Copyright Warning & Restrictions

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

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

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

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

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

ABSTRACT

COMPUTATIONAL INTELLIGENCE IN STEGANOGRAPHY: ADAPTIVE IMAGE WATERMARKING

by
Xin Zhong

Digital image watermarking, as an extension of traditional steganography, refers to the process of hiding certain messages into cover images. The transport image, called marked-image or stego-image, conveys the hidden messages while appears visibly similar to the cover-image. Therefore, image watermarking enables various applications such as copyright protection and covert communication. In a watermarking scheme, fidelity, capacity and robustness are considered as crucial factors, where fidelity measures the similarity between the cover- and marked-images, capacity measures the maximum amount of watermark that can be embedded, and robustness concerns the watermark extraction under attacks on the marked-image. Watermarking techniques are often trade-offs between these factors; for example, a high capacity usually implies more modification on the cover-images and thus lowers the fidelity, and the robustness often applies redundancy and lowers capacity.

Traditional image watermarking schemes place the watermark on the trivial portions of cover images to enable the invisibility; however, the hiding can be easily revealed by statistical analysis. Hence, during recent years, researchers have proposed different image watermarking schemes aiming at improvements from various perspectives, such as embedding the watermark into the frequency spectrum for high fidelity and high security, extending the capacity via iterative embedding, enhancing the undetectability by maintaining the image statistics and improving the robustness applying statistical features like image histogram. But the adaptation to varying, flexible or multi-purposed situations remains a challenge in existing watermarking

methods due to the randomness of the cover image contents. In addition, fewer attempts have been reported to level the trade-off when two or more controversial watermarking factors are required. Moreover, although computational intelligence has grown rapidly in the past decade, applying its adaptation ability in image watermarking remains a gap.

In this dissertation, some adaptive image watermarking schemes are presented. First, to achieve content adaptation on the spatial domain, a novel salient region detection model is presented to automatically segment the cover images into regions-of-interests (ROIs) and region-of-noninterests (RONI). The ROIs containing the most representative information are kept intact during the embedding and the RONI is collated for watermarking. Second, an intelligent image watermarking scheme based on the ROI detection is presented. A novel reversible watermarking algorithm that achieves high capacity and low distortion is firstly introduced. It is then followed by partitioning algorithms to bridge the gap between ROIs based image watermarking schemes and watermarking embeddings on frequency domain. Partition ranking schemes based on entropy as well as swarm intelligence are proposed, not only to optimize the overall watermark embedding, but also to provide flexibility that the watermarking purpose can be determined by the end user. Third, to conquer the robustness issue, a robust image watermarking scheme based on the ROIs detection is presented. With a robust watermarking algorithm based on contrast modulation, it matches the segmented ROIs between the marked-image and the distorted image to rectify the attacks. Finally, an image watermarking system using deep learning is introduced, where the rules of watermark embedding and extraction are learned and generalized in an unsupervised manner, which is fully adaptive to image contents and features.

COMPUTATIONAL INTELLIGENCE IN STEGANOGRAPHY: ADAPTIVE IMAGE WATERMARKING

by
Xin Zhong

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirement for the Degree of
Doctor of Philosophy in Computer Science

Department of Computer Science

December 2018

APPROVAL PAGE

COMPUTATIONAL INTELLIGENCE IN STEGANOGRAPHY: ADAPTIVE IMAGE WATERMARKING

Xin Zhong

Dr. Frank Y. Shih, Dissertation Advisor Professor of Computer Science, NJIT	Date
Dr. Yun Q. Shi, Committee Member Professor of Electrical Computer Engineering, NJIT	Date
Dr. Reza Curtmola, Committee Member Associate Professor of Computer Science, NJIT	Date
Dr. Hai Nhat Phan, Committee Member Assistant Professor of Informatics, NJIT	Date
Dr. Qiang Tang, Committee Member Assistant Professor of Computer Science, NJIT	Date

BIOGRAPHICAL SKETCH

Author: Xin Zhong

Degree: Doctor of Philosophy

Date: December 2018

Undergraduate and Graduate Education:

- Doctor of Philosophy in Computer Science,
  New Jersey Institute of Technology, Newark, NJ, 2018

- Master of Science in Integrated Science and Technology,
  Southeastern Louisiana University, Hammond, LA, 2014

- Bachelor of Management in Electronic Commerce
  Wuhan University of Technology, Wuhan, P. R. China, 2007

Major: Computer Science

Publications:

F.Y. Shih and X. Zhong, "Intelligent watermarking for high-capacity low-distortion data embedding," International Journal of Pattern Recognition and Artificial Intelligence, vol. 30, no. 05, pp.1654003, 2016.

F.Y. Shih and X. Zhong, "High-capacity multiple regions of interest watermarking for medical images," Information Sciences, vol. 367, pp.648-659, 2016.

F.Y. Shih and X. Zhong, "Achieving image watermarking robustness by geometric rectification," International Journal of Pattern Recognition and Artificial Intelligence, vol. 31, no.04, pp.1754007(16 pages), 2017.

F.Y. Shih and X. Zhong, "Automated counting and tracking of vehicles," International Journal of Pattern Recognition and Artificial Intelligence, vol. 31, no.12, pp.1750038(12 pages), 2017.

F.Y. Shih, X. Zhong, I.C. Chang and S.I. Satoh, "An adjustable-purpose image watermarking technique by particle swarm optimization," Multimedia Tools and Applications, vol. 77, no.2, pp.1623-1642, 2018.

X. Zhong and F.Y. Shih, "A high-capacity reversible watermarking scheme based on shape decomposition for medical image," International Journal of Pattern Recognition and Artificial Intelligence, vol.23, no.1, Jan. 2019.

X. Zhong and F.Y. Shih, "An efficient saliency detection model based on wavelet generalized lifting," International Journal of Pattern Recognition and Artificial Intelligence, vol.23, no. 2, pp.1954006, Feb. 2019.

Manuscripts Under Review:

X. Zhong, F. Y. Shih and X. Guo, "Automatic image pixel clustering based on mussels wandering optimization," under submission to Multimedia Tools and Applications.

F. Y. Shih, Y. Shen and X. Zhong, "Development of deep learning framework for mathematical morphology," under submission to International Journal of Pattern Recognition and Artificial Intelligence.

X. Zhong and F. Y. Shih, "Robust multibit image watermarking based on contrast modulation and affine rectification," under submission to International Journal of Pattern Recognition and Artificial Intelligence.

X. Zhong and F. Y. Shih, "A robust and blind image watermarking system based on deep neural networks," under submission to IEEE transaction on Circuits and Systems for Video Technology.

ACKNOWLEDGMENT

Firstly, I would like to express my sincere gratitude to my advisor Prof. Frank Y. Shih for his continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. He has provided me with a fountain of ideas and wisdom throughout the years. His guidance has helped me in all the time of research and writing of this dissertation. I could not have imagined having a better advisor and mentor for my PhD study.

Besides my advisor, I would like to thank the rest of my dissertation committee: Prof. Yun Q. Shi, Prof. Reza Curtmola, Prof. Hai Nhat Phan and Prof. Qiang Tang, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives. I would like to thank the NJIT faculty for providing me a firm base to build my computer science research career.

My sincere gratitude also goes to my labmates for the stimulating discussions and for all the fun we have had in the last years.

Last but not the least, I would like to thank my wife Shuang Yuan. She has supported me in every way throughout writing this dissertation.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

An enormous number of digital images are produced and distributed online with the rapid growth of digital imaging devices. Digital content protection has become a crucial demand as multimedia data are widely spread over Internet. Image watermarking has been adopted as an efficient tool to safeguard the digital properties with various applications of copyright protection, image authentication, data privacy, broadcast monitoring and covertly communication [1-3]. Generally, image watermarking refers to the process of embedding some information (i.e., the watermark) into a cover image to form a marked image. The original image content is protected by only allowing the marked image to be publicly accessible. Only the owners are able to extract or detect the existence of the watermark information.

Although the boundaries between image watermarking and image steganography are sometimes fuzzy, image watermarking is more often treated as an extension of traditional steganography. Because image steganography often highlights the undetectability that the existence of the watermark should be hardly revealed, the capacity indicating the maximum amount of embedded data as well as the reversibility that the cover image can be reconstructed, while image watermarking considers many other issues such as fidelity and robustness.

In a watermarking scheme, fidelity, undetectability, capacity and robustness are often considered as crucial factors, where fidelity measures the similarity between the cover- and marked-images, undetectability guarantees that the watermark is inaccessible without the knowing the algorithms, capacity measures the maximum

amount of watermark that can be embedded, and robustness concerns the watermark extraction under variant attacks on the marked-image. It is often a trade-off between the factors; for example, a higher capacity usually implies more modification on the cover-images and thus lowers the fidelity.

Early image watermarking research focuses on single-bit watermark extraction [7], thus the output determines whether an image contains a watermark. Currently, it is expanded to multi-bit watermarking that requires the extraction of the entire secret message with the demand of various watermark applications.

Based on the extraction demands, watermarking schemes can be divided into non-blindness where the extraction requires the information of cover image along with the marked image, and blindness that only demands the marked image. Usually, image watermarking schemes requires a key in the extracting process for enhanced security. The watermark data can be encrypted [4, 5] by different purposes, such as increasing the perceivable randomness for additional security and decreasing the impact of noise for watermark integrity under attacks.

Based on the domain in which the watermark is inserted, image watermarking can be categorized into spatial-domain and frequency-domain methods. Traditional spatial-domain watermarking schemes place the watermark on the least significant bits (LSB) via substitutions or some mathematical operations [6]. Although the trivial replacement enables the invisibility, LSB-based methods can be easily revealed by statistical analysis. Frequency-domain watermarking methods starts with the purpose of enhancing the undetectability using the spectrum spread scheme [7] to distribute the embedded data around the entire image by inserting the watermark as noise-like data in the low-frequency components. The frequency transformations include discrete cosine transform (DCT), discrete Fourier transform (DFT), and discrete

wavelet transform (DWT) are often applied for this purpose. However, the capacity of the frequency-domain methods is often lower than that of the spatial-domain methods. Too much inserted data in the frequency domain will degrade the image quality significantly.

Based on the watermark embedding algorithms, image watermarking can be categorized into: substitution, modification, and modulation. Substitution schemes replace a part of the cover-image with the watermark. The classic LSB is the most representative one. Modification schemes make additive or multiplicative changes corresponding to the watermark on the cover-image, where the spectrum spread is the representation. Modulation schemes tune the cover-image according to the watermark. Quantization index modulation (QIM) watermarking [8], which tunes the quantized index in orthogonal image transforms for watermark insertion, represents this category.

Existing image watermarking proposals consider the enhancement for one or two watermarking factors. However, the adaptation to varying and multi-purposed situations remains a challenge due to the randomness of the cover image contents in the real world, while the rapid growth of digital imaging requires a flexibility of image watermarking systems. Moreover, less attempts have been reported to level the trade-off when two or more controversial watermarking factors are required. For example, a robust watermarking system often applies redundancy, and hence lowers the capacity. Technically, the computational intelligence, such as optimization, fuzzy logic and deep learning, has enabled a lot of applications in the past decade; but applying it in image watermarking remains a gap. To address these issues in image watermarking systems, this dissertation proposes some adaptive, flexible, and intelligent image watermarking schemes.

1.2 Contributions and Outline of this Dissertation

The contributions of this dissertation can be summarized into a four-fold. Having the purpose of image content adaptation on the spatial domain, a novel salient region detection model is presented to automatically segment the cover images into regions-of-interests (ROIs) and region-of-noninterests (RONI) in Chapter 2. The ROIs containing the most representative information are kept intact during the embedding and the RONI is collated for watermarking. This strategy not only ensures the watermarking fidelity by preserving major contents, but also facilitates the robustness by performing a ROI matching. Chapter 3 presents an intelligent image watermarking scheme based on the ROI detection. A novel reversible watermarking algorithm that achieves high capacity and low distortion is firstly introduced. Using an iterative strategy on the magnitudes of image frequency domain, it embeds a large amount of information without visible degradation to the cover image. Since frequency transforms cannot perform on a concave RONI where the holes are the ROIs. partitioning algorithms are proposed to bridge the gap between ROIs based image watermarking schemes and watermarking embeddings on frequency domain. The embedding operates frequency transforms on the rectangular partitions instead of the entire concave image. Having these partitions, partition ranking schemes based on entropy as well as swarm intelligence are proposed, not only to optimize the overall watermark embedding, but also to provide flexibility that the embed purpose can be determined by the end user. To conquer the affine distortions in watermarking systems, a robust image watermarking scheme based on the ROIs detection is presented in Chapter 4. Accompanied with a robust watermarking algorithm based on contrast modulation, it matches the segmented ROIs between the marked-image and the distorted image to evaluate the distortion parameters of translation, rotation,

scaling and shearing, so that the affine attacks can be automatically rectified. In Chapter 5, an image watermarking system using deep learning is introduced, where the rules of watermark embedding and extraction are learned and generalized by convolutional neural networks in an unsupervised manner. Hence, the scheme is fully adaptive to image contents and features. The robustness is achieved without any prior knowledge of possible attacks and distortions. A challenging application of watermark extraction on camera-captured images is also presented to demonstrate the practicality in the chapter. Finally, conclusions and possible future directions are discussed in Chapter 6.

CHAPTER 2

ROI IDENTIFICATION USING SALIENT REGION DETECTION

2.1 Background

Having the purposes of cover-image crucial information preservation as well as human vision system (HVS) fidelity stabilization, image watermarking schemes applying ROIs have been developed, where the ROIs containing extremely important information in an image must be maintained during watermarking and the watermark is embedded into the RONI. Researchers have proposed various watermarking schemes [9-13] of this kind focusing on medical image authentication. However, the state-of-the-art image watermarking schemes use end-user's prescribed ROIs, i.e., every cover-image is manually annotated. This will lower the computational efficiency.

In this dissertation, ROIs identification without requiring human annotation or prescription is proposed, which not only speeds up the process when it comes to a batch, but also provides intelligence to the system by adapting the content of cover-images. Concretely, a novel bottom-up saliency region detection model is proposed in the system to segment the conspicuous areas on cover images as ROIs.

Human vision system (HVS) can extract distinctive objects from multiple distracters in a scene image rapidly and accurately. The process of extracting all these visually salient locations from various backgrounds is visual saliency detection. In computer vision, automated saliency detection remains a challenging problem, and has received tremendous interests during recent years. Saliency detection has widespread applications, including scene understanding, image perception, content-based image retrieval, and seam carving Visual attention studies have been

categorized into two processes: top-down and bottom-up, based on the causes of attention attraction [14]. Top-down attention requires prior-knowledge and focuses on high-level cognitive factors, such as target-oriented tasks and purpose-oriented tasks. In contrast, saliency detection mainly focuses on bottom-up attention, which is cognitively low-level and is a process of rendering certain pixels which are more conspicuous than the others. Bottom-up saliency has been studied in fixation points prediction, which produces density or probabilistic saliency maps predicting eye gaze patterns in free-viewing tasks of images or videos. Like the traditional figure-ground separation process, salient object detection treats the segmentation task as a binarization problem that labels the salient object regions. However, in contrast to fixation prediction that aims to generate a saliency map with a confidence value at each pixel, the salient object detection algorithms often construct a saliency map and then produce an object mask that overlaps the salient regions.

Researchers have proposed many salient region detection algorithms based on different disciplines. Motivated by an early discovery of the significance of phase spectrum in signals [15], Hou *et al*. proposed a spectrum residual (SR) model [16], which finds that the difference between the original and the smoothed Fourier amplitude spectrum can be used to obtain the saliency map. Then an object map is obtained by empirically thresholding the saliency map. With the similar idea, Hou *et al.* [17] presented an image signature model that produces the saliency map using only the sign of cosine transform coefficients. Achanta *et al*. [18] proposed a frequency tuned (FT) model, which combines several Gaussian band-pass filtered outputs to obtain the saliency map, and then averages the saliency map with mean-shift segmentation. A maximum symmetric surround (MSS) model was proposed to modify the global mean used in FT to the local mean of a most probable symmetrical

surrounding region in the saliency map generation, and then obtain an object map via graph-based segmentation [19]. Cheng *et al.* [24] proposed a global contrast (GC) based model, which uses histogram and region based contrast for the saliency map, and names GrabCut to refine the object mask by a fixed thresholding. A low-level wavelet transform based model was developed to capture local features using a Daubechies wavelet that has five vanishing moments and global features by computing the distribution of the local feature map [20].

Several current models involve Gaussian low-pass filtering in their saliency map creation schemes. This is by based on the assumption that a salient object is spatially clustered in a local region. However, it has been pointed out in [17] and [19] that the selection of the standard deviation of the Gaussian kernel is subjective. The standard deviation should be proportional to the size of the salient objects for a complete detection. Otherwise, the saliency map will only emphasize the objects on the edges. This will end up with having implicit assumptions and prior-knowledge of the object size before the detection, that is contrary to the idea of bottom-up saliency. In addition, physiological and psychophysical evidences have shown that multi-resolution analysis (MRA) is an indispensable factor for HVS attention [21] while several current models analyze the input image in single scale and produce down-sampled saliency maps. Down-sizing the input image suppresses the background by removing higher spatial frequency information but simultaneously eliminates the details of large salient objects. Specifically, FT and MSS (a local version of FT) broaden the spatial frequency range in multi-scale so that produce full-resolution saliency map with an emphasis of the largest salient object.

In this paper, we propose an efficient bottom-up salient object detection model based on wavelets lifting. Wavelets domain is employed to obtain saliency maps for

its multi-resolution properties. A nonlinear wavelet filter bank is designed through generalized lifting [22, 23] for wavelet coefficients enhancement. A saliency map is then obtained through a combination of wavelet coefficients in different color feature channels. An object mask is constructed by developing a simple adaptive thresholding scheme given the saliency maps.

The advantages of the proposed model can be summarized into three-fold, as shown in Figure 2.1. First, the saliency map contains a wide range of spatial frequency information because of the wavelets multiple scale derivation. Second, the proposed model produces full-resolution saliency maps that uniformly highlight multiple salient objects of different sizes and shape. Third, the proposed model uses no kernels, in which the parameters involve implicit assumptions and prior-knowledge.

Figure 2.1 An example of the proposed salient region detection. (a) An input image (b) Our saliency map (c) Our object mask (d) Ground-truth.

2.2 Algorithm

Figure 2.2 illustrates the overall process of the proposed model. Firstly, an input image (normalized to [0, 1]) is transformed into a set of broadly-tuned color feature channels for early visual feature extraction. These channels are constructed by considering the opponency of colors in HVS according to the detection mechanism in the cortex and neurons [21]. Let *r*, *g*, and *b* denote the red, green, and blue channels of

an image respectively. This broadly-tuned color feature set contains: (1) The intensity image, (2) $R = r - (g + b) / 2$ for red color tuned channel, (3) $G = g - (r + b) / 2$ green color tuned channel, (4) $B = b - (g + r) / 2$ for blue color tuned channel, (5) $Y = (r + g) / 2 - |r - g| / 2 - b$ for yellow color tuned channel, and additional two color feature spaces defined by $R - G$ and $B - Y$ for the opponent color. Secondly, saliency maps are computed in each channel by developing a wavelet filter bank via generalized lifting scheme, and a comprehensive saliency map is obtained by a linear combination of the maps in each channel. Finally, an object mask is constructed through the saliency map by the adaptive thresholding scheme.

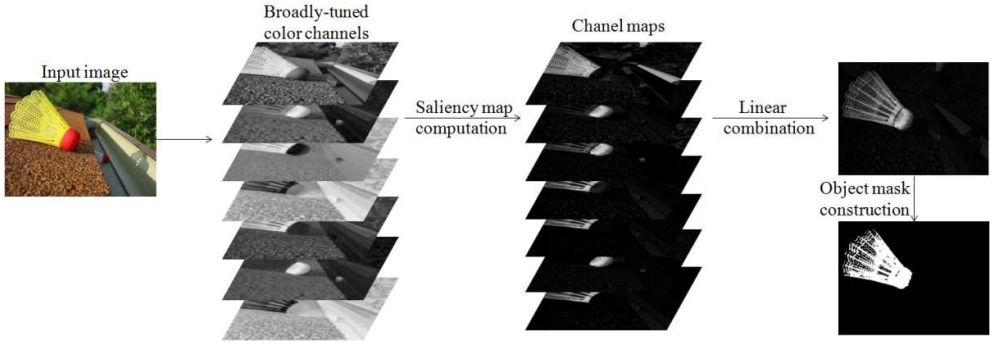


Figure 2.2 The pipeline of the proposed salient region extraction model.

2.2.1 Saliency Map Computation

A wavelet transform captures localized signal information with a zooming procedure that gradually reduces the scale. In order to isolate discontinuities in signals, the wavelet transform enables large temporal supports for lower frequencies while maintaining short temporal widths for higher frequencies. This multi-resolution property extends spatial-frequency analysis into spatial-scale analysis. To decompose

a signal into its projections to subspaces, wavelet transform selects the coarsest scale V_L and the finest scale V_0 that form a chain as

$$V_L \subset V_{L-1} \subset \cdots \subset V_1 \subset V_0 \tag{2.1}$$

and obtain V_0 by

$$V_0 = V_L \oplus_{s=1}^{L} W_s \tag{2.2}$$

where W_s is a subspace containing the differences between successive scales V_s and V_{s+1} L is the total number of scales and $\oplus$ denotes the direct summation.

To obtain *V* and *W*, a scale function and a wavelet function are constructed through translation and dilation of a prototype wavelet basis, where Fourier methods play a key role [25]. However, the translation and dilation of a single basis function imposes the constraints that limit the utility of the multi-resolution idea at the core of wavelet transform. To extend the utility of wavelet methods, Sweldens [22] introduced the second generation wavelets via lifting. The main feature of the lifting scheme is that it provides a spatial interpretation of the wavelet transform which removes the necessity of Fourier analysis. This allows the adaptive customizations of discrete biorthogonal wavelets. Given a signal $X(n)$, a single lifting step involves three basic operations:

Split: Split $X(n)$ into a disjoint polyphase representation. A common sampling scheme is the lazy wavelet, extracting the even and odd polyphase components of $X(n)$.

Dual lifting: Predict the components in phase *i* based on a linear combination of samples of another phase *j*. Then it replaces the components in phase *j* by the difference between the components in phase *i* and the predicted value. The dual lifting operation is also referred to as the prediction. It can be formulated as

$$X_j(n)^{new} = X_i(n) - P(X_j(n)) \tag{2.3}$$

where *P( )* denotes the linear combination operations used in the prediction, *P( )* returns the predicted value.

Primal lifting: Update the components in phase *i* based on a linear combination of the difference samples produced from the dual lifting. The primal lifting operation is also referred to as the update. Mathematically, it is expressed as

$$X_i(n)^{new} = X_i(n) + U(X_j(n)^{new}) \tag{2.4}$$

where *U( )* is the linear combination operations that return the values for the update. The inverse transform is simply the inverse of Equations (2.3) and (2.4).

