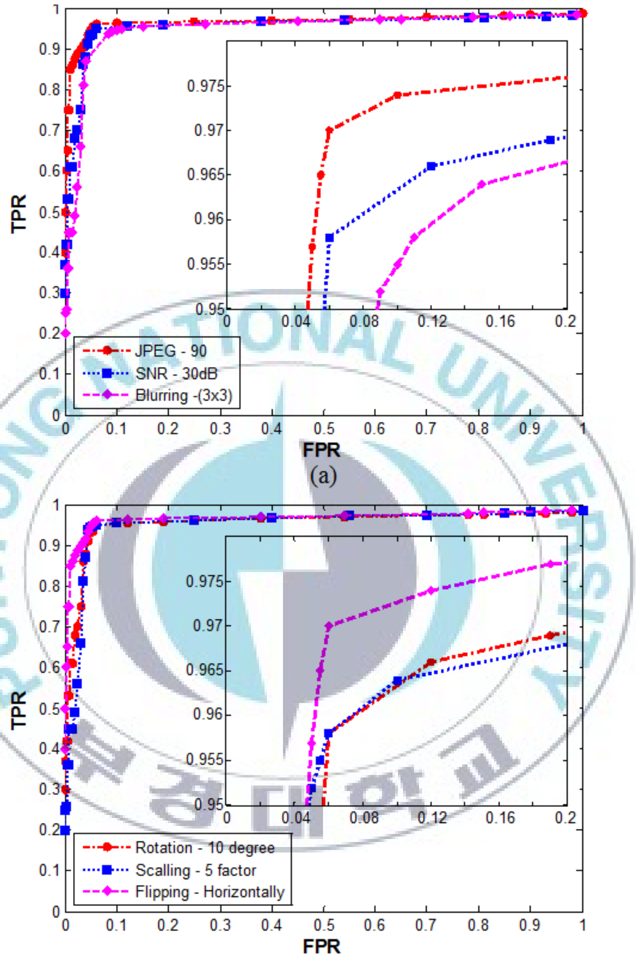
b) The robustness against post-processing operations. The advantage of the proposed method is that it can resist against geometrical and image processing operations. In order to test the efficiency and robustness of our method further, we test all images from Benchmark dataset. For each image, a random sized region was copied then pasted onto a non-overlapping position, while the copied regions are distorted by different mixture post-processing operations. For instance, as shown in Figure 4.7, the copied region is distorted by scaling with horizontal flipping, rotation with scaling, multi-copy with $JPEG$, and blurring with scaling, respectively. From the results we show that the forged regions can be detected accurately. Figure 4.8 presents the detection results of our method on various kinds of individual post-processing operations. As can be seen, we are able to attain quite high accuracies at low false positive rates in selection of higher rate values. In the case of blurring, it can be seen that the resistance of such operation is lower than other post-processing operations.

c) The performance comparisons. The overall average performance comparisons of our method with other related work are performed more precisely in this section. Some invariant feature extraction methods for copy-move forgery are presented in Fridrich [28], Huang *et al.* [29], and Li *et al.* [30]. As shown in Figure 4.9(a-b), the forged images are contaminated with additive Gaussian noise ($5dB \leq SNR \leq 35dB$). Fridrich's method has the lowest TPR than other methods, when less than $10dB$, the TPR is approximate to zero. Observation of TPR in our method achieves higher TPR among other methods. For FPR, Fridrich's method has lower FPR value, that cannot detect any forged region, when the FPR is less than $15dB$. However, such method quickly leads to higher FPR when the SNR level is higher, which indicates it is sensitive to noise adding. Our method have a better performance with Li *et al.*'s method, however with lower FPR.

In case of blurring, the forged regions are blurred by a Gaussian blurring filter ($w = 5$, $\sigma = 1$ to 7). Figure 4.9(c-d) shows the TPR curve of our method has better performance followed by Li *et al.*'s method, however, the TPR curves of Fridrich and Huang *et al.* are drop significantly, when the blurring radius increased. In FPR, our method has the lowest value, even increased the larger blurring radius.

4.7 Discussion

In this work, we proposed a robust copy-move forgery detection method for a suspicious image. To extract an invariant robust features of a given image, we applied dual-transform. The extracted features are represented by lexicographically ordered DCT coefficients on the frequency domain from the Radon space, that each overlapped image blocks are projected by the columns of a matrix with the number of the defined angles ϑ_n on the Radon domain. Experimental results supported that the proposed method was appropriated to identify and localize the copy-move forgery even when though the forged region had been manipulated intentionally. The main contribution of our work is a method capable of easily detecting traces of various attacks. We concerned the geometrical and image processing operations, and any of their arbitrary combinations. The detection performance of our method is satisfactory enough and meets the robustness criteria.



(a)

(b)

Figure 4.8: Detection results with various attacks. (a) Image processing operations, and (b) Geometrical operations.

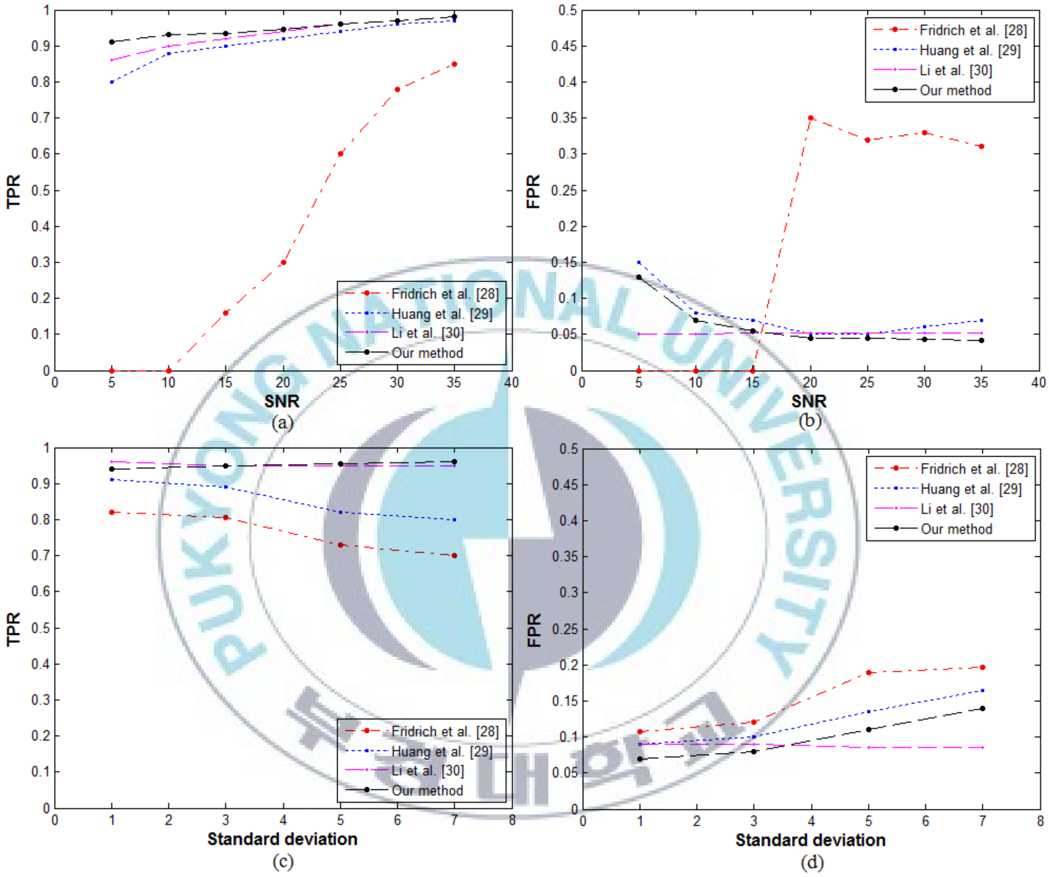


Figure 4.9: Detection results with TPR/FPR curves. The performance comparisons (a-b) with different the SNR levels ($5dB \leq SNR \leq 35dB$), and (c-d) with Gaussian blurring (w $= 5,\ 1 \leq \sigma \leq 7$).

Chapter 5. A Counter-Forensics Method for SIFT-Based Copy-Move Forgery Detection

5.1 Introduction

Digital imaging has experienced tremendous growth in recent decades, and digital camera images have been used in a growing number of applications. With such increasing popularity and the availability of low-cost image editing software, the integrity of digital image content can no longer be taken for granted. Therefore, research on digital image forensics and tamper detection has gained ground [31].

