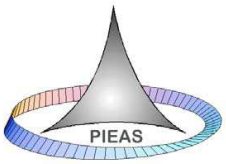
Digital Watermarking Using Machine Learning Approaches

Imran Usman

PhD Thesis



Pakistan Institute of Engineering and Applied Sciences

Department of Computer and Information Sciences,
Islamabad, Pakistan

In the name of Allah, the most Beneficent, the most Merciful

Digital Watermarking Using Machine Learning Approaches

By

Imran Usman

A dissertation submitted for the degree of
Doctor of Philosophy in Computer and Information Sciences

To
The Department of Computer and Information Sciences,
Pakistan Institute of Engineering and Applied Sciences,
Islamabad, Pakistan
2010

To the holy prophet (PBUH),
my loving parents,
and my wonderful wife,
whose watermarks no attacker can remove from my soul.

Abstract

In recent years digital watermarking has gained substantial attraction by the research community. It promises the solution to many problems such as content piracy, illicit manipulation of medical/legal documents, content security and so on. Watermarked content is usually vulnerable to a series of attacks in real world scenario. These attacks may be legitimate, such as common signal processing operations, or illegitimate, such as a malicious attempt by an attacker to remove the watermark. A low strength watermark usually possesses high imperceptibility but weak robustness and vice versa. On the other hand, different set of attacks are associated with distinctive watermarking applications, which pose different requirements on a watermarking scheme. Therefore, intelligent approaches are needed to adaptively and judiciously structure the watermark in view of the current application.

In addition, traditional watermarking techniques cause irreversible degradation of an image. Although the degradation is perceptually insignificant, it may not be admissible in applications like medical, legal, and military imagery. For applications such as these, it is desirable to extract the embedded information, as well as recover the sensitive host image. This leads us to the use of reversible watermarking. An efficient reversible watermarking scheme should be able to embed more information with less perceptual distortion, and equally, be able to restore the original cover content. Therefore, for reversible watermarking, capacity and imperceptibility are two important properties. However, if one increases the other decreases and vice versa. Hence, one needs to make an optimum choice between these two properties for reversible watermarking.

The research in this work is two-fold. Firstly, we develop intelligent systems for making optimum robustness versus imperceptibility tradeoffs. The performance of the existing watermarking approaches is not up to the task when we consider watermark structuring in view of a sequence of attacks, which is much desirous in real world applications. In order to resist a series of attacks, we employ intelligent selection of both the frequency band as well as strength of alteration for watermark embedding using Genetic Programming. To further enhance the robustness of the watermarking system, Support Vector Machines and Artificial Neural Networks are applied to adaptively modify the decoding strategy in view of the anticipated sequence of attacks at the watermark extraction phase.

Secondly, we devise an intelligent system capable of making optimum/ near optimum tradeoff between watermark payload and imperceptibility. In the context of reversible watermarking, we propose an intelligent scheme which selects suitable coefficients in different wavelet sub-bands and yields superior capacity versus imperceptibility tradeoff. Experimental results show that machine learning approaches are very promising in state of the art watermarking applications.

This work is accomplished under the kind supervision of

Dr. Asifullah khan

Associate Professor

Department of Computer and Information Sciences,

Pakistan Institute of Engineering and Applied Sciences,

Islamabad Pakistan

This work is funded by Higher Education Commission (HEC) Govt. of Pakistan

through indeginous-5000 PhD program,

Grant No. 17-5-1(Cu-204)/HEC/Sch/2004.

DECLARATION

The research described in this dissertation was carried out by the author between 2006 and 2009. Except as indicated in the text, the contents are entirely original and have not been previously submitted or approved for the award of a degree by this or any other university.

Imran Usman

It is certified that the work in this thesis is carried out and completed under my supervision.

Supervisor:
Dr. Asifullah Khan
Associate Professor
DCIS, PIEAS, Islamabad.

Co-Supervisor:
Dr. Abdul Majid
Associate Professor
DCIS, PIEAS, Islamabad.

Acknowledgements

First of all I would like to bow my head in front of Allah Almighty, the most Beneficent, and most Merciful who has given me strength, persistence and knowledge to execute this research. Also, Whose countless mercies bestowed upon me sufficient opportunities, highly qualified and talented teachers, praying parents and loving friends without whom this work would have never been possible.

I would like to express my gratitude and am highly indebted to my loving parents whose prayers follow me wherever I go and whatever endeavors I do. I am also thankful to my wife Mrs. Mah-e-Mubeen Imran, my sisters, and all the family members and friends who have been very patient with me during this research activity.

This thesis contains research work conducted at Department of Computer and Information Sciences (DCIS), Pakistan Institute of Engineering and Applied Sciences (PIEAS), Nilore, Islamabad from August 2006 till January 2010 under the kind supervision of Associate Professor Dr. Asifullah Khan to whom I am highly indebted for his countless efforts, advice and encouragement. He has always been a source of motivation and inspiration for me throughout this research activity. His experience and expertise was a source of training during this period and I would thank Allah Almighty once again being a part of such a great team.

I am very grateful to Dr. Anwar Majid Mirza for his kind support, appreciation, advice and valuable comments from time to time during this research.

I would also like to thank Dr. Anila Usman, Dr. Mutawarra Hussain, Dr. Abdul Jalil and Dr. Asmatullah Chaudhry for their best support, valuable comments and

encouragement during this research. I would take this dissertation as an opportunity to thank Dr. Abdul Majid, Dr. M. Arif, Dr. Javaid Khurshid and Mr. Arshad Iqbal for sharing their knowledge and providing me with all the technical and administrative support during this research. I am very grateful to the current and former staff at the Department of Computer and Information Sciences, PIEAS who helped me in day-to-day activities.

I appreciate and thank for the help of my friends especially Rafiullah, Saeed Ehsan, Fayyaz Afsar, Zia, Adnan, Matiullah Shah, Manzoor, Sajjad and all the rest, too numerous to list here.

In the last, I would like to thank Higher Education Commission (HEC), Government of Pakistan, for the financial support provided through Indigenous 5000 PhD program with reference to the award letter number 17-5-1(Cu-204)/HEC/Sch/2004.

Imran Usman

List of Publications

The following papers have been presented, published/submitted in refereed conferences, journals or books during this research, and contain material based on the content of this thesis.

JOURNAL PUBLICATIONS

2010 **Imran Usman**, Asifullah Khan, and Rafiullah Chamlawi, Employing Intelligence in the Embedding and Decoding Stages of a Robust Watermarking system, *International Journal of Electronics and Communications*, (accepted) *impact factor 0.371*

2010 **Imran Usman** and A. Khan, "BCH coding and intelligent watermark embedding: employing both frequency and strength selection", *Applied Soft Computing*, Elsevier, vol. 10, No. 1, *impact factor: 1.909*.

2010 Rafiullah Chamlawi, Asifullah Khan, and **Imran Usman**, Authentication and Recovery of Images using Multiple Watermarks, *Computers and Electrical Engineering*, Elsevier, Vol. 36, No. 3, *impact factor 0.284*.

2009 **Imran Usman**, Asifullah Khan, Asad Ali, and Tae-Sun Choi, "Reversible Watermarking based on Intelligent Coefficient Selection and Integer Wavelet Transform", *International Journal of Innovative Computing, Information and Control*, Vol. 5, No. 12, 2009, *impact factor 2.79*.

2009 **Imran Usman**, Asifullah Khan and Anwar M. Mirza "Genetic Perceptual Shaping of a Digital Watermark: Enhancing Resistance Against a Cascade of Attacks", *Journal of Heuristics* (in review), *impact factor 1.064*.

2009 Asifullah Khan, Asad Ali, M. Tariq Mehmood, **Imran Usman**, and Tae-Sun Choi, " Reversible Hiding of Depth Maps for 3D Cameras", *Information Sciences*, Elsevier (submitted), *impact factor 3.095*.

2008 A. Khan, **Imran Usman**, Fahad Tahir, Labiba Gilani, "Intelligent Attack Adaptive Watermarking", *Computational Intelligence*, Wiley (in review), *impact factor 3.31*.

2007 **Imran Usman**, A. Khan, Rafiullah, Abdul Majid "Image Authenticity and Perceptual Optimization via Genetic Algorithms and a Dependence Neighborhood", *International Journal of Applied Mathematics and Computer Science (IJAMCS)*, vol. 4, no. 1, pp. 37-42.

CONFERENCE PUBLICATIONS

2009 Rafiullah Chamlawi, **Imran Usman**, and Asifullah Khan, Dual Watermarking Method for Secure Image Authentication and Recovery, *IEEE International Multitopic Conference* INMIC, pp. 270-273, December, 2009

2009 Rafiullah Chamlawi, Chang-Tsun Li, **Imran Usman**, and Asifullah Khan, Authentication and Recovery of Digital Images: Potential Application in Video Surveillance and Remote Sensing, *IEEE, international Conference on Consumer Electronics*, Las Vegas, USA, January, 2009

2008 A. Khan, M. Tariq Mahmood, Asad Ali, **Imran Usman**, and Tae-Sun Choi, "Hiding Depth Map of an Object in its 2D Image: Reversible Watermarking for 3D Cameras", 27th Inter. Conference on Consumer Electronics 2009, 10-14 Jan. 2009, Las Vegas, USA.

2008 **Imran Usman**, A. Khan, "Reversible Watermarking based on Intelligent Coefficient Selection and Variable Thresholds using Integer Wavelet Transform", *2008 International Symposium on Intelligent Informatics (ISII2008)***,** December 12-13, 2008, Kumamoto, Japan.

2008 **Imran Usman**, Asifullah Khan, Rafiullah Chamlawi and Tae-Sun Choi, "A Generalized Approach for Embedding Watermark in Digital Images using LDPC Codes and Genetic Programming", *WSC 2008 Online World Conference on Soft Computing in Industrial Applications*, 10-21 November 2008.

2008 Asifullah Khan, Asad Ali, M. Tariq Mahmood, **Imran Usman**, and Tae-Sun Choi, "Variable Threshold Based Reversible Watermarking: Hiding Depth Maps For Subsequent 3-D Analysis", *2008 IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications*, October 12-15, 2008, Beijing, China.

2007 **Imran Usman**, A. Khan, R. Chamlawi and A. Majid, "Towards a Better Robustness-Imperceptibility Trade-off in Digital Watermarking", *IEEE International Joint Conferences on Computer, Information and Systems Sciences and Engineering*, Dec 3rd – 12th, 2007 Bridgeport, USA.

BOOK CHAPTERS

2009 **Imran Usman**, Asifullah Khan, Rafiullah Chamlawi and Tae-Sun Choi, "Perceptual Shaping in Digital Image Watermarking using LDPC Codes and Genetic Programming" in *Applications of Soft Computing: From Theory to Praxis*, Springer-Verlag Berlin

/Heidelberg, 2009. DOI-10.1007/978-3-540-89619-7_50

2008 A. Khan, **Imran Usman**, "Intelligent Perceptual Shaping in Digital Watermarking" in *Artificial Intelligence in Multi-media Data Hiding* , Springer-Verlag.