One restriction in conventional lifting scheme structure is that the dual lifting and primal lifting consider only linear operations. This may result in unsuccessful decorrelation for some complex signals. In order to break out this limitation, Rojals [23] presented an improved method, which absorbs linear or nonlinear operations used in the dual and primal lifting into some bijective mappings. Mathematically, this generalized lifting revises the dual lifting as in Equation (2.3) and primal lifting as in Equation (2.4) to become Equation (2.5) and Equation (2.6), respectively

$$X_j(n)^{new} = P(X_i(n), X_j(n)) \tag{2.5}$$

$$X_i(n)^{new} = U(X_i(n), X_j(n)) \tag{2.6}$$

where *P* and *U* are the mappings used in the prediction and update, respectively. If the mappings arise and arrive on finite sets, *P* and *U* are considered to be injective. In order to guarantee the invertibility of the entire scheme, *P* and *U* must be invertible.

The saliency map computation intends to find the salient regions taking advantage of the spatial-frequency and spatial-scale properties of wavelet transform. If W_s (s = 1, …, L) subspace in Equation (2.2) appropriately collects the local details in each scale, we find that the salient regions can be obtained by reconstructing the input signal without the coarsest scale V_L. This saliency map combines local surround details to a wide range of spatial frequency information due to the multiple scale derivation as in Equation (2.1). Hence, it has high performance for salient object detection. We summarize our computation of saliency map *S* as

$$S = \bigoplus_{s=1}^{L} W_s \tag{2.7}$$

The direct sum $\bigoplus$ is specified as the dual and primal mappings as in Equations (2.5) and (2.6). We will show that this saliency map computation is both analytically and experimentally confirmed. The core problem now in our saliency computation is to find an appropriate local spatial frequency information collector in wavelet transform. Given an image $X(n)$, unlike other applications such as image compression in which a better representative approximation is the goal, we aim to develop a wavelet filter bank that emphasizes the wavelet subspaces for decorrelating

$X(n)$. Therefore, we restrict the dual and primal mapping design to the following criteria:

1. Full reconstruction. This is guaranteed by the invertibility of the dual and primal mappings.

2. Compact support. This is defined by the length of the filters. Compact wavelet support ensures the isolation of singularities, thus enables us to filter out salient regions that possess peculiarities.

3. Smoothness. It is determined by the number of primal or dual vanishing moments. The primal vanishing moments determine the smoothness of reconstruction. The dual vanishing moments determine the convergence rate of subspace projections. Increasing the dual vanishing moments not only decreases the magnitude of the wavelet coefficients and produces a sparser wavelet subspace, but also increases the wavelet support which will reduce the number of large wavelet coefficients produced by isolated singularities [25, 26]. Hence in our case, we need less number of dual vanishing moments.

In this dissertation, we develop a single round nonlinear wavelet filter bank in terms of generalized lifting framework. The input signal $X(n)$ is assumed to be normalized into the range of [0, 1].

1. Split. Starting with a lazy wavelet, we decompose $X(n)$ into even components $X_e(2n)$ and odd components $X_o(2n+1)$.

2. Dual lifting. $X_o(2n+1)$ is used to predict $X_e(2n)$ in scale *s*, and the dual mapping *P* is defined as

$$P(X_o^s, X_e^s) = ||e^{-|X_o^s - X_e^s|}|| \tag{2.8}$$

For collecting a wide range of spatial information that uniformly highlights salient regions, we apply an exponential decay function to emphasize the difference of phase components by weighting more on small differences and less on large differences. Then L2 norm is applied to eliminate the over-sparsity of wavelet sub-bands. It is obvious that the mapping *P* is not only injective but also bijective; hence,

it satisfies criterion 2 of compact support with a narrow filter. To guarantee the dual invertibility, the inverse dual mapping P^{-1} for scale *s* is correspondingly developed as

$$P^{-1}(X_o^{s+1}, X_e^s) = log(||X_o^{s+1}||^{-1}) + X_e^s \tag{2.9}$$

3. Primal lifting. Primal lifting is the preparation of the next scale's dual lifting. $X_e(2n)$ is updated by $X_o(2n+1)$ in such a way indicating the average information of neighboring components. The update mapping *U* and its inverse mapping U^{-1} are respectively defined as

$$U(X_o^{s+1}, X_e^s) = X_e^s + log(||X_o^{s+1}||^{-1})/2 \tag{2.10}$$

$$U^{-1}(X_o^{s+1}, X_e^{s+1}) = X_e^{s+1} - log(||X_o^{s+1}||^{-1})/2 \tag{2.11}$$

For the criteria, full reconstruction and compact support are guaranteed in the scheme construction described above. We then prove that the proposed nonlinear lifting scheme satisfies the smoothness criterion because the dual vanishing moment is one, which is the minimum in any wavelet schemes. A function $\psi(x)$ has *M* vanishing moments if

$$\int x^m \psi(x)\, dx = 0, \quad \text{for } 0 \le m < M \tag{2.12}$$

$$\int x^m e^{cx}\, dx = e^{cx} \sum_{i=0}^{m} (-1)^{m-i} \frac{m!}{i!\, c^{m-i+1}} x^i \tag{2.13}$$

Equation (2.13) is zero if and only if $m = 0$ since only the first term of the summation is zero. Thus, the dual lifting has exactly one vanishing moment that enables large wavelet coefficients for singularities emphasis.

Having this generalized lifting scheme, we are able to obtain the channel saliency maps based on Equation (2.7), as defining the converging condition when the signal has only one component. Thus, we define the coarsest scale L to be the most general information. We summarize the computation process as Algorithm 2.1.

Algorithm 2.1: Saliency Map Computation

Input: One map C in the broadly-tuned color feature set (normalized to [0,1]).

Output: Corresponding saliency map S.

1. Recursively decompose C and the primal lifting result of C using P and U as in Equation (2.8) and Equation (2.10), and converge the decomposition when the primal lifting result has only a single element. Note that for a matrix C, the 2D decomposition at each round is a row decomposition plus a column decomposition.

2. Discard current single element and recursively reconstruct the map using the dual lifting results in each scale using the inverse mappings as described in Equation (2.9) and Equation (2.11).

3. Normalize the reconstructed map into [0, 1] to obtain the result S.

To integrate all the channel maps for the final saliency map, we weight the maximum of all the channel maps by an average (see Figure 2.3). The maximum map contains the most salient information of the input image and highlights both the salient region and background patterns, while the average map conveys more general information of all the channel maps that balance the spatial frequency. The elementwise multiplication of these two maps weights the maximum map in such a way diminishing the spatial frequency information over-highlighted, and therefore suppressing background patterns. We scale the integrated saliency map into [0, 1] for the convenience of object mask threshold.

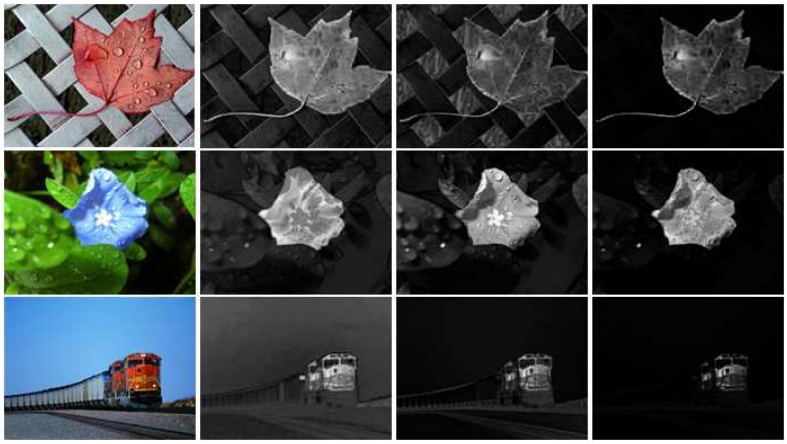


Figure 2.3 Examples of saliency map integration. (a) The input image, (b) the maximum saliency map, (c) the average saliency map, (d) the integrated saliency map.

2.2.2 Object Mask Computation

The effectiveness of a saliency map is usually application-oriented. In the proposed model, we focus on a core application of content-based image processing, i.e., the salient object segmentation. Hence, a binarization of the saliency map is necessary. The goal is to produce a binary mask that overlaps the human labeled salient regions. Some state-of-the-art salient object detection methods used fixed or empirical thresholds for this task, and the result is hence restricted on the threshold selection. Some also combined image segmentation approaches to assist threshold selection; however, their accuracies heavily rely on the segmentation outputs. In this paper, we propose an efficient binarization method that adapts to different saliency maps automatically. It does not require empirical threshold selection, and independently takes the saliency map as the input.

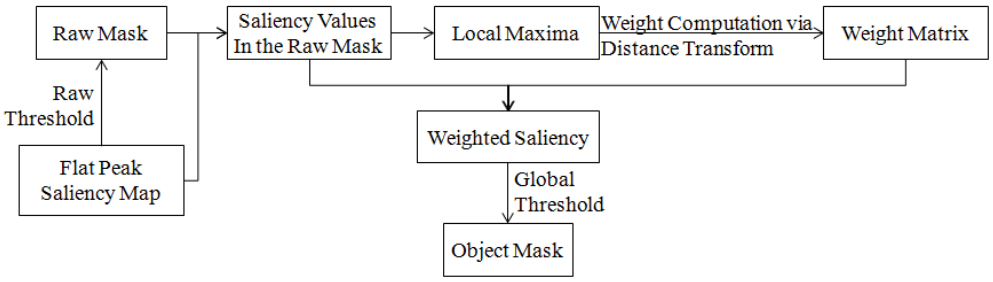


Figure 2.4 The overall process of object mask construction.

Figure 2.4 shows the overall process of object mask construction. There are three main steps in the object mask construction: raw thresholding, weight computation, and global thresholding. The proposed saliency map binarization process is presented as follows.

1. A morphological opening by reconstruction is applied to the saliency maps for the flatness of high saliency values. It contains an erosion process that eliminates undesirable local peaks, followed by a morphological reconstruction that dilates the shape back. The result flattens peaks at high salient regions. We then threshold this flat peak saliency map by the global mean to obtain a raw mask, as shown in the third row of Figure 2.5.

2. The raw mask contains noisy regions that do not belong to the objects. Hence, we define a weight matrix to diminish those noises. The saliency values at a raw mask are collected by the elementwise multiplication between the raw mask and the flat peak saliency map. Then the local maxima of the saliency values at the raw mask are used as the seeds towards a distance transform. Remarkably, local maxima regions with the average saliency less than the global mean of the flat peak saliency map are discarded to eliminate noises. The weight matrix *WD* is defined as

$$WD = ||com(DT)|| \tag{2.14}$$

where *com* denotes the complement operation of an image and *DT* is the result of the distance transform [27]. The weight matrix *WD* contains the regions to be highlighted (as shown in bright areas of the fourth row of Figure 2.5) and the regions to be suppressed (as shown in dark areas of the fourth row of Figure 2.5).

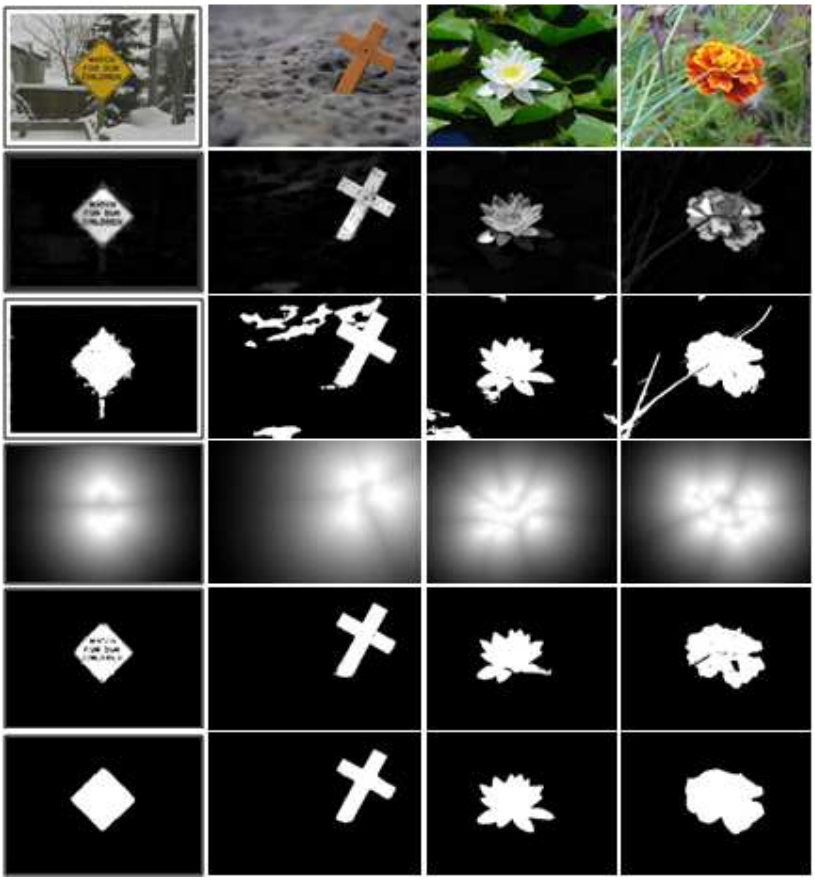


Figure 2.5 Examples of object mask construction. First row: Input images, second row: saliency maps, third row: raw masks, fourth row: the weight matrices via distance transform, fifth row: object masks, and sixth row: the human labeled ground truths.

3. Saliency values at the raw mask are then weighted using the weight matrix *WD*. The next task is to threshold this weighted saliency map for a final object mask. Our normalized and weighted saliency map produces most of values close to zero indicating the background, and a small amount lager values indicating the salient objects. Figure 2.6 shows the average histogram of our saliency maps of 1000 images [18], from which we observe an obvious drop corresponding to the background. To locate this drop, we apply the adaptive global threshold method, which consists of four steps [27]. (1) Initial the threshold T as the global mean, (2) calculate the two mean values μ_1 and μ_2 within the two groups of pixels after thresholding at T, (3) calculate the new threshold $T = (\mu_1 + \mu_2)/2$, (4) converge the algorithm if T does not change.

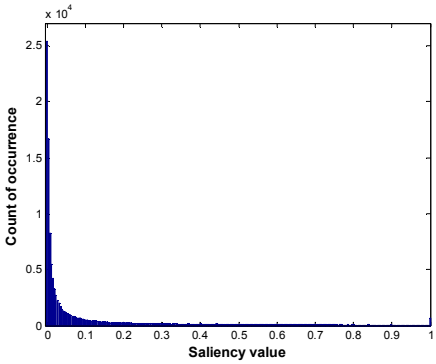


Figure 2.6 The average histogram of our saliency maps of 1000 images.

We summarize the object mask construction process as Algorithm 2.2.

Algorithm 2.2: Object Mask Construction

Input: A flat peak saliency map *S*.

Output: The corresponding object mask *O*.

1. Compute the global mean *M*of the input saliency map *S*.

2. Threshold *S* with *M* to obtain the binarized raw mask *R*.

3. Compute the elementwise multiplication $SR = S \times R$.

4. Find the eight-connectivity local maxima *LM* in *SR*, and discard maxima components with average saliency value less than *M* in *LM*.

5. Compute the weight matrix *WD* in Equation (2.14), and use *LM* as the seeds in the distance transform.

6. Calculate the weighted saliency $WS = WD \times SR$.

7. Threshold *WS* to obtain the final object mask *O* using the adaptive global threshold method mentioned above.

2.3 Experiments

To measure the performance, we present quantitative evaluations and comparisons of the proposed model against various existing methods on two benchmark datasets: Microsoft and FT. The quantitative performance is the evaluation of consistency between the algorithm-produced object masks and the ground truths. We first apply spread-over thresholds to evaluate the performance of saliency maps and then focus on the accuracy of object masks in different approaches.

2.3.1 Saliency Map Validation

In the first experiment our goal is the saliency map validation; hence, Microsoft benchmark dataset [20] is applied. It includes 5000 color images categorized in nine subjects as well as the corresponding human annotated attention regions as the ground truths. Therefore, we provide a variety of challenging situations and a more comprehensive measurement to saliency detection algorithms. Then we compare our saliency maps against the following representative algorithms: the classic Itti saliency method [21] (denoted IT), the spectral residual model [16] (denoted SR), the frequency-tuned salient object detection model [18] (denoted FT), and the low-level feature collection via wavelets transform method [20] (denoted WT). The reasons for selecting these competitors are threefold:

(1) Citation. They are frequently cited saliency detection papers in the state-of-the-art.

(2) Variety. Each model has the uniqueness in terms of the domains. For instances, IT uses spatial domain, SR utilizes Fourier domain, FT exploits Gaussian pyramid, and WT uses wavelet domain.

(3) Relevance. Both SR and FT are considered as frequency-based methods, and WT also uses wavelet domain.

Some examples of the saliency maps produced from these models are shown in Figure 2.7.

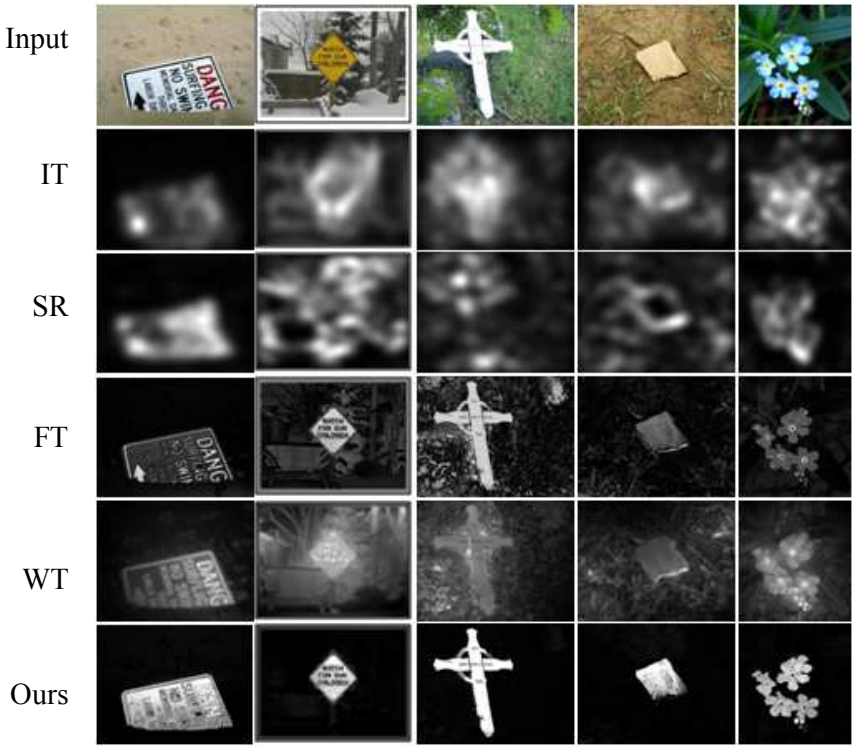


Figure 2.7 Saliency maps of various methods.

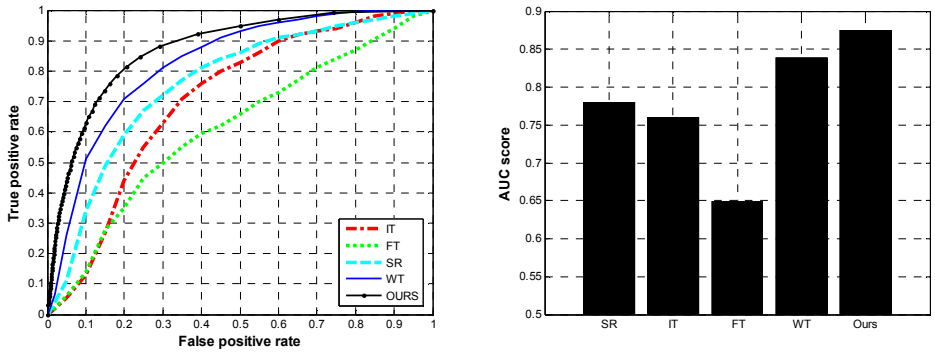
In order to visualize quantitative evaluations of saliency maps in different models, we plot the ROC (receiver operating characteristics) curves and compute their corresponding AUC (area under the curve) scores. In this process, we threshold every saliency map in aforementioned models into a binary mask. This mask is the classification result which segments the positive samples (i.e. salient region) from the negative samples (i.e., non-salient region). At a given threshold T, the true positive rate TPR is defined as the percentage of the positive samples from the binary mask overlapping the positive samples from the corresponding ground truth. The false positive rate FPR is the percentage of the positive samples from the binary mask

overlapping the negative samples from the ground truth. *TPR* and *FPR* can be mathematically described as

$$TPR = TP/(TP + FN) \tag{2.15}$$

$$FPR = FP/(FP + TN) \tag{2.16}$$

where *TP*, *FN*, *FP*, and *TN* denote the number of true positive samples, false negative samples, false positive samples, and true negative samples, respectively. Computing each pair of *TPR* and *FPR* using spread-over thresholds produces the ROC curves in Figure 2.8(a), where the AUC scores are shown in Figure 2.8(b) to measure the effectiveness of a saliency map when the salient objects or regions are segmented. The AUC score is between the chance level 0.5 and the perfect segmentation 1. A larger ROC curve beneath area (i.e., the curve closer to the upper-left corner) or a larger AUC score indicates better performance of the saliency maps in a model. FT shows the lowest AUC score, indicating its object mask relies heavily on its special binarization scheme. Saliency maps in the proposed model outperform those in the reviewed papers by showing the highest 0.8754 of average AUC score of 5000 test images.



(a) (b)

Figure 2.8 ROC and AUC. (a) Average ROC curves in various models, (b) the corresponding AUC scores.

2.3.2 Object Mask Evaluation

We then focus on the object mask evaluation in the second experiment by applying the FT benchmark dataset [18] that is specially developed for salient object detection. It consists of 1000 color images with various sizes and shape of salient objects and human perceived object masks as the ground truths. In addition to IT, SR, and FT models, the global contrast based salient region detection model [24] (denoted GC) is added for comparisons concentrating on the salient object detection. Note that we evaluate the object masks in all the models using their published binarization approaches. In other words, we measure the ultimate output of each salient object detection scheme. Average values of the F-measure are computed based on the ground truths. F-measure is the harmonic mean of precision and recall defined as

$$F = \frac{Recall \times Precision(1+\beta^2)}{Recall + \beta^2 Precision} \tag{2.17}$$

where β^2 is set to be 0.3 to weight precision over recall as suggested in [18]. Precision is defined as the proportion of true positive samples to all positive samples in the algorithm-produced object mask, and recall is the same as *TPR*. The average precision, recall, and F-measure values of various methods are reported in Figure 2.9. Clearly, the object mask in the proposed model shows the highest performance with an averaging F-measure value 0.8452 of 1000 test images.

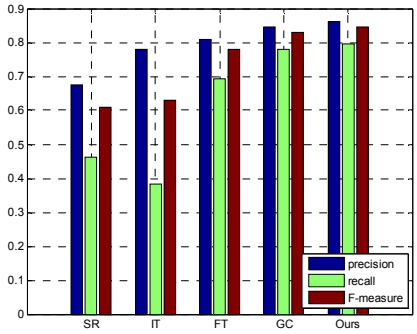


Figure 2.9 Bars of average F-measure for different models.