However, every image forensics tool has assumed that the image forger has not taken any countering measure to remove its trace. In reality, like every information security field, vulnerabilities in existing forensics tools will be exploited, and modified images will not only fool our eyes, but also pass safely though detection programs. Thus there is urgent need to re-evaluate all existing forensics tools to take countering measure into account [32]. In the field of forensics sciences, countermeasures to the investigation activities are known under the name of *counter-forensics* or *anti-forensics*. The counter-forensics aims at concealing the traces introduced by processing tools when the user edits or tampers an image content. Harris [33] defines counter-forensics techniques as *any attempt to compromise the availability or usefulness of evidence* *to the forensics process.* Under this interpretation, the simple wiping-off of fingerprints from a crime scene can be considered as a counter-forensics act. In a similar way, multimedia counter-forensics involves all those means that allow covering traces of image manipulation, or, more precisely, to make manipulation invisible to the existing detection methods. Hence, the study of counter-forensics methods to mislead forensics techniques by tamper hiding or concealing traces of manipulations, is becoming a hot research topic [34].

Most of the tamper hiding algorithms are borrowed from steganalysis research [35], which try to achieve undetectability by preserving as many image properties as possible. Yet, steganography and tamper hiding differ in the amount and source of information to hide, and the extent to which an image can be altered. Most steganographic methods are designed to embed a given message by minimizing the number of changes to the cover (hence, keep its semantic) while tamper hiding conceals the mere information that larger parts of the original medium have been modified with the aim to change its semantic [36]. Nevertheless, counter-forensics techniques do not have the requirement to transmit a message, so the modification is more flexible.

When designing counter-forensics methods, we simultaneously consider the presence of, at least, two players: *the forensics analyst* and *the adversary.* The goal of the forensics analyst is to devise a method (detector) that is able to tell apart untouched images from those that have undergone some (usually very specific) processing. But, the adversary has a different goal that he wants to produce a processed image, having some desired characteristics, and do that in such a way that forensics analyst's tools will misclassify it as original. In this sense, these methods can be considered as attacks against the investigation

[34]. One of the most common types of image forgeries is the copy-move forgery, where a region from one part of an image is copied and pasted onto another part in same image, thereby concealing the image content in the latter region. Image forensics literature offers several examples of detector for such manipulation and among them, the most recent and efficient ones are based on SIFT [19, 38, 39]. The capability of SIFT to discover correspondences between similar visual content, in fact, allows the forensics analysis to detect even very accurate and realistic copy-move forgery.

The remainder of this chapter is organized as follows: A SIFT-based copy-move forgery detection method is presented in Section 5.2. About SIFT and its scale-spaces are presented in Section 5.3. Section 5.4 introduces review on semantically admissible distortion. Our intention is presented in Section 5.5. The experimental results are provided in Section 5.6. Discussion is drawn in Section 5.7.

5.2 A SIFT-Based Copy-Move Forgery Detection Method [43]

In this section, a brief review of the SIFT technique and of the approach for detecting copy-move forgeries is drawn. Given an image, SIFT features [41] are detected at different scales by using a scale space representation implemented as an image pyramid. The pyramid levels are obtained by Gaussian smoothing and sub-sampling of the image resolution while interest points are selected as local extrema (min/max) in the scale-space. These points (usually called keypoints) are extracted by applying a computable approximation of the

Laplacian of Gaussian (*LoG*) called Difference of Gaussians (*DoG*). In particular, the SIFT algorithm approximates *LoG* by iteratively computing the difference between two nearby scales in the scale-space. Once these keypoints are detected, SIFT descriptors are computed at their locations in both image plane and scale-space. Each descriptor consists in a histogram of 128 elements, obtained from a 16 × 16 pixels area around the corresponding keypoint. The contribution of each pixel is obtained by calculating image gradient magnitude and direction in scale-space and the histogram is computed as the local statistics of gradient directions (8 bins) in 4 × 4 sub-patches of the 16 × 16 area. The procedure in which interest points are localized ends with a list of N keypoints each of which is completely described by the following information: $x_i = \{x, y, \sigma, o, f\}$, where (x, y) are the coordinates in the image plane, σ is the scale of the keypoint (related to the level of the image pyramid used to compute the descriptor), o is the dominant orientation (used to achieve rotation invariance) and f is the final SIFT descriptor. After SIFT features are extracted the copy-move forgery detection is performed in the SIFT space among the f_i vectors of each keypoint to identify similar local patches in the test image. The best candidate match for each keypoint x_i is found by identifying its nearest neighbor from all the other $(n - 1)$ keypoints of the image, which is the keypoint with the minimum Euclidean distance in the SIFT space. In order to decide if two keypoints match the ratio between the distance of the closest neighbor to that of the second-closest one is used, and then this ratio is compared with a threshold T (typically fixed to 0.6). For the sake of clarity, given a keypoint we define a similarity vector $D = \{d_1, d_2, ..., d_{n-1}\}$ that represents the sorted euclidean distances with respect to the other descriptors.

The keypoint is matched only if this constraint is satisfied: $d_1/d_2 < T$, where $T \in (0, 1)$. Finally, by iterating over keypoints in X, we can obtain the set of matched points which, at this stage, already provides a draft idea of the authenticity of the image and of the presence of duplicated areas Procedures of segmentation and clustering can successively be adopted to better individuate manipulated patches.

5.3 Scale Invariant Feature Transform (SIFT)

The research community has recently started to approach SIFT-based copy-move forgery detection from the perspective of the attacker, whose goal is to hide the features causing similar blocks or keypoints to match. In presence of a copy-move forgery, the extracted SIFT keypoints from the copied and the original regions have similar descriptor vectors (keypoints). Therefore, matching among SIFT features adopted to detect if an image has been tampered with and, subsequently, localize such forgery. In this sense, the investigation of the attacker is considered on the keypoints (detection) extraction of SIFT descriptor.

Lowe [41] has presented a powerful framework to recognize or retrieve objects. The SIFT approach can be viewed as a texture descriptor composed by four major stages:

1. Scale-space extrema detection
2. Keypoint localization
3. Orientation assignment

4. Keypoint description

Our main intention is investigated in this method by avoiding the local extrema in the scale-space extrema detection stage. At first, we introduce a scale-spaces for the extraction of SIFT descriptor, and then we present our proposed method. Our method modifies the selection of local extrema on DoG space by using semantically admissible distortion. Our activity is countermeasure against the exact detection of feature points in digital image.

5.3.1 Gaussian Scale-Space

The SIFT detector and descriptor [41] are constructed from the Gaussian scale-space of the source image $I(x, y)$, which is defined as a function $L(x, y, \sigma)$. This is produced from the convolution of $I(x, y)$ with a variable-scale Gaussian $G(x, y, \sigma)$:

$$L(x, y, \sigma) = G(x, y, \sigma) \star I(x, y) \tag{5.1}$$

where $\star$ is the convolution operation in x and y, and

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{5.2}$$

where $G(x, y, \sigma)$ is an isotropic Gaussian kernel of variance σ^2, x and y are the spatial coordinate and σ is the scale coordinate.

Since the scale-space $G(x, y, \sigma)$ represents the same information (the image $I(x, y)$) at different levels of scale, it is sampled in a particular way to reduce redundancy as shown in Figure 5.1(a). The domain of the variable σ is discretized in logarithmic steps arranged in O octaves. Each octave is further subdivided in S sub-levels. The distinction between octave and sub-level is

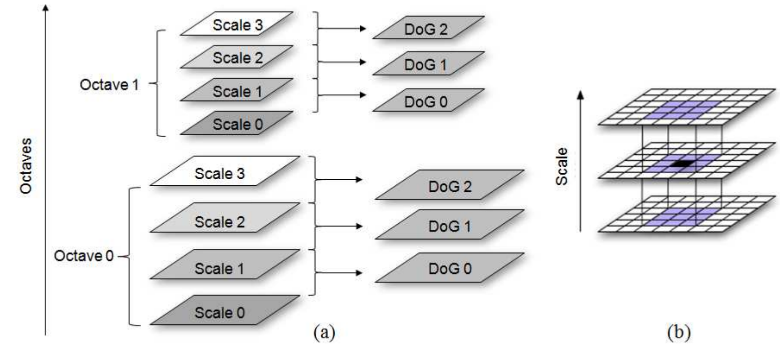


Figure 5.1: Scale-space representation. (a) Gaussian scale-space, (b) Scale-space extrema detection.

important because at each successive octave the data is spatially downsampled by half. Octaves and sub-levels are identified by a discrete *octave index* ø and *sub-level index* s respectively. An example of Gaussian scale-space representation is illustrated in Figure 5.2.

The octave index ø and the sub-level index s are mapped to the corresponding scale σ by the formula,

$$\sigma(\o, s) = \sigma_0 2^{\o+s/S}, \quad \o \in \o_{min} + [0, ..., O-1], \quad s \in [0, ..., S-1] \tag{5.3}$$

where $\sigma_0 \in \mathbb{R}_+$ is the base scale level, $S \in N$ is the scale resolution. Note that it is possible to have octaves of negative index.