2007 **Imran Usman**, A. Khan, Rafiullah and Abdul Majid, "Towards a Better Robustness-Imperceptibility Trade-off in Digital Watermarking", in *Innovations and Advanced Techniques in Systems, Computing Sciences and Software Engineering*, Springer Science, 2008.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AAW	Attack Adaptive Watermarking
ANN	Artificial Neural Networks
ASSW	Additive Spread Spectrum Watermarking
BCH	Bose-Chaudhury-Hocquengum
BCR	Bit Correct Ratio
CDF	Cohen-Daubechies-Fauraue wavelet
CI	Computational Intelligence
DCT	Discrete Cosine Transform
DSSS	Direct Sequence Spread Spectrum
DVD	Digital Versatile Disc
ECC	Error Correcting Codes
EPR	Electronic Patient Records
GGD	Generalized Gaussian Distribution
GP	Genetic Programming
GWSS	Genetic Watermark Shaping Scheme
HH	High-High: corresponds to diagonal frequencies in wavelet domain
HL	High-Low: corresponds to vertical frequencies in wavelet domain
HVS	Human Visual System
IARW	Intelligent Attack Resistant Watermarking

LDPC	Low Density Parity Check
LH	Low-High: corresponds to horizontal frequencies in wavelet domain
LL	Low-Low: corresponds to approximations in wavelet domain
LSB	Least Significant Bit
MLD	Machine Learning Decoding
MSE	Mean Squared Error
MSS	Mean Squared Strength
pdf	Probability Density Function
PRS	Pseudo Random Sequence
PSF	Perceptual Shaping Function
PSNR	Peak signal to Noise Ratio
S-VTF	Visual Tuning Function evolved using strength selection only
SP-VTF	Visual Tuning Function evolved using strength as well as position selection
SP-VTF (BCH)	Visual Tuning Function evolved by employing BCH coding along with strength and position selection
SSIM	Structure Similarity Index Measure
SVM	Support Vector Machines
TD	Threshold Decoding
VTF	Visual Tuning Function
WDR	Watermark to Document Ratio
wPSNR	Weighted Peak signal to Noise Ratio

1. Introduction

1

"Fantasies are more than substitutes for unpleasant reality;
They are also dress rehearsals, plans.
All acts performed in the world begin in the imagination."
Barbara G. Harrison

1.1. Motivation

In recent years the ease of creation, storage, manipulation and communication of digital data has stimulated the research in the field of digital watermarking, to protect digital information against illegal manipulation, usage, duplication and distribution. Digital watermarking is viewed as the practice of imperceptibly embedding the secret information, or the *watermark*, into the original data, or the *cover work*, to embed information about the same data [1]; in contrast to traditional methods that often require the transmission of additional metadata. It has a broad range of applications from copyright protection, device control, fingerprinting to data authentication and media forensics.

Watermarking can be further categorized into robust watermarking, fragile watermarking, semi fragile watermarking, and reversible watermarking. Robust watermarking is based on embedding a transparent watermark in the cover work for the purpose of copyright protection or identifying the sender/receiver of the data. In this type

of watermarking, the integrity of the watermark is of prime importance, i.e. the watermark should sustain all attacks unless the cover work is substantially degraded and becomes of little or no use. On the other hand, fragile watermarking deals with embedding a transparent watermark in the cover work for the purpose of content authentication. In this type of watermarking, the integrity of the cover work is of prime importance. That is, the watermark should be sensitive to the slightest modification in the cover work and raise alarms to signal and localize any tampering attempts. Semi fragile watermarking, on the other hand, allows the watermark to survive minor transformations, such as lossy compression, but must be destroyed by major changes, or *illegitimate distortions*. Reversible watermarking deals with the ability of a watermarking system to restore the original cover content completely after the extraction of watermark.

Watermarking Systems have three characteristics, namely:

i. Robustness,
ii. Imperceptibility, and
iii. Capacity.

These essential characteristics contradict one another, i.e. if one increases the other decreases. Therefore, one needs to make a tradeoff between these characteristics according to the application requirements. In the first phase of this work, the aim is to make an optimum trade-off between watermark robustness and imperceptibility, while keeping the capacity constant. For this purpose, machine learning techniques are employed on the encoding as well as decoding side of a robust watermarking system capable of surviving a sequence of attacks. In the second phase of this work, the goal is to

achieve an optimum tradeoff between watermark capacity and imperceptibility while devising an intelligent reversible watermarking system.

1.2. Watermarking in Historical Perspective

For thousands of years, people have sought secure ways to communicate. *The Histories* [2] of Herodotus reveal the following story which occurred in 480 BC, almost twenty five hundred years back. Where, once a slave was shaved and a secret message was tattooed in his scalp. His hairs were let grow again to conceal the message. He was then sent by his master, Histiaeus, to the Ionian city of Miletus. Upon arriving at Miletus, he was shaved again to reveal the message to the city's regent, Aristagoras; and this message encouraged him to start a revolt against the Persian occupier. In this ancient Greek story, the tattoo is the *mark* and the slave acted as a *cover work*.

Chinese invented paper more than a thousand years ago. But paper watermarking actually started in Italy around 1282 AD, when thin wire patterns were added to paper moulds for generating watermarks [1]. However, it was eighteenth century when paper watermarks became popular and were widely used in Europe and America. They were primarily used for making trademarks, to record the manufacture date, and as anti-counterfeiting measures on money and other documents.

The term *watermark* was probably originated from the German term *wassermarke* in eighteenth century which means a distinctive mark on paper. The mark is not produced by water, but probably so called because it looks like a wet spot. The first watermarking example similar to digital methods used nowadays appeared in 1954 when Muzak

Corporation filed a patent for watermarking musical work, which was used until around 1984 [3, 4].

The first research publication focused on digital watermarking appeared in 1990 by Tanaka et al. [5]. The term *digital watermarking* first appeared in 1993 when Tirkel et al. [6] presented two techniques of data hiding in digital images. These techniques were based on manipulations of the least significant bit (LSB) of the pixel values. It was after this that the field of digital watermarking began to mushroom.

In 1996 Cox et al. [1] introduced the concept of spread spectrum based modulation of a watermark signal to provide robustness. This work brought a major enhancement to the field of digital watermarking and introduced possibilities of new, diverse application areas. After a year, Piva et al. [7] proposed the idea of blind watermarking in which the original cover work is not required for detecting the watermark signal [1, 8]. Much of the research then concentrated on the development of novel algorithms by exploiting these two concepts of spread spectrum based embedding and blind detection. A much detailed discussion can be found in [1, 8, 9]. The third major development in watermarking originated from the work of Chen and Wornell [10], who used the concept of quantization index modulation based embedding of a watermark signal. The recent advancements in watermarking approaches include the idea of informed embedding and coding based watermarking [1, 8], reversible watermarking, fragile watermarking for authentication purposes, and semi-fragile watermarking approaches.

1.3. Robust Image Watermarking

Different applications pose different requirements on the watermark design [9], since they are heavily dependent upon the set of attacks that are likely to be mounted on them. These anticipated attacks may include intentional and/or unintentional manipulation of the marked cover work; thus, reducing the watermark detection and decoding performance. A universal watermarking system that can withstand all attacks and at the same time fulfill all other desirable requirements is almost impossible to be developed [1, 9]. Therefore, while designing a watermarking system, its intended application and thus the corresponding set of conceivable attacks are of prime importance.

An efficient watermarking system allows one to embed the watermark while minimizing the distortion of the watermarked work with respect to the original, and making it robust against any illegitimate attacks. But robustness is usually achieved at the cost of watermark imperceptibility (and vice versa), since these two properties contradict each other. Therefore, one needs to make a delicate balance between these two properties in accordance to the intended application while designing a watermarking system. A watermark is shaped according to the cover work by exploiting properties of Human Visual System (HVS) for properly hiding it into the cover work. To fulfill this requirement, perceptual models [11-14] that are in frequent usage by image compression community, are utilized. These perceptual models make a tradeoff between robustness and imperceptibility according to the cover image. However, they do not take into consideration the watermark application and, thus, the anticipated attack information. For

instance, consider wireless medical applications, where compression and transmission of medical data, especially images, has to be performed in view of context aware medical networks [15]. Similarly, print-to-web technology, broadcast monitoring, secure digital camera, etc., are such watermarking applications that mostly encounter a series of attacks. Other pertinent examples where watermarking system needs to be robust against a set of anticipated attacks exist in literature [1, 9, 16-18]. In these set of circumstances, it is prudent to use intelligent approaches, such as those being recently introduced by Laskov et al. [19], for not only structuring the watermark at the embedding phase, but also change the decoding strategy accordingly.

A variety of watermarking schemes are proposed in literature including transform domain techniques based on Discrete Cosine Transform (DCT) [20], Discrete Fourier Transform [21], Discrete Wavelet Transform [22], Vector Quantization schemes [23] and spatial domain methods [24, 25]. These schemes make use of a predetermined set of coefficients, mostly in the middle frequency range, to serve as a tradeoff for watermark embedding. The disadvantages of such a selection are discussed in detail by Shieh et al. [17], who also proposed a scheme in which suitable embedding positions in a block based DCT domain watermarking are selected using Genetic Algorithm. Nevertheless, they have not taken into consideration the selection of suitable strength of alteration and watermark application. On the other hand, Khan and Mirza [16] have employed Genetic Programming (GP) for the selection of suitable strength of alteration to each selected DCT coefficient for watermark embedding. However, to the best of our knowledge, there is no watermarking approach reported so far that performs both selection of optimal

frequency band as well as strength of alteration simultaneously, in view of HVS and hostile application environment.

Similarly, as regards the decoding stage of a watermarking system, fixed watermark extraction strategies may not be suitable for the widespread and dynamic applications of watermarking, where different types of attacks are expected at different instances. Therefore, an intelligent watermark extraction strategy is needed, which may adapt itself in accordance to the new watermarking application. In this context, most of the existing watermark extraction strategies [26, 27] do not consider the existence of attacks during the training phase and thus are not adaptive. In the same way, watermarking approaches [28, 29] that do not exploit Computational Intelligence (CI), generally, use simple Threshold Decoding (TD) and thus, are also not adaptive as regards the attack on the watermark. These approaches neither consider the alterations that may incur to the features nor exploit the individual frequency bands; rather treat all the frequency bands collectively. Consequently, these approaches use a single feature for extraction of a message bit. Recently, Khan et al. [30] have proposed CI based decoding; however, their system does not employ intelligent embedding. In contrast, this work presents an innovative scheme of utilizing CI based decoding strategies in conjunction with intelligent embedding. For this purpose, firstly, we perform GP based intelligent embedding. This helps us in exploiting only the GP selected frequency bands, individually through CI techniques. Considering the available frequency bands in an 8×8 block DCT, this in turn also acts as a feature reduction strategy. Secondly, we exploit the learning capabilities of Support Vector Machine (SVM) and Artificial Neural Networks (ANN) to gain knowledge of the distortion that might have incurred varyingly on the

different selected frequency bands due to the sequence of attacks. Lastly, to perform this efficiently and in an adaptive manner, we devise a new feature extraction strategy for watermark extraction in presence of attacks.