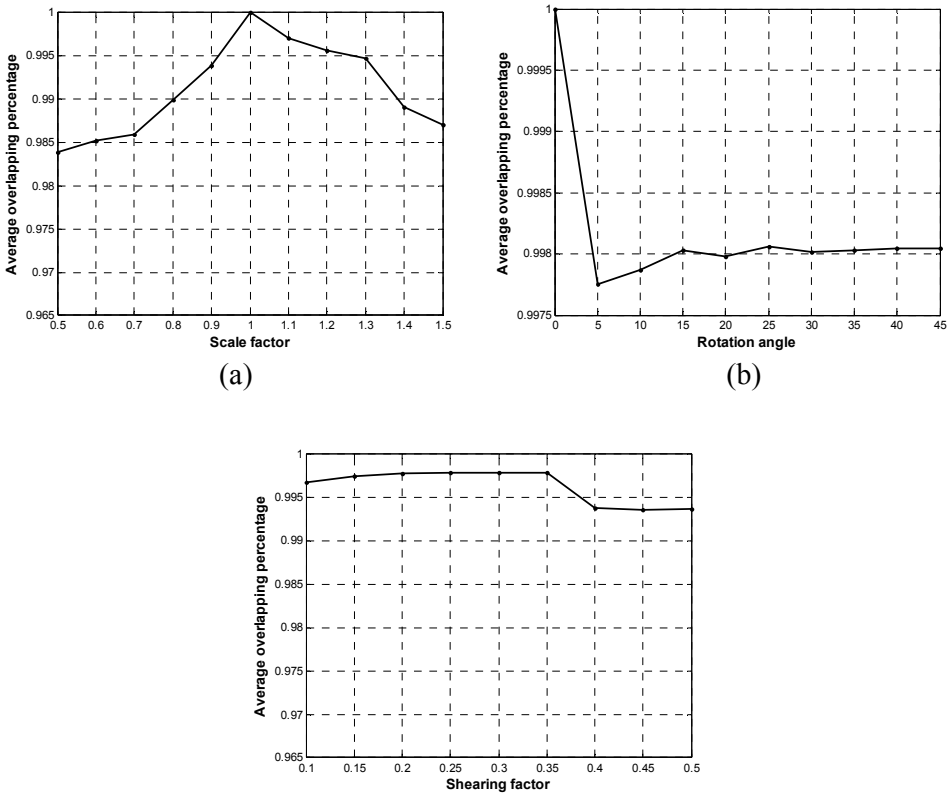
Conclusively, in the proposed model the saliency maps are validated based on the ROC curve and AUC scores, object masks are evaluated using the F-measure. Regarding these criteria, the proposed model yields better results with respect to the relevant state-of-the-art algorithms.

Since our purpose of the salient region detection is to facilitate the ROIs automatic extraction in the adaptive image watermarking, we evaluate robustness and invariance of the proposed model under different affine transformations, including scaling, rotation, and shearing, which simulate the viewpoint changes of the HVS. Commonly, a stable region detector is able to identify similar regions while changing the viewpoints. Similar with HVS, the detected regions should be invariant to image

transformations and covariantly transform with these transformations. The proposed detection model should have covariant detection results due to the covariance characteristics of wavelet domain. We confirm this empirically, and some examples are shown in Figure 2.10, where the distortions consist of a scaling of 0.7, rotation 85° counterclockwise, and a shearing of 0.2 both horizontally and vertically.

Figure 2.10 Saliency maps under affine distortions. (a) Original input images, (b) saliency maps of the original image, (c) affine-distorted input images, (d) saliency maps of the distorted image.

There are three main steps in this confirmation. First, the saliency maps of both the original and the affine distorted images are computed. Two object masks are obtained by thresholding both saliency maps using the proposed model. We then affine transform the object mask of the original image with the same parameters and compute the overlapping percentage between the transformed original mask and the mask computed from the distorted image. We sweep over the scaling factor from 0.5 to 1.5, the rotation angle from 1° to 45°, and the shearing factor from 0.1 to 0.5, respectively, and plot the average overlapping percentage versus the varying factor as shown in Figure 2.11. Based on experimental results, we conclude that the average overlapping percentage of the proposed model in the tested datasets reaches higher than 98%, and therefore the region detection is affine invariant or robust against affine transformations.



(a) (b)

(c)

Figure 2.11 Distortions vs overlapping. (a) Scaling vs overlapping percentage, (b) rotation vs overlapping percentage, (c) shearing vs overlapping percentage.

CHAPTER 3

HIGH-CAPACITY AND INTELLIGENT IMAGE WATERMARKING SHCEME BASED ON THE ROI DETECTION

3.1 Background

Capacity is an important factor in image watermarking. To maximize the amount of watermark bits that a cover image could convey, researchers have been proposing various methods. Among them, image watermarking based on difference expansion [28] is considered as the most significant break-through towards the capacity. By utilizing the nature of high correlation between local pixels, the difference expansion approach can embed multiple bits per pixel. Later, researchers have proposed some improved versions of difference expansion. For example, difference expansion using generalized integer transform [29] and difference expansion embedding in the least significant bits [30]. However, the state-of-the-art proposals based on the difference expansion still suffer from the drawbacks of spatial domain that the embedding can be revealed easily via basic steganalysis.

On the other hand, existing polygon and arbitrarily-shaped ROI-based image watermarking methods insert the secret information into the spatial domain of the RONI, which is less secure than the frequency-domain embedding. Methods applying frequency-domain embedding only use rectangular bounding boxes for ROI enclosure, while the RONI often contains a concave region with arbitrarily-shaped ROIs. Based on the above reasons, the current image watermarking techniques have not achieved the optimization of embedding capacity.

In this chapter, a novel image watermarking scheme for achieving high data capacity and high image quality in frequency domain is presented, where the embedding exploits the nature of images. The high capacity is achieved via an

iterative modification of image magnitudes in the frequency domain inspired by the ideas of the difference expansion in the spatial domain. Hence, the security advantage of frequency-domain based watermarking, that the watermark is spread over the entire image, has been preserved while the watermarking capacity is significantly enhanced.

To bridge the gap between ROIs based image watermarking schemes and watermarking embeddings on frequency domain, partitioning algorithms are proposed so that the embedding operates only frequency transforms on the rectangular partitions instead of the entire concave image. Furthermore, intelligent partition ranking schemes based on entropy as well as swarm optimization are proposed, which not only optimizes the overall watermark embedding, but also provides flexibility that the embed purpose can be determined by the end user.

Figure 3.1 shows the general of the image watermarking scheme in this Chapter. A cover image is firstly segmented into region-of-interests (ROIs) and region-of-noninterests (RONI) using the method in Chapter 2. The RONI is decomposed into non-overlapping (i.e., partitioning) rectangles to facilitate frequency-domain based watermarking algorithms. Each RONI rectangle is then transformed into frequency domain for watermark insertion. The watermark, considered as an arbitrary bit stream, can be encoded for various reasons such as added security and error correction. Each RONI rectangle is embedded with some portions of the watermark to obtain the embedded rectangle. Finally, the embedded rectangles and the intact ROIs are concatenated to generate the marked-image.

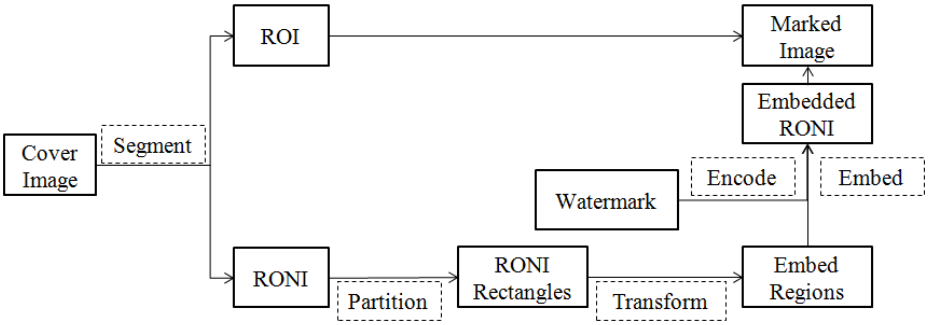


Figure 3.1 General pipeline of the proposed intelligent watermarking scheme.

3.2 Algorithm

3.2.1 A Reversible and High-capacity Image Watermarking Algorithm

From the frequency and statistics perspective, natural images follow the power law [31, 32], low frequency components convey most of the energy of an image. When viewing an image, human vision would focus on those most energetic parts. Especially for medical images which are viewed as piecewise smooth or even piecewise constant, the magnitude spectrums share almost the same distribution that most of the energy concentrates on low frequencies. Figure 3.2 shows an example of the magnitude spectrum of an image in log scale.

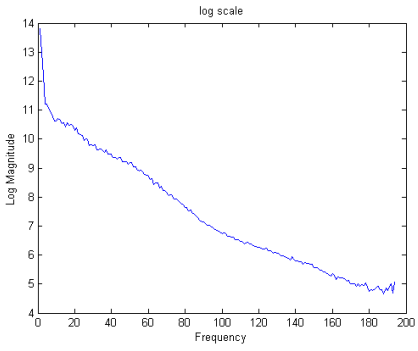


Figure 3.2 The magnitude spectrum.

We exploit the similarity of the magnitude distribution for image watermarking by inserting the watermark into middle frequency components (MFC) of images. MFC avoids the most important parts (low frequency) for human vision, while significantly protecting the watermark from being removed in compression and noises (high frequency). To capture the magnitudes, we adopt DCT since it is more computationally effective and tends to have more of its energy concentrated in a small number of coefficients when compared to other transforms like the DFT, which makes it easier to break up the frequency bands for watermarking. Through DCT, low-frequency components are placed at the upper-left corner and high-frequency components at the lower-right corner. We set low-frequency area as a left triangle with two sides being one-half of the image's width and height respectively, and high-frequency area as a right triangle with two sides being four-fifths of the image's width and height respectively. The remainder is MFC as shown in Figure 3.3(a) white area. Figure 3.3(b) shows a medical image, and Figure 3.3(c) is the reconstructed image after MFC removal. It is observed that removing the MFC does not bring much distortion to human vision.

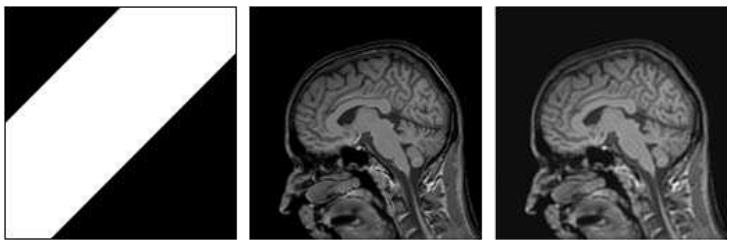


Figure 3.3 MFC of an image.

The watermark insertion process of the proposed high capacity algorithm is to iteratively modify the magnitude of each harmonic wave at MFC. The modification is

inspired by difference expansion [28]. Different from difference expansion, our modification on the frequency domain can be increasive (i.e., to expand the magnitude) or decreasive (i.e., to shrink the magnitude). In addition, we introduce the embedding iteration time t to determine the modification level. Reconstructing the image with the modified magnitude produces a marked image. Figure 3.4 shows an illustration of both increasive and decreasive modification on a harmonic wave.

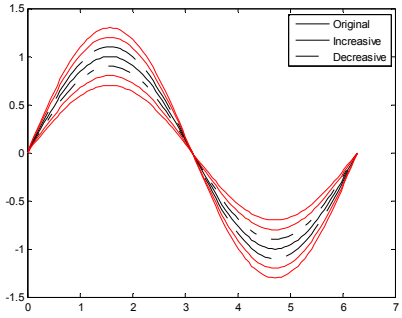


Figure 3.4 Illustration of watermark insertion on a harmonic wave.

Officially, the proposed watermark embedding algorithm with the increasive strategy is presented as Algorithm 3.1.

Algorithm 3.1: Watermark embedding

Input: MFC, watermark bits *S*, and embedding iteration time *t*.

Output: Embedded MFC' and a key *K*.

1. Compute the mean value *M* of MFC. Store *M* in key *K*.

2. For each coefficient *V* in MFC , compute *A* as the arithmetic mean of *V* and *M*. Store *A* in key *K*. Note that we can perform encryptions on *K* for enhanced security.

3. Compute the embedding length *L* by

$$L = \lfloor log(|A - M|) \rfloor \quad (3.1)$$

Note that if L= 0 or $|A - M|$ < the logarithm base, one bit will be embedded.

4. Obtain the embedding integer value E by taking L bits of S.

5. If $V > 0$, update V as V + E
   Else, update V as V – E

6. Repeat steps 1 through 5 by t times.

After embedding, the marked image is obtained by performing IDCT (inverse discrete cosine transform) on the combination of MFC' and the original L&HFA (low and high frequency area). Figure 3.5 presents an example of a medical image before- and after- embedding.

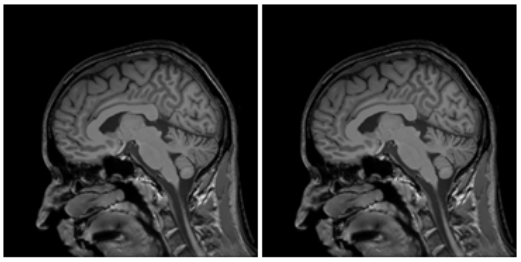


Figure 3.5 An example of cover and marked image.

The corresponding watermarking extraction algorithm is presented in Algorithm 3.2. The original image can be restored so the proposed watermarking scheme is reversible.

Algorithm 3.2: Watermark extracting and MFC restoration

Input: MFC' of a marked image, key K and the embedded iteration time t.

Output: Original MFC and the watermark bits B.

1. For each coefficient V' in the MFC', compute the embedded value E' by

$$E' = |V' - A| \tag{3.2}$$

2. Compute the embedded length L' by Equation (3.1). Note that if L'= 0, one bit will be extracted.

3. Obtain the embedded bits B' by converting E' with length L'.

4. For each coefficient V' in the MFC', update V' by

$$V' = 2A - M \tag{3.3}$$

5. Obtain the original MFC by repeating steps 1 to 4 by *t* times and obtain the whole bits *B* by stacking B'. The original cover image is obtained by performing IDCT on the combination of restored MFC and the original L&HFA.

Experimentally, we evaluate the watermarking capacity as bits per pixel (*BPP*) and the fidelity as the peak signal-to-noise ratio (*PSNR*). *BPP* is computed as the ratio between the total number of bits embedded and the total number of image pixels. *PSNR* is defined as

$$PSNR(f, g) = 10 \times \log_{10} \frac{(2^k - 1)^2}{MSE} \tag{3.4}$$

where *k* is the bit depth, *f* and *g* are the cover and marked image respectively, and MSE is mean square error:

$$MSE = \frac{1}{mn}\sum_{1}^{m}\sum_{1}^{n} \| f(i,j) - g(i,j) \|^2 \tag{3.5}$$

We first explore the relationship between *BPP* and *PSNR* for the proposed algorithm, the 256 × 256 MRI brain image shown as Figure 3.4 with the embed area 36,293 pixels is used. We set the embedding iteration time *t* from 1 to 5. Table 3.1 lists the results, as *t* increases, *BPP* increases; however, *PSNR* decreases. We can continue increasing *t* for higher *BPP* while compromising *PSNR*. It is observed that BPP increases linearly whereas PSNR declines exponentially. So, keeping increasing BPP would not compromise PSNR a lot.

Table 3.1 *PSNR*, *BPP* and Capacity (*t* from 1 to 5) of the Proposed Algorithm

t	1	2	3	4	5
PSNR (dB)	48.53	45.00	43.66	43.07	42.78
BPP	1.63	3.06	4.40	5.71	7.03
Total embedded bits	59,138	110,956	159,552	207,336	255,052

Secondly, we compare the proposed technique against some reviewed state-of-the-art techniques [29, 33, 34, 28, 35, 36, 37, 38]. The results are listed in Tables 3.2 and 3.3. By limit *t* from 1 to 5 as compared to [29], the proposed technique can improve PSNR from 46.36% to 66.03% and BPP from 120.27% to 850%. As compared to [33], the proposed technique can improve PSNR from 12.58% to 27.71% and BPP from 63% to 603%. As compared to [34], the proposed technique can improve PSNR from 45.56% to 65.12% and BPP from 65.65% to 610.10%. As compared to [28], the proposed technique can improve PSNR from 35.90% to 54.16%

and BPP from 232.65% to 1334.69%. As compared to [35], the proposed technique can improve PSNR from 91.32% to 117.04%. Although BPP is 18.50% lower when *t* is 1, it could be increased up by 251.5% when increasing *t* to 5. As compared to [36], the proposed technique can improve PSNR from -16.51% to -5.28% and BPP from 201.85% to 1201.85%. Although our PSNR is at most 16.51% lower, the capacity is increased up to 1201.85%. As compared to [37], the proposed technique can improve PSNR from 34.95% to 53.07% and BPP from 53.77% to 563.21%. As compared to [38], the proposed technique can improve PSNR from -4.17% to 8.71% and BPP from 640.91% to 3095.45%. Although after $t > 3$, we have a lower PSNR, but the capacity is increased by 3095%. These comparisons indicate that the proposed technique achieves high quality and high capacity.

Table 3.2 PSNR Comparisons of the Proposed Technique against Some Techniques

Method	PSNR (dB)	Increment compare to our PSNR in the t range from 48.53 dB down to 42.78 dB Note: Increment is negative means our PSNR is lower				
		t				
		1	2	3	4	5
[29]	29.23	66.03%	53.95%	49.37%	47.35%	46.36%
[33]	38.00	27.71%	18.42%	14.89%	13.34%	12.58%
[34]	29.39	65.12%	53.11%	48.55%	46.55%	45.56%
[28]	31.48	54.16%	42.95%	38.69%	36.82%	35.90%
[35]	22.36	117.04%	101.25%	95.26%	92.62%	91.32%
[36]	51.24	-5.28%	-12.17%	-14.79%	-15.94%	-16.51%
[37]	31.70	53.09%	41.96%	37.73%	35.87%	34.95%
[38]	44.64	8.71%	0.81%	-2.19%	-3.51%	-4.17%

Table 3.3 BPP Comparisons of the Proposed technique against Some Techniques

Method	BPP	Increment compare to our BPP in the t range from 1.63 down to 7.03 Note: Increment is negative means our BPP is lower				
		t				
		1	2	3	4	5
[29]	0.74	120.27%	313.51%	494.59%	671.62%	850.00%
[33]	1.00	63.00%	206.00%	340.00%	471.00%	603.00%
[34]	0.99	65.65%	209.09%	344.44%	476.76%	610.10%
[28]	0.49	232.65%	524.49%	797.96%	1065.31%	1334.69%
[35]	2.00	-18.50%	53.00%	120.00%	185.5%	251.5%
[36]	0.54	201.85%	466.67%	714.81%	957.41%	1201.85%
[37]	1.06	53.77%	188.68%	315.09%	438.68%	563.21%
[38]	0.22	640.91%	1290.90%	1900.00%	2495.45%	3095.45%

3.2.2 RONI Partitioning Algorithms

Frequency transforms cannot be performed on a concave RONI with multiple holes inside. Hence, in this section a task of watermarking on a cover image with multiple ROIs is considered. These ROIs are assumed to contain crucial information of the cover image content; hence they should be kept intact and only the remainder (i.e., RONI) can be used for watermark embedding.

First, we consider that ROIs are identified by rectangular bounding boxes, a binary ROI mask can be simply generated by filling “1” inside each ROI rectangle and “0” for outside RONI part. Then connected component analysis can be used to locate the position of each ROI rectangle by finding each connected area of “1”. Figure 3.6(a) shows an image with two preselected ROIs, and Figure 3.6(b) shows the generated ROI mask. The upper-left corners of both rectangles are (24, 102) with

width 96 and height 129 for the left ROI and (142, 45) with width 95 and height 128 for the right ROI, respectively. The location of each ROI is saved.

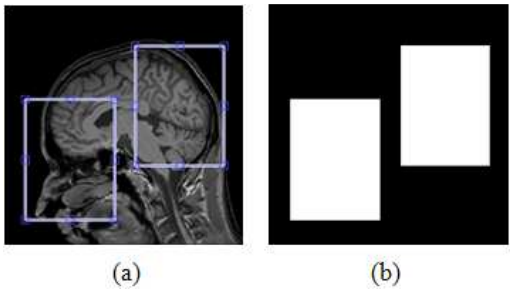


Figure 3.6 An example of rectangular ROI mask.

The concave RONI part shown in black color in Figure 3.6(b) is partitioned into the minimum number of non-overlapping rectangles to facilitate the transformation to frequency domain. It has been proved in [39] that the minimum number of rectangles that a rectilinear polygon can be partitioned is defined by its number of vertices, number of caves and number of collinear concave lines. It can be mathematically stated as:

$$MP = \frac{n}{2} + h - 1 - C \tag{3.6}$$

where *MP* denotes the minimum partition number, *n* denotes the total number of vertices of both the caves' and polygon's, *h* is the number of caves and *C* denotes the maximum cardinality of a set *S* of concave lines, no two of which are intersected.

Inspired by this definition, we develop a novel algorithm for partitioning the concave RONI into the minimum number of rectangles with coordinates of all vertices. This algorithm first draws horizontal and vertical concave lines respectively

for each rectangular ROI, and then selects the result of minimum rectangles as the partition. Let $[(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)]$, respectively denote the coordinates of upper-left, upper-right, lower-left, and lower-right corners of each ROI. The algorithm can be stated as Algorithm 3.3

Algorithm 3.3: RONI partitioning into rectangles

Input: $[(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)]$ of each ROI

Output: Minimum number of rectangles partition

1. While $x_1, x_3 \neq$ image margin
$x_{1\_end} = x_1 - 1; \quad x_{3\_end} = x_3 - 1$
If (x_{1_end}, x_{3_end} go through other ROI) break;

2. While $x_2, x_4 \neq$ image margin
$x_{2\_end} = x_2 + 1; \quad x_{4\_end} = x_4 + 1$
If (x_{2_end}, x_{4_end} go through other ROI) break;

3. Draw lines from point (x_{1_end}, y_1) to point (x_{2_end}, y_2) and from point (x_{3_end}, y_3) to point (x_{4_end}, y_4) to generate horizontal partitioned results.

4. While $y_1, y_2 \neq$ image margin
$y_{1\_end} = y_1 - 1; \quad y_{2\_end} = y_2 - 1$
If (y_{1_end}, y_{2_end} go through other ROI) break;

5. While $y_3, y_4 \neq$ image margin
$y_{3\_end} = y_3 + 1; \quad y_{4\_end} = y_4 + 1$
If (y_{3_end}, y_{4_end} go through other ROI) break;

6. Draw lines from point (x_1, y_{1_end}) to point (x_3, y_{3_end}) and from point (x_2, y_{2_end}) to point (x_4, y_{4_end}) to generate vertical partitioned results.

7. Select the result from steps 3 and 6 that generate fewer rectangles as the final partition result.

Figure 3.7 shows an example of partitioning the RONI concave rectangle into the minimum number of rectangles. Figure 3.7(a) shows the possible horizontal partition, Figure 3.7(b) shows the possible vertical partition, and Figure 3.7(c) shows the corresponding result. There are no collinear lines in this example, and we use the horizontal partition as default. In addition, each RONI rectangle is represented as having value "1" in the binary partitioned mask. Note that the white areas in Figure 3.7(c) are RONI partitioned rectangles for embedding, while the white areas in Figure 3.6(b) indicating ROI for preservation. According to Equation (3.6), we have $n = 12$, $h = 2$, and $C = 0$. Therefore, $MP = 7$ is the optimal partition. Figure 3.8 shows another example of partitioning a three-ROI image with two collinear lines. In this case, we have $n = 16$, $h = 3$, and $C = 2$, and therefore $MP = 8$.