The spatial coordinate x and y are sampled on a lattice with a resolution which is a function of the octave. We denote $x_\o$ and $y_\o$ the spatial index for octave ø; this index is mapped to the coordinate x and y by

$$x = 2^\o x_\o, \quad y = 2^\o y_\o, \quad x_\o \in [0, ..., M_\o - 1], \quad y_\o \in [0, ..., N_\o - 1], \quad \o \in \mathbb{Z}. \tag{5.4}$$

where $(N_\o, M_\o)$ is the spatial resolution of octave ø. If (M_0, N_0) is the the resolution of the base octave $\o = 0$, the resolution of the other octaves is

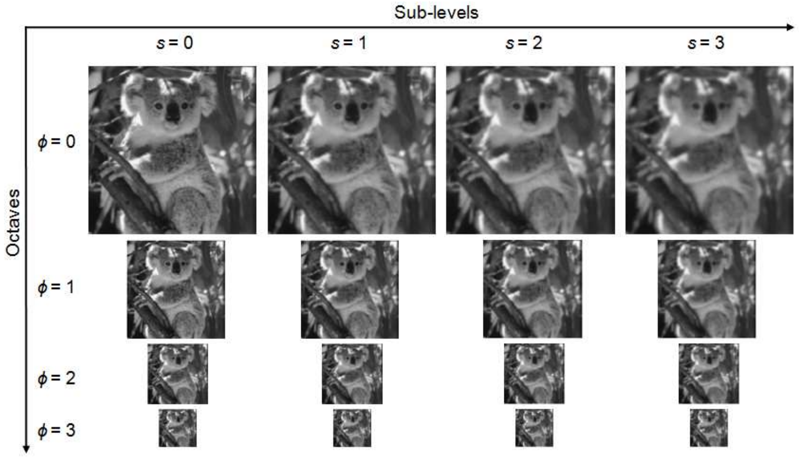


Figure 5.2: An example of a Gaussian scale-space representation.

obtained as

$$M_{\o} = \lfloor \frac{M_0}{2^{\o}} \rfloor, \quad N_{\o} = \lfloor \frac{N_0}{2^{\o}} \rfloor \tag{5.5}$$

It will be useful to store some scale levels twice, across different octaves. We do this by allowing the parameter s to be negative or greater than S. Formally, we denote the range of s as $[s_{min}, s_{max}]$. We also denote the range of the octave index $\o$ as $[\o_{min}, \o_{min} + O - 1]$, where $O \in \mathbb{N}$ is the total number of octaves. Table 5.1 for a summary of these symbols used in this paper.

5.3.2 Difference of Gaussian Scale-Space

To efficiently detect stable keypoint locations in scale-space, the algorithm make use of another scale-space too, called difference of Gaussian (DoG), which is, coarsely speaking, the scale derivative of the Gaussian scale-space $G(x, y, \sigma)$ along the scale coordinate σ, as shown in Figure 5.1(a). The difference of Gaussian pyramid is generated from a single input image. The output is a pyramid of several images, each being a unique difference of Gaussians. To

Table 5.1: Scale-space parameters.

Symbols	Descriptions
$G(x, y, \sigma)$	Gaussian scale-space
$D(x, y, \sigma)$	DoG scale-space
$*(\cdot, \sigma(ø, \cdot))$	Octave data
σ_0	Base scale offset
$\sigma(ø, s) = \sigma_0 2^{ø+s/S}$	Scale coordinate formula
$ø \in [ø_{min}, ø_{min} + O - 1]$	Octave index and range
$s \in [s_{min}, s_{max}]$	Scale index and range
$M_0, \quad N_0$	Base spatial resolution (octave $ø = 0$)
$M_ø = \lfloor \frac{M_0}{2^ø} \rfloor, \quad N_ø = \lfloor \frac{N_0}{2^ø} \rfloor$	Octave lattice size formulas
$x = 2^ø x_ø, \quad y = 2^ø y_ø$	Spatial coordinate formula
$x_ø \in [0, ..., M_ø - 1], \quad y_ø \in [0, ..., N_ø - 1]$	Spatial indexes and ranges
$\pi_B(x, y)$	A random permutation of the indices belonging to the B-th block
$\Delta(x, y)$	A *i.i.d* random variables uniformly distributed in the interval $[-\Delta_{max}, \Delta_{max}]$

generate the pyramid, the input image is repeatedly blurred; the difference between consecutive blur amounts is then output as one octave of the pyramid. One of the blurred images is downsampled by a factor of two in each direction, and the process occurs again with output in a different size. It is given by

$$D(x, y, \sigma(s, ø)) = (G(x, y, \sigma(s + 1, ø)) - G(x, y, \sigma(s, ø))) \star I(x, y) \qquad (5.6)$$

Lowe's [41] implementation uses the following parameters:

$$\sigma_n = 0.5, \qquad \sigma_0 = 1.6 \cdot 2^{1/S}, \qquad ø_{min} = -1, \qquad S = 3.$$

In order to compute the octave $ø = -1$, the image is doubled by bilinear interpolation (for the enlarged image $\sigma_n = 1$). In order to detect extrema at all scales, the difference of Gaussian scale-space has $s \in [s_{min}, s_{max}] = [-1, S + 1]$. Since the difference of Gaussian scale-space is obtained by differentiating the Gaussian scale-space, the latter has $s \in [s_{min}, s_{max}] = [-1, S + 2]$. The parameter O is set to cover all octaves (i.e. as big as possible.)

The feature points are chosen from the local maxima or minima in the DoG space. Each point in $D(x, y, \sigma(s, \o))$ will be compared with its 26 neighbouring pixels, of which 8 pixels located in current scale image and others located in the scale above and below. As shown in Figure 5.1(b), the candidate pixel in black is compared with those other 26 pixels in white. The candidate pixel will be considered to be a feature point and its coordinate pixel value is larger than all those 26 pixels values or smaller than them.

In order to impede the detection of local maxima or minima, we applied semantically admissible distortion on the DoG space. As a results, the detected keypoints are found on totally different positions by effect of the keypoint localizatioan and the orientation assignment processes.

5.4 Review on Semantically Admissible Distortions

The random pre-warping must be strong enough to avoid that registration techniques can undo the warping and, in the meantime, it must guarantee the invisibility of the distortion. For this reason gathering information about the subset of semantically admissible geometric distortions is a vital requirement. In a general case of a geometric distortion can be seen as a transformation of the position of the pixels in the image. It is possible to distinguish between global and local geometric distortions.

A global transformation, in fact, is defined by a mapping analytic function that relates the points in the input image to the corresponding points in the output image. It is defined by a set of operational parameters and performed over all the image pixels.

Local distortions, in fact, refer to transformations affecting in different ways the position of the pixels of the same image or affecting only part of the image. A general model which comes to mind to do this is a distortion according to which each pixel of the image is assigned a random displacement vector $\Delta(x, y) = (\Delta_h(x, y), \Delta_v(x, y))$, where $\Delta_h(x, y)$ and $\Delta_v(x, y)$ are *i.i.d* random variables uniformly distributed in the interval $[-\Delta_{max}, \Delta_{max}]$. The main problem in a so defined transformation is that it does not take into account the way the Human Visual System (HVS) perceives geometrical distortions. In the following models to treat geometric transformations are sketched. The models are analyzed by means of visual inspection under semantic constraint.

5.5 Attacks or Local Distortions on *DoG* Space

In this section, we describe an attack scenario to impede a SIFT-based copy-move forgery detection methods. Our goal is to take into account the HVS to find a perceptually admissible subset of the possible distortions that can be applied to the *DoG* space.

As explained above, a generic local distortion can be described, for example, by a permutation of the position of pixels on *DoG* space. Of course this kind of distortion introduces an annoying degradation. A way to overcome this problem could be to fix a maximum displacement of the position of pixels, i.e. to perform a block-based local permutation.

5.5.1 Block-Based Local Permutation ($\boldsymbol{B}$-$\boldsymbol{LP}$)

This model consists in partitioned the $b \times b$ blocks on DoG space and obtaining the distorted DoG space by allowing random permutations within each block. Here, the size of the partitioned block should be smaller, which provides a semantic constraint.

Each spatial coordinates x and y on DoG space (base $D(x, y, \sigma(s, ø)))$ is tiled by non-overlapping blocks a size of $b \times b$, $(b = 3)$ pixels. Blocks are horizontally slid by b pixels rightwards starting with upper left corner and ending with the bottom right corner. The total number of non-overlapping blocks for each spatial coordinates of $M_ø \times N_ø$ pixels are $B_ø = (M_ø/b) \times (N_ø/b)$, $ø \in ø_{min} + [0, ..., O - 1]$.

Let $D(x, y, \sigma(s, ø))$ be a generic pixel of the distorted DoG space belonging to the B-th block in $ø$-th octave, then

$$D(x, y, \sigma(s, ø)) \leftarrow D(x, y, \sigma(s, ø)) \cdot \pi_B(x, y) \tag{5.7}$$

where $\pi_B(x, y)$ is a random permutation of the indices belonging to the B-th block. Increasing the size of image allows to consider a larger number of transformations but, at the same time, affects the image quality leading to increasingly annoying artifacts.