1.4. Reversible Watermarking

Traditional watermarking techniques cause irreversible degradation of an image. Although the degradation is perceptually sparse, it may not be admissible in applications like medical, legal or military imagery. For applications such as these, it is desirable to recover the embedded information as well as the sensitive host image. For example, in a database of medical images marked with patient information, it is desired to extract the patient information along with recovery of the original host signal for proper diagnosis. Unlike robust watermarking, in reversible watermarking the original image is completely restored from the watermarked image, thus, preserving the originality of the cover work.

An efficient reversible watermarking scheme should be able to embed more information with less perceptual distortion, and equally, be able to restore the original cover work content. Watermark capacity and imperceptibility are two contradicting properties. If one increases, the other decreases and vice versa. Therefore, one needs to make an optimum choice between them.

There are a variety of reversible data hiding methods proposed in literature. These include Chen and Kao's DCT based zero replacement reversible watermarking [31], Ni et al.'s [32] histogram manipulation based reversible watermarking technique, Xuan et al.'s [33] spread spectrum based watermarking method and Tian's [34] lossless data hiding technique based on difference expansion transform of a pair of pixels. Alattar [35]

extended Tian's work to the difference expansion of a vector of several pixels to achieve larger capacity. These approaches are simple and efficient but do not make an efficient tradeoff with respect to imperceptibility. Xuan et al. [36] proposed a reversible watermarking scheme using integer wavelet transform and threshold embedding. However, they have used a fixed threshold for all of the coefficients in different sub-bands of integer wavelet transform. Xuan et al. [37] also proposed a bitplane compression based technique which losslessly compresses one or more middle bitplanes to save space for data embedding. In this work we propose an intelligent scheme which selects suitable coefficients in different wavelet sub-bands. By selecting the appropriate coefficients for embedding, it is possible to make a good choice between the watermark payload and imperceptibility. Considering this as an optimization problem and employing an intelligent technique provides an optimum payload/imperceptibility tradeoff, thus, the margin of improvement. We use GP to exploit the hidden dependencies of the wavelet coefficients in different sub-bands. GP belongs to the class of biologically inspired optimization techniques. Experimental results demonstrate that the proposed reversible watermarking scheme is more efficient compared to prior works.

1.5. Research Objectives and Contributions

The objective of this research is to use machine learning techniques for making suitable tradeoff in the contradicting, but essential, properties of watermarking systems. We apply our intelligent approaches in two domains of watermarking, i.e. robust watermarking (phase I) and reversible watermarking (phase II). This research contributes in the following manner:

Phase I: Robust watermarking

1. We exploit the properties of HVS in order to find all the possible functional dependencies which were not taken care of previously. We also investigate the problem with respect to the robustness properties and constraints of a watermarking system.

2. We use GP to visually tune the watermark with respect to HVS in order to achieve robustness against a series of attacks instead of a single attack, which is much more desired scenario.

3. In our proposed GP based watermark embedding, we employ the concept of wrapper for carrying out both frequency band as well as optimal strength selections of the transformed domain coefficients.

4. We further enhance the robustness of the proposed scheme by employing machine learning based adaptive decoding strategy in addition to adaptive visual tuning at the embedding side.

5. Both the above attributes of the proposed scheme make it extremely difficult for an adversary to gain the secret knowledge of the system by analyzing the embedding and/or decoding space.

6. We introduce another layer of robustness in the system by the suitable use of error correcting codes in our watermarking system. We also propose a method for increased security of the watermark by incorporating lossless compression and RSA based encryption before the embedding of the watermark.

Phase II: Reversible watermarking

7. We analyze threshold based reversible watermarking techniques and devise the use of intelligent coefficient selection instead for watermark embedding. Using machine learning and choosing the appropriate coefficients based upon coefficient value, its neighborhood mean, its block number etc. makes a good choice between watermark payload and imperceptibility.

8. We use GP to make an optimum choice of coefficients in different wavelet sub bands such as to embed maximum payload with little or no compromise on imperceptibility.

1.6. Structure of the Thesis

Chapter 2 discusses some preliminaries of digital watermarking with main focus on robust watermarking and reversible watermarking. In the context of robust watermarking, spread spectrum based additive watermarking technique is explained and basics of HVS modeling are examined. It also discusses some preliminary topics on GP and its use in digital watermarking. In addition, a brief introduction to SVM and ANN is provided at the end of this chapter.

Chapter 3 starts with the contributions in this research. It explains the evolution of a simple Visual Tuning Function (VTF) through the use of GP which is resistant against a series of attacks. It then explains the evolution of a more sophisticated VTF that uses the concept of wrapper during its evolution in order to select both the strength as well as

position of watermark embedding in transform domain coefficients. The use of error correcting codes to further enhance robustness follows next in the chapter.

Chapter 4 describes the use of intelligent embedding as well as decoding of a watermark in the presence of a series of attacks. The GP based intelligent embedding is explained in the context of attacks and HVS. It is followed by CI based intelligent decoding, where SVM and ANN are used as feature reduction and extraction techniques.

Chapter 5 discusses the second phase of this research. It describes the use of GP to select the appropriate coefficients for embedding in integer wavelet based reversible watermarking scheme. It also discusses various dependencies taken into consideration while evolving optimal expressions using GP for making a tradeoff between watermark payload and imperceptibility.

Chapter 6 presents conclusions of this research by highlighting the major achievements. It addresses various challenges posed in robust and reversible image watermarking and different methodologies adapted in this research to counteract them. It also discusses some of the aspects still needing considerations in the future research on the topic.

2. Digital Watermarking and Machine Learning: Some Preliminaries

2

"O, what may man within him hide,
Though angel on the outward side!"
William Shakespeare

2.1. Digital Watermarking

For centuries, man has sought secure ways to communicate. Digital Watermarking is one such effort. With the abundance of computers and digital devices, and prevalence of growing networks to connect them, digital information exchange, production, and sharing has increased like never before. This also prompted the efforts towards copyright violation and illegitimate manipulation of digital data.

A digital watermark is a piece of information that is hidden directly in media content, in such a way that it is imperceptible to a human observer, but easily detected by a computer. The principle advantage of this is that the watermark is inseparable from the content and it removes any need to store separate, associated metadata, such as cryptographic signatures. Broadly speaking, digital watermarking can be categorized into two major categories namely:

1. *Visible Watermarking*, and
2. *Invisible Watermarking*.

Visible watermarking is the type of watermarking in which the watermark is visible to the end user after embedding, mostly as a translucent mark in images, videos and documents. Logos displayed at one corner of different television channels is an example of such type of watermarking. A majority of disadvantages are associated with this type, which include perceptible degradation of the cover work, ease of mounting attacks such as cropping, and decrease in usefulness of the cover work. On the other hand, invisible watermarking is the type of watermarking in which the watermark is not visible to the end user of the work. Rather, it is imperceptibly embedded into the cover work by altering spatial amplitudes or transform-domain coefficients. Some of the key advantages of such watermarking type include high perceptual quality, difficulty in mounting attacks, increase in usefulness of the marked work, and a large number of applications areas. For the rest of this work, digital watermarking simply refers to invisible digital watermarking.

Likewise, there are two different approaches for watermark detection/ extraction as follows:

1. *Blind Detection*, and
2. *Informed Detection*.

In blind detection, the original cover work is not needed for watermark detection and consequently, watermark extraction. On the other hand, original cover work is required for informed watermark detection. Systems that use blind detection are often referred as *public watermarking systems* whereas those that use informed detection are called *private watermarking systems*. In terms of different applications and constraints on the watermarking systems, digital image watermarking can be subdivided into four types, namely:

1. *Robust Watermarking*,
2. *Reversible Watermarking*,
3. *Fragile Watermarking*, and,
4. *Semi-Fragile Watermarking*.

First, a brief description of these watermarking types is described as follows and afterwards, different defining properties of digital watermarking are discussed.

Robust Watermarking

Robust watermarking is based on embedding a transparent watermark in the cover work for the purpose of copyright protection or identifying the sender/receiver of the data. In this type of watermarking, the integrity of the watermark is of prime importance, i.e. the watermark should sustain all attacks unless the cover work is substantially degraded and becomes of little or no use.

Reversible Watermarking:

Reversible watermarking deals with the ability of a watermarking system to restore the original cover content completely after the extraction of watermark. It provides suitable methods for suppressing induced distortions from different transformations during embedding, detection, and decoding stages of watermarking. In this type of watermarking, usually high capacity is required at a higher level of watermark imperceptibility.

Fragile Watermarking:

This type of watermarking employs the embedding of a transparent watermark in the cover work for the purpose of content authentication. In this type of watermarking, the integrity of the cover work is of prime importance. That is, the watermark should be sensitive to the slightest modification in the cover work and raise alarms to signal and localize any tampering attempts.

Semi-fragile Watermarking:

Semi-fragile watermarking allows the watermark to survive minor transformations and legitimate distortions, such as lossy compression, but destroyed by illegitimate distortions. Since a fragile watermarking system is very sensitive to slightest modification in the marked work, it cannot be used in majority of applications in which a marked work is susceptible to common signal processing and channel noise. Semi-fragile watermarking systems provide some flexibility which, indeed, makes it very suitable for a variety of applications.

Watermarking systems can be characterized by a number of different properties. The relative importance of each property is dependent on the requirements of the application and the role the watermark will play. There are three properties typically associated with a watermark embedding process: *robustness*, *imperceptibility*, and *capacity* (also known as payload).

2.1.1. Robustness

Watermark robustness means the capability of the hidden data to survive host signal manipulation including both malicious and non malicious manipulations. Malicious manipulations precisely aim at damaging the hidden information, and non-malicious manipulations do not explicitly target removing the watermark or at making it unreadable. The exact level of robustness that the hidden data must possess can not be specified without taking into account a particular application.

Achieving watermark robustness in presence of a series of attacks is one of the main challenges watermarking community is facing. Even without going into many details, we can say that robustness against signal distortions, constituting a series, is better achieved if the watermark is placed in perceptually significant parts of the signal. Lossy compression algorithms operate on the similar lines by discarding perceptually insignificant data without affecting the quality of the compressed image. On the other hand, watermarks hidden within perceptually insignificant regions are likely not to survive compression attacks.

In real world applications the attacker intentionally tries to destroy the watermark, often through a sequence of a variety of attacks which may include geometric distortions or additive noise attacks. In case of image watermarking, the resistance to geometric manipulations such as rotation, resizing, translation, and cropping is still an open area of research. But mostly, there are some specific attacks related to a particular watermarking application. If one can anticipate them at the time of design of the watermarking scheme, it is possible to counter most of them.

2.1.2. Imperceptibility

Imperceptibility is one of the most important requirements of a watermarking system which deals with the perceptual transparency of the watermark, independent of the application and purpose of the watermarking system. The watermark should not be noticeable to the viewer nor should the watermark degrade the quality of the content. Artefacts introduced through a watermarking process are not only annoying and undesirable, but may also reduce or destroy the commercial value of the watermarked data. It is therefore important to design marking methods which exploit effects of HVS in order to maximize the energy of the watermark under the constraint of not exceeding the perceptible threshold. Two problems are related to this issue. The first one is the reliable assessment of the introduced distortion. The second problem occurs when processing is applied to the watermarked data. For example, the visibility of the watermark may increase if the image is scaled.