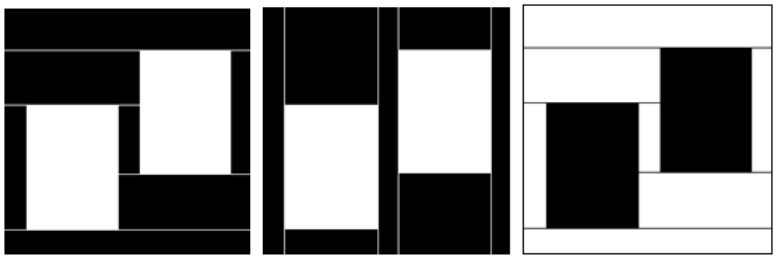


Figure 3.7 An example of partitioning.

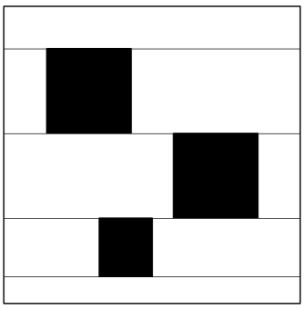


Figure 3.8 An example of partitioning a three-ROI image with two collinear lines.

Secondly, instead of using bounding boxes, we consider a RONI partition algorithm to deal with arbitrarily-shaped ROI inside (see an example in Figure 3.9). Adopting similar concept as in MAT (medial axis transforms), we develop a square-production algorithm. The MAT is a block-based scheme representing connected components in an image by the medial axis and radii [40]. Instead of applying the MAT for skeletonization, we apply MAT to produce different sized squares for watermark embedding.

Figure 3.9 An arbitrarily-shaped ROI example. (a) the cover image, (b) the ROI.

The pixels of the arbitrarily-shaped ROI are used as the seeds to generate distance transform of the image. Then we choose the pixels which are the 3 × 3 local maxima of the distance transform to be the medial axis. The redundant squares containing unions or connected to others are removed by selecting the first scanned

square. Small blocks are eliminated intentionally for the resistance of watermark embedding noises. After each square-production round, we can continuously generate more squares in the remaining RONI area using the arbitrarily-shaped ROI along with the squares produced in the previous round as the new seeds for the distance transform. In general, there are three types of distance measures often used in image processing: Euclidean, city-block, and chessboard. Since the RONI area is decomposed into several square-shaped RONI parts, the 8-neighbor chessboard distance is adopted. The proposed square-production algorithm is presented as Algorithm 3.4.

Algorithm 3.4: RONI partitioning into squares

Input: a binary image with the ROI region masked as "1," denoted as *R*

Output: a squared bounding box image *R_S*

1. Perform the chessboard distance transform on *R* using the arbitrarily-shaped object region (i.e. ROI) as the seeds to obtain *CDT* (the chessboard distance transformed ROI).

2. Upon scanning *CDT*, if a radius is the local maxima (i.e. maximum of 3 × 3 window), keep the distance value; otherwise, reset the pixel's distance to be zero.

3. If the nonzero output pixels are connected (i.e., they have the same distance value), choose the first scanned pixel to obtain the squares so the redundancy in MAT is removed.

4. Mark the above selected squares to one inside the bounding box *R* to form a new bounding box *R_N*. Remove squares whose radii are smaller than a prescribed threshold *TS*.

5. Use *R_N* as new seeds to repeat steps 1 to 4.

6. The algorithm is terminated if no more squares can be found. Collect those produced squares as well as the object region as *R_S*.

The step-by-step results in the first iteration of Algorithm 3.4 are shown in Figure 3.10, where the non-overlapping white blocks are the squares produced for watermark embedding. Continuing the iteration number *N* will repeatedly produce the squares. The continuous square production is shown in Figure 3.11. In this case, the square production algorithm stops at the fourth round.

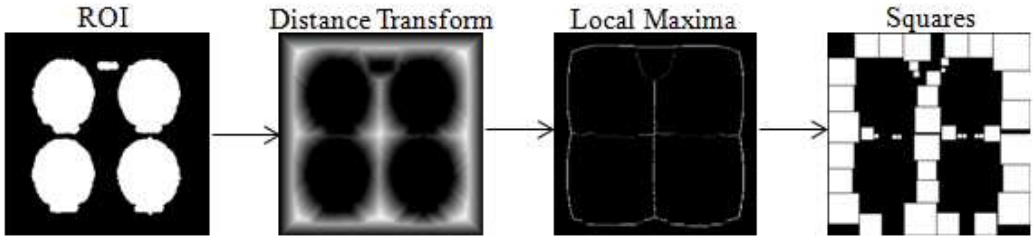


Figure 3.10 The step-by-step results in the first iteration of RONI decomposition.

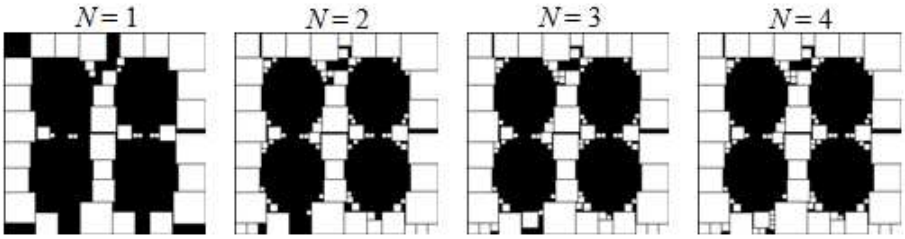


Figure 3.11 Square production for each iteration number *N*.

Note that a different threshold value *TS* will cover the RONI area by a different size of squares. If the threshold is one, the RONI area inside will be completely covered. Increasing the threshold value will approximate the coverage of the RONI area by eliminating small squares. Figure 3.12 shows the squares produced by increasing the threshold value *TS*.

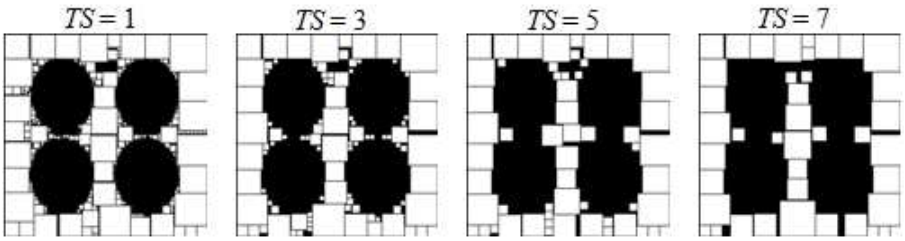


Figure 3.12 Square production of the cover-image by different thresholds.

Without loss of generality, we summarize more details of our low distortion and high capacity watermarking scheme as the pipeline in Figure 3.13. A cover image is firstly segmented with multiple ROIs indicated by either rectangular bounding boxes or arbitrarily-shaped areas. Either algorithm 3.3 or 3.4 is selected to partition the RONI into rectangles according to the shape or ROIs. Each RONI rectangle is transformed into frequency domain with watermark inserting at the magnitude of MFC based on algorithm 3.1. The Marked image is obtained by combining the intact ROIs and the embedded RONI rectangles.

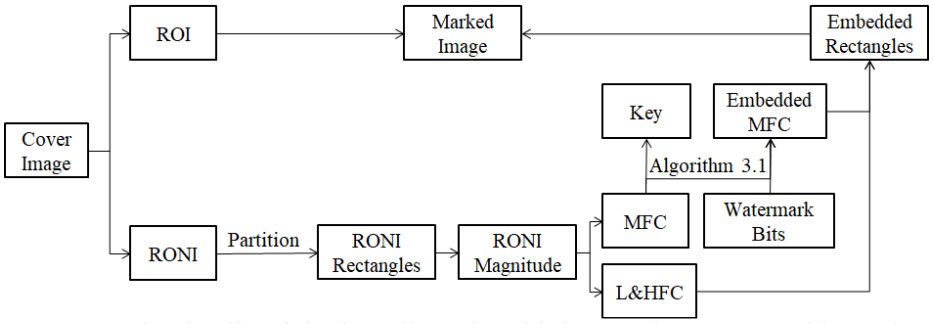


Figure 3.13 The details of the low distortion, high capacity watermarking scheme.

The proposed image watermarking scheme is tested on the MRI medical image dataset from OASIS [41], which consists of a cross-sectional collection of 416 subjects aged from 18 to 96. This dataset is selected because it includes MRI images

of different sizes with multiple arbitrarily-shaped ROIs. Some images from the dataset are shown in Figure 3.14.

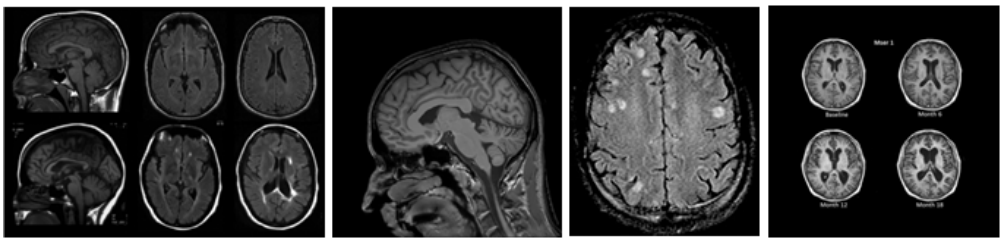


Figure 3.14 Some images from OASIS dataset.

The ratio of the ROI areas with respect to the entire cover-image is computed. Figure 3.15 visually compares the results of applying the arbitrarily-shaped RONI decomposition scheme (Algorithm 3.3) and the ROI bounding box partition method (Algorithm 3.4). The increment is the extra pixels for embedding generated by the arbitrarily-shaped RONI decomposition divided by the pixels for embedding using a bounding box partition. If the ratio of the ROI versus the cover-image is small, the increment is small too. However, if the ratio is large, a significantly high amount of increment will be obtained. Varying the ROI proportion from 30% to 70%, we plot the curves of increments using the threshold value *TS* to be 1, 3, 5, 7, and 9 in Figure 3.16. It is observed that the increments grow exponentially as the ROI proportions increase linearly for all variant thresholds. Note that the squares of radii smaller than *TS* will be removed. If $TS = 1$, all the produced squares of size 1×1 (i.e., a single pixel whose radius is zero) are eliminated.

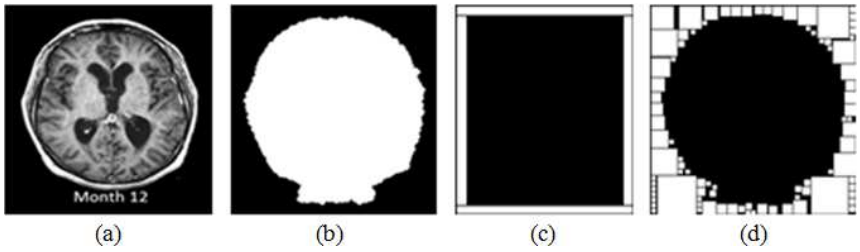


Figure 3.15. Comparisons of increment (a) The cover-image, (b) the ROI, (c) the ROI enclosure with a bounding box, (d) the squares produced for embedding using the decomposition scheme.

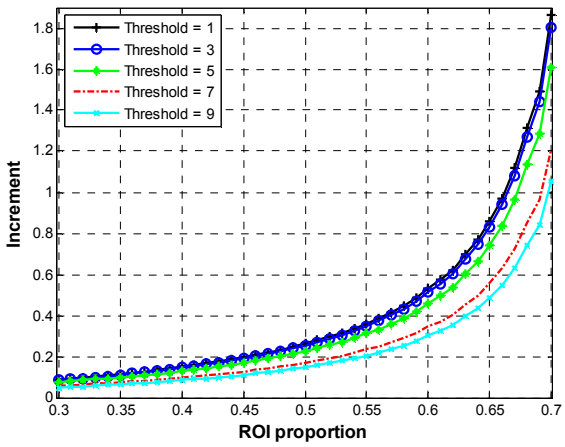


Figure 3.16 ROI proportion vs. increment with varying threshold values.

We compare the proposed scheme against five state-of-the-art, blind, fragile, and ROI-based image watermarking algorithms [37, 42-45]. The following categories are used for comparisons: the embedding domain, the embedding capacity in *BPP*, the ROI annotation method, the ROI enclosure shape, ROI lossless, and *PSNR*. Table 3.4 lists the results, from which we observe that all existing methods need manual annotation of ROI on medical images. Moreover, the existing methods allow arbitrarily- or polygon-shaped ROIs to be embedded into the spatial domain, whereas our methods enable both rectangular and arbitrarily-shaped ROIs to be embedded into

the frequency domain. Besides, the proposed schemes have significant increments in the watermark capacities.

Table 3.4 Comparisons of the Proposed Scheme against Five Existing Schemes

Scheme	Domain	Capacity (BPP)	ROI annotation	ROI shape	ROI lossless	PSNR (dB)
[13]	Spatial	Authentication data only	Manual	Rectangle	Near lossless	Threshold at 32.
[14]	Spatial	0.75	Manual	Polygon	Yes	Related to the selection of polygons
[15]	Spatial	0.39 to 0.89 for ROI extent 30% to 5%	Manual	Arbitrary	Yes	Around 43.06
[16]	Spatial	0.46 to 0.50	Manual	Polygon	Yes	36.71 to 85.50 on 16-bit images.
[17]	Spatial	0.5 when ROI extent is 5%	Manual	Polygon	Yes	Only focus on the extracted and original cover-image.
Ours (Rectangular)	frequency	>= 1.63	Manual	Rectangle	Yes	48.53. Drop exponentially when increasing capacity.
Ours (Arbitrary)	frequency	1.13 to 5.09	Manual	Arbitrary	Yes	56.22

3.2.3 RONI Ranking for Embedding Optimization and Purpose Adjustment

Having the RONI rectangles partitioned, in order to achieve an intelligent and optimized image watermarking scheme, we answer the question “How much information can we place in each rectangle?”. We first address this problem by

considering the appearance of each RONI rectangle, to preserve the fidelity, those rectangles look complex should convey less information than those look flat. Image entropy is adopted to evaluate the complexity of appearance for each RONI rectangle. Image entropy is a statistical measure of randomness, representing the energy or complexity of an image. It can be used to put image regions into an order according to human vision subjective perception of complexity [46]. In this proposal, the entropy of each RONI rectangle is computed by

$$E = -\sum plog_2(p) \tag{3.7}$$

where E is the entropy, p denotes the probability of each pixel obtained from an image histogram. After computing the entropy of all RONI rectangles, we rank them by applying a fuzzy membership using a sigmoid function [47], where the independent value is the entropy and the dependent value is the embedding iteration time t in Algorithm 3.1. Note that t determines the modification level of the magnitude, thus implying the embedding amount. Figure 3.17 shows the sigmoid curve in the proposed scheme.

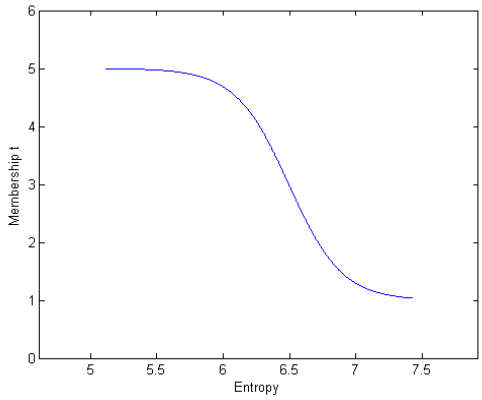


Figure 3.17 The sigmoid membership in the proposed scheme.

Mathematically, we obtain t by

$$t = \left\lfloor \frac{1}{1 + \exp\{a(E - c)\}} (t_h - t_l) + t_l \right\rfloor \tag{3.8}$$

where a is the function acceleration, c denotes the function center, t_l and t_h constitute the range of the membership. According to experiment, a is set to be 2.5, c is the mean value of entropy E, and t_l and t_h are set to be 1 and 5 respectively. After this computing, the embedding iteration time for each RONI rectangle depends on its complexity of appearance.

Figure 3.18 presents an example of ranking the RONI rectangles on a natural image with three prescribed ROIs. Figure 3.18(a) shows the 450 × 300 image with three prescribed ROIs bounding boxes. Figure 3.18(b) is the corresponding RONI partition, where we use Algorithm 3.3 for rectangular ROIs to produce nine RONI rectangles. The t value for each rectangle in Figure 3.18(b) column-wisely is shown in Table 3.5. The embedding amount is larger for rectangles with lower entropy.

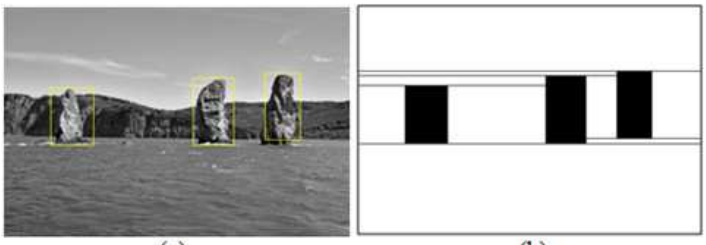


Figure 3.18 An example of embedding a sample watermark image. (a) A natural image with three ROIs. (b) RONI rectangle partition.

Table 3.5 Embedding Iteration Time *t* for Each Rectangle in Figure 5.2(b)

t	5	4	5	1	4	1	2	1	2

Secondly, we enhance the ranking idea by considering both the appearance of the cover image and the watermarking capacity simultaneously via swarm intelligence. Swarm intelligence is a promising computing technology developed from effective processing mechanisms and desired characteristics of biological evolutions. Its effectiveness in global optimizations has been deeply studied and applied. We adopt particle swarm optimization (PSO) [48] in this proposal for its enhanced capability of global optimization. PSO iteratively optimizes a problem by improving existing solutions that represented by particles moving in the search space at each round.

In the proposed image watermarking context, each particle represents an embedding solution of a RONI rectangle with an iteration time *t*. The algorithm initializes a group of particles and computes the fitness to evaluate them. The update process involving the personal best position as well as the global best particle guides the evolution to a better direction. Any updated positions should not exceed the search range, avoiding the particle going to an undesirable field. PSO in the proposed watermarking scheme can be divided into the following steps:

Randomly initialize the position (i.e. an embedding with *t*) and velocity of a group particles.

Compute the fitness for each particle, record the global best particle at current iteration *G* and personal best position of each particle *PB*.

Update the position P' and the new velocity V' for each particle by

$$P' = P + mV \tag{3.9}$$

$$V' = V + \left(c_1 rand()(G - P)\right) + (c_2 rand()(PB - P)) \tag{3.10}$$

where P is current position of a particle, *rand()* is a 0 to 1 random number draw from uniform distribution, V is current velocity of a particle and m, c_1, c_2 are the updating pace factors. We set $m = 0.1$, $c_1 = 1$ and $c_2 = 1$ according to the suggestion in [48].

Update G and PB according to the fitness function.

Stop the algorithm if a termination condition is satisfied, otherwise go to the next iteration.

The fitness function in the proposed scheme, is designed to achieve high fidelity and higher capacity simultaneously after inserting a watermark image into a cover image. We consider the universal image quality index [49] for fidelity evaluation and the number of bits embedded for capacity. The fitness function is computed as

$$fitness = w_1 Q + w_2 C \tag{3.11}$$

where C is the ratio between the number of bits capable to be embedded into a RONI rectangle with a t, and total embedding bits. Considering a host image I_1 and a marked image I_2 of size $M \times N$, the universal image quality matrix Q can be computed by

$$Q = \frac{\sigma_{I_1 I_2}}{\sigma_{I_1}\sigma_{I_2}} \times \frac{2\overline{I_1}\,\overline{I_2}}{\overline{I_1}^2 + \overline{I_2}^2} \times \frac{2\sigma_{I_1}\sigma_{I_2}}{{\sigma_{I_1}}^2 + {\sigma_{I_2}}^2} \tag{3.12}$$

where $\overline{I_1}$ and $\overline{I_2}$ are the global mean of I_1 and I_2, respectively, ${\sigma_{I_1}}^2$ and ${\sigma_{I_2}}^2$ are the variance of I_1 and I_2, respectively, and $\sigma_{I_1 I_2}$ is the cross correlation between I_1 and I_2. The first term of Q is the correlation coefficient, which measures the correlation loss between I_1 and I_2. With a range [-1, 1], the best value 1 indicating both standard deviations are the same. The second term measures the luminance loss with a range [0, 1], the best value 1 indicating the mean luminance are the same. The third term measures the contrast loss with a range [0, 1], the best value 1 indicating no loss of contrast. In overall, the multiplication produces a Q in the range [-1, 1], with the best value 1 indicating images I_1 and I_2 are identical to each other.

Remarkably, two weighting factors w_1 and w_2 are set up in the fitness function. They respectively determine the importance of fidelity and capacity during watermarking to make our scheme adjustable-purpose to the end users. For instance, if quality preservation is the primary purpose for a embedding, a user can set $w_1 > w_2$. Note that $w_1, w_2 \in [0,1]$, and $\sum_{i=1}^{2} w_i = 1$.

Figure 3.19 illustrates the RONI ranking with PSO with an example of embedding a sample watermark image into a natural image. Figure 3.19(a) shows a natural image of size 500 × 300 with four ROIs selected. Figure 3.19(b) is the corresponding RONI partition, Algorithm 3 produces thirteen rectangles in this case. Figure 3.19(c) shows the watermark logo. We use a binary image as the watermark bits. Figure 3.19(d) is the marked image with *PSNR* 57.33. The watermark is embedded into each RONI rectangle using Algorithm 3.1. Figure 3.19(e) is the difference between the cover and marked images, from which we can observe that the watermark embedded amount for each RONI rectangle varies. RONI rectangles with less texture convey more watermark bits according to the results of swarm intelligence.

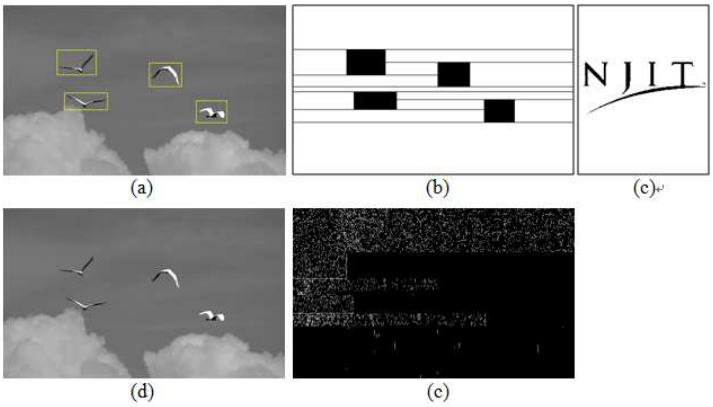
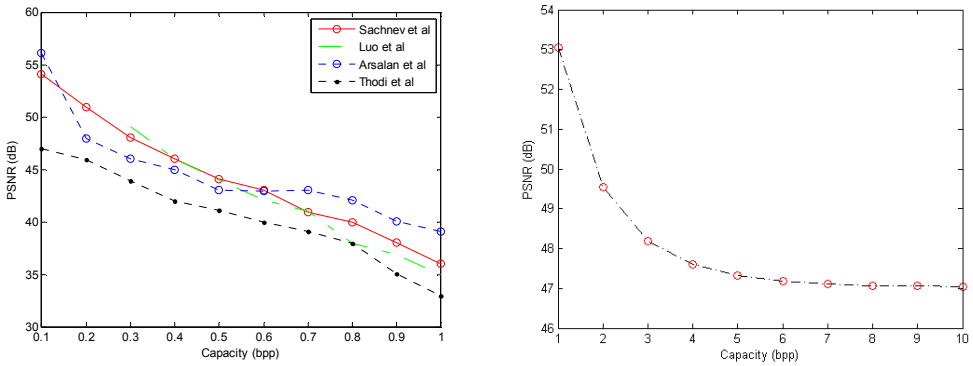


Figure 3.19 An illustration of the RONI ranking with PSO. (a) A natural image with ROIs. (b) RONI rectangle partition. (c) The watermark logo image. (d) The marked image. (e) The difference between original and marked images (amplified by 100 times).