Hence we permuted the element of the base levels on DoG space, the detection of the local maxima or minima is chosen in the different locations. Thus, the block-based local permutation allows to impede the detection of local invariant features with eliminating or creating a local features under lower rate value. This property provides the hiding traces left in an image counter-forensics area.

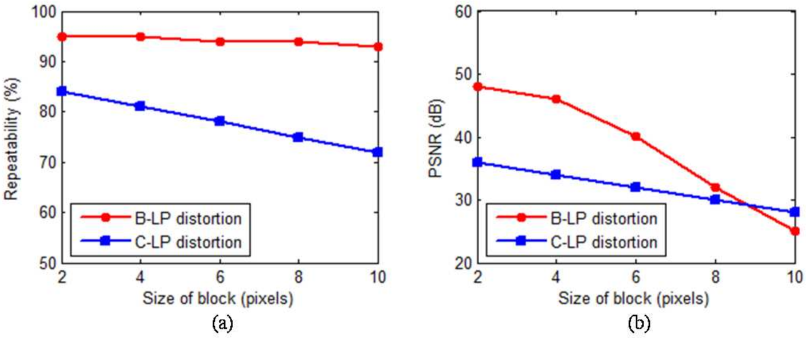


Figure 5.3: The repeatability (a) and the perceptual quality (b) measures for the block- and the cancellation- based local permutations.

5.5.2 Cancellation-Based Local Permutation (*C-LP*)

In this model, we add to the previous one the possibility of duplicating and canceling sample values so that it is also possible to model local expansions and shrinkings. Furthermore in this way we allow for a larger number of possible distortions. Let $D(x, y, \sigma(s, \o))$ is a generic pixel of the distorted *DoG* space, we have

$$D(x, y, \sigma(s, \o)) \leftarrow D(x + \Delta_h, y + \Delta_v, \sigma(s, \o)) \tag{5.8}$$

where Δ_h and Δ_h are sequences of *i.i.d* integer random variables uniformly distributed in the interval $[-\Delta_{max}, \Delta_{min}]$.

Important property of invariant feature is measured by repeatability measure. The same feature can be found in several images despite geometric and photometric transformations. We test the *B-LP* and the *C-LP* distortions, respectively, in order to alter the detection of SIFT keypoints as shown in Figure 5.3. As a results, the *B-LP* distortion can provide more stable and repeatability properties than the *C-LP* distortion during the increases of the size for divided blocks. From this results, we have chosen the *B-LP* distortion

Figure 5.4: Examples of Benchmark data (3888×2592).

under semantic constraint to impede the detection of keypoints well.

5.6 Experimental Results

In this section, we extensively evaluate the proposed counter-forensics method in a realistic scenario, and show that it yields good results in hiding traces while retaining a high image quality for the attacked image.

We simulated our method under a PC with 3.2G Hz Core i5 CPU, 8G RAM, and Windows 8 platform. The simulation was carried out using Matlab version R2008a. We test our method on commonly used 8 gray-scale images of size 512×512 pixels for performance evaluation (e.g., Lena, Barbara, Baboon, etc.) and Benchmark data for image copy-move detection dataset including 120 authentic and 124 forged color images of size 3888×2592 pixels with different outdoor scenes for copy-move forgery, as shown in Figure 5.4. For Benchmark data, the authentic images were taken by different digital cameras. All tampered images in this dataset are generated from the authentic images by crop-and-paste operation using Adobe Photoshop CS3 version 10.0.1 on Windows XP. The tampered regions are from the same authentic image. In the

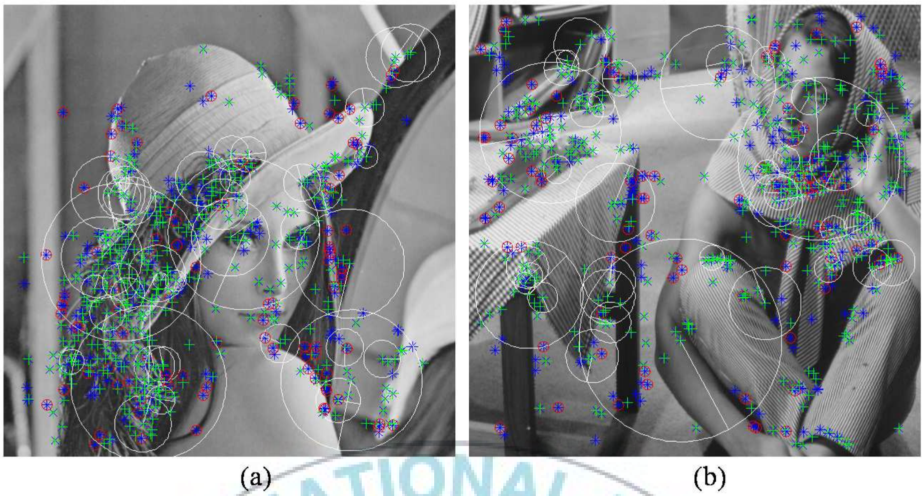


Figure 5.5: Results of keypoint detection. (a) For Lena, $PSNR$=43.02dB (b) For Barbara, $PSNR$=41.31dB. Size of block (3×3) on DoG space.

following tests, the keypoints have been computed by means of VLFeat, Vedaldi and Fulkerson's implementation of SIFT [42]. (*DoG* peak and edge thresholds set to 4 and 10, respectively). The threshold for keypoint matching is fixed to 0.6, as suggested by Lowe in [41].

5.6.1 Efficiency for Tamper Hiding and Impeding Keypoint Matching

We present an analysis on the efficiency of the proposed procedure for impeding keypoint matching. The experimental tests carried out to check the keypoint detection of the proposed method. In Figure 5.5, the number of original keypoints (blue) are detected by VLFeat algorithm, and the detected keypoints (green) after the proposed processing tool (adversary) are described on the Lena and Barbara images, respectively.

Table 5.2: The performance evaluation for the number of eliminated and created keypoints after the effect of processing tool (adversary).

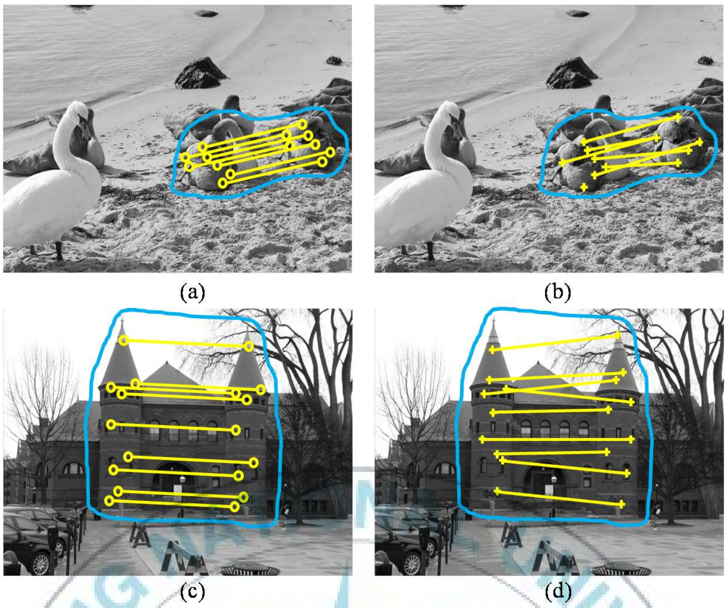
Test Images	KP (#) detected	KP (%) removed	KP (#) removed	KP (%) created	KP (#) created	α rate	KP (#) attacked	PSNR (dB)
Lena	1218	11%	134	14%	167	0.8	1251	41.56
Baboon	3124	17%	531	13%	405	1.3	2998	39.56
Barbara	1825	13%	244	12%	224	1.1	1805	41.31
House	1136	6%	73	11%	122	0.6	1185	36.83
Pepper	2889	12%	346	13%	389	0.9	2932	38.33
Boat	2351	14%	278	13%	312	1.1	2385	40.38

During the procedures, the processing tool can eliminate some keypoints by an effect of semantically admissible distortion. However, similar keypoints are generated as such effect, also the removed percents of keypoints (KP) are almost equal to the created percents of keypoints (around ±10%), as shown in Table 5.2. This means our method can provide the tamper hiding scenario, that hiding traces left by processing tool. In other words, the forensics analist can not detect exact position of the keypoints, that the processed keypoints are altered by the keypoint removal or creation procedures. In our impeding method, the keypoints are eliminated while created, and α rates for two procedures are almost equal to 1 (α=removed/created ≈ 1). Here we note that α is measure of the tamper hiding, which defined by comparison between the keypoint removal and creation rates.

5.6.2 Analysis for Copy-Move Forgery Detection

In this section, we report some experimental results on images where a copy-move attack has been performed by taking into account the context.

a) Evaluation of the Detection Accuracy. In Figure 5.6, detection



(a) (b) (c) (d)

Figure 5.6: An examples of a SIFT-based copy-move forgery detection method [35] is pictured in (a,c), and corresponding detection results of our attacking is reported in (b,d).

results are pictured by presenting on the tampered images for (a, c) a SIFT-based copy-move forgery detection method [43] and (b, d) the corresponding one, where matched keypoints and clusters, attacked by our processing tool, are highlighted. As a result, an interesting situation concerns that our method can impede the (similarity) keypoint matching process and to make a false matching results.