Theoretically, the watermark should be invisible to a human eye, even on the highest quality equipment. Although visible watermarks are usually more robust, for most of the applications it is advantageous for the embedded mark to be indiscernible to the human eye. To date, researchers have attempted to conceal the watermark in such a way that it is not possible to be noticed. However, this constraint contradicts with other requirements such as robustness.

2.1.3. Capacity

The capacity of a channel or medium is defined as the maximum achievable data rate for that channel. In the context of digital watermarking, capacity may be referred as

the maximum allowable data to be embedded in the form of a watermark. Capacity is a major requirement in a number of watermarking applications where high embedding rate is highly desirous. This higher rate may be due to large amount of information to be embedded, associated extra information, or the redundancy in the watermarking algorithm to achieve superior robustness.

Capacity has a direct influence on the robustness of a watermark. The higher the capacity, the less likely it is to destroy the watermark without substantial degradation of the cover work. On the other hand, high capacity degrades the imperceptibility of the watermark. Therefore, one needs to make an optimum choice between capacity and imperceptibility.

One way to achieve optimum capacity-imperceptibility trade-off is by intelligently selecting the appropriate locations in an image where data is not easily perceptible. The second phase of this work focuses on this approach and uses an intelligent algorithm to embed high capacity watermark in a reversible watermarking application.

2.1.4. Robustness versus Imperceptibility

There is an intrinsic relation between the two most important, but contradicting properties of a watermarking system; robustness and imperceptibility. For instance, if we look at an example where one wants to conceal a watermark so that it is well hidden from an observer, then perceptually insignificant areas are selected to fulfil this imperceptibility requirement. On the other hand, a watermarked image may undergo lossy compression, which tends to remove less visible high-frequency components and

keeps only the lower ones (i.e. perceptually significant ones). This, in other words, decreases robustness. If we try to improve the watermark imperceptibility, robustness decreases and vice versa. Consequently, one needs to make a tradeoff, in terms of watermark strength distribution, according to the application domain. For this purpose, different methods, both in spatial as well as transformed domain, have been used to tailor a watermark according to the cover image [25, 38]. These watermarking systems are known to be image adaptive. On the other hand, most of the earlier approaches are not image adaptive and use a global watermarking strength for all the selected coefficients [7].

2.1.5. Imperceptibility versus Capacity

Watermark imperceptibility and capacity are another very important but contradicting properties in a watermarking system. A watermark is usually not embedded in all of the frequency bands, rather perceptually insignificant bands are selected to fulfill this purpose. Mostly, these include different coefficients in the middle and high frequency bands. This limits the amount of information that could be embedded, and hence, reduces the payload. That is, the higher the capacity, the lower is the imperceptibility and vice versa. Therefore, an efficient watermarking system should be able to embed a high capacity watermark while minimizing any perceptual distortions.

2.1.6. Watermarking Applications

Watermarking has a wide range of applications. Generally, a watermarking scheme is designed in view of its application, as the application poses certain

requirements to be fulfilled [9]. Watermarking has found a large number of applications recently due to its advantages over the possible alternative technologies. A few examples of watermarking applications are:

Copyright Protection:

Copyright protection is probably the most prominent application of watermarking today. The objective is to embed information about the source, and thus typically the copyright owner, of the data in order to prevent other parties for claiming the copyright on the data. Thus, the watermarks are used to resolve rightful ownership. This application requires a very high level of robustness.

Data Authentication:

When watermarking is used for authentication purpose, the objective is to detect modifications of the data. This can be achieved with fragile watermarks that have a low robustness to certain modifications like compression, but are impaired by other modifications. Furthermore, the robustness requirements may change depending on the data type and application.

Copy protection:

A desirable feature in multimedia distribution systems is the existence of a copy protection mechanism that disallows unauthorized copying of the media. Copy protection is very difficult to achieve in open systems; in closed or proprietary systems, however, it is feasible. In such systems it is possible to use watermarks indicating the copy status of the data. An example is the Digital Versatile Disc (DVD) system where the data contains

copy information embedded as a watermark. A compliant DVD player is not allowed to copy data that carry a “copy never” watermark.

Fingerprinting for traitor tracking:

There are some applications where the objective is to convey information about the legal recipient rather than the source of digital data, mainly in order to identify single distributed copies of the data. This is useful to monitor or trace back illegally produced copies of the data that may circulate, and is very similar to serial numbers of software products. This type of application is usually called *fingerprinting* and involves the embedding of a different watermark into each distributed copy.

2.1.7. Objective Measures for Robustness, Imperceptibility and Payload

Watermarking systems are implemented in different feature domains and, thus, rely on diverse embedding and decoding techniques. Therefore, for comparing performance, a set of general and objective measures are defined. For this purpose, robustness is generally measured in terms of Bit Correct Ratio (BCR) [8, 9], bit error rate and false alarm probability. Whereas, imperceptibility of a watermark is measured in terms of Structural Similarity Index Measure (SSIM) [39], weighted Peak Signal to Noise Ratio (wPSNR) [38] and Watermark to Document Ratio (WDR) [40]. The capacity of a watermarking system is mostly measured in terms of bits per pixel values.

Watermark power can also be used as a measure of robustness. For this reason, Mean Squared Strength (MSS) may be used as an objective measure representing estimated robustness:

$$MSS = \frac{1}{N_b N_c} \sum_{x_1=1}^{N_b} \sum_{x_2=1}^{N_c} \alpha(x_1, x_2)^2 \tag{2.1}$$

where, N_b is the total number of 8×8 blocks in the cover image and N_c is the number of selected DCT coefficients in a block. x_1 and x_2 are the respective indices.

BCR is the second objective measure for robustness and represents Bit Correct Ratio. It is the ratio between the number of correctly extracted watermark bits to the actual number of bits embedded. It is defined as:

$$BCR(M, M') = \sum_{i=1}^{L_m} \overline{(m_i \oplus m_i')} \Big/ L_m \tag{2.2}$$

where M is the original, while M' represents the decoded message. L_m is the length of the message and $\oplus$ represents exclusive-OR operation.

SSIM is an imperceptibility measure and denotes the structural similarity index measure of the marked image at a certain level of estimated robustness. This image quality assessment relies on the assumption that human visual system is highly adapted to extract structural information. The work by Wang et al. [39] shows that a measure of change in structural information can provide a good approximation for perceived distortion. The *SSIM* between two images **x** and **y** is provided as:

$$SSIM(\boldsymbol{x}, \boldsymbol{y}) = \left[l(\boldsymbol{x}, \boldsymbol{y})\right]^{\alpha} \cdot \left[c(\boldsymbol{x}, \boldsymbol{y})\right]^{\beta} \cdot \left[s(\boldsymbol{x}, \boldsymbol{y})\right]^{\gamma}, \tag{2.3}$$

where, $l(\boldsymbol{x}, \boldsymbol{y})$ is the luminance comparison function, $c(\boldsymbol{x}, \boldsymbol{y})$ represent the contrast comparison function, and $s(\boldsymbol{x}, \boldsymbol{y})$ is the structure comparison function. α, β and γ are the

parameters used to adjust the relative importance of the three components following the constraints $\alpha > 0$, $\beta > 0$, and $\gamma > 0$. Further details on *SSIM* are provided in [39].

w*PSNR*, which is a modified version of Peak Signal-to-Noise Ratio (PSNR), is the second objective measure for watermark imperceptibility. It uses a texture masking function as a penalization factor and is defined as:

$$wPSNR = 10\log_{10}\left(\frac{255}{RMSE \times NVF}\right)^2 , \tag{2.4}$$

where *NVF* is the Noise Visibility Function proposed by Voloshynovskiy et al. [38] and is an arbitrary texture masking function. Its value ranges from 0, for extremely textured areas, and up to 1, for clear smooth areas of an image. *RMSE* is the Root Mean Squared Error between original and the watermarked work given by:

$$RMSE = \sqrt{(\mathbf{y} - \mathbf{x})^2} \ . \tag{2.5}$$

2.2. Human Visual System Modeling

Development of adaptive watermarking schemes to structure a watermark requires the understanding of the cover image in the context of HVS. In spatial-domain, this understanding means knowing the distribution of smooth and textured areas in a cover image. In transform-domain, it means knowing the distribution of low, mid and high frequency components of the cover image. Thus, in order to hide the watermark, the watermark is tuned/shaped using perceptual models that exploit sensitivities/insensitivities of HVS. The better a perceptual model is, the better is the perceptual shaping and hence imperceptibility of the watermark. Perceptual model can be

viewed as a VTF, providing maximum allowed alteration to a pixel value (or DCT coefficient) that is not observed by a human observer.

Perceptual models, usually based on empirical studies, are used to exploit sensitivities/insensitivities of HVS for assigning a perceptual slack, $\alpha(i, j)$ to each term of the cover work expressed in some domain. Many perceptual models have been proposed in literature including Watson's Perceptual Model [11, 12], originally proposed by Ahumada and Peterson [41], which assigns a slack to each term in block DCT domain. In context of watermark tuning, in the sequel, we use Perceptual Models and PSF interchangeably. Other proposed PSF for images assign slacks to frequencies in Fourier and Wavelet domains [42], or pixels in the spatial domain [38]. Similarly, Kutter and Winkler's [43] model is based on local isotropic measure and a masking function, while that of Lambrecht and Farrell's [13] model is based on Gabor filters. In phase I of this work, the proposed genetic watermark embedding scheme is based on the modification of Watson's PSF.

Consider an image matrix **x** in spatial domain. The image is transformed to matrix $\mathbf{X}$ by applying 8×8 block DCT. According to the Watson' perceptual model, the visibility threshold $T(i, j)$ for every (i, j) DCT coefficient of 8x8 block is defined as follows:

$$\log T(i,j) = \log\left(\frac{T_{\min}\,(f_{i,o}^2 + f_{o,j}^2)^2}{(f_{i,o}^2 + f_{o,j}^2)^2 - 4(1-r)\,f_{i,o}^2 f_{o,j}^2}\right) + u\left(\log\sqrt{(f_{i,o}^2 + f_{o,j}^2)} - \log f_{\min}\right)^2, \qquad (2.6)$$

where $f_{i,o}$ and $f_{o,j}$ denotes the vertical and horizontal frequencies (cycles/degree) of the DCT basis functions respectively. $T_{\min}$ is the minimum value of $T(i, j)$ corresponding to

the minimum frequency $f_{\min}$. The rest of the parameters are set empirically [11, 16]. The effect of luminance sensitivity is considered by correcting this threshold corresponding to average luminance of each block:

$$T'(i,j) = T(i,j)\left(\frac{X_{o,o}}{\bar{X}_{o,o}}\right)^{a_T}, \tag{2.7}$$

where, $X_{o,o}$ is the DC coefficient of each block and $\bar{X}_{o,o}$ represents the average screen luminance =1024 (for an 8-bit image). The effect of contrast masking is incorporated by the following relation:

$$T^*(i,j) = \max[T'(i,j)\ , \left|T'(i,j)\right|^{1-\omega} X(i,j)^{\omega}], \tag{2.8}$$

where, $X(i,j)$ is AC DCT coefficient of each block and ω has been empirically set to a value of 0.7. These allowed alterations represent the perceptual mask denoted by $\boldsymbol{\alpha}$.