The computational complexity of the proposed algorithm is analyzed as follows. Most evolutionary algorithms have, at each iteration, a complexity of O($n*p + Cof*p$), where n is the dimension of the problem, p is the population size, and Cof is the cost of the objective function. Furthermore, assume that an evolutionary algorithm performs FEs/p iterations, where FEs is the maximum amount of function evaluations allowed. Thus, the complexity cost becomes O($n*FEs + Cof*FEs$). The second term tends to dominate the time complexity, and this complexity is determined by the cost of evaluating the objective function and the amount of evaluations performed. Therefore, the algorithm complexity is measured by the amount of evaluations it performs.

The method in Figure 3.13 is improved with RONI ranking. It is compared with an intelligent watermarking scheme that involves evolution, [50] (Arsalan *et al.*), a reversible watermark scheme with interpolation [51] (Luo *et al.*) and prediction error expansion based image watermarking schemes with sorting [52] (Sachnev *et al.*) or without sorting [34] (Thodi *et al.*). The evaluation criteria include capacity (*BPP*)

and *PSNR* (dB). The results are as shown in Figure 3.20. Figure 3.20(a) shows the *PSNR* vs. *BPP* for the state-of-the-art algorithms and Figure 3.20(b) presents the *PSNR* vs. *BPP* for the proposed scheme.



(a) (b)

Figure 3.20 Comparisons between the proposed watermarking scheme and the state-of-the-art. (a) *PSNR* vs. *BPP* for the state-of-the-art. (b) *PSNR* vs. *BPP* for the proposed scheme.

All the competitors have at most the capacity of *BPP* = 1.0, so we control the *BPP* from 0.1 to 1.0 and observe their quality (*PSNR*). For a fair comparison, we adjust the parameters w_1 and w_2 in our adjustable-purpose scheme to achieve similar *PSNR*, however the capacity in our scheme is at least 1.0, thus the *BPP* is improved at least ten times if we compare the *PSNR* on the same unit. Hence, we conclude that the proposed intelligent image watermarking significantly improves the capacity while preserving a high fidelity.

CHAPTER 4

ROBUST IMAGE WATERMARKING SCHEME BASED ON THE ROI DETECTION

4.1 Background

One crucial challenge in image watermarking systems is the embedded data synchronization when variant distortions are applied on the marked-image, namely the robustness issue. The variant distortions include some sorts of image-processing attacks and geometric attacks. Therefore, a solution often consists of a robust watermarking strategy and a geometric resynchronization.

Robust watermarking strategies are usually constructed to conquer the image-processing attacks on the marked-image, such as filtering, cropping, compression and noising. For example, focusing on the JPEG compression, the quantization-index-modulation (QIM) watermarking [8] tunes the quantized index in orthogonal image transforms for watermark insertion. On the other hand, there exists a few image watermarking strategies aiming to solve both the image-processing and geometric attacks via a single algorithm. For example, the histogram shape-based watermarking scheme [53] that applys the local contrast of image histograms. However, it suffers from a low-payload drawback that only a few bits can be embedded to ensure the robustness.

The approaches to watermark resynchronization under geometric and affine distortions can be blind or non-blind. For non-blind methods, the problem can be addressed through effective search between the distorted and original images due to the availability of the undistorted image [54]. The more challenging blind solution, in which the original image is not available during the watermark extraction, is categorized into three types: invariant domain-based, normalization-based, and

rectification-based. Invariant domain-based approaches embed the watermark into a rotation, scale, translation (RST) invariant domain. For example, Lin *et al.* [55] uses a Mellin transform to convert a rotation into a translation in log-polar coordinates, and apply a translation-invariant transform, such as Fourier transform, to achieve geometric invariance. However, the inverse log-polar mapping, in which the interpolation is involved, introduces deviation and error. Moreover, shearing distortion, which is an important atomic operation in affine transforms, can be hardly solved using domain-based approaches. Normalization-based approaches derive geometric image statistics, such as central moments [56] and Zernike moments [57], to spatially transform both the original and distorted images into a standard status, so the invariance can be achieved. But normalization is global and vulnerable to local changes, such as lossy coding. The rectification-based watermark resynchronization approaches [58] aim to determine and invert the distortion parameters through some reference patterns. One way is to intentionally embed some templates [59], in which rectification is achieved by searching between the original and distorted templates. However, the embedded template not only compromises the payload, but also causes failed synchronization since certain attacks interfere the detection of the template itself. The self-referencing or feature-based rectification methods, in which invariant features on both the original and distorted images are identified, have some advantages because of the avoidance of additional embedding data [60, 61].

In this chapter, a multibit and robust image watermarking scheme by using an improved embedding strategy as well as a synchronization approach is presented. By modulating image contrast information, a watermark embedding strategy that achieves high robustness and high payload simultaneously is developed. The self-referencing rectification approach is used for watermark resynchronization under

affine transformation, in which the centers of mass in affine covariant regions identified by bottom-up saliency detection are extracted and matched to estimate the parameters for affine distortions.

4.2 Algorithm

Figure 4.1 illustrates the pipeline of watermark embedding and extraction processes of the proposed scheme. A watermark image is first converted into a binary sequence, denoted as W, and then encoded for security and robustness enhancement. The encoded watermark is embedded into the cover-image CI by a contrast modulation embedding strategy. The marked-image MI is segmented using the saliency detection described in Chapter 2, for reference region identification. The reference points on the region mask from MI are compared to the corresponding reference points on the mask from the possibly distorted marked-image MI' for rectification. We can compress the region mask by some coding methods (for instance, chain coding) since the mask is Boolean map.

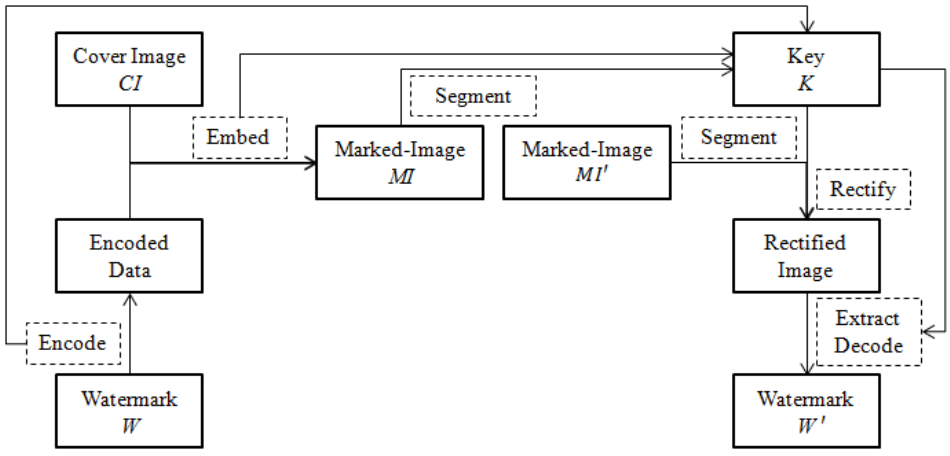


Figure 4.1 The pipeline of the robust ROI-based watermarking scheme.

Watermark extraction involves synchronization since the marked-image MI' is possibly distorted. All the information needed for the extraction is stored in the key K, which includes the encoding parameters, the embedding position and length, and the segmentation mask from MI. The key K is encrypted via Advanced Encryption Standard (AES) during the transition from the encoder to the decoder. The affine covariant reference regions on MI' are segmented and matched by the regions on MI. The distortion parameters are determined by the corresponding centers of mass in these regions. An extracted watermark W' is obtained through the watermark extraction on the rectified image followed by the decoding.

4.2.1 Watermark Encoding

As the proposed scheme deals with robust extraction of a multibit watermark, the watermark bits are encoded for higher performance. The first watermark encoding step involves utilization of error correction code (ECC) to recover error bits extracted from a distorted marked image, although some embedding payload is sacrificed. Bose-Chadhuri-Hocquenghem (BCH) encoding [62] is adopted for this purpose since it provides a variety of code block lengths and error correction rate. Different BCH encoding schemes can be selected to balance the tradeoff between total payload and robustness. For example, a BCH(31,6,7) encoding (i.e., it tolerates 7 random error bits in a codeword of 31 bits.) has a higher correction rate, yet more redundant bits than a BCH(7,4,1) encoding.

The second watermark encoding step applies random permutation to combat the burst errors caused by the cropping and jitter attacks, in which the error bits are grouped in some areas of an image. When the burst errors are occurred, a large number of error bits that exceed the correction capability will be placed within one codeword. Those errors cannot be simply removed by adjusting the BCH coding

parameters. Instead, we randomly permute the BCH encoded data, in such a way that the error bits are spread out while inversing the permutation. Therefore, each codeword will have to correct a less amount of errors, so as to improve the correction rate (also see [59]). The BCH encoding parameters and the random permutation order are stored in the key *K* for watermark extraction.

4.2.2 Watermark Embedding and Extraction Algorithm

A robust watermark embedding strategy aims to find a stable factor on the cover-image for data insertion. Ideally, this factor survives under modifications in various image-processing attacks. This is somehow a dilemma that a robust watermarking scheme requires a factor strong enough to resist different pixel changes, and at the same time weak enough for a higher visual fidelity. Noise-like trivial insertion strategies emphasize on minor modification for marked image quality but can hardly survive under different pixel modifications. Hence, the watermark embedding strategy based on modulation is used, in which some features of the cover-image are adjusted to indicate the watermark. Specifically, we propose an efficient strategy to modulate image contrast features in order to facilitate the tradeoff between robustness and fidelity.

According to Weber's law, a fluctuation is perceivable only when its proportion to the initial stimuli surpasses a threshold. In image context, Weber's contrast is stated as

$$C_{Weber} = \Delta I / I \tag{4.1}$$

where ΔI denotes the changes and I stands for the background stimuli. This physical property implies that a watermark stays invisible when the embedding change

fluctuates within a threshold. At the same time, insignificant pixel modification does not affect the steadiness of the contrast. Thus, we can achieve both imperceptibility and robustness in the watermarking scheme.

In the proposed algorithm, a binary watermark is embedded into a grayscale cover-image by modulating the contrast in an image area. In a similar fashion with root-mean-square (RMS) contrast [63], the mean of the area is used to approximate the background stimuli, and the fluctuation is computed by the difference of intensity values between a reference point and the background stimuli. We modulate the sign of the fluctuation to +1 and -1 according to the watermark bit 1 and 0, respectively. Utilizing the sign function as well as the mean for background approximation introduces tolerance towards pixel modifications. Therefore, the embedded data is robust as long as the modifications do not change the contrast information.

Concretely, a watermark is embedded by modulating a reference point p_r in the low frequency band of integer-to-integer wavelet transform [64] (IWT) by:

$$p_r{}' = \begin{cases} T_{up} & \text{if } W = 1 \text{ and } p_r < T_{up} \\ T_{low} & \text{if } W = 0 \text{ and } p_r > T_{low} \\ p_r & \text{otherwise} \end{cases} \tag{4.2}$$

where W is the watermark, and T_{up} and T_{low} respectively are the upper and lower thresholds. Setting up these two thresholds enables the modulation that the reference point p_r is larger when the embedding bit is 1, and smaller when the embedding bit is 0. The low frequency band of IWT is applied to avoid blocking artifacts, to generate locations in spatial and frequency domains, to contain inherent scaling, to better identify regions that are sensitive to human vision, as well as to achieve high robustness [65]. T_{up} and T_{low} are defined based on the mean intensity:

$$T_{up} = \left\lfloor \frac{\sum_{\forall i \in LW, i \neq r} p_i}{N-1} \times (1+\alpha) \right\rfloor \tag{4.3}$$

$$T_{low} = \left\lfloor \frac{\sum_{\forall i \in LW, i \neq r} p_i}{N-1} \times (1-\alpha) \right\rfloor \tag{4.4}$$

where LW is a slide window specifying an area on the cover-image with N points inside, p denotes the intensity values of pixels in LW, and α is the embedding strength. The reference point p_r is selected in a certain position for the watermark extraction. Figure 4.2 shows a watermark embedding example of a 2×2 window LW. A cover-image is firstly decomposed to four bands using IWT. And at the low frequency band in each LW, the upper-left pixel p_r is selected as the reference point and modulated according to the mean intensity of its three neighbors p_1, p_2, p_3, and the watermark. A higher embedding strength α has a higher tolerance to distortion at the price of a lower fidelity.

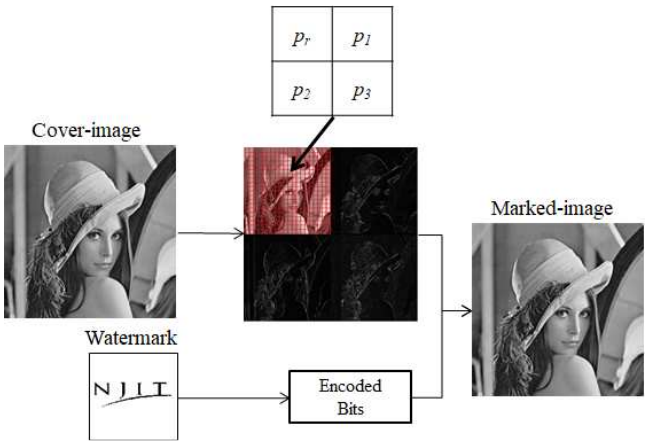


Figure 4.2 An example of watermark embedding.

In the watermark extraction process, a marked-image is firstly decomposed using IWT, and only the modulated reference point $p_r{}'$ and the local mean in LW are compared to determine a watermark bit. The embedding strength α is not required since we can quantize the relationship between $p_r{}'$ and its local neighbors for a binary watermark using a sign function. The watermark extraction for W' can be mathematically stated as

$$W' = \begin{cases} 1, & \text{if sign}\{p_r{}' - \text{mean}(\forall p \in LW)\} = +1 \\ 0, & \text{if sign}\{p_r{}' - \text{mean}(\forall p \in LW)\} = -1 \end{cases} \tag{4.5}$$

The extractions are combined in each slide window LW to produce the entire extracted watermark. Both the position of reference point p_r and the size of each slide window LW can be randomized for enhancing security.

4.2.3 Affine Rectification

Another important factor in a robust watermarking scheme is the watermark resynchronization under geometric or affine distortions. In this Chapter, we focus on general affine distortion, i.e., a set of linear transformations that simulate the view-point change in computer vision. Major atomic affine transformations include rotation, scaling, translation (RST), and shearing. By including the challenging shearing, we can process more general image distortions in observer's view-point changes than merely including RST. In digital images, two-dimensional affine transformations can be mathematically represented by

$$I' = \begin{bmatrix} A & C & 0 \\ B & D & 0 \\ E & F & 1 \end{bmatrix} I \tag{4.6}$$

where I is the original image and I' is the distorted image. Different combinations of A to F represent different operations on I. The parameters of RST transformation are listed in Table 4.1.

Table 4.1 Parameters in Atomic Affine Transformation

Affine Transform	Atomic Matrix	Notation
Rotation	$\begin{bmatrix} cos(\theta) & sin(\theta) & 0 \\ -sin(\theta) & cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$	θ is the rotation angle.
Scaling	$\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$	s_x and s_y are the scaling factor along horizontal and vertical dimensions.
Shearing	$\begin{bmatrix} 1 & sh_y & 0 \\ sh_x & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	sh_x and sh_y are the shearing factor along horizontal and vertical dimensions.
Translation	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t_x & t_y & 1 \end{bmatrix}$	t_x and t_y are the offset along horizontal and vertical dimensions.

Given parameters A to F, an inverse matrix (i.e., the rectification) transforming I' back to I can be deduced as

$$I = \begin{bmatrix} D/N & -C/N & 0 \\ -B/N & A/N & 0 \\ BF - DE & CE - AF & 1 \end{bmatrix} I' \tag{4.7}$$

where $N = AD - BC$. However, in practical applications, parameters A to F are often unavailable, so they need to be estimated. Let $[x', y']$ and $[x, y]$ be the location of pixels on I' and I, respectively. There are three unknowns along each dimension, including A, B, E to convert from x to x' and C, D, F to convert from y to y'. A unique solution of these parameters requires at least three pairs of corresponding reference points $[x_i, y_i]$ and $[x'_i, y'_i]$ $(i = 1,2,3)$.

Center of mass (COM) is selected in affine covariant regions as the reference point for its consistent interrelation with the image. The algorithms discussed in Chapter 2 is applied for the affine invariance. The COM of each segmented region is then computed as the reference point for rectification. Figure 4.3 shows a region mask before and after affine distortions, with the spot inside indicating the covariant COM.

Figure 4.3 Region masks before and after affine distortions with COM.

To warrant at least three pairs of covariant points, we split the region and identify the COM in each subregion to produce sufficient reference points. An example of illustrating the split process is shown in Figure 4.4, where the object is detected on the marked-image *MI*. The major and minor axes of the ellipse having the same second moment with the identified region are used to split the region. The COM is computed on each divided subregion. In this example, we use "*" to plot the general COM of the entire region, "x" to plot the COM of subregion's from major axis split, and "o" to plot the COM of subregion's from minor axis split.

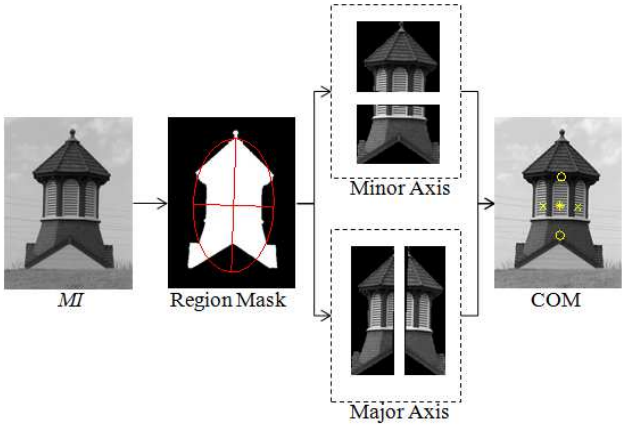


Figure 4.4 Identification of COMs on divided subregions.

The same process is followed to identify the covariant COMs on the distorted marked image MI', as shown in Figure 4.5. COMs on the distorted and the undistorted images are paired for rectification.

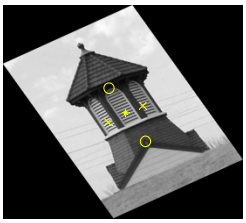


Figure 4.5 Identification of COMs on an affine distorted image.

A rectification could fail when a segmented region is relatively symmetric, where a single split will produce the reference points nearly collinear. Using collinear points to conduct the rectification along a single dimension cannot fully correct the distortion. Figure 4.6 shows a problematic rectification, in which the nearly collinear two "o" points from minor axis split and the "*" point are used as the reference points. Due to the similarity of horizontal positions of the reference points, the resulting image is not horizontally rectified, even if the distorted image MI' is just a simple

rescale to 90% of the original image *MI*. Therefore, the region is split twice to produce a total of five points to prevent from such a failure. We select the general COM "*" along with one of the COMs "o" from minor axis split and one of the COMs "x" from major axis split to produce three triangularly located reference points. In some cases of less than three covariant regions detected, it is necessary to split some regions when the number of noncollinear COMs is less than three.

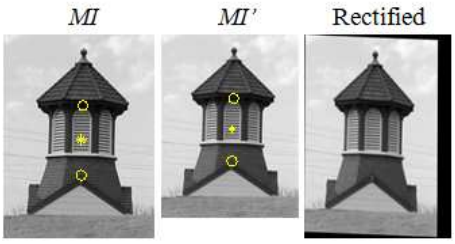


Figure 4.6 A problematic rectification.

After obtaining enough reference points, we associate the same region before and after affine distortion, so that the COMs can be matched. We compute the affine moment invariants (*AMI*), which is a moment-based region descriptor invariant under general affine transformations for each region. The details of comprehensive derivation from classic algebraic invariants to *AMI* can be found in [66]. We compute the *AMI* up to the fourth order to distinguish similarly-shaped regions for matching.

$$AMI^1 = (\mu_{20}\mu_{02} - {\mu_{11}}^2)/{\mu_{00}}^4 \tag{4.8}$$

$$AMI^2 = \begin{pmatrix} {\mu_{30}}^2{\mu_{03}}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30}{\mu_{12}}^3 \\ +4\mu_{03}{\mu_{21}}^3 - 3{\mu_{21}}^2{\mu_{12}}^2 \end{pmatrix} / {\mu_{00}}^{10} \tag{4.9}$$

$$AMI^3 = \begin{pmatrix} \mu_{20}(\mu_{21}\mu_{03} - {\mu_{12}}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) \\ +\mu_{02}(\mu_{30}\mu_{12} - {\mu_{21}}^2) \end{pmatrix} / {\mu_{00}}^7 \tag{4.10}$$

$$AMI^4 = \begin{pmatrix} {\mu_{20}}^3{\mu_{03}}^2 - 6{\mu_{20}}^2\mu_{11}\mu_{12}\mu_{03} - 6{\mu_{20}}^2\mu_{02}\mu_{21}\mu_{03} \\ +9{\mu_{20}}^2\mu_{02}{\mu_{12}}^2 + 12\mu_{20}{\mu_{11}}^2\mu_{21}\mu_{03} \\ +6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} - 18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} \\ -8{\mu_{11}}^3\mu_{30}\mu_{03} - 6\mu_{20}{\mu_{02}}^2\mu_{30}\mu_{12} + 9\mu_{20}{\mu_{02}}^2{\mu_{21}}^2 \\ +12{\mu_{11}}^2\mu_{02}\mu_{30}\mu_{12} - 6\mu_{11}{\mu_{02}}^2\mu_{30}\mu_{21} + {\mu_{02}}^3{\mu_{30}}^2 \end{pmatrix} / {\mu_{00}}^{11} \tag{4.11}$$

where μ_{ij} denotes the central moment of order $i + j$ and is computed by

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j r(x, y) \tag{4.12}$$

where $[\bar{x}, \bar{y}]$ are the coordinates of the COM in $r(x, y)$. Let AMI_l ($l \in [1,2,3, \dots]$) be an array containing AMI^1 to AMI^4 of the l-th subregion on the undistorted image *MI*, and AMI'_{ld} ($ld \in [1,2,3, \dots]$) be an array containing AMI^1 to AMI^4 of the *ld*-th subregion on the distorted image MI'. The l-th subregion on *MI* is matched with *ld*-th subregion on MI' by

$$\underset{ld}{argmin}\{ \sum (AMI_l - AMI'_{ld})^2 \} \tag{4.13}$$

the corresponding COM in the matched regions can be associated after *MI* and MI' are matched.

Figure 4.7 shows the rectification process of a single-object and a multiple-object. The COM of salient regions is first detected on both the marked image *MI* and the distorted marked image MI'. If less than three noncollinear COMs are detected,

we split and select three matched reference points for the rectification; otherwise, we select three matched COMs from different objects.