In order to quantify the accuracy of detection, the true positive ratio (TPR) and the false positive ratio (FPR) are employed, as follows:

$$TPR = \frac{|\Omega_1 \bigcap \Omega_2| + |\overline{\Omega_1} \bigcap \overline{\Omega_2}|}{|\Omega_1| + |\overline{\Omega_2}|}, \qquad FPR = \frac{|\Omega_1 \bigcup \Omega_2| + |\overline{\Omega_1} \bigcup \overline{\Omega_2}|}{|\Omega_1| + |\overline{\Omega_1}|} - 1 \quad (5.9)$$

where Ω_1 and Ω_2 are the original copied region and the detected copied re-

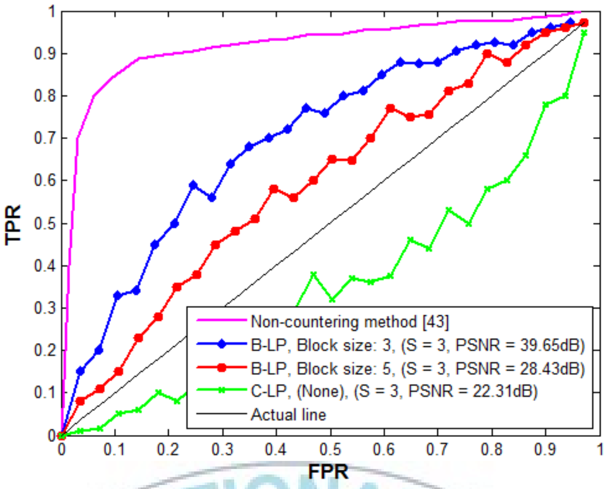


Figure 5.7: Results of the detection accuracy for copy-move forgery.

gion, while $\overline{\Omega_1}$ and $\overline{\Omega_2}$ are the forged region and the detected forged region, respectively.

Our goal is to minimize the TPR while maintaining a higher the FPR. The horizontal axis corresponds to the false positive rate (incorrectly labeling an image as altered) and the vertical axis corresponds to the true positive rate (correctly labeling an image as altered). We applied two different processing tools in order to avoid the detection accuracy of similarity matching for the copy-move forgery, such as the *B-LP* and the *C-LP* distortions on the *DoG* space, respectively. Figure 5.7 shows the small sized *B-LP* can achieve higher the image quality while maintaining degraded the detection accuracy compared with non-countered method. For example, with a false positive rate of 0.1, we achieve a true positive rate of 0.34. But, in the increased size of *B-LP*, the detection accuracy is reduced significantly. For the *C-LP* distortion, the detection accuracy is lower than other two cases, because, the cancellation is strongly affected to change the value of point on the *DoG* space, while also can

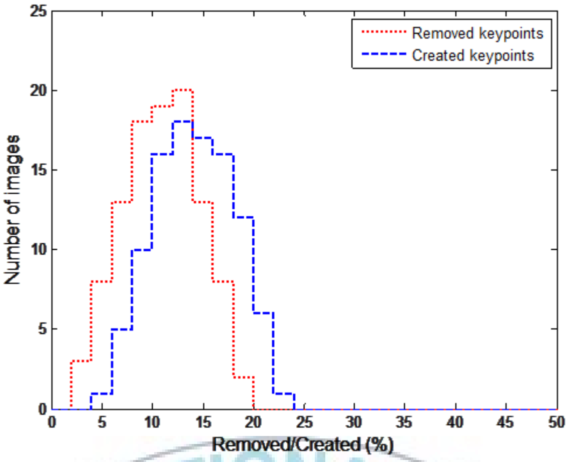


Figure 5.8: Efficiency of our method respect to the tamper hiding. Curves correspond to the envelopes of removal and creation rate histogram, obtained by analyzing the manipulated Benchmark data.

decrease the image quality ($PSNR$=22.31dB). As a results of the detection accuracy, the small sized *B-LP* approach is efficient to impede the similarity matching methods without higher rate of changes for the image quality.

b) Efficiency for Tamper Hiding. We calculated the removal and creation rates, respectively, actually achieved on each test images of Benchmark dataset (N = 124 images), as shown in Figure 5.8. Each histograms of the removal and creation rates are concentrated on around a value of ±10%. For this case, our method can also successfully provide efficient the tamper hiding scenario on the number of images for Benchmark dataset.

5.6.3 Comparison with Keypoint Forging Techniques

In the approach of Do *et al.* [44], the authors focused on a SIFT for content based image retrieval scenario and devised a number of interesting attacks.

These attacks were successful in deluding the system can simultaneously remove keypoints and forge new keypoints in the images to be concealed. Although these attacks are efficient, the quality of an attacked image is also significantly reduced as shown in Table 5.3. The first attack is Removal with Minimum Local Distortion (RMD). The idea behind this technique is to calculate a small patch ϵ that added to the neighborhood of a keypoint allows its removal. The coefficients of ϵ are chosen in such a way to reduce the contrast around the keypoint computed at the DoG level, thus invalidating the check performed by the SIFT algorithm on all potential keypoints. Moreover, it is requested that the coefficients locally introduce the minimum visual distortion.

Next attack is Global Smoothing (GS). Performing a smoothing on the whole image reduces the number of keypoints while avoiding the creation of new ones, as it does not introduce strong discontinuities. Experiments show this global smoothing is quite effective even with a Gaussian kernel of small variance. The value of σ is set empirically by conducting many experiments. $\sigma = 1.3$ is found as a good trade-off between visual and score (or number of unchanged keypoints) of the attacked images. A greater σ would in turn remove more keypoints, but the quality of the resulting images would be worse. Almost all keypoints not removed by the GS attack have high absolute DoG value.

GS attack is efficient to remove keypoints in the image. However, there are some regions/images which do not have any keypoints or have a sparse density of keypoints. Smoothing such regions/images is not necessary: few keypoints are affected while much visual distortion is introduced. It is more efficient to only smooth regions presenting a high density of keypoints. A variant version

Table 5.3: The performance comparison with other attacks.

Test Image	Attacks	KP (%) removed	KP (#) removed	KP (%) created	KP (#) created	α rate	KP (#) attacked	PSNR (dB)
	RMD	90.48%	1102	72.91%	888	1.2	1004	27.78
Lena	*GS*	91.30%	1112	27.83%	339	3.3	445	31.17
(1218)	*DS*	82.51%	1005	37.36%	455	2.3	668	33.95
	Our	11.00%	134	14%	167	0.8	1251	41.56

of the GS attack which takes into account the keypoints density and is referred to as Density Smoothing (DS). The image is segmented to dense regions and non-dense regions by sliding a window of size 50 × 50. A smaller window makes DS more time consuming. A bigger window makes DS less efficient in spotting high keypoint density regions. A region is considered as dense if the number of keypoints in the window is more than 60. It means that there are approximately one keypoint per 7×7 pixels square. Dense regions are smoothed by the Gaussian kernel with $\sigma = 1.3$, like the GS attack.

As a results of the performance comparison, our method can provide higher the visual quality than other attack scenarios. The average $PSNR$ after the DS attack is 33.95dB, but not suitable for tamper hiding scenario ($\alpha > 1$). For RMD attack, the visual quality is degraded by higher degree compared with other methods.

5.7 Discussion

In this work, we proposed a targeted counter-forensics method for SIFT-based copy-move forgery detection by applying semantically admissible distortion in processing tool. Our activity is countermeasure against the exact detection of feature points in digital image. In presence of a copy-move manipulation, the extracted SIFT keypoints from the copied and the original regions have similar descriptor vectors. Therefore, matching among SIFT features adopted to detect if an image has been tampered with and, subsequently, localize such forgery. In this sense, the investigation of the attacker is considered on the keypoints extraction of tampered image. Our proposed processing tool is considered on the *DoG* space of SIFT algorithm, where we applied the semantically admissible distortion in order to alter the detected keypoints under semantic constraint. The proposed method allows the attacker to delude a similarity matching process and conceal/hide traces left by modification of SIFT keypoints, while maintaining a high fidelity between processed and original images under semantic constraint.

Chapter 6. Conclusion

Digital image forensics can be defined as the science that tries, by analysing a digital image content, to give an assessment on such a content and to extract information that can be useful to address and support an investigation linked to the scene represented in that specific digital image. Image forensics has to be able to develop efficient instruments to deal with the disparate digital devices that can generate images and, above all, with the different processing tools that allows also an unskilled user to manipulate digital image contents. In response to these challenges, digital image forensics involves the study and development of techniques to determine the authenticity, processing history, and origin of digital image content without relying on any information aside from the digital content itself.