Watson's VTF, although fair enough to give us imperceptible alterations, is not an optimum VTF. This is because some effects like spatial masking in frequency domain are ignored as well as many of the constants are set empirically. Based on this, this work proposes development of a system that can deliver an effective VTF for a given watermarking application.

2.3. Additive Spread Spectrum Watermarking (ASSW)

Let us represent an image in the spatial domain as a discrete 2-D sequence $\mathbf{x}$ and its 8×8 block DCT transform as $\mathbf{X}$. The watermark that is being added to $\mathbf{X}$, generating watermarked DCT image $\mathbf{Y}$, is viewed as a 2-D DCT signal $\mathbf{W}$. Let M be a message,

which is mapped to a codeword vector (see figure 2.1). At the decoding stage, we have to retrieve this message M.

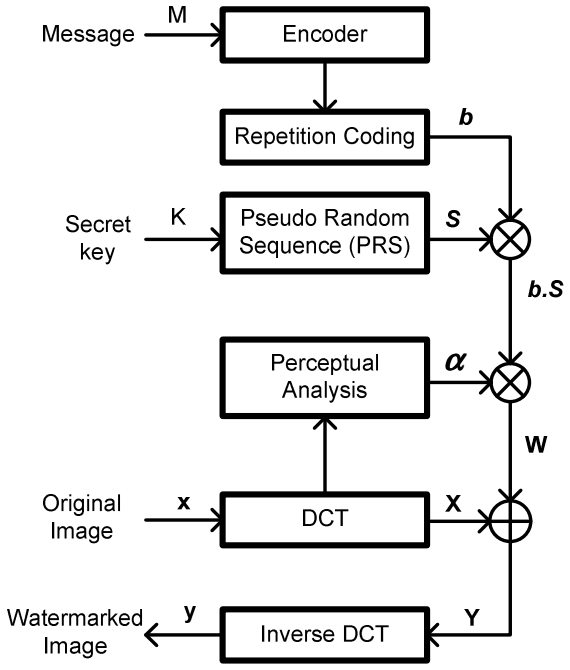


Fig.2.1 Additive Spread Spectrum Watermark Embedding

The codeword vector is then expanded to generate $\mathbf{b}$, with each element b_i repeated over a selected set of DCT coefficients. This redundancy bolsters robustness. The resulting signal $\mathbf{b}$ is further direct sequence spread spectrum (DSSS) modulated. The DSSS modulation is performed using 2-D pseudo random sequence (PRS) denoted by $\mathbf{S}$. The PRS behaves as spread sequence taking values ± 1 and has zero mean. Next, to shape the resultant signal according to the cover image, it is multiplied with perceptual mask $\boldsymbol{\alpha}$ (obtained by applying the VTF to the cover image in DCT domain) to produce the watermark $\mathbf{W}$:

$$\mathbf{W} = \boldsymbol{\alpha} \cdot \mathbf{S} \cdot \mathbf{b} \tag{2.9}$$

Adding this watermark to the original image, coefficient by coefficient thus, performs the embedding:

$$\mathbf{Y} = \mathrm{X} + \mathrm{W} \tag{2.10}$$

Here the watermark $\mathbf{W}$ is our desired signal, while the cover image $\mathbf{X}$ acts as an additive noise.

2.3.1. Watermark Extraction

In the ASSW approach [28], it is assumed that the probability density function (pdf) of the original coefficients remains the same even after embedding. Based on this idea, Hernandez et al. [28] have obtained expressions for maximum likelihood decoder structures. The zero mean generalized Gaussian pdf is given as follows:

$$f_X(x) = A e^{-|\beta x|^c} \tag{2.11}$$

where, both A and β are expressed as a function of the unknown parameters c and standard deviation σ:

$$\beta = \frac{1}{\sigma}\left(\frac{\Gamma(3/c)}{\Gamma(1/c)}\right)^{1/2} \quad , \quad A = \frac{\beta c}{2\Gamma(1/c)} \tag{2.12}$$

with Γ being the gamma function. The unknown parameters c and σ should be estimated from the received image at the decoding stage. Figure 2.2 illustrates the watermark extraction process.

2.3.2. Generalized Gaussian based Maximum Likelihood decoder:

In the verification process, our job is to obtain an estimate of the hidden message M from the marked image (figure 2.2). We know that, in cases where there is less or no knowledge of the priori probabilities of the classes/hypotheses, priori probabilities are selected that make the classes equally likely. Hence if we assume that the messages are equiprobable, then it is fair to consider maximum likelihood test. The estimated message should satisfy:

$$\ln \frac{f(Y/\boldsymbol{b}_l)}{f(Y/\boldsymbol{b}_m)} > 0 \quad , \quad \forall\ m \neq l \tag{2.13}$$

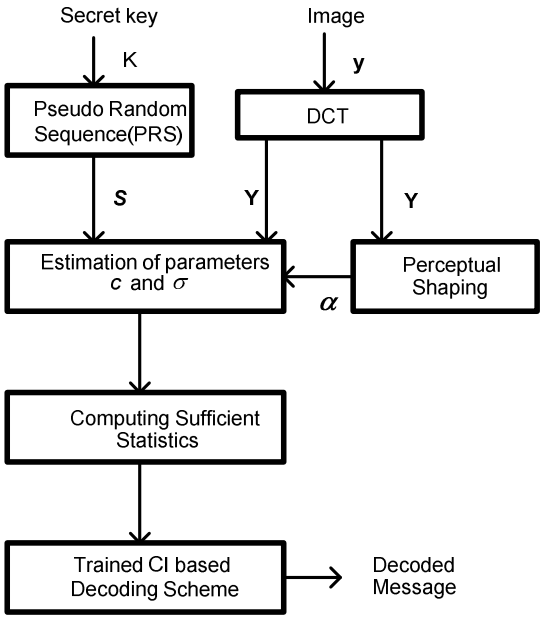


Fig.2.2 Additive Spread Spectrum Watermark Extraction

The sequences that are generated by considering each (i, j) DCT coefficient of all 8x8 blocks are assumed to behave like generalized Gaussian and are statistically

independent. For notational clarity, let us denote 2-D indices by $\boldsymbol{k}$ and the 2-D sequences by $Q_{i,j}[\boldsymbol{k}]$, obtained as follows:

$$Q_{l_1,l_2}[k_1,k_2] \triangleq X[8k_1+l_1, 8k_2+l_2] \quad , \quad l_1,l_2 \in \{0,\cdots,7\} \tag{2.14}$$

The maximum likelihood is equivalent to finding index $l \in \{1,2,\cdots\cdots,L\}$ that obey

$$\sum_{\boldsymbol{k}} \frac{\left|Y[\boldsymbol{k}]-W_m[\boldsymbol{k}]\right|^{c[\boldsymbol{k}]} - \left|Y[\boldsymbol{k}]-W_l[\boldsymbol{k}]\right|^{c[\boldsymbol{k}]}}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}} > 0 \quad , \quad \forall\ m \neq l \tag{2.15}$$

We should remember that our bit can be +1 or –1 only i.e. a bipolar signal. If now G_i sample vector denote all the DCT coefficients of different 8×8 blocks that correspond to a single bit *i*, then the sufficient statistics of this sample vector is given by:

$$r_i \triangleq \sum_{\boldsymbol{k}\in G_i} \frac{\left|Y[\boldsymbol{k}]+\alpha[\boldsymbol{k}]S[\boldsymbol{k}]\right|^{c[\boldsymbol{k}]} - \left|Y[\boldsymbol{k}]-\alpha[\boldsymbol{k}]S[\boldsymbol{k}]\right|^{c[\boldsymbol{k}]}}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}} \tag{2.16}$$

For bipolar signal i.e. $\boldsymbol{b} \in [-1,1]$, the hidden bits are then estimated using '0' as a threshold:

$$\hat{b}_i = \mathrm{sgn}(r_i) \qquad \forall\ i \in \{1,2,\cdots,N\} \tag{2.17}$$

Details of this ASSW approach are well documented by Hernandez et al. [28].

2.4. Error Correcting Codes and Digital Watermarking

Error Correcting Codes (ECC) are gaining more popularity in terms of their capability to enhance watermark robustness. Alattar and Alattar [44] have used spread

spectrum technique and Bose-Chaudhury-Hocquengum (BCH) coding for watermarking electronic documents. The work by Miaou et al. [45] demonstrates significant improvement in robustness by using BCH coded watermarks in error-prone transmission of MPEG video. Other pertinent examples include the use of low-density parity check codes (LDPC) [46], and Reed-Solomon codes [47] as ECCs to enhance robustness in digital watermarking.

2.5. Attacks and their Countermeasures

A watermarked data can be attacked in a variety of different ways. However, each application usually has to deal with a specific sequence of distortions. Cox et al. [1] and Barni et al. [8] have discussed in detail the types and levels of robustness that might be required for a particular watermarking application. They have discussed some of the attacks as well as their countermeasures. Keeping in view the expected distortions, several strategies are implemented to make a watermark system reliable. Few examples include; redundant embedding, selection of perceptually significant coefficients, spread spectrum modulation, and inverting distortion in the detection phase.

Voloshynovsky et al. [48, 49] have classified attacks into four basic categories: removal and interference attacks, geometrical attacks, cryptographic attacks and protocol attacks. Intentional tempering, as opposed to the common signal processing attacks are difficult to survive. However, watermark attacks as well as their countermeasures are complex and still a topic of research. Therefore, in evaluating the potential of a watermarking technique to meet the robustness requirements, many assumptions are made especially about the attacker. For example, does the attacker know the

watermarking algorithm, does he possess a detector that he can modify, what tools are available to him etc. Once the watermarking system is specified publicly, an attacker usually has more freedom as compared to a watermarker because the attacker is free to develop extra and more intricate attacks, while the watermarker can no longer amend it [1, 8, 9].

2.6. Genetic Programming and the Basic GP Algorithm

Behavioral/Cognitive models are often very difficult and time consuming to design and implement. On the other side, GP is a type of evolutionary algorithms which mimic biological evolution through selection. It is mostly used in optimization problems, by using a quality norm to measure and compare candidate solutions in a stepwise enhancement process. This constitutes the basic principle of biological evolution, where the comparison of the candidates and their subsequent selection for breeding characterize the main concept of enhancement through evolution. Figure 2.3 illustrates the flow chart of a typical GP search mechanism.

In GP, a candidate solution is represented using a data structure such as a tree. Initially, a random population of such candidate solutions is created. Every candidate solution is evaluated and scored using application dependent fitness function. The survival of fittest is implemented by retaining the best individuals. The rest are deleted and replaced by the offspring of the best individuals. The retained ones and the offspring make a new generation. Some offspring may have high score than their parents in the previous generation.