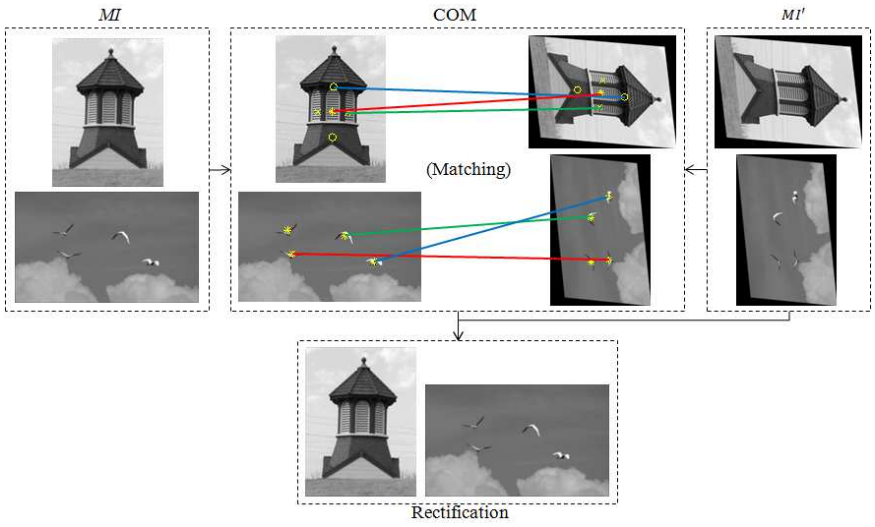


Figure 4.7 Examples of the rectification process.

4.3 Experiments

This Chapter presents a robust image watermarking system with an embedding capacity of on bit per slide window. Increasing the size of the slide window trades the payload for fidelity by decreasing the number of reference points modulated.

4.3.1 Parameter and Tolerance Range

Increasing the embedding strength α can improve the robustness, while the fidelity is compromised by increasing the modulation magnitude. The bit error rate (BER), which is the ratio of the number of different bits between the original and the extracted watermarks to the watermark length, is applied to evaluate the robustness of the syste. A 3×3 LW with p_r being at the center is used. We adopt 120 images from Bruce and Tsotsos [5] as the cover media to include the variations in size and salient region. Some images are shown in Figure 4.8.

Figure 4.8 Sample cover images from Bruce and Tsotsos dataset.

For a varying α, the average BER, *PSNR*, as well as the extraction of the sample watermark under both image-processing and affine attacks are analyzed to illustrate the relationship between robustness and fidelity. For example, the results under salt-and-pepper noise (5%) are given and Figure 4.9(a). The results under a non-linear scaling (0.9 horizontally and 1.1 vertically) are shown Figure 4.9(b). It can be observed that the descending BER and *PSNR* with the ascending α, and BER drops faster than *PSNR* does.

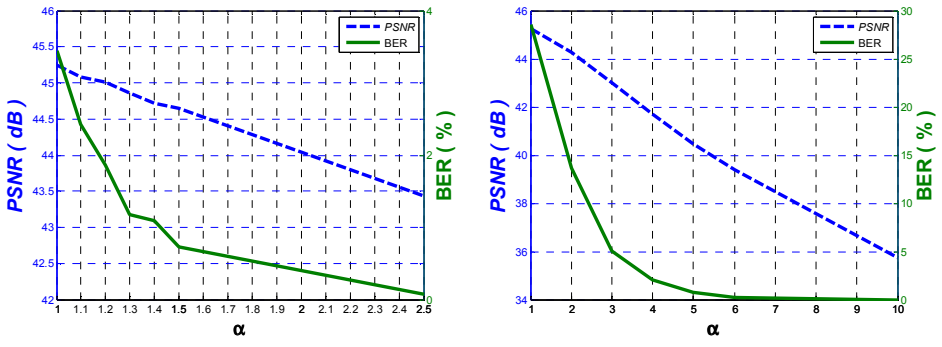


Figure 4.9 Examples of BER and *PSNR* image with varying α. (a) salt-and-peppered (b) scaled.

The tolerance range of the proposed scheme is analyzed by evaluating its responses towards various parameters in different attacks. These parameters determine the strength of certain attacks, such as the percentage of a cropping area. Some distortions on the Lena image and the extracted watermarks using the proposed scheme are shown in Figure 4.10. The distortion parameters vs average BER for the challenging attacks, including cropping, jittering (random removal of rows and columns), rotation, and shearing, are presented in Figure 4.11. The *BER* shows a severe degradation when the marked image is cropped 50%, jittered 30%, or sheared 30%. The BER in each case is larger than 20%. The *BER* under rotation has less fluctuation since the interference of varying rotation angles in the marked image is relatively stable.

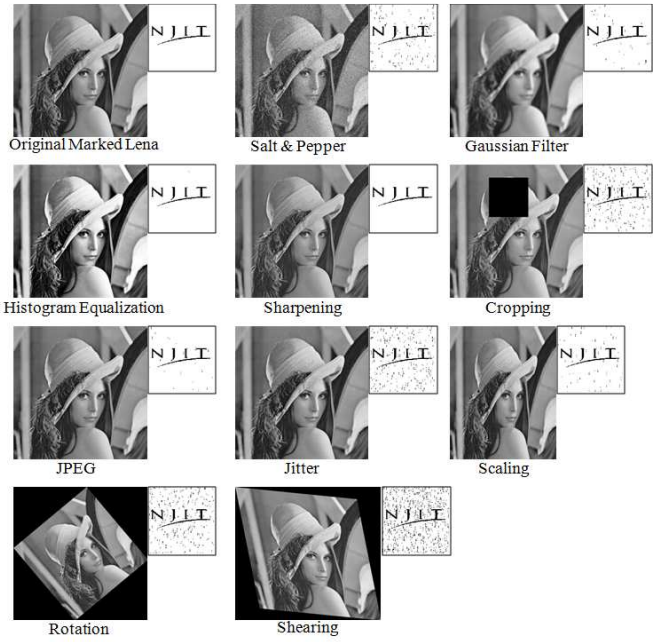


Figure 4.10 Sample attacks and the corresponding sample extractions.

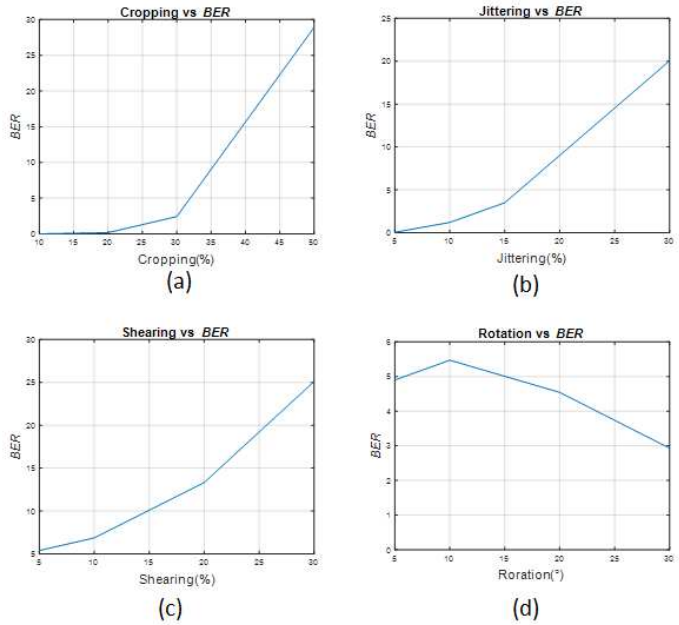


Figure 4.11 Distortion parameters vs *BER*. (a) Cropping, (b) Jittering, (c) Shearing, (d) Rotation.

The proposed scheme shows promising performance with high tolerance range of the challenging distortions. It can tolerate shearing up to 30%, while the template-based rectification method in [59] can only tolerate shearing up to 5%.

4.3.2 Comparative Study

The proposed scheme is compared against four existing multibit robust image watermarking methods [8, 53, 67, 68], where [68] and [53] are statistical feature-based, [8] is modulation-based, and [67] is spectrum-based. The embedding strength α in the proposed scheme is adjusted to achieve close visual fidelity for a fair comparison.

Table 4.2 lists the comparison results of the proposed scheme and the methods in [68] and [53], where 'n/a' (not applicable) means failure to extract or inapplicable in the watermarking scheme. It is observed that the proposed scheme outperforms others by having the lowest BER in most of the attacks and the highest tolerance

range in cropping, jittering, and shearing attacks. Furthermore, the proposed scheme can take care of the histogram equalization and sharpening attacks, whereas the statistical and histogram feature-based methods cannot. Note that in shearing 10%, our scheme shows a higher BER than the method in [53], which used the general histogram shape to be less affected by small pixels loss caused in shearing. However, when an image has a larger pixel change with an increased shearing factor (i.e. shearing 30%), the general histogram shape is destroyed; however, the proposed scheme can still extract a large portion of the watermark.

Table 4.2 BER (%) Comparison of the Proposed Scheme and the Methods in [53, 68]

Attack	[53]	[68]	Ours
Histogram equalization	n/a	n/a	< 0.1
Sharpening	n/a	n/a	< 0.1
Cropping 20%	1.93	16.00	0.13
Cropping 30%	2.73	16.20	1.98
Cropping 50%	n/a	n/a	25.76
JPEG	9.75	16.30	5.73
Jitter 1%	1.67	6.80	< 0.1
Jitter 10%	n/a	n/a	1.75
Salt & Pepper	1.17	6.60	< 0.1
Filtering	2.97	9.60	2.31
Rotation 30°	4.43	10.60	2.93
Scaling	3.83	8.65	0.99
Shearing 10%	3.10	8.40	6.07
Shearing 30%	n/a	n/a	22.87

Table 4.3 shows the comparison of the proposed scheme and the methods in [67] and [8]. The proposed scheme obtains the lowest BER under the filtering and affine distortions, whereas the methods in [67] and [8] achieve lower BERs in JPEG compression since they used the quantized frequency index.

Table 4.3 BER (%) Comparison of the Proposed Scheme and the Methods in [59, 19]

Attack	[59]	[19]	Ours
JPEG	1.45	3.51	5.73
Filtering	3.60	6.25	2.31
Rotation 0.5°	< 0.1	43.67	2.03
Rotation 30°	n/a	n/a	2.93
Scaling	n/a	n/a	0.99
Shearing 10%	n/a	n/a	6.07

In addition, the proposed scheme outperforms the competitors regarding the payload. Under the same *PSNR* around 42.00, the embedding capacity in [53] is fixed at 64 bits, and the capacity in [8] is fixed at 256 bits, while the embedding capacity in the proposed method is linear in terms of cover-image size the and more than 7k bits are embedded in these experiments.

CHAPTER 5

ROBUST IMAGE WATERMARKING USING DEEP LEARNING

5.1 Background

Incorporating deep neural networks with image watermarking has attracted increasing attentions during recent years. Compared to the significant achievements in steganalysis [69, 70], less attempts of applying deep learning in the processes of watermark embedding and extraction are reported. Instead of manual determination of the LSB, some methods in [71-73] use neural networks to assign the significance of the bits for each pixel. Tang *et al*. [74] proposed a variant of a generative adversarial network to determine the embedding position and the strength on cover images. Kandi *et al*. [75] used two deep auto-encoders for non-blind binary watermark extraction, where in the marked image, the pixels produced by the first auto-encoder represent bit zero and the pixels produced by the second auto-encoder indicate bit one. However, the neural networks used in all aforementioned methods do not learn the rule of watermark embedding and extraction. To fully apply the fitting ability of deep learning systems to image watermarking, Baluja [76] proposed a variant of auto-encoders to form a blind scheme aiming at high fidelity as well as high capacity. Li *et al*. [77] embedded the watermark into discrete cosine domain and used convolutional neural networks to cooperate the extraction. Either the embedding or extraction rule is generalized by neural networks; however, due to fragility of neural networks [78], the robustness issue remains a challenge since inputting a modified marked image to a pretrained deep learning system can cause failure in watermark extraction. Like the idea in adversarial networks, Mun *et al*. [79] proposed to resolve this issue by including attack simulation in the training. But the attacks included in the training set

show much higher robustness than those did not. This significantly limits the applications because of the difficulties in enumerating possible attacks in practice.

Different from the state-of-the-art, in this Chapter, we introduce a robust and blind image watermarking scheme using deep convolutional neural networks to generalize the rule of watermark embedding and extraction. The advantages of the proposed model can be summarized into three-fold. First, the proposed system achieves robustness without any prior knowledge of possible attacks and distortions. Second, under the novel construction of the network structure as well as the computation of the loss, the proposed system increases the watermarking capacity as comparing against other robust image watermarking systems. Last, experimental results confirm that the proposed system has an enhanced tolerance range towards common attacks.

5.2 Model

The proposed system is compatible with the deep auto-encoders [80], in which a latent space is learned through a bottleneck to compress the dimensionality of an image. It is structured as two nested auto-encoders, where the outer watermark encoder-decoder network learns a latent space of the binary watermark as the watermark code, and the inner embedder-extractor network controls the visual appearance of the watermark code by referencing the cover image. As a result, an intermediate latent space of the watermark code is obtained as the marked image that appears visually similar to the cover image, while simultaneously contains some information of the watermark. Instead of training for dimensionality compression, the proposed system learns over-complete representation to secure accurate extraction as well as to achieve robustness.

The overall architecture of the proposed image watermarking system is presented in Figure 5.1. First, the encoder network prepares the binary watermark for enhanced robustness and added security. Note that the watermark can be color, grayscale, or binary images, which are converted into binary codes. The marked image is then generated by the embedder network considering both the watermark code and the cover image. An invariance layer provides invariant representation of the marked image, so that its extraction of the watermark tolerates noises and distortions, and therefore provides robustness. Finally, the extractor network restores the watermark code, and the decoder network reconstructs the binary watermark. The proposed system is considered as blind since the watermark extraction only takes from the marked image. The entire system is trained as a single deep network, while we separately describe each component in detail. For illustration purposes, we present the embedding of a 1,024-bit (32 × 32) watermark into a 128 × 128 × 3 color cover-image. One can adapt the sizes by slightly modifying the structure, for example, adding a fully-connected layer with 1,024 neurons to receive an arbitrarily-sized binary watermark.

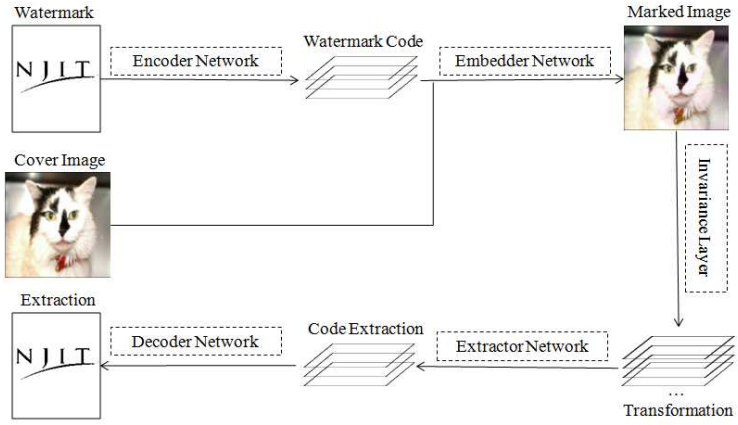


Figure 5.1 The overall architecture of the proposed deep watermarking system.

5.2.1 Watermark Encoder and Decoder Networks

The encoder and decoder network learns an over-complete latent space of binary watermark as watermark code. A 32 × 32 watermark is input into the encoder network and successively increased into 24 channels and 48 channels by convolutional blocks. The 48-channel feature map is the watermark code to be used in the embedder network. On the other hand, the decoder network receives a 32 × 32 × 48 watermark code and restores it to a 32 × 32 watermark. Figure 5.2 shows the structure of the encoder and decoder networks.

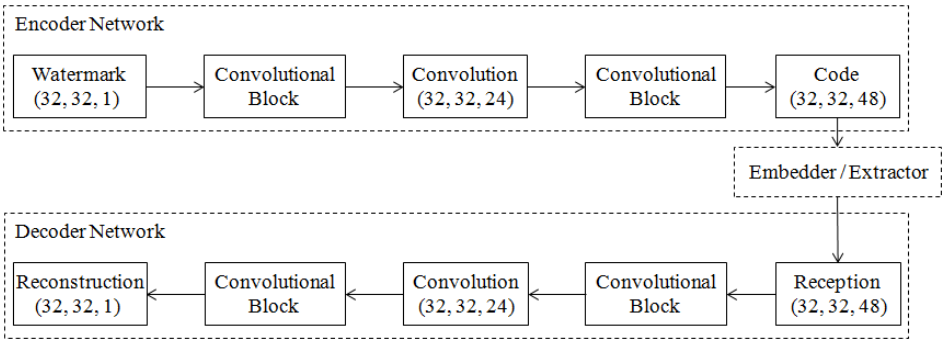


Figure 5.2 Structure of the encoder and decoder networks.

To increase the channels, the inception residual block [81] is adopted to extract crucial features as well as to preserve gradient flow. The block consists of a 1 × 1, a 3 × 3, and a 5 × 5 convolution, and a residual connection which sums up the feature maps and the input itself, so that various perception fields are applied. Each convolution has 32 filters, and the 5 × 5 convolution is replaced by two 3 × 3 convolutions for efficiency. Applying the convolutions will partition the patterns in the binary watermark into different channels, where important features can be duplicated. The convolutional block is shown in Figure 5.3.

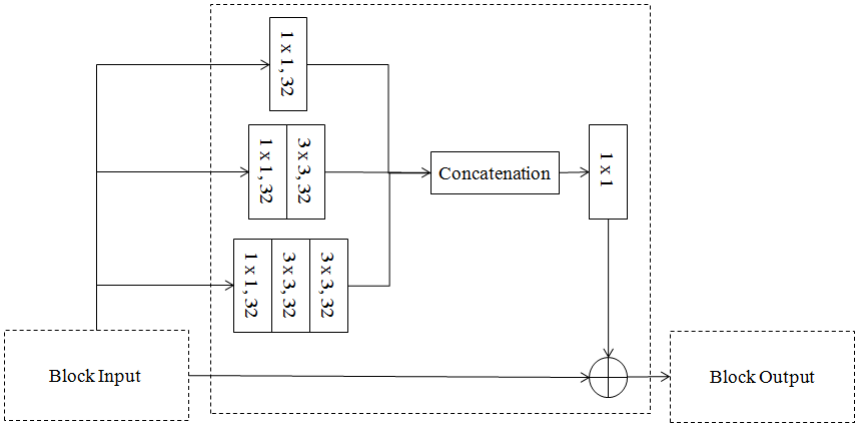


Figure 5.3 The convolutional block.

Encoding the single-channel watermark into 48 channels produces a latent space with redundancy, decomposition, and encryption for adding security and enhancing robustness to the proposed system. As a result, this channel-increased latent space along with the invariance layer accounts for high tolerance range of robustness, so that cropping 65% of the marked image yields only 8% errors in the extraction. A few 32×32 binary watermarks and their corresponding $32 \times 32 \times 48$ watermark codes (reshaped to $128 \times 128 \times 3$ for display) are shown in Figure 5.4, from which we observe the perceivable randomness.

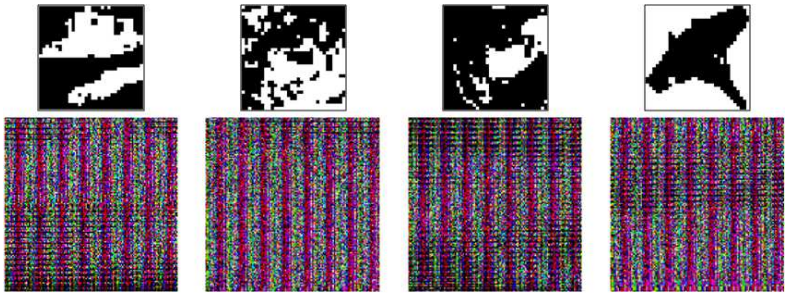


Figure 5.4 Some examples of the watermark code. First row: sample binary watermark, and second row: the corresponding watermark code.

5.2.2 Embedder and Extractor Networks

Having the watermark code along with the cover image, the embedder and extractor network learns a latent space of the watermark code as the marked image, in which its visual appearance must be similar to the cover image and its feature must correlate with the watermark code. The detailed structure of the embedder and extractor networks is shown in Figure 5.5. A reshaped watermark code of size $128 \times 128 \times 3$ is input into the embedder network. A convolutional block is first used to extract features from the watermark code, so that the code details are included in the marked image to facilitate later explanation of the loss. Then, the second convolutional block creates the $128 \times 128 \times 3$ marked image by depth concatenation of the transformed code feature and RGB channels of the cover image as considering different perception fields. The extractor network reversely restores the watermark code from the marked image.

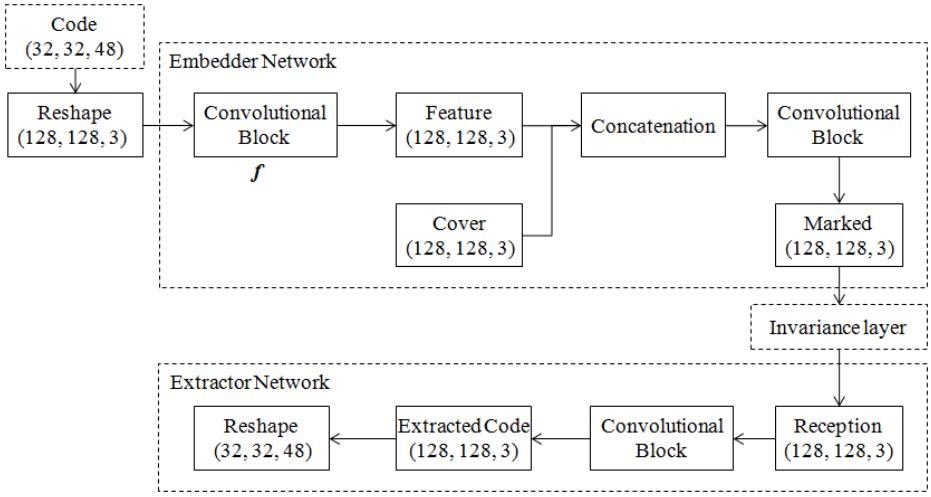


Figure 5.5 Detailed structure of the embedder and extractor networks.

By inputting two images, the embedder and extractor network encodes the watermark code into the least noticeable components of the cover image. Some

examples of the embedding are shown in Figure 5.6. Human vision can hardly tell the differences between marked- and cover-images in spatial domain, while the convolutional blocks in the extractor are able to find the watermark code in feature maps.

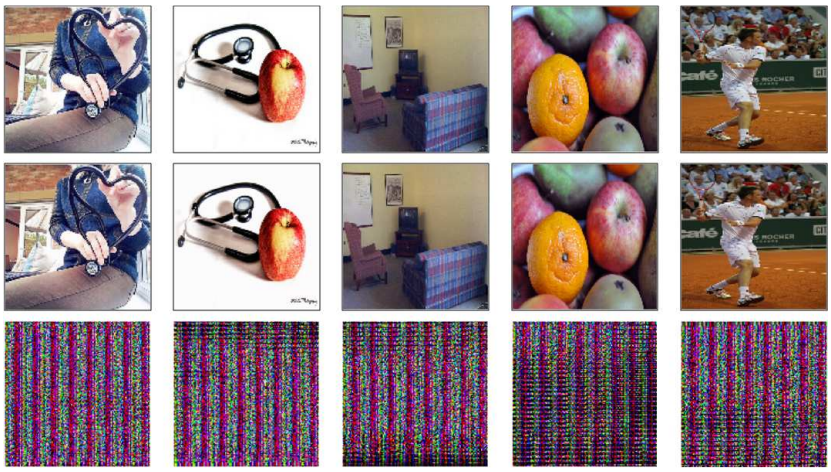


Figure 5.6 Some examples of the embedding. 1st row: the cover images, 2nd row: the marked image, and 3rd row: the embedded and extracted watermark code.