In this thesis, principles and motivations of digital image forensics have been discussed and new methods in image forgery detection for content-changing manipulations have been presented. Additionally, we considered the problem of multimedia security from the forgers point of view. Though existing digital image forensics techniques are capable of detecting several standard digital image manipulations, they do not account for the possibility that counterfeiting operations designed to hide traces of manipulation may be applied to digital content. All the proposed techniques can be sketched as a forensics tool that extracts, from the considered data, some digital fingerprints, and that, by exploring some properties of such contents, is able to make a decision based on either classification or estimation procedure.

Finally, an image forensics is just at its infancy stage, there is still much work to be done and some ideas can be borrowed from other research areas, like techniques developed for camera identification. Also, knowledges from computer vision, signal processing, computer graphics, pattern recognition and imaging process will be needed for further analysis.

References

[1] Worth1000 home page. (http://www.worth1000.com)

[2] Photoplasty Archive. (http://www.cracked.com/photoplasty)

[3] W. Chen, Y.Q. Shi, W. Su, ”Image Splicing Detection Using 2-D Phase Congruency and Statistical Moments of Characteristic Function,” *SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents*, San Jose, CA, USA, 2007.

[4] L. Li, S. Li, H. Zhu, ”An Efficient Scheme for Detecting Copy-Move Forged Images by Local Binary Patterns,” *Journal of Information Hiding and Multimedia Signal Processing*, vol. 4, no. 1, pp. 46-56 2013.

[5] B. Mahdian, S. Saic, ”Blind Authentication Using Periodic Properties of Interpolation,” *IEEE Transsctions on Information Forensics and Security*, vol. 3, no. 3, pp. 529-538, 2008.

[6] M. Stamm, S. Tjoa, W. Lin, K. Liu, ”Undetectable Image Tampering Through JPEG Compression Anti-Forensics,” *In Proceedings of ICIP 2010, IEEE International Conference on Image Processing*, pp. 2109-2112, 2010.

[7] M. Stamm, K. Liu, ”Blind Forensics of Contrast Enhancement in Digital Images,” *In Proceedings of ICIP 2008, IEEE International Conference on Image Processing*, pp. 3112-3115, 2008.

[8] G. Cao, Y. Zhao, R. Ni, ”Detection of Image Sharpening Based on Histogram Aberration and Ringing Artifacts,” *In International Conference on Multimedia and Expo*, New York, 2009.

[9] M. Kirchner, J. Fridrich, ”On Detection of Median Filtering in Digital Images,” *SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents*, San Jose CA, USA, 2010.

[10] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, ”Determining Image Origin and Integrity Using Sensor Noise,” *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 74-90, 2008.

[11] V. Conotter, Active and Passive Multimedia Forensics, PhD thesis, University of Trento, 2011.

[12] H.T. Sencar, N. Memon, ”Overview of State-of-the-Art in Digital Image Forensics,” *Part of Indian Statistical Institute Platinum Jubilee Monograph series titled ”Statistical Science and Interdisciplinary Research,”* World Scientific Press, pp. 325-347, 2008.

[13] T. Lindeberg, ”Edge Detection and Ridge Detection with Automatic Scale Selection,” *Technical report ISRN KTH/NA/P96/06SE, International Journal of Computer Vision*, vol. 30, no. 2, 1998.

[14] H. Hu, G. Haan, ”Low Cost Robust Blur Estimator,” *International Conference on Image Processing*, pp. 617-620, 2006.

[15] CASIA Tampered Image Detection Evaluation Database, (http://forensics.idealtest.org:8080/), 2009.

[16] D. Duda, P. Hart, "Pattern Classification and Scene Analysis," John Wiley and Sons, New York, NY, 2nd edition, 1973.

[17] H. Zhongwei, L. Wei, S. Wei, "Improved Run Length Based Detection of Digital Image Splicing," *IWDW 2011. LNCS 7128*, pp. 349-360, 2012.

[18] J. Dong, W. Wang, T. Tan, Y.Q. Shi, "Run-length and Edge Statistics Based Approach for Image Splicing Detection," *IWDW 2008. LNCS 5450*, pp. 76-87. Springer, Heidelberg, 2009.

[19] H. Farid, "A Survey of Image Forgery Detection," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 16-25, 2009.

[20] V. Chrislein, R. Riess, J. Jordan, E. Angelopoulou, "An Evaluation of Popular Copy-Move Forgery Detection Approaches," *IEEE Transactions on Information Forensics And Security*, vol. 7, no. 6, pp. 1841-1854, 2012.

[21] Y. Huang, W. Lu, W. Sun, D. Long, "Improved DCT-Based Detection of Copy-Move Forgery in Images," *Journal of Forensic Science International*, vol. 206, no. 13, pp. 178-184, 2011.

[22] S. Khan, A. Kulkarni, "Reduced Time Complexity for Detection of Copy-Move Forgery Using Discrete Wavelet Transform," *International Journal of Computer Application*, vol. 6, no. 7, pp. 31-36, 2010.

[23] B. Mahdian, S. Saic, "Detection of Copy-Move Forgery Using a Method Based on Blur Moment Invariants," *Journal of Forensic Science International*, vol. 171, no. 27, pp. 180-189, 2007.

[24] S.J. Ryu, M.J. Lee, H.K. Lee, ”Detection of Copy-Rotate-Move Forgery Using Zernike Moments,” *In Proceedings of the 12th International Conference on Information Hiding*, pp. 51-65, 2010.

[25] G.J. Liu, J.W. Wang, S.G. Lian, Z.Q. Wang, ”A Passive Image Authentication Scheme for Detecting Region Duplication Forgery with Rotation,” *Journal of Network and Computer Application*, vol. 34, no. 5, pp. 1557-1565, 2011.

[26] M.A. Fiffy, ”The Radon Transform and Some of Its Applications,” *Journal of Modern Optics*, vol. 32, no. 1, pp. 3-4, 1985.

[27] S.A. Khayam, ”The Discrete Cosine Transform (DCT): Theory and Application,” *Journal of Information Theory and Coding*, pp. 1-31, 2003.

[28] A. Fridrich, ”Detection of Copy-Move Forgery in Digital Images,” *Proceedings of the Digital Forensic Research Workshop*, Cleveland OH, USA, 2003.

[29] Y. Huang, W. Lu, W. Sun, D. Long, ”Improved DCT-Based Detection of Copy-Move Forgery in Images,” *Journal of Forensic Science International*, vol. 206, no. 13, pp. 178-184, 2011.

[30] L. Li, S. Li, H. Zhu, ”An Efficient Scheme for Detecting Copy-Move Forged Images by Local Binary Patterns,” *Journal of Information Hiding and Multimedia Signal Processing*, vol. 4, no. 1, pp. 46-56, 2013.

[31] J. Redi, W. Taktak, J.-L. Dugelay, ”Digital Image Forensics: A Booklet for Beginners,” *Multimedia Tools and Applications*, vol. 51, pp. 133162, 2011.

[32] L. Chen, S. Wang, S. Li, J. Li, "Countering Universal Image Tampering Detection with Histogram Restoration," *International Workshop on Digital Forensics and Watermarking, IWDW 2012, LNCS 7128*, 2012.

[33] R. Harris, "Arriving at an Anti-Forensics Consensus: Examining How to Define and Control the Anti-Forensics Problem," *Digital Investigation, vol. 3 (Supplement 1)*, pp. 44-49, 2006.

[34] R. Böhme, M. Kirchner, "Counter-Forensics: Attacking Image Forensics," *in Digital Image Forensics, ed. by H.T. Sencar, N. Memon, Springer*, New York, pp. 327-366, 2013.

[35] Y.Q. Shi, C. Chen, G. Xuan, W. Su, "Steganalysis Versus Splicing Detection," *In IWDW 2007, LNCS 5041*, pp. 158-172, 2008.

[36] M. Kirchner, R. Böhme, "Tamper Hiding: Defeating Image Forensics," *In Information Hiding, LNCS 4567*, pp. 326-341, 2007.

[37] M. Kirchner, R. Böhme, "Hiding Traces of Resampling in Digital Images," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 4, pp. 582-592, 2008.

[38] A. Fridrich, "Detection of Copy-Move Forgery in Digital Images," *Proceedings of the Digital Forensic Research Workshop*, Cleveland OH, USA, 2003.

[39] V. Christlein, R. Riess, E. Angelopoulou, "On Rotation Invariance on Copy-Move Forgery Detection," *IEEE International Workshop on Information Forensics and Security*, pp. 1-6, 2010.

[40] B. Mahdian and S. Saic, ”Blind Methods for Detecting Image Fakery,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 25, no. 4, pp. 18-24, 2010.

[41] D. Lowe, ”Distinctive Image Features from Scale-invariant Keypoints,” *International Journal Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[42] A. Vealdi, B. Fulkerson, ”VLFeat: An Open and Portable Library of Computer Vision Algorithms,” (http://www.vlfeat.org/), 2008.