The whole process is repeated for the subsequent generations with the scoring and selection procedure in place. Every new generation, on average, has a slightly higher

score than the previous one. The process is stopped when any of the stopping criteria is met, which may include the number of generations, fitness value or time limit. In this way, the solution space is refined generation by generation and thus converges to the optimal/near optimal solution. For a detailed study, one may refer to [50, 51]. In this work, we use GP to evolve superior VTFs that are able to make superior tradeoff between robustness-imperceptibility and imperceptibility-capacity with respect to the existing tradeoff techniques, along with taking into consideration the information pertaining to a cascade of conceivable attacks.

Traditionally, GP uses a *generational* evolutionary algorithm. In generational GP, there exist well-defined and distinct generations. Each generation is represented by a complete population of individuals. The newer population is created from and then replaces the older population. The following are the four preliminary steps in a GP run, followed by the execution cycle of the generational GP algorithm.

Preliminary steps:

1. define the terminal set
2. define the function set
3. define the fitness function
4. Define parameters such as population size, maximum individual size, crossover probability, selection method, and termination criterion (e.g., maximum number of generations, time limit, fitness limit etc.).

Execution cycle:

5. initialize the population
6. Evaluate the individual programs in the existing population. Assign a numerical rating or fitness to each individual.
7. until the new population is fully populated, repeat the following steps:
    - Select an individual or individuals in the population using selection algorithm.
    - Perform genetic operations (crossover, mutation, replication etc.) on the selected individual or individuals
    - Insert the result of the genetic operations into the new population.

Fig.2.3 GP search mechanism

8. If the termination criterion is fulfilled, then continue. Otherwise, replace the existing population with the new population and repeat steps 6-8.

9. Present the best individuals in the population as the output from the algorithm.

2.6.1. Genetic Programming Program Structures

The functions and terminals are the primitives with which a program in GP is built. Functions and terminals play different roles. In general, terminals provide a value to the system while functions process a value already in the system. They are discussed as follows.

The Terminal Set:

The terminal set is comprised of the inputs to the GP program (features from the learning domain with which to conduct learning), the constants supplied to the GP program, and the zero-argument functions with side-effects executed by the GP program. Input, constants and other zero-argument nodes are called terminals or leafs because they terminate a branch of a tree in tree-based GP.

The Function set:

The function set is composed of the statements, operators, and functions available to the GP system. The function set may be application specific and be selected to fit the problem domain. The range of available functions is very broad. Some examples include Boolean functions, arithmetic functions, trigonometric functions, conditional statements, and logarithmic functions.

2.6.2. Fitness Function:

In a generation, every candidate solution is evaluated and scored using the fitness function, which is application dependent. The user, according to the application, defines the fitness function. The better an individual is performing at this function; the better its survival is, and thus, better is its chances of producing children for the next generation. The survival of fittest is implemented by retaining the best individuals. The rest are deleted and replaced by the offspring of the best individuals. Together (the retained ones

and the offspring) make a new generation. Some offspring may have high score than their parents in the previous generation. Offspring are produced by applying genetic operators which are discussed in the coming section.

2.6.3. Genetic Operators

An initialized population usually has very low fitness. Evolution proceeds by transforming the initial population by the use of genetic operators. In machine learning terms, these are the search operators. The three principle genetic operators are:

Crossover:

The crossover operator combines the genetic material of two parents by swapping a part of one parent with a part of the other. In fact, crossover tries to mimic recombination and sexual reproduction. It mainly helps converging onto an optimal solution.

Mutation:

Mutation operates only on one individual. Normally, after crossover has occurred, each child produced by the crossover undergoes mutation with a low probability. This small random change often brings diversity in the solution space and helps to avoid trapping in local minima/maxima. The probability of mutation is a parameter of the run. A separate application of crossover and mutation is also possible and provides another reasonable procedure.

Replication:

The replication operator is straightforward. An individual is selected. It is copied, and the copy is placed into the population. There are now two versions of the same individual in the population.

3. Attack Resistant Visual Tuning

3

"In the struggle for survival, the fittest win out at the expense of their rivals; because they succeed in adapting themselves best to their environment."

Charles Darwin

In the previous chapter, some basics related to digital watermarking and machine learning have been discussed. This chapter starts with the first phase of this research and explains the proposed Intelligent Attack-Resistant Watermarking (IARW) which is able to tune a watermark with respect to HVS and a cascade of attacks. This intelligent tuning is performed at the embedding stage. The proposed watermarking scheme utilizes GP and a DCT based watermarking approach [28]. The watermark is embedded in DCT domain with GP providing the delicate balance between robustness and imperceptibility by developing application specific VTF.

The proposed IARW technique is developed and enhanced in different steps. First, we present an initial approach which intelligently selects the appropriate strength and position of coefficients for watermark embedding. It is further enhanced with the development of a more sophisticated GP fitness function which uses the concept of wrapper for selection of appropriate DCT bands. The evolved VTF for selecting appropriate strengths and positions inside DCT bands using wrapper function as SP-VTF. The aspect of robustness is further improved, with no compromise on imperceptibility, by

the use of error correction codes (or more specifically, BCH coding). The VTF evolved by the use of BCH coding is named as SP-VTF(BCH). We use these notations in the rest of the chapter while referring to different VTFs evolved at different steps.

3.1. Attack Resistant Visual Tuning: The Basic Architecture

Figure 3.1 illustrates the basic architecture of the IARW scheme. It consists of two phases. The training phase and the testing phase. In the training phase, a number of images are used to train the GP based algorithm in order to *evolve* a suitable attack resistant VTF. Initially, a population of candidate VTFs is taken randomly in GP simulation. Each VTF is then scored using its mean fitness specified by a fitness function. The best ones are retained as parents for the next generation. The outcome of the training phase is the evolution of an optimum VTF. The evolved VTF is then used by the testing phase to validate the generalization of the VTF for real world applications. Figure 3.2 illustrates the detailed block diagram of the IARW scheme. Theoretical description and implementation details of different modules are elaborated in the following sections.

3.2. Watermark Embedding

Consider a cover image which is transformed using 8×8 block DCT transformation. In an 8×8 block DCT image, each position inside the block can be considered as a communication channel characterized by a frequency band. Thus, even a single attack on the watermarked image may affect different frequency bands differently.

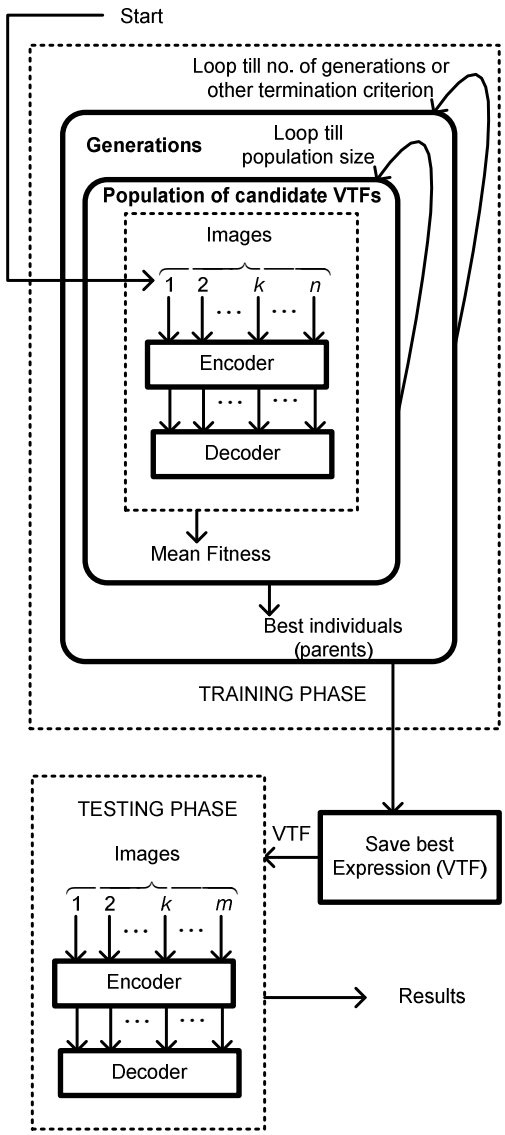


Fig.3.1 Basic Architecture of Intelligent Attack Resistant Watermarking scheme

Consequently, for watermark insertion in any given image, VTF is supposed to decide both; *which frequency band to select?,* as well as, *how much alteration we can perform for this frequency band?* If X[**k**] and Y[**k**] represents the original and watermarked images respectively, then watermark embedding is given by:

$$Y[\mathbf{k}] = X[\mathbf{k}] + W_m[\mathbf{k}], \tag{3.1}$$

where, **k** represents 2D indices, and

$$W_m[\mathbf{k}] = S[\mathbf{k}] \cdot b[\mathbf{k}] \cdot \alpha_G[\mathbf{k}], \quad (3.2)$$

is the watermark. S[**k**] is a pseudo random sequence and b[**k**] is the repetition-based expanded code vector, while α_G[**k**] is the genetic perceptual mask obtained from genetic VTF (figure 3.2). The watermark extraction is performed by assuming that the frequency bands follow Generalized Gaussian Distribution (GGD), and then using maximum likelihood based estimation, the hidden message is decoded. The embedding and extraction phases of this GGD based spread spectrum watermarking scheme have been discussed earlier in chapter 2.

3.2.1. BCH Encoding

BCH codes belong to the class of multilevel, variable length and cyclic error correction codes. They are isomorphic to Hamming codes and allow multiple random error correction. Binary BCH codes can be constructed with parameters (*n*, *k*, *t*), where *n* represent the length of the codeword, *k* is the message size and *t* represent the number of random bits this code can correct.

For any arbitrary integers $m \geq 3$ and $t < 2^{m-1}$, there exists a primitive BCH code with the parameters $n = 2^m - 1$, $n - k \leq mt$ and $d_{\min} \geq 2t + 1$. Where, $d_{\min}$ denotes the minimum Hamming distance between two vectors. This code can correct *t* random errors over a span of $2^m - 1$ bit positions. Further information on BCH codes can be found in [52]. In this work, we use BCH (31, 16, 3) encoding. It is employed before repetition coding module in the watermark embedding stage as shown in figure 3.2.

3.2.2. Strength selection

In literature [1, 11, 16], VTF should exploit the characteristics of HVS. Its functional dependencies must include the parameters for luminance sensitivity, frequency sensitivity, and contrast masking. Therefore, the functional form of a VTF is represented as follows:

$$\alpha_G(k_1,k_2) = f\left(X_{0,0},\ X(i,j), T(i,j)\right), \tag{3.3}$$

where, $X_{0,0}$ is the DC coefficient and represents dependence of VTF on Luminance sensitivity, $X(i,\ j)$ is AC coefficient and represents dependence of VTF on contrast masking, and *T(i,j)* represents frequency sensitivity. These are treated as independent variables (*terminal set*) in the GP simulation. The magnitude of α_G represents the strength of the alteration performed to the DCT coefficient. To help GP converge to an optimal VTF quickly, instead of *T(i,j),* we provide the Watson's VTF (equation 2.8) as an independent variable. Thus, the GP terminal set consists of the current value of Watson's VTF 'α', '$X_{0,0}$', and'$X(i,\ j)$'. Therefore, equation 3.3 is replaced by the following:

$$\alpha_G(k_1,k_2) = f\left(X_{0,0},\ X(i,j), \alpha(i,j)\right). \tag{3.4}$$

However, if $\xi_3 = f(\xi_2, \xi_1, X)$ denotes the information pertaining to the distortions caused by the set of three attacks performed in sequence, then, the functional dependency of the genetic VTF on both the characteristics of HVS and the distortion due to attacks can be represented as follows:

$$\alpha_G(k_1,k_2) = f\left(X_{0,0},\ X(i,j),\ \alpha(k_1,k_2),\ \xi_3\right). \tag{3.5}$$

Through its feedback and implicit learning mechanism, GP is able to learn and exploit information regarding ξ_3. Similar approach is adopted for more attacks in a sequence.