5.2.3 Invariance Layer

For the robustness, the transformer layer learns invariant representation of the marked image. As shown in Figure 5.7, it converts the three-channel marked image into *M*-channel over-complete representation with a fully-connected layer, where *M* is the redundant parameter. Note that the higher *M* is, the higher robustness it can achieve.

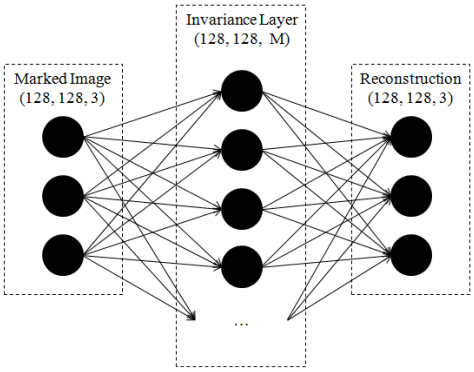


Figure 5.7 The invariance layer.

Referring to the contractive auto-encoder [82], this layer enables the robustness of learned representation of the marked image by a regularization term that is obtained by the fully-connected layer's Frobenius norm of Jacobian matrix with regards to the training input. Mathematically, the penalty term P is given as the partial derivative of the layer to its inputs

$$P = \sum_{i,j} \left(\frac{\partial h_j(X)}{\partial X_i}\right)^2 \tag{5.1}$$

where X_i (i = 1, 2, 3, …) denotes the i-th input and h_j denotes the output of the j-th hidden unit. Similar to the common gradient computation in neural networks, the partial derivative can be written as

$$\frac{\partial h_j(X)}{\partial X_i} = \frac{\partial \alpha(W_{ji}X_i)}{\partial W_{ji}X_i} W_{ji} \tag{5.2}$$

where α is an activation function and W is the weight of the layer.

In practice, we treat each channel as a single neuron for high computational efficiency, structure preservation, and enhanced robustness. Concretely, treating one pixel as an input neuron means 49,152 inputs for a 128 ×128 × 3 marked image. Thus, having 147,456 units in the fully-connected layer requires at least 7,247,757,312 parameters if we only set the redundant parameter *M* as 3, which is not practical in most of current graphical computation units and significantly lowers the efficiency. On the other hand, treating one channel as an input unit considers only 3 input units for the RGB marked image, which enables faster computation as well as a much larger *M* as hundreds for an enhanced robustness.

We propose to apply hyperbolic tangent as the non-linear activation function of the invariance layer for strong gradients as well as bias avoidance [83]. With α assigned as the hyperbolic tangent, *P* can be defined as

$$P = \sum_{j} (1 - h_j^2)^2 \sum_{i} (W_{ji}^T)^2 \tag{5.3}$$

Minimizing term *P* is essentially rendering the weights in the hidden layer unchangeable towards all the inputs *X*. However, placing it as a penalty in the total loss function enables the layer to preserve only useful information while rejecting all other noises and irrelevant information to achieve the invariance.

5.2.4 Loss and Error Propagation

The proposed system is trained as a single and deep network by minimizing the loss function *L*

$$L = \mu_1 || w, e || + \mu_2 || m, c || + \mu_3 || \theta(m, w) || + \mu_4 P \tag{5.4}$$

where *w, e, m, c* denote the watermark, the watermark extraction, the marked image, and the cover image, respectively, θ denotes the correlation computation function, *P* is the regularization term as in Equation (5.3), and $\mu_i, i = (1,2,3,4)$ is the weight controlling the contribution of each term.

The error propagation of each term is presented in Figure 5.9. $||\ w, e\ ||$ is the cyclic term that ensures the similarity between the extraction and the original watermarks. All the components in the system apply this error term during their weights update. The $||\ m, c\ ||$ guarantees the visual similarity between cover- and marked-images by comparing their contents. To explain the error term $||\theta(m, w)||$, we annotate f_1 and f_2 (shown in Figure 5.8) for the convolutional block *f* of Figure 5.5. Under this annotation, $\mu_3||\theta(m, w)||$ can be computed as

$$\mu_3||\theta(m, w)|| = \frac{\mu_3}{2}(||\ g(f_1(cw)),\ g(f_1(m))|| + ||\ g(f_2(cw)),\ g(f_2(m))||) \quad (5.5)$$

where cw is the watermark code and g denotes the gram matrix. Besides the watermark code, the convolutional block *f* also extracts features on the marked image. The feature maps of the marked image must correlate with those of the watermark code. The correlation is maximized by minimizing the distance between the gram matrices. Remarkably, $||\ m, c\ ||$ and $||\theta(m, w)||$ are only applied to the weights of encoder and embedder networks. All the components contain the information of the watermark under this error propagation.

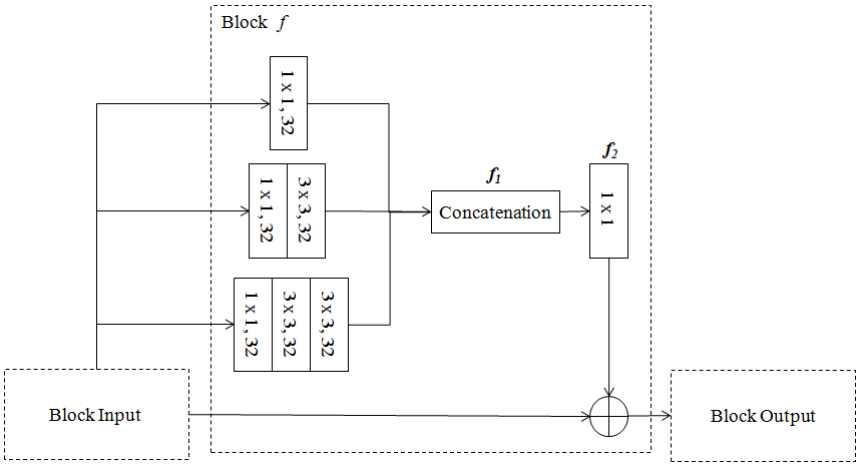


Figure 5.8 f_1 and f_2 in the convolutional block f

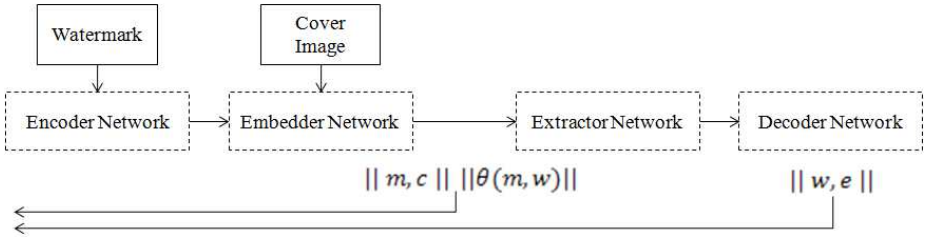


Figure 5.9 The error propagation of the proposed system.

5.3 Experiments

5.3.1 Training, Testing and on Synthetic Images

This chapter presents a robust image watermarking system based on deep neural networks, with a fixed watermark payload of 1,024 bits. The proposed system is trained using ImageNet [84] (rescaled to 128 × 128) as the cover image and the binary version of CIFAR [85] (32 × 32) as the watermark. Both datasets include more than millions of images to introduce a large scope to the system. The ADADELTA [86] optimizer that applies a moving window in gradient updates is adopted for its ability of continuous learning after large epochs. Figure 5.10 shows the value of loss L during 200 epochs. The loss L drops smoothly and converges below 1%. The distance

measures in L are set as the mean absolute error to highlight the overall performance over a few outliers. The $\mu_i, i = (1,2,3,4)$ is set to be one for equal highlight of each error term. All the layers in the system apply the rectified linear unit (ReLU) as the activation function except that the marked image and watermark extraction use sigmoid to limit the output range into (0, 1) and the invariance layer uses hyperbolic tangent.

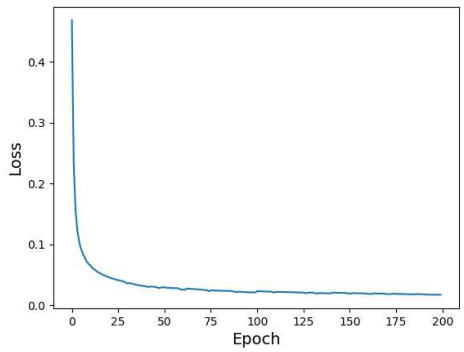


Figure 5.10 Training loss versus epoch.

The testing is performed on 10,000 images of Microsoft COCO dataset [87] as the cover image, and 10,000 images of the testing division of the binary CIFAR as the watermark. Both the testing cover images and testing watermarks are not used in the training. It demonstrates that the proposed system generalizes the watermarking rules without over-fitting to the training samples. The peak signal-to-noise ratio (*PSNR*) and bit-error-rate (*BER*) are respectively used to quantitatively evaluate the fidelity of the marked image and the quality of the watermark extraction. The *PSNR* is defined as in Equation (3.4) and the *BER* is computed as the percentage of error bits on the binarized watermark extraction. In the testing, the BER is very close to zero, indicating that the original and the extracted watermarks are identical. The testing *PSNR* is 39.72 dB, meaning a high fidelity of the marked images, so that the hidden

information cannot be noticed by human vision. Some examples of the watermark embedding with various image content and color are shown in Figure 5.11. The residual errors showing the absolute difference in each RGB channel between the marked and the cover images are also displayed, from which we observe that the watermark is dispersed over the marked image. It provides added security to the marked image even if the cover image has leaked. But subtracting it from the marked image does not reveal the watermark information. Ranging the pixel values between 0 and 255, we compute the mean of residual errors for each RGB channel, averaging along the testing results of 2.57, 2.10, and 1.63, respectively. Similarly, the maxima of residual errors are 14.11, 24.79, and 17.08. These numbers indicate that there are some relatively spiky modifications for the extraction, but on average the watermark insertion does not alter channels a lot.

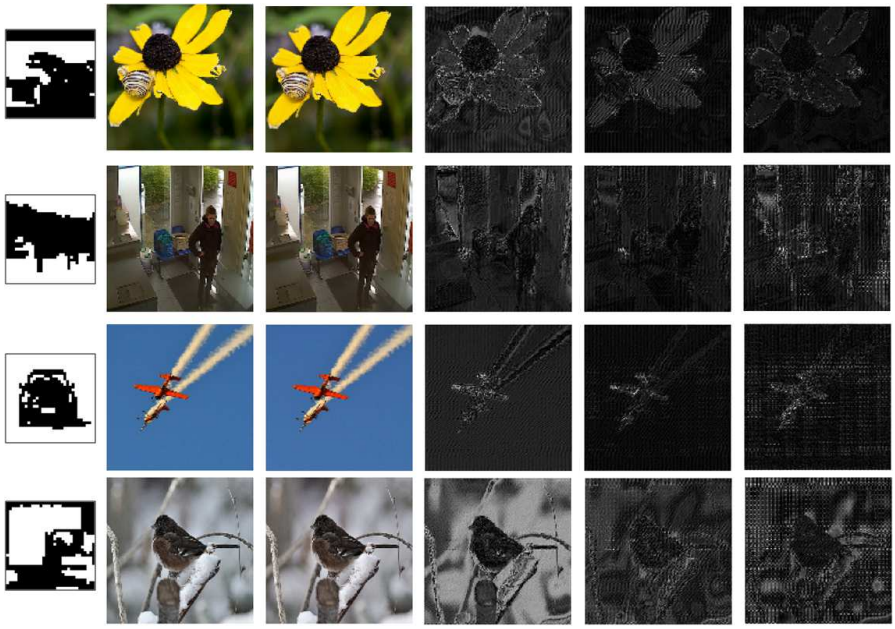


Figure 5.11 A few testing examples. 1st column: the embedded and extracted watermark, 2nd column: the cover image, 3rd column: the marked image, and 4th, 5th, and 6th columns: the respective residual errors of R, G, and B channels between the marked and the cover images.

We explore the proposed system to some extreme cases using synthetic images for further estimation. In particular, the synthetic situations that are not included during the training process are analyzed, and the results involving blank and random-noise images are presented here.

Figure 5.12 shows the results of embedding real watermarks into some blank cover images of black, white, red, green, and blue, where the residuals are amplified 10 times. Although the blank cover images are not included in the training, the proposed system provides acceptable results in the cases. The residual errors display more green color and the blank green marked image displays relatively more noticeable noises than those in other colors, implying that the proposed system modifies the green color slightly more.

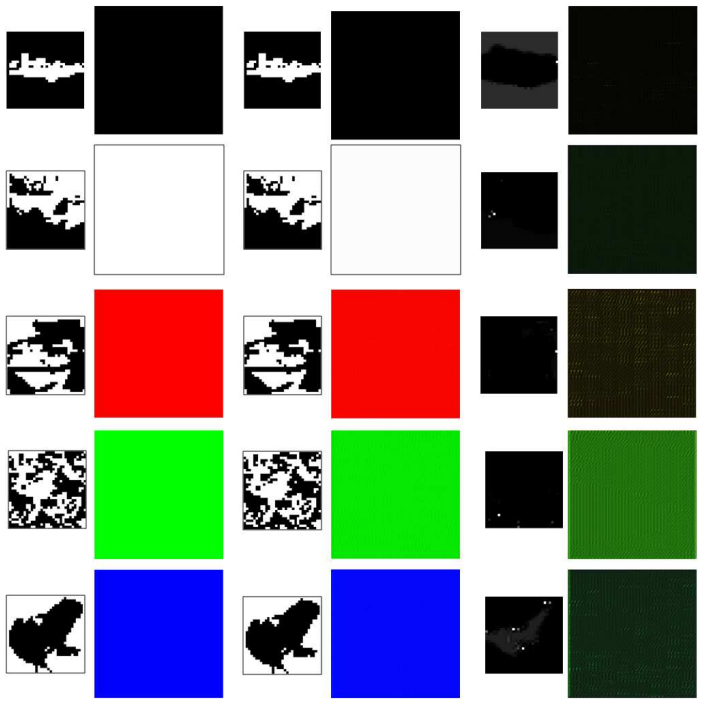


Figure 5.12 Embedding watermarks into blank covers. 1st column: the watermark, 2nd column: the blank cover image, 3rd column: the extracted watermark, 4th column: the marked image, and 5th and 6th columns: the residual errors.

Figure 5.13 presents the results of embedding a random binary image into a natural cover image, as well as embedding a real watermark into a random color-spotted cover. Applying a random binary image as the watermark displays good results. Although the general shape of extraction is recognizable, there are obvious distortions on the extraction when it comes to embedding a watermark into random noises. In practice, hiding a watermark into random noises indicates that the appearance of the marked media is noisy and meaningless, so the encryption methods mapping a watermark into random patterns could be used.

Figure 5.13 Embedding involving noise images. 1st column: the watermark, 2nd column: the cover image, 3rd column: the extracted watermark, 4th column: the marked image, and 5th and 6th columns: the residual errors.

5.3.2 Robustness

The robustness of the proposed system against different distortions on the marked image is evaluated by analyzing the tolerance range towards the attacks. Figure 5.14 shows visual comparisons between the marked images and their distortions as well as between the original watermarks and the extractions from the distorted marked images.

Figure 5.14 Visual comparisons of distortions. 1st row: the marked image, 2nd row: the distorted marked image, from left to right the operations are histogram equalization, Gaussian blur, random noise, salt-and-pepper noise, and cropping, 3rd row: the original watermark, and 4th row: the extraction from the distorted marked image.

Quantitatively, swept-over distortion parameters which control the strength of the attacks are applied on the testing datasets, and the average *BER* are recorded. Since the geometric distortions, such as translation, rotation, and scaling, are rectified, we focus on the responses of the proposed system against image processing attacks. Figure 5.16 presents the results of some common but challenging situations. The extracted watermarks from the proposed system respectively have 11%, 8.1%, 31%, 8.2%, 42%, and 5.1% bits errors when the distortions are a Gaussian blur with mean 0 and variance 85%, a cropping discarding 65% percent of the marked image, a Gaussian additive noise mean 0 and variance 20%, a JPEG compression with quality factor 10, a 20% random noise, and a 30% salt-and-pepper noise. The proposed system shows a high tolerance range on these challenges especially for cropping, salt-and-pepper noise, JPEG compression, and the noises that randomly fluctuate the pixel values through image channels show higher *BER* such as Gaussian additive noise and random modificative noise. However, a 20% Gaussian noise or a 20% random noise destroys almost content of the marked image as shown in Figure 5.15. The proposed

system responds acceptable performances given a decent distortion parameter of these attacks, such as 16% *BER* on 10% Gaussian noise.

Figure 5.15 Extreme cases. Left: the original marked image. Middle: Gaussian additive noise with variance 20%. Right: 20% Random modificative noise.

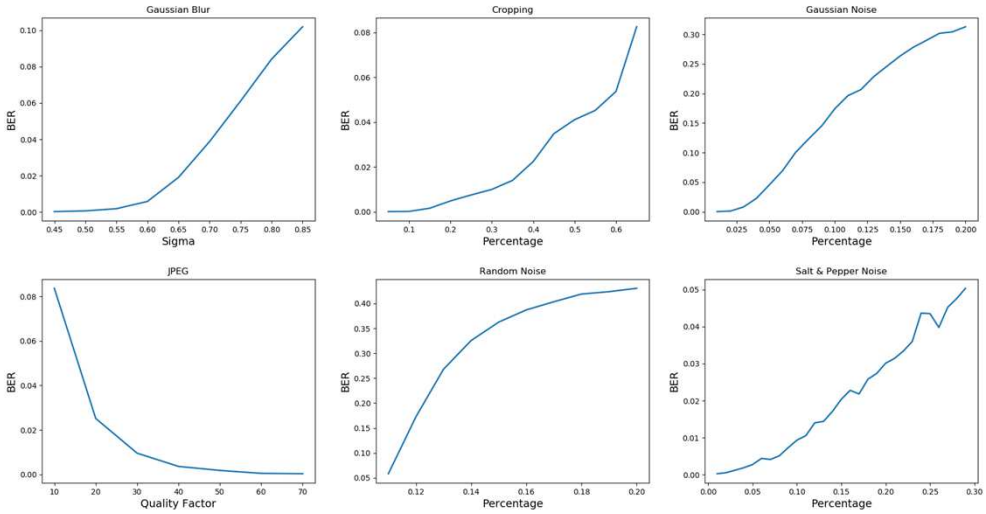


Figure 5.16 Distortion parameters vs *BER*.

5.3.3 Comparative Study

First, the proposed method is analytically compared against several state-of-the-art image watermarking methods that incorporate deep neural networks as shown in Table 5.1. To the best of our knowledge, Kandi *et al*. [75] is the first image watermarking scheme utilizing convolutional neural networks. But it is a non-blind scheme for achieving robustness. Embedded by increasingly changing an image block to represent a watermark bit, the system in [79] is trained to extract the watermark bits from their corresponding blocks with attack simulation and achieves both blindness

and robustness. However, it requires the distortions to be included in the training phase for robustness, it is difficult to predict and enumerate all encountering attacks in practice. On the other hand, the proposed system not only achieves blindness and robustness simultaneously, but also is trained without the requirement of any prior knowledge of attack information, and hence has a wider range of applications.

Table 5.1 Comparison of the Proposed System against State-of-the-art Image Watermarking Methods

Method	Function of the deep neural networks	Blindness	Robustness	Concentration
[74]	determine the embedding position and probability	no	no	undetectability
[75]	embedding and extraction	no	yes	robustness
[76]	embedding and extraction	yes	no	capacity
[77]	extraction	yes	no	Undetectability
[79]	extraction	yes	yes	Robustness
Ours	embedding and extraction	yes	yes	Robustness

Second, the proposed system is compared against several blind and robust competitors quantitatively. The selection of the competitors mainly considers variation. Mun *et al*. [79] applied convolutional neural networks, and Zong *et al*. [53], Zareian and Tohidypour [8], and Ouyang *et al.* [88] are classic, traditional, and robust methods with different image domains in recent years, such as histogram domain adopting statistical image features, frequency domain, and log-polar domain with summarized image features. For a fair comparison, the testing is performed on the same image sets reported in the references. The crucial results are presented in Table 5.2, where "/" denotes not applicable, S&P is the salt-and-pepper noise, and GF is Gaussian filtering. The proposed method shows clear advantages by covering more general distortion categories as well as lower *BER* under the same distortion parameters. For instance, the traditional methods such as manipulating the image histogram cannot tolerate the histogram equalization attack. In addition, the proposed

method has a higher tolerance range; for example, [79] and [88] can only extract the watermark with high JPEG quality of 80 to 90, while the proposed method covers as low as 10 although the method in [8] focuses on the JPEG having a higher performance. Remarkably, the competitors tolerate cropping 20% to 30%, while the *BER* is as high as 8.2% if 65% of the marked image is cropped. Finally, under a close *PSNR*, the proposed method outperforms the existing methods by simultaneously achieving the highest robustness and the highest capacity.

Table 5.2 Quantitative Comparison Between the Proposal and Some Blind and Robust Competitors

Method	*BER (%)*					*PSNR* (dB)	Capacity
	HE	JPEG 10	Cropping 20%	S&P 5%	GF 10%		
[79]	/	/	6.61	7.98	4.81	38.01	1 bit per block
[53]	/	17.50	7.06	3.51	6.33	46.63	25 bits
[8]	/	2.15	/	4.94	0.21	41.00	256 bits
[88]	/	/	7.51	9.41	27.91	36.77	24 bits
Ours	0.43	8.16	0	0.97	0	39.93	1,024 bits

5.3.4 Application on Camera-captured Images

We present one of the core applications using the proposed system to extract watermark from a camera-captured image. Success in this problem has the potential of many useful applications, such as connecting the virtual world and the real world to serve as the low-level interface for the internet of things. Watermark detection and extraction on a camera resampled image remains a challenge. Its difficulty is mainly caused by the comprehensive combination of noises [89], including geometric distortions, optical tilt, quality degradation, compression, and lens distortions. Researchers and engineers have been trying to address this issue from various perfectives. For example, Pramila *et al*. [90] proposed to extract the watermark from a

phone-camera capture of an image printed on blank paper. However, it has restricted applications since the presence of the cover image is required.

We apply the proposed blind system towards this issue. Instead of using printing, a phone camera is used to capture a marked image displayed on a laptop screen. The distortions are brought by the camera, the resolution, brightness, refresh rate, and frame rate, etc. and could be challenging to robustness issue. The scenario is related to the widely-used Quick-Response (QR) code, where the users are directed to online resource via a scan. What dissimilar to the traditional QR code in our system is that the end users just need to scan the content image for more information and the code/watermark becomes completely invisible. As shown in Figure 5.17, the provider distributes the marked image obtained with the cover image and the watermark information via our marking app containing the trained encoder and embedder networks. The user installs our scanning app containing the trained decoder and extractor networks, and scans the image displayed on a screen for the hidden information.