[43] I. Amerini, L. Ballan, R. Caldelli, A.D. Bimbo, S. Serra, ”A SIFT-based Forensic Method for Copy-move Attack and Transformation Recovery,” *IEEE Transactions on Information Forensics And Security*, vol. 6, no. 3, pp. 1099-1110, 2011.

[44] T.-T. Do, E. Kijak, T. Furon, L. Amsaleg, ”Deluding Image Recognition in SIFT-Based CBIR Systems,” *In Proceedings of the 2nd ACM Workshop on Multimedia Forensics, Security and Intelligence, MiFor'10*, pp. 7-12, 2010.

[45] T.-T. Ng and S.-F. Chang, ”Identifying and Prefiltering Images: Distinguishing Between Natural Photography and Photorealistic Computer Graphics,” *IEEE Signal Processing Magazine*, vol. 2, no. 26, pp. 49-58, 2009.

[46] S. Bayram, H. T. Sencar, N. Memon, and I. Avcibas, ”Source Camera Identification Based on CFA Interpolation,” *IEEE International Conference on Image Processing*, pp. 69-72, 2005.

[47] H. Cao and A. C. Kot, "Accurate Detection of Demosaicing Regularity for Digital Image Forensics," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 899-910, 2009.

[48] Z. Geradts, J. Bijhold, M. Kieft, K. Kurosawa, K. Kuroki, and N. Saitoh, "Methods for Identification of Images Acquired with Digital Cameras," *In Proceedings of SPIE*, no. 4232, pp. 505-512, 2001.

[49] A. E. Dirik, H. T. Sencar, and N. Memon, "Digital Single Lens Reflex Camera Identification from Traces of Sensor Dust," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 539-552, 2008.

[50] J. Fridrich, "Digital Image Forensics: Introducing Methods to Estimate and Detect Sensor Fingerprint," *IEEE Signal Processing Magazine*, vol. 2, no. 26, pp. 26-37, 2009.

[51] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining Image Origin and Integrity Using Sensor Noise," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 74-90, 2008.

[52] A. E. Dirik, S. Bayram, H. T. Sencar, and N. Memon, "New Features to Identify Computer Generated Images, " *International Conference on Image processing*, pp. 433-436, 2007.

[53] S. Dehnie, T. Sencar, and N. Memon, "Digital Image Forensics for Identifying Computer Generated and Digital Camera Images," *International Conference on Image processing*, pp. 2313-2316, 2006.

[54] S. Lyu and H. Farid, "How Realistic is Photorealistic? *IEEE Transactions on Signal Processing*," vol. 53, no. 2, pp. 845-850, 2005.

[55] T.-T. Ng, S.-F. Chang, J. Hsu, L. Xie, and M.-P. Tsui, "Physics-motivated Features for Distinguishing Photographic Images and Computer Graphics," *In 13th ACM International Conference on Multimedia*, pp. 239-248, 2005.

[56] G. Sankar, H. V. Zhao, and Y.-H. Yang, "Feature Based Classification of Computer Graphics and Real Images," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1513-1516, 2009.

[57] A.C. Popescu, H. Farid, "Exposing Digital Forgeries by Detecting Traces of Resampling," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 758-767, 2005.

[58] S. Bayram, H.T. Sencar, and N. Memon, "A Survey of Copy-move Forgery Detection Techniques," *IEEE Western New York Image Processing Workshop*, 2008.

[59] M. Chen, J. Fridrich, M. Goljan, J. Lukas, "Determining Image Origin and In- tegrity Using Sensor Noise," *IEEE Transactions on Information Forensics and Security 3*, pp. 74-90, 2008.

[60] G. Chierchia, S. Parrilli, G. Poggi, C. Sansone, L. Verdoliva, "PRNU-based Detection of Small Size Image Forgeries," *In 17th International Conference on Digital Signal Processing (DSP)*, pp. 1-6, 2011.

[61] C. Zhang, H. Zhang, "Exposing Digital Image Forgeries by Using Canonical Correlation Analysis," *International Conference on Pattern Recognition (ICPR)*, pp. 838-841, 2010.

[62] W. Li, Y. Yuan, N. Yu, "Passive Detection of Doctored JPEG Image via Block Artifact Grid Extraction," *IEEE Transactions on Signal Processing*, vol. 89, no. 9, pp. 1821-1829, 2009.

[63] C.-T. Li, "Detection of Block Artifacts for Digital Forensic Analysis," *In 2nd International Conference on Forensics in Telecommunications, Information and Multimedia 8*, pp. 173-178, 2009.

[64] H. Farid, "Exposing Digital Forgeries from JPEG Ghosts," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 4, pp. 154-160, 2009.

[65] T.-T. Ng, S.-F. Chang, "A Model for Image Splicing," *International Conference on Image Processing 2*, pp. 1169-1172, 2004.

[66] W. Wang, J. Dong, and T. Tan, "Effective Image Splicing Detection Based on Image Chroma," *IEEE International Conference on Image Processing*, pp. 1257-1260, 2009.

[67] M.K. Johnson, H. Farid, "Exposing Digital Forgeries by Detecting Inconsistencies in Lighting," *In Proceedings of ACM Multimedia and Security Workshop*, pp. 19, 2005.

[68] F. Schroff, A. Criminisi, A. Zisserman, "Object Class Segmentation Using Random Forests," *In Proceedings of British machine vision conference*, 2008.

[69] M.K. Johnson and H. Farid, "Detecting Photographic Composites of People," *In 6th International Workshop on Digital Watermarking*, pp. 19-33, 2007.

[70] W. Zhang, X. Cao, Z. Feng, J. Zhang, and P. Wang, ”Detecting Photographic Composites Using Two-view Geometrical Constraints,” *IEEE International Conference on Multimedia and Expo*, pp. 1078-1081, 2009.

[71] W. Zhang, X. Cao, J. Zhang, J. Zu, and P. Wang, ”Detecting Photographic Composites Using Shadows,” *IEEE International Conference on Multimedia and Expo*, pp. 1042-1045, 2009.

[72] M. Goljan, J. Fridrich, and M. Chen, ”Sensor Noise Camera Identification: Countering Counter-forensics,” *In SPIE Conference on Media Forensics and Security*, 2010.

[73] G. Cao, Y. Zhao, R. Ni, and H. Tian, ”Anti-forensics of Contrast Enhancement in Digital Images,” *In Proceedings of MMSec 2010, 12th ACM workshop on Multimedia and security (MMSec 10)*, 2010.

[74] M.C. Stamm and K.J. Liu, ”Blind Forensics of Contrast Enhancement in Digital Images,” *In Proceedings of ICIP 2008, IEEE International Conference on Image Processing*, pp. 3112-3115, 2008.

[75] M.C. Stamm, S.K. Tjoa, W.S. Lin, and K.J. Liu, ”Anti-forensics of JPEG Compression,” *In Proceedings of ICASSP 2010, IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1694-1697, 2010.

[76] M. Stamm, S. Tjoa, W. Lin, and K. Liu, ”Undetectable Image Tampering Through JPEG Compression Anti-forensics,” *In Proceedings of ICIP 2010, IEEE International Conference on Image Processing*, pp. 2109-2112, 2010.

[77] M. Wirth, M. Fraschini, M. Masek, and M. Bruynooghe, ”Performance Evaluation in Image Processing,” *EURASIP Journal on Advances in Signal Processing*, vol. 2006, pp. 14, 2006.

[78] W. Luo, J. Huang, and G. Qiu, ”Robust Detection of Region- Duplication Forgery in Digital Images,” *In Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006)*, pp. 746-749, 2006.

[79] S. Bravo-Solorio and A. Nandi, ”Passive Forensic Method for Detecting Duplicated Regions Affected by Reflection, Rotation and Scaling,” *In Proceedings of the 17th European Signal Processing Conference (EUSIPCO 2009)*, pp. 824-828, Glasgow, Scotland, UK, 2009.

[80] S. Bayram, H. Sencar, and N. Memon, ”An Efficient and Robust Method for Detecting Copy-Move Forgery,” *In Proceedings of the 34th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009)*, pp. 1053-1056, 2009.

[81] S. Bravo-Solorio and A. K. Nandi, ”Exposing Duplicated Regions Affected by Reflection, Rotation and Scaling,” *In Proceedings of the 36th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*, pp. 1880-1883, 2011.

[82] H. Lin, C. Wang, and Y. Kao, ”Fast Copy-Move Forgery Detection,” *WSEAS Transactions on Signal Processing*, vol. 5, no. 5, pp. 188-197, 2009.

[83] X. Kang and S. Wei, ”Identifying Tampered Regions Using Singular Value Decomposition in Digital Image Forensics,” *In Proceedings of the Interna-*

tional Conference on Computer Science and Software Engineering (CSSE 2008), pp. 926-930, 2008.

[84] M. Bashar, K. Noda, N. Ohnishi, and K. Mori, ”Exploring Duplicated Regions in Natural Images,” *IEEE Transactions on Image Processing*, 2010.