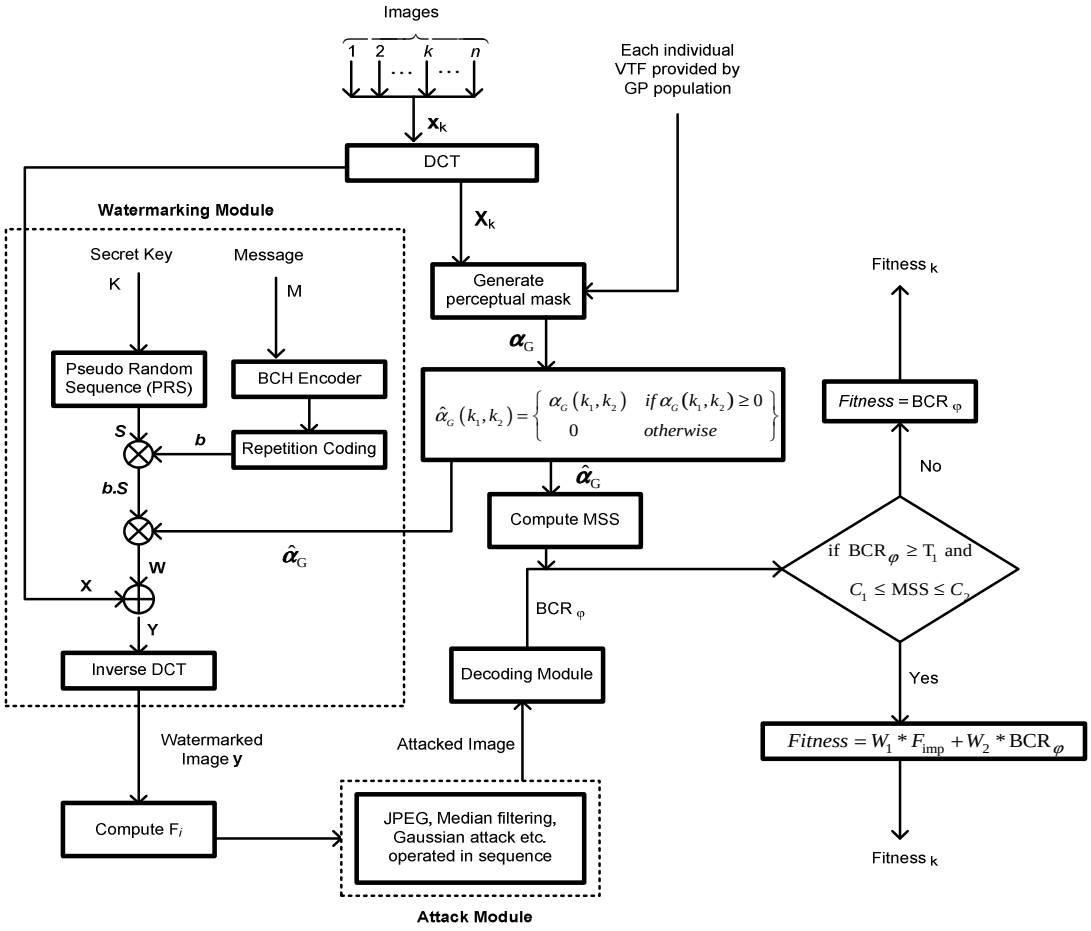


Fig.3.2 Detailed representation of encoding, decoding and fitness evaluation processes of the IARW scheme

3.2.3 Frequency band selection

Still with the provision of all the above mentioned information, optimal exploitation of both HVS and distortion caused by a set of attacks is a difficult challenge. To further help GP in developing optimal VTF, we let GP exploit frequency band information as well; both in view of HVS and distortion caused by a set of attacks. For this purpose, we modify Eq. (3.5) as follows:

$$\alpha_G(k_1,k_2) = f\left(X_{0,0},\ X(i,j),\ \alpha(k_1,k_2),\ \xi_3, i, j\right), \tag{3.6}$$

where *i, j* represent block indices.

However, we observed from our GP simulations that just by including the block indices *i,j* as independent variables does not help GP in finding optimal VTF. This was because GP had not been able to make decision efficiently as to whether the frequency band in question is suitable for watermark insertion or not. Consequently, to enable GP to perform the selection of frequency bands both in view of HVS and distortion caused by a set of attacks, we use the concept of wrapper as follows:

$$\hat{\alpha}_G(k_1,k_2) = \left\{ \begin{array}{cc} \alpha_G(k_1,k_2) & \text{if } \alpha_G(k_1,k_2) \geq 0 \\ 0 & \text{otherwise} \end{array} \right\}. \tag{3.7}$$

By doing so, GP is also able to select appropriate embedding positions in the 8×8 DCT block for optimum tradeoff in subsequent generations, in addition to selecting proper strength of alteration. It is to be noted that equations 3.5 and 3.6 only depict the functional dependencies of a VTF. Their realization appears later in equation 3.14.

3.3. Generating Candidate VTF in GP Domain

In order to carry out GP simulation to evolve an optimum VTF, one needs to define GP control parameters, initial population, GP functions and terminals, and termination criteria for the simulation. These are elaborated as follows:

3.3.1. GP Control Parameters

The control parameters in our GP simulation are the number of generations, population size, selection type, the probabilities of performing the genetic operations, the maximum size for programs, and termination criteria etc.

3.3.2. Initial Population

At the start of our GP simulation initial population of candidate VTFs is created by randomly generating trees representing mathematical expressions. These expressions cater for both the dependencies on HVS and deformities in the 2-D watermarked image signal caused by a sequence of attacks. Ramped half and half strategy is used to create this initial population.

3.3.3. Basic Functions used for GP simulation

In our GP simulations, the set of functions available to the GP system is simple functions, including four binary floating arithmetic operators (+, -, *, and protected division), *LOG, EXP, SIN, COS, MAX and MIN*.

3.3.4. Terminals used for GP simulation

To exploit the search space representing different possible forms of dependencies of the watermark shaping function, we consider genetic VTF as watermark shaping function and the characteristics of HVS as independent variables. Therefore, the current value of Watson's VTF, DC and AC DCT coefficients of 8x8 block are provided as variable terminals (equation 3.4). Also, the zigzag indices of the current DCT coefficients are used as variable terminals (equation 3.6). Random constants in the range [-1, 1] are used as constant terminals.

3.3.5. Termination Criterion

The GP simulation is ceased when the generation count reaches maximum number of generations, or when a program surpasses a threshold fitness level, or when a predefined time limit exceeds. If the termination criterion is accomplished, then the best-evolved solution is saved. Otherwise, we replace the existing population with the new population and the simulation is continued. In the implementation of our proposed technique, we have taken 100 generations as the GP termination criterion.

3.4. Watermark Extraction

In order to compute the performance of a given VTF in terms of robustness, one needs to extract the watermark after the watermarked work is presented to a sequence of different real world attacks. These attack sequences include a series of legitimate and illegitimate distortions on the watermarked work and are discussed later in detail in section 3.8. In the ASSW approach [28], it is assumed that the pdf of the original

coefficients remains the same even after embedding. Based on this idea, one needs to obtain expressions for maximum likelihood decoder structures as presented in Section 2.3. The unknown parameters *c* and σ should be estimated from the received image at the decoding stage. With the help of sufficient statistics, the hidden bits are then estimated using '0' as a threshold (section 2.3).

Once the watermark is extracted, it can be used for computing results of different objective tests for watermark robustness.

3.5. Evaluating Each Potential VTF

Once the watermark is embedded, we can calculate the values of different objective tests for imperceptibility. After these calculations, the watermarked work is presented to the attack module, where a sequence of attacks is applied. These attacks mimic the possible distortions of the real world scenarios and are application dependent. The attacked work is then used to compute the second constraint, i.e. robustness, of the optimization problem at hand. The objective measure used for robustness is BCR.

We define the *fitness function* as*:*

$$Fitness = W_1 * F_{imp} + W_2 * BCR_{\varphi}, \tag{3.8}$$

where F_{imp} is the imperceptibility related fitness and is given by:

$$F_{imp} = w_1 * SSIM + w_2(wPSNR / 46.0) \tag{3.9}$$

It represents watermark imperceptibility, whereas SSIM denotes the Structural Similarity Index Measure [39], and wPSNR represents weighted Peak Signal-to-Noise Ratio [53], of the watermarked image. $W_{[.]}$ and $w_{[.]}$ depict the corresponding weightage of the different terms used in the fitness.

Structural similarity based image quality assessment relies on the assumption that human visual system is highly adapted to extract structural information. Therefore, a measure of change in structural information can provide a good approximation for perceived distortion. The SSIM index measure between two images, **x** and **y**, is given as:

$$\mathrm{SSIM}(\boldsymbol{x},\boldsymbol{y}) = \left[l(\boldsymbol{x},\boldsymbol{y})\right]^{\alpha} . \left[c(\boldsymbol{x},\boldsymbol{y})\right]^{\beta} . \left[s(\boldsymbol{x},\boldsymbol{y})\right]^{\gamma}, \tag{3.10}$$

where, $l(\boldsymbol{x},\boldsymbol{y})$ is the luminance comparison function, $c(\boldsymbol{x},\boldsymbol{y})$ represent the contrast comparison function, and $s(\boldsymbol{x},\boldsymbol{y})$ is the structure comparison function as discussed in [39]. $\alpha > 0, \beta > 0,$ and $\gamma > 0$ are parameters used to adjust the relative importance of the three components.

The wPSNR is a modified version of PSNR that uses a texture masking function as a penalization factor and is defined as follows:

$$\mathrm{wPSNR} = 10\log_{10}\left(\frac{255}{\mathrm{RMSE}\times\mathrm{NVF}}\right)^2, \tag{3.11}$$

where, RMSE is the root mean squared error and NVF is the Noise Visibility Function proposed by Voloshynovskiy *et al.* [53], and is an arbitrary texture masking function. The value for *NVF* ranges from 0, for extremely textured areas, and up to 1, for clear smooth areas of an image.

A fair enough quality of the resultant watermarked image in terms of wPSNR that we can expect is approximately 46.0 db, while the maximum value that SSIM can achieve is 1.0. As a result, in $\mathrm{Fitness}_{\mathrm{imp}}$, we divide the wPSNR by 46.0 in order to scale

its value to 1.0 as well. If W_1 and W_2 are set to 1.0, while w_1 and w_2 are set 0.5 each, the fitness (Eq. (3.8)) attains a maximum value near to 2.0.