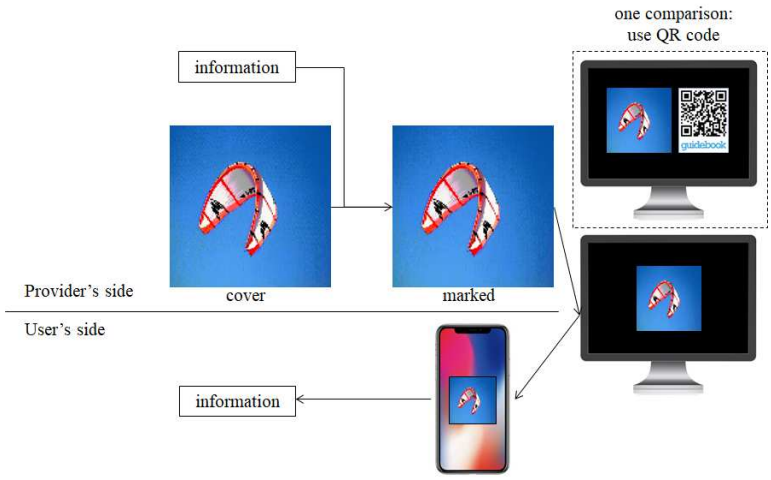


Figure 5.17 The process of the application.

In practice, error correction coding can be used for further protection to the watermark. For instance, a classic Reed Solomon (RS) code [91] can correct up to 30% error in the extraction. In this paper, we present the raw results for visual comparison as well as for simplicity. To test the system under realistic situation, a protype is developed and a raw 32 × 32 binary watermark (see Figure 5.18) is used for its clear structure. Five volunteers are asked to take a few pictures of some marked images displayed 425px × 425px on a 2,560 × 1,440 screen, with the camera of a mobile phone. Two rules are told to the users. First, as shown in the user's interface, the entire image should be placed as large as possible inside the region of interest (ROI). As a prototype of demonstrating purpose, this rule facilitates our segmentation that the largest contour inside the ROI is the marked image, so that we can focus on the test of the proposed system instead of some complicated segmentation algorithms. In addition, placing the image largely in the ROI helps us to capture desired details and features for the extraction. Second, the camera should be kept as still as possible. Although the proposed system tolerates some blurring effects, it is not designed to extract watermark in a high-speed motion.

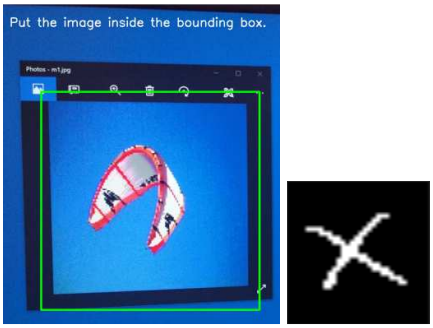


Figure 5.18. The prototype. Left: the appearance of the prototype, Right: the sample binary watermark.

The prototype only analyzes the ROIs, and hand-taken pictures can hardly be completely parallel to the screen. Therefore, there exist some geometrical, affine, and

perspective distortions. As the proposed system does not include the robustness towards these attacks, the image registration technique in [92, 93] is adopted. To simplify the prototype as in Figure 5.19, four corners of the largest contour inside the ROI are used as the reference points. The contoured content is mapped on the bird view plane, and the watermark is extracted from the rectified image.

Figure 5.19 presents a few extractions and their corresponding ROIs. The *BER*s from left to right are 3.71%, 4.98%, 1.07%, 4.30%, and 8.45%. We observe that the closer the picture is taken, the lower the error is. The more parallel between the camera and the screen, the lower the error is. Note that the tolerance limit is around 30° in this test. Also, the flash light brings more errors since it may over- and underexpose some image areas. We may turn off the flash lights in this case since the screen has backlit. In total from the user test we have 20 images, and the average *BER* is 5.13%. This is the raw result without the error correction code and can be considered as acceptable since the RS code can correct 30% errors theoretically. Moreover, the scanning app extracts the watermark within one second as it only applies the pretrained weights in the extractor and decoder networks to the marked image rectification.

Figure 5.19 A few extractions and the ROIs.

In summary, experimental results confirm that the proposed system extracts the watermark well on a rectified marked image with the registration techniques.

Hence, it can be used in solving the challenging problem of watermark extraction on camera-captured images.

CHAPTER 6

CONCLUSION

Focusing on the computational intelligence in steganography, adaptive digital image watermarking systems that can be applied to varying, flexible and multi-purposed watermarking situations are presented in this dissertation.

By proposing a nonlinear mapping through wavelet generalized lifting to obtain a saliency map as well as an efficient adaptive thresholding scheme, a novel salient region detection model is proposed to segment the cover-image into ROIs and RONI. The ROIs containing the most representative information of the images are kept intact during and the RONI is for watermarking. Hence image watermarking adaptive to visual contents is achieved. Concentrating on salient object segmentation at the core of content-based image watermarking, the proposed model not only produces full-resolution saliency maps that highlight multiple salient objects, but also requires no kernels with implicit assumptions and prior-knowledge. Experimental results have shown the reliability and high performance of the proposed model. Extending the proposed model for the applications of video saliency detection as well as depth map generation for 2D to 3D Conversion and including more feature maps at the saliency map computation stage, such as the texture, orientation and added psychological patterns can be considered at the next step.

Secondly, an intelligent image watermarking scheme based on the ROI detection is presented. It is a novel technique for image frequency-domain watermarking by exploiting the phase spectrum of the original image. Applying an iterative strategy on the magnitudes of image frequency domain, a novel reversible watermarking algorithm is proposed to embed a large amount of information without

visible degradation to the cover image. Partitioning algorithms are proposed to facilitate the embedding of the watermark into only RONIs, and therefore, the crucial information is undistorted. With the help of swarm intelligence, the proposed scheme allows an optimal watermarking solution with an option of adjusting different weights of capacity and quality to satisfy user's need.

A robust multibit image watermarking scheme based on the ROIs using an improved embedding strategy and the synchronization approach is also presented. A novel contrast modulation-based watermark embedding strategy is developed to achieve high robustness and high payload simultaneously. A self-referencing rectification approach is designed for watermark resynchronization under affine transformations, by which the proposed scheme offers a high tolerance range on parameters in affine distortions. In the future work, handling more comprehensive attacks such as random bending, and extend both the embedding strategy and the rectification into three-dimensional contexts can be the directions.

Finally, a robust and blind image watermarking system using deep learning is introduced. In an unsupervised manner, a novel structure of applying deep convolutional neural networks is proposed to learn the watermark embedding and extraction rules with the constraint of loss function. The robustness is achieved without any prior knowledge of possible attacks and distortions. Comprehensive evaluations are presented to confirm the superiority and a challenging application of watermark extraction from camera-capture image is introduced to validate the practicality of the proposed system. In the future work, statistical features such as the probability density functions could be merged into the deep neural networks, so that the embedding and extraction can consider the image distribution for a higher robustness.

REFERENCES

[1] H. Berghel and L. O'Gorman, "Protecting ownership rights through digital watermarking," *Computer*, vol. 29, no. 7, pp. 101-103, 1996.

[2] J. Dittmann and F. Nack, "Copyright-copywrong," *IEEE Transaction on Multimedia*, vol. 7, no. 4, pp. 14-17, 2000.

[3] F. Y. Shih, "Digital watermarking and steganography: fundamentals and techniques," Boca Raton, Florida: *CRC Press*, 2017.

[4] R. Caldelli, F. Francesco and R. Becarelli, "Reversible watermarking techniques: An overview and a classification," *EURASIP Journal on Information Security*, no. 134546, 2010.

[5] B. Gunjal and R. R. Manthalkar, "An overview of transform domain robust digital image watermarking algorithms," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, no. 1, pp. 37-42, 2010.

[6] A. A. Tamimi, A. M. Abdalla and O. Al-Allaf, "Hiding an image inside another image using variable-rate steganography," *International Journal of Advanced Computer Science and* Applications, vol. 4, no. 10, 2013.

[7] I. J. Cox, J. Kilian, F. T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transaction on Image Processing*, vol. 6, no. 12, pp. 1673-1687, 1997.

[8] M. Zareian and H. R. Tohidypour, "Robust quantisation index modulation-based approach for image watermarking," *IEEE Transaction on Image Processing*, vol. 7, no. 5, pp. 432-441, 2013.

[9] J. M. Zain, L. P. Baldwin and C. M. Larke, "Reversible watermarking for authentication of DICOM images," in *Proc. IEEE International Conference on Engineering in Medicine and Biology*. pp. 3237-3240, 2004.

[10] X. Guo and T. G. Zhuang, "A region-based lossless watermarking scheme for enhancing security of medical data," *Journal of Digital Imaging*, vol. 2, no. 1, pp. 53-64, 2009.

[11] S. Das and M. K. Kundu, "Effective management of medical information through ROI-lossless fragile image watermarking technique," *Computer Methods and Programs in Biomedicine*, vol.111, no.3, pp. 662-675, 2013.

[12] O. M. Al-Qershi and B. E. Khoo, "Authentication and data hiding using a hybrid ROI-based watermarking scheme for DICOM images," *Journal of Digital Imaging*, vol. 24, no. 1, pp. 114-125, 2011.

[13] Y. Liu, X. Qu and G. Xin, “A ROI-based reversible data hiding scheme in encrypted medical images,” *Journal of Visual Communication and Image Representation*, vol. 39, pp. 51-57, 2016

[14] A. Borji, D. N. Sihite and L. Itti, “Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study,” *IEEE Transaction on Image Processing*, vol. 22, no.1, pp. 55-69, 2013.

[15] V. A. Oppenheim and L. S. Jae, “The importance of phase in signals,” in *Proc. IEEE*, vol. 69*,* no. 5, pp. 529-541,1981.

[16] X. Hou and L. Zhang, “Saliency detection: A spectral residual approach,” in *Proc. IEEE International. Conference on Computer Vision and Pattern Recognition*, 2007.

[17] X. Hou, H. Jonathan and K. Christof, “Image signature: Highlighting sparse salient regions,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 34, no.1, pp. 194-201, 2012.

[18] R. Achanta, S. Hemami, F. Estrada and S. Susstrunk, “Frequency-tuned salient region detection,” in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition,* pp. 1597-1604, 2009.

[19] R. Achanta and S. Sabine, "Saliency detection using maximum symmetric surround," in *Proc. IEEE International Conference on Image Processing*, 2010.

[20] N. Imamoglu, W. Lin and Y. Fang, “A saliency detection model using low-level features based on wavelet transform,” *IEEE Transaction on Multimedia*, vol.15, no. 1, pp. 96-105, 2013.

[21] L. Itti, C. Koch and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 20, no.11, pp. 1254-1259, 1998.

[22] W. Sweldens, “The lifting scheme: A construction of second generation wavelets,” *SIAM Journal on Mathematical Analysis*, vol. 29, no.2, pp. 511-546, 1998.

[23] J. S. Rojals, “Optimization and generalization of lifting Schemes: application to lossless image compression,” *Universitat Politècnica de Catalunya*, 2006.

[24] M. M. Cheng, N. J. Mitra, X. Huang, P. H. Torr and S. M. Hu, "Global contrast based salient region detection," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569-582, Mar. 2015.

[25] S. Mallat, "A Wavelet Tour of Signal Processing: The Sparse Way, third edition," Cambridge, Massachusetts: Academic Press, 2008.

[26] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *IEEE Transaction on Signal Processing*, vol. 40, no. 9, 2207-2232, 1992.

[27] F. Y. Shih, "Image processing and pattern recognition: fundamentals and techniques," Hoboken, New Jersey: Wiley-IEEE Press, 2010.

[28] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 890-896, 2003.

[29] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE Transaction on Image Processing*, vol. 13, no. 8, pp. 1147-1156, 2004.

[30] O. M. Al-Qershi and B. E. Khoo, "Two-dimensional difference expansion (2D-DE) scheme with a characteristics-based threshold," *Signal Processing*, vol. 93, no. 1, pp. 154-162, 2013.

[31] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S. C. Zhu, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 17-33, 2003.

[32] D. L. Ruderman and W. Bialek, "Statistics of natural images: scaling in the woods," *Physical Review Letters*, vol. 73, no. 6, pp. 814-822, 1994.

[33] F. Y. Shih and Y. T. Wu, "Robust watermarking and compression for medical images based on genetic algorithms," *Information Sciences*, vol. 175, no. 3, pp. 200-216, 2005.

[34] D. M. Thodi and J. J. Rodríguez. "Expansion embedding techniques for reversible watermarking," *IEEE Transaction on Image Processing*, vol. 16, no. 3, pp. 721-730, 2007.

[35] A. Wakatani, "Digital watermarking for ROI medical images by using compressed signature image," in *Proc. the 35th IEEE Annual Hawaii International Conference on System Sciences*, pp. 2043-2048, Jan. 2002.

[36] Z. H. Wang, C. F. Lee and C. Y. Chang, "Histogram-shifting-imitated reversible data hiding," *Journal of Systems and Software*, vol. 86, no. 2, pp. 315-323, 2013.

[37] J. M. Zain, L. P. Baldwin and C. M. Larke, "Reversible watermarking for authentication of DICOM images," in *Proc. IEEE International Conference on Engineering in Medicine and Biology*. pp. 3237-3240, 2004.

[38] Z. Zhao, H. Luo, Z. M. Lu and J. S. Pan, "Reversible data hiding based on multilevel histogram modification and sequential recovery," *AEU-International Journal of Electronics and Communications*, vol. 65, no. 10, pp. 814-826, 2011.

[39] R. Chadha and D. Allison, "Decomposing rectilinear figures into rectangles," *Technical Report, Department of Computer Science, Virginia Polytechnic Institute and State University*, pp. 1-34, 1988.

[40] A. Rosenfeld and A. C. Kak, "Digital picture processing," vol. 2, New York, New York: Academic Press, 1982.

[41] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris and R. L. Buckner, "Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *Journal of Cognitive Neuroscience*, vol. 19, no. 9, pp. 1498-1507, 2007.

[42] O. M. Al-Qershi and B. E. Khoo, "Authentication and data hiding using a hybrid ROI-based watermarking scheme for DICOM images," *Journal of Digital Imaging*, vol. 24, no.1, pp. 114-125, 2011.

[43] S. Das and M.K. Kundu, "Effective management of medical information through ROI-lossless fragile image watermarking technique," *Computer Methods and Programs in Biomedicine*, vol. 111, no.3, pp. 662-675, 2013.

[44] X. Guo and T.G. Zhuang, "A region-based lossless watermarking scheme for enhancing security of medical data," *Journal of Digital Imaging*, vol. 22, no. 1, pp. 53-64, 2009.

[45] Y. Liu, X. Qu, and G. Xin, "A ROI-based reversible data hiding scheme in encrypted medical images," *Journal of Visual Communication and Image Representation*, vol. 39, pp. 51-57, 2016.

[46] D. Faur, I. Gavat and M. Datcu, "Mutual information based measure for image content characterization," in *Current Topics in Artificial Intelligence*, Berlin Heidelberg: Springer pp. 342-349, 2006.

[47] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *From Natural to Artificial Neural Computation*, Berlin Heidelberg: Springer, pp. 195-201, 1995.

[48] J. Kennedy, "Particle swarm optimization." in *Encyclopedia of Machine Learning*, New York, New York: Springer, pp. 760-766, 2010.

[49] Wang Z and Bovik AC, "A universal image quality index," *IEEE Signal Processing Letters,* vol. 9, no. 3, pp. 81-84, 2002.

[50] M. Arsalan, S. A. Malik and A. Khan, "Intelligent reversible watermarking in integer wavelet domain for medical images," *Journal of Systems and Software*, vol. 85, no. 4, pp. 883-894, 2012.

[51] L. Luo, Z. Chen, M. Chen, X. Zeng and Z. Xiong, "Reversible image watermarking using interpolation technique." *IEEE Transaction on Information Forensics and Security*, vol. 5, no. 1, pp. 187-193, 2010.

[52] V. Sachnev, H. J. Kim, J. Nam, S. Suresh and Y. Q. Shi, "Reversible watermarking algorithm using sorting and prediction." *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 19. no. 7, pp. 989-999, 2009.

[53] T. Zong, Y. Xiang, I. Natgunanathan, S. Guo, W. Zhou and G. Beliakov, "Robust histogram shape-based method for image watermarking," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 25, no. 5, pp. 717-729, 2015.

[54] G. W. Braudaway and F. Minter, "Automatic recovery of invisible image watermarks from geometrically distorted images," in *Proc. SPIE: Security and Watermarking of Multimedia Contents I*, vol. 3971, CA, 2000.

[55] C. Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, M. L. Miller and Y. M. Lui, "Rotation, scale and translation resilient public watermarking for images," in *Proc. SPIE Security Watermarking Multimedia Contents II*, vol. 3971, pp. 90-98, 2000.

[56] P. Dong, J. G. Brankov, N. P. Galatsanos, Y. Yang and F. Davoine, "Digital watermarking robust to geometric distortions," *IEEE Transaction on Image Processing*, vol. 14, no. 12, pp. 2140-2150, 2005.

[57] H. Kim and H. Lee, "Invariant image watermark using Zernike moments," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 8, pp. 766-775, 2003.

[58] M. Kutter, S. K. Bhattacharjee and T. Ebrahimi, "Towards second generation watermarking schemes," in *Proc. IEEE International Conference on Image Processing*, Kobe, Japan, pp. 320-323, 1999.

[59] X. Kang, J. Huang, Y. Q. Shi and Yan Lin, "A DWT-DFT composite watermarking scheme robust to both affine transform and JPEG compression," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 13, no. 8, Aug. 2003.

[60] M. Kutter, F. Jordan and F. Bossen, "Digital watermarking of color images using amplitude modulation," *Journal of Electronic Imaging*, vol. 7, no. 2, pp. 326-332, Feb. 1998.

[61] X. Gao, C. Deng, X. Li and D. Tao, "Geometric distortion insensitive image watermarking in affine covariant regions," *IEEE Transaction on Systems, Man, and Cybernetics, Part C*, vol. 40, no. 3, pp. 278-286, 2010.

[62] J.G. Proakis, "Digital Communication, Fourth Edition," New York, New York: *McGraw-Hill*, 2000.

[63] E. Peli, "Contrast in complex images," *Journal of the Optical Society of America*, vol. 7, no. 10, pp. 2032-2040, 1990.

[64] A. R. alderbank, I. Daubechies, W. Sweldens and B. L. Yeo, "Lossless image compression using integer to integer wavelet transforms." In *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. 596-599. 1997.

[65] A. Shaamala, S. M. Abdullah and A. A. Manaf, "Study of the effect DCT and DWT domains on the imperceptibility and robustness of genetic watermarking," *International Journal of Computer Science Issues*, vol. 8, issue 5, no. 2, pp. 220-225, 2011.

[66] J. Flusser and T. Suk, "A moment-based approach to registration of images with affine geometric distortion," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 32, no. 2, pp. 382-387, 1994.

[67] N. Bi, Q. Sun, D. Huang, Z. Yang and J. Huang, "Robust image watermarking based on multiband wavelets and empirical mode decomposition," *IEEE Transaction on Image Processing*, vol. 16, no. 8, pp. 1956-1966, 2007.

[68] S. Xiang, H.J. Kim and J. Huang, "Invariant image watermarking based on statistical features in the low-frequency domain," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 18, no. 6, pp. 777-790, 2008.

[69] Y. Qian, J. Dong, W. Wang and T. Tan, "Deep learning for steganalysis via convolutional neural networks," *Media Watermarking, Security, and Forensics*, vol. 9409, International Society for Optics and Photonics, 2015.

[70] L. Pibre, J. Pasquet, D. Ienco and M. Chaumont, "Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover source mismatch," *Electronic Imaging*, vol. 1, no. 11, 2016.

[71] H. Sabah and B. Haitham, "Artificial neural network for steganography," *Neural Computing and Applications*, vol. 26, no.1, pp. 111-116, 2015.

[72] S.B. Alexandre and C.J. David, "Artificial neural networks applied to image steganography," *IEEE Latin America Transactions*, vol. 14, no. 3, pp. 1361-1366, 2016.

[73] J. Robert, V. Eva and K. Martin, "Neural network approach to image steganography techniques," *Mendel*, pp. 317-327. Springer, 2015.

[74] W. Tang, S. Tan, B. Li and J. Huang, "Automatic steganographic distortion learning using a generative adversarial network," *IEEE Transaction on Signal Processing Letters*, vol. 24, no. 10, pp. 1547-1551, 2017.

[75] H. Kandi, D. Mishra and S.R.S Gorthi, "Exploring the learning capabilities of convolutional neural networks for robust image watermarking," *Computers & Security*, vol. 65, pp. 247-268, 2017.

[76] S. Baluja, "Hiding images in plain sight: deep steganography," *Advances in Neural Information Processing Systems*, 2017.

[77] D. Li, L. Deng, B.B. Gupta, H. Wang and C. Choi, "A novel CNN based security guaranteed image watermarking generation scenario for smart city applications," *Information Science*, 2018.

[78] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z.B. Celik and A. Swami, "The limitations of deep learning in adversarial settings," *IEEE European Symposium on Security and Privacy*, 2016.

[79] S.M. Mun, S.H. Nam, H.U. Jang, D. Kim and H.K. Lee, "A robust blind watermarking using convolutional neural network," *arXiv preprint arXiv:1704.03248*, 2017.

[80] G.E. Hinton and R.R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," S*cience*, 313, no. 5786, pp. 504-507, 2006.

[81] C. Szegedy, S. Ioffe, V. Vanhoucke and A.A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI*, vol. 4, pp. 12, 2017.

[82] S. Rifai, P. Vincent, X. Muller, X. Glorot and Y. Bengio, "Contractive auto-encoders: Explicit invariance during feature extraction," in *Proc. 28th International Conference on Machine Learning*, pp. 833-840, Omnipress, 2011.

[83] Y. LeCun, L. Bottou, G.B. Orr and K.R. Müller, "Efficient backprop," *Neural Networks: Tricks of the Trade*, pp. 9-50. Berlin, Heidelberg: Springer, 1998.

[84] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein and A.C. Berg, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211-252, 2015.

[85] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," *Technical Report*, University of Toronto, vol. 1, no. 4, pp. 7, 2009.

[86] M.D Zeiler, "ADADELTA: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.

[87] T.Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár and C. L.Zitnick, "Microsoft coco: Common objects in context," in *Proc. European Conference on Computer Vision*, pp. 740-755. Berlin, Heidelberg: Springer, 2014.

[88] J. Ouyang, G. Coatrieux, B. Chen and H. Shu, "Color image watermarking based on quaternion Fourier transform and improved uniform log-polar mapping," *Computers & Electrical Engineering*, vol. *46*, pp. 419-432, 2015.

[89] A. Pramila, A. Keskinarkaus and T. Seppänen, "Camera based watermark extraction-problems and examples," in *Proc. Finnish Signal Processing Symposium*, 2007.

[90] A. Pramila, A. Keskinarkaus and T. Seppänen, "Increasing the capturing angle in print-cam robust watermarking," *Journal of Systems and Software*, vol. 135, pp. 205-215, 2018.

[91] I.S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no.2, pp. 300-304, 1960.

[92] L.G. Brown, "A survey of image registration techniques," *ACM Computing Surveys (CSUR)*, vol. 24, no. 4, pp. 325-376, 1992.

[93] S. Zokai and G, Wolberg, "Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations," *IEEE Transaction on Image Processing*, vol. 14, no. 10, pp. 1422-1434, 2005.