[85] X. Pan and S. Lyu, ”Region Duplication Detection Using Image Feature Matching,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 857-867, 2010.

[86] B.L. Shivakumar and S. Baboo, ”Detection of Region Duplication Forgery in Digital Images Using SURF,” *International Journal of Computer Science Issues*, vol. 8, no. 4, pp. 199-205, 2011.

[87] H. Huang, W. Guo, and Y. Zhang, ”Detection of Copy-Move Forgery in Digital Images Using SIFT Algorithm,” *In Proceedings of the IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application (PACIIA 2008)*, pp. 272-276, 2008.

[88] J.S. Beis and D.G. Lowe, ”Shape Indexing Using Approximate Nearest-Neighbour Search in High-Dimensional Spaces,” *In Proceedings of the 10th IEEE Computer Society Conference on Computer Vision and Pattern Recogognition (CVPR 1997)*, pp. 1000-1006, 1997.

[89] J.H. Friedman, J.L. Bentley, and R.A. Finkel, ”An Algorithm for Finding Best Matches in Logarithmic Expected Time,” *ACM Transactions on Mathematical Software*, vol. 3, no. 3, pp. 209-226, 1977.

[90] Image manipulation. (http://en.wikipedia.org/wiki/photo_manipulation)

Acknowledgement

First of all, I would like to express my deepest gratitude to my advisor, Prof. Kyung-Hyune Rhee, for his guidance and support during my graduate study at the Pukyong National University (PKNU). He has always encouraged me to be novel and engage in creative thinking and reasoning. His knowledge, vision, patience, enthusiasm, and endless pursuit of excellence have influenced me with lifetime benefits. I deeply appreciate him for helping me reach this milestone in my life and advising me to be a real researcher and professional.

I would like to thank the chairman of my dissertation committee, Prof. Man-Gon Park, and other members, Prof. Sang-Uk Shin, Prof. Song Ha-joo, and Prof. Woen Shin, for their valuable comments, and assistance that have greatly enhanced my thesis. Besides, I would like to appreciate the other professors and secretaries of the department of Information Security and the department of IT Convergence and Application Engineering in PKNU for their help.

My research works contained in this thesis would not have been finished without the supports, encouragements and assistances of LISIA (Lab. of Information Security and Internet Application). My interaction with members of LISIA has certainly made me a better professional. My deep acknowledges to the following previous and current colleagues in LISIA: Dr. Chae-Duk Jung, Dr. Yang Ou, Young-Shin Park, Huaqing Wen, Wang Jing, Sug-Ja Lee, Won-Jun Jo, Nkenyereye Lewis, Otieno Mark Brian, Kouayep Carole Sonia, Nininahazwe Sheilha, Myeong-Hak Heo, Su-Hyun Lee, Hyun-Woo Kim, Kyeong-Deok Kwon, Jung-Hee Kim.

Special thanks to Dr. Chul Sur, and Dr. Yong-Ho Park for their variable discussions, guidance, and helps since beginning of my research. Their thoughtful comments and encouragement on my research are highly appreciated.

I am very grateful to my sincere friends, especially Dorjpurev Batbaatar, Tsegmed Dulbadrakh, Turmunkh Myagmarsuren, Gochoosuren Batjargal, Otgonchimeg Sugarbayar, who took care of me in many ways during my stay in Korea. While the space here does not permit to list all the friends I made in Korea, I appreciate and cherish their friendship.

Lastly, I wish to express my heartfelt gratitude to my parents and family. They have been giving me unconditional support and full understanding during my whole journey of study. My father, Gaanjuur Doyoddorj, is the person who inspired my interests in engineering since childhood, and constantly encouraged me to pursue my dreams and focus on realizing my goals. My mother, Sandag Miyasuren, for my upbringing. Through her endless love and understanding, she instilled in me a sense of diligence and discipline, which have been vital during my life. She emphasized the virtues of honesty, sincerity, and integrity. I am also grateful to my mate, Zandan Enkhzaya, who has kept exemplary patience while I completed my thesis. She is always the wind beneath my wings. This thesis is dedicated to them.

Munkhbaatar Doyoddorj
December 30, 2013
LISIA, Busan

Publications

Journals

1. **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”A Feature-Based Robust Watermarking Scheme Using Circular Invariant Regions,” *The Journal of Korea Multimedia Society (KMMS)*, vol. 16, no. 5, pp. 591-600, May 2013.

2. **Munkhbaatar Doyoddorj**, Youngho Park, Kyung-Hyune Rhee, ”A Robust Data Hiding with Large Embedding Capacity and High Visual Quality,” *The Journal of Korea Multimedia Society (KMMS)*, vol. 15, no. 7, pp. 891-902, July 2012.

3. **Munkhbaatar Doyoddorj**, Chul Sur, Youngho Park, Kyung-Hyune Rhee, ”An Improved Data Hiding Scheme Using Extra Space Modulation for Color Palette Image,” *The International Journal of Database Theory and Application (IJDTA)*, vol. 6, no. 2, pp. 97-107, April 2012.

International Conferences

1. Sonia Carole Kouayep, **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”A Copy-Move Forgery Detection Method Based on SIFT-LBP Features”, *The International Workshop on Advanced Image Technology (IWAIT2014)*, January 2014.

2. **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”Robust Copy-Move

Forgery Detection Based on Dual-Transform,” *The 5th International Conference on Digital Forensics & Cyber Crime, (ICDF2C2013)*, LNICST series, September 2013.

3. **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”A Blind Forgery Detection Scheme Using Image Compatibility Metrics,” *The 22th IEEE International Symposium on Industrial Electronics (ISIE2013)*, May 2013.

4. **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”Design and Analysis of a Fragile Watermarking Scheme Based on Block-Mapping,” *The International Cross-Domain Conference and Workshop on Availability, Reliability, and Security (CD-ARES2012)*, LNCS 7465, pp. 654-668, August 2012.

5. **Munkhbaatar Doyoddorj**, Man-Gon Park, Kyung-Hyune Rhee, ”A Block-Mapping Based Fragile Watermarking Scheme for Optimized Tampering Localization,” *The International Conference and Call for Paper, Information and Communication Technology in Education (ICCP2012)*, May 2012.

6. **Munkhbaatar Doyoddorj**, Chul Sur, Youngho Park, Kyung-Hyune Rhee, ”Reversible Data Hiding with Enhanced Hiding Capacity Based on the Color Palette Image,” *The 1st International Conference on Advanced Signal Processing (ASP2012)*, SERSC series, March 2012.

7. **Munkhbaatar Doyoddorj**, Youngho Park, Kouichi Sakurai, Kyung-Hyune Rhee, ”Reversible Watermarking with Large Capacity Based on the Differences of Center Pixels,” *The 29th Symposium on Cryptography and Information Security (SCIS2012)*, pp. 114-116, February 2012.

8. **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”A Novel Secure Image Hashing Based on Reversible Watermarking for Forensic Analysis,” *The International Cross-Domain Conference and Workshop on Availability, Reliability, and Security (CD-ARES2011)*, LNCS 6908, pp. 286-294, August 2011.

9. **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”A Region-Based Robust 3D Face Recognition,” *The 8th International Conference on Computer Application (ICCA2011)*, pp. 159-166, 2011.

Domestic Conferences

1. **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”Image Splicing Detection Using Edge Inconsistency for Image Forensics”, *The Conference on Information Security and Cryptography (CISC2013)*, December 2013.

2. **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”Image Forgery Detection Using a Noise Dependent Watershed Transformation”, *The 39th Conference on Korea Information Processing Society (KIPS2013)*, vol. 20, no. 1, pp. 667-670, 2013.

3. **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”A Feature Based Robust Watermarking Scheme Using a Circular Region”, *The Conference of Korea Multimedia Society (KMMS)*, vol. 15, no. 2, 2012.

4. **Munkhbaatar Doyoddorj**, Youngho Park, Kyung-Hyune Rhee, ”A Perceptual Quality Preserved Reversible Data Hiding Scheme Enhancing Hiding Capacity”, *The Conference of Korea Multimedia Society (KMMS)*, vol. 14, no. 2, pp. 8-12, 2011.

5. **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”A Novel Secure Image Hashing Based on FASHION Model for Forensic Analysis,” *The Conference of Korea Multimedia Society (KMMS)*, vol. 14, no. 1, pp. 64-67, 2011.

6. **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”A Secure 3D Face Recognition System Based on Signatures of Sliced Bit-Planes on the Facial Shape”, *The Korea Institute of Information Security & Cryptology (KIISC)*, pp. 194-203, 2010.

7. **Munkhbaatar Doyoddorj**, Kyung-Hyune Rhee, ”A Privacy Enhancing Scheme Based on Cancellable Secure Biometric Templates for ROI in Face Recognition”, *The Conference of Korea Multimedia Society (KMMS)*, vol. 13, no. 1, pp. 64-67, 2010.