We assume watermark power to depict robustness at the embedding stage of GP simulation. For this purpose, we use MSS as an objective measure representing estimated robustness of individuals in a GP population given by:

$$\mathrm{MSS} = \frac{1}{N_b N_c} \sum_{x_1=1}^{N_b} \sum_{x_2=1}^{N_c} \alpha(x_1, x_2)^2 \ , \tag{3.12}$$

where, N_b is the total number of 8×8 blocks in the cover image and N_c is the number of selected *DCT* coefficients in a block. x_1 and x_2 are the respective indices.

On the other hand, BCR_φ representing Bit Correct Ratio; after the set of conceivable attacks are carried out, is given as:

$$BCR(M, M')_\varphi = \sum_{i=1}^{L_m} \overline{(m_i \oplus m_i')} \Big/ L_m \ , \tag{3.13}$$

where, M represents the original, while M' represents the decoded message; L_m is the length of the message and $\oplus$ represents exclusive-OR operation. It should be noted that $(1 - BCR)$ represents bit incorrect ratio.

Thus, each individual genetic VTF of a GP population is scored using Eq. (3.8) as a fitness function. The greater the fitness is, the better the individual has performed. Figure 3.2 illustrates the detailed representation of encoding, decoding and fitness evaluation modules of the proposed scheme, as indicated by the innermost block of the training phase in figure 3.1. It is elaborated as follows. Let

$S = \{x_k = (i_k, j_k);\ k = 1, \ldots\ldots, n\}$ represent the training set with n being the number of training images. We consider a function space $\mathcal{F}$ consisting of VTFs in the form of Eq. 3.6 to represent the decision space. Note that information regarding a set of attacks (ξ_3) is implicit and learned through the repeated feedback mechanism of GP. An initial random population $\mathcal{P}_1$ of size z is created such that $\mathcal{P}_1 \subset \mathcal{F}$. We represent an individual from this population as β_t, where $t = 1, \ldots\ldots, z$. The algorithm for training a GP population is as follows:

FOR each $\beta_t : t = 1, \ldots\ldots, z$, *and* $\beta_t \in \mathcal{F}$

FOR each image $x_k : k = 1, \ldots\ldots, n$ *and* $x_k \in S$

Step 1. Compute 8×8 block DCT

Step 2. Compute perceptual mask $\alpha_{t,k}$ (Eq. (3.7))

Step 3. Embed message M using $\alpha_{t,k}$ (Eq. (3.2))

Step 4. Compute MSS (Eq. (3.12))

Step 5. Compute F_{imp} (Eq. (3.9))

Step 6. Perform the selected sequence of attacks

Step 7. Perform decoding to retrieve the decoded message M'.

Step 8. Compute BCR_φ (Eq. (3.13))

Step 9. Calculate fitness based on the following rule

IF $BCR_\varphi \geq T_1$ *and* $C_1 \leq MSS \leq C_2$

$Fitness_{t,k} = W_1 \cdot F_{imp} + W_2 \cdot BCR_\varphi$

ELSE

$Fitness_{t,k} = BCR_\varphi$

END

END

Step 10. Compute average fitness score for β_t using $Fitness_t = \sum_{k=1}^{n} Fitness_{t,k} \Big/ n$

END

At the end of the GP simulation, the best expression, or the evolved VTF, is saved. It is used in the testing phase in order to watermark a number of test images from the test data set as illustrated in figure 3.1.

3.6 Implementation details

Simulations for the proposed IARW technique are carried out in MATLAB. To employ GP, we have used MATLAB-based GPLAB toolbox version 2.1 [54]. Initial population in GP simulation is generated through Ramped Half-and-Half method. Initial operator probabilities are set to 'variable', while the selection method for reproduction is set to 'tournament'. The remaining parameters used are as default in the software and summarized in table 3.1. The proposed technique is implemented on Intel Core 2 Duo 2.4 GHz processor, and on the average, each training simulation takes 6-8 hours.

Ten images, such as, Lena, Baboon, Trees, Couple, etc., of size 256x256 are used as cover images during the training phase of GP simulation. These images are shown in figure 3.3. The fitness of each GP individual is actually the average of the fitness corresponding to all training images. The number of bandpass DCT coefficients within an 8x8 block, N_d is set to 25 (5 to 30 in zigzag order). Message size is kept equal to 64 bits. To assign bonus fitness (first term on the right hand side of equation 3.8), we have used T_1, C_1, C_2 , W_1 and W_2 as 0.97, 1.0, 120.0, 1.0 and 1.0 respectively. w_1 and w_2 are set 0.5 each. The values of T_1, C_1 and C_2 are set empirically. Specifically the values of C_1 and

C_2 are set by observing the MSS values of Watson's VTF for the training images, while that of T_1 is set in view the severity of the attacks for a given watermark application. W_1 and W_2, on the other hand are set to dictate the overall fitness dependence of an individual solution on imperceptibility and robustness respectively. w_1 and w_2 weight parameters can be adjusted by the user so as to which imperceptibility measure, wPSNR or SSIM, should be given more importance. MATLAB based BCH encoding and decoding modules are used for employing BCH(31,16, 3) codes.

Table 3.1 GP Parameters Settings

Objective:	**To perform intelligent watermark structuring**
Function Set:	+, -, *, protected division, *SIN, COS, LOG, MAX* and *MIN*
Terminal Set:	Constants: *random constants in range* of [-1, 1] Variables : $X_{0,0}/1024,\ \lvert X(i,j)\rvert,\ \alpha(k_1,k_2), i, j$
Fitness :	$\text{Fitness} = (W_1 * \text{SSIM} + W_2 * \text{wPSNR}/46) + \text{BCR}_\varphi$
Selection:	Generational
Population Size:	160
Initial max.Tree Depth	8
Initial population:	Ramped half and half
Operator prob. type	Variable
Sampling	Tournament
Expected no. of offspring method	rank89
Survival mechanism	Keep best
Real max level	30
Termination:	Generation 40

In the testing phase, for each of the test image, grid search with a step of 0.001 is applied to find the watermark strength needed to produce a resultant image of same image distortion in terms of SSIM measure. A total of about 10 GP simulations for each set of conceivable attacks are carried out. The best expression of all the simulations is selected for the test phase. In the testing phase, the watermarking scheme using the best-evolved genetic VTF spends about 5 sec to watermark a single image.

3.7 Potential applications of the IARW approach

In this section we discuss some of the potential applications of the proposed GP based IARW scheme. The first potential application is in medical information handling and sharing, with applications ranging from cooperative working sessions, telediagnosis to telesurgery [55]. The high security risks involved in medical data such as Electronic Patient Records (EPRs), and concomitantly, a higher desired visual quality such as in telediagnosis and telesurgery, urges for the use of intelligent watermarking systems to embed the watermark in the most efficient manner. In applications such as these, a high capacity watermark is not always functional to produce high robustness. Rather, exploitation of the hidden dependencies in HVS in respect of the robustness and imperceptibility is a more feasible option.

The next generation of digital cameras is taking the photographer's copyright to a level further by acquiring and embedding biological information of a photographer in a digital image. One of the examples include a recent patent application by Cannon [56] which reveals the use of a photographer's iris information for watermarking digital images. In applications like these, a high imperceptibility is required since the watermark

is being embedded at the image capturing stage, along with no compromise on robustness later on. Our proposed watermarking technique may be implemented in the camera hardware and a pre-stored user profile based on iris information may be utilized to watermark batch files after each photographic session.

Similarly, the proposed adaptive visual tuning scheme may be deployed on a chip for broadcast monitoring and device control after generating appropriate VTFs in view of the application environment, and hence the conceivable attacks.

E-democracy [57], Mobile context aware medical networks [58] and mobile context aware multimedia stories [59] are some of the recent trends in mobile usage. The use of intelligent watermarking is highly desirable in applications where imperceptibility and robustness requirements constantly change. Furthermore, it is worthwhile in applications where high robustness is required with no compromise on imperceptibility. An appropriate example of such an application could be watermarking a database of faces in a face recognition system [60].

3.8 Experimental results and discussion

In the proposed IARW scheme, during the training phase, a random population is created with each individual representing potential VTF using Eq. 3.6. Every candidate is evaluated with a fitness function based on objective measures for watermark robustness and imperceptibility. VTFs are evolved using 10 standard training images as displayed in figure 3.3. Their performance is tested in the light of a typical transform domain watermarking scheme [28], by using 300 test images from everyday life. In order to demonstrate its effectiveness, we consider two attack sequences: (A); the watermarked

image is communicated through a channel characterized by Gaussian noise ($\sigma = 20$). The received image is operated upon with a lowpass filter to reduce noise and its intensity range is scaled to 0-255, representing a type of value-metric attack. (B); a watermarked image is first compressed to reduce its size. A malicious attacker somehow manages to perform median filtering, before it is transmitted via a channel characterized by noise. These attack sequences are displayed in figure 3.4.

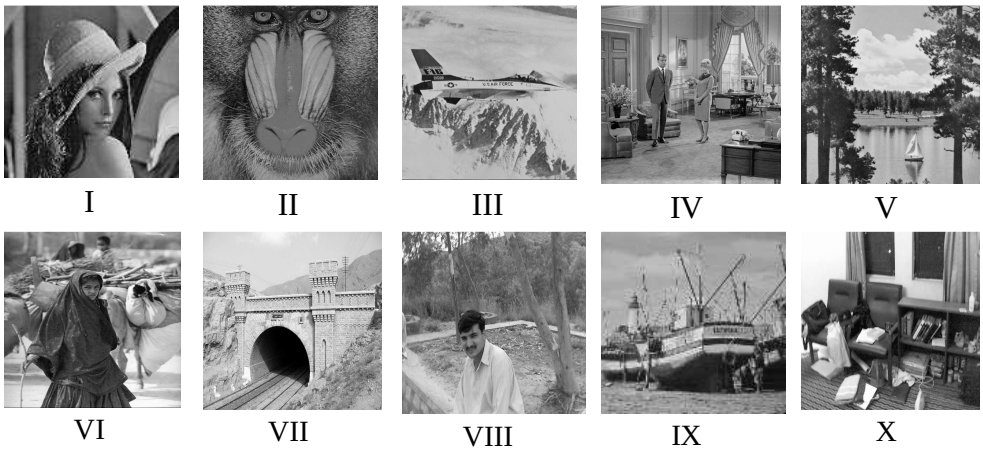


Fig.3.3 Training images used during GP simulation

The best evolved expression representing the VTF for attack sequence B is as follows:

$$\alpha_G(k_1, k_2) = Max\left(a\ , log\left(b * c\right)\right) * X_{0,0} * cos(X_{0,0}), \tag{3.14}$$

where,

$$a = Max\left(j\ , \alpha\left(k_1 * k_2\right)\right) - j, \tag{3.15}$$

$$b = -sin(0.34168)\ exp\ log(j), \text{ and} \tag{3.16}$$

$$c = X(i, j) - exp\left(\alpha\left(k_1 * k_2\right)\right). \tag{3.17}$$

First we discuss the performance of GP based watermark structuring alone. Then, we discuss robustness comparison using combination of GP based watermark structuring and BCH coding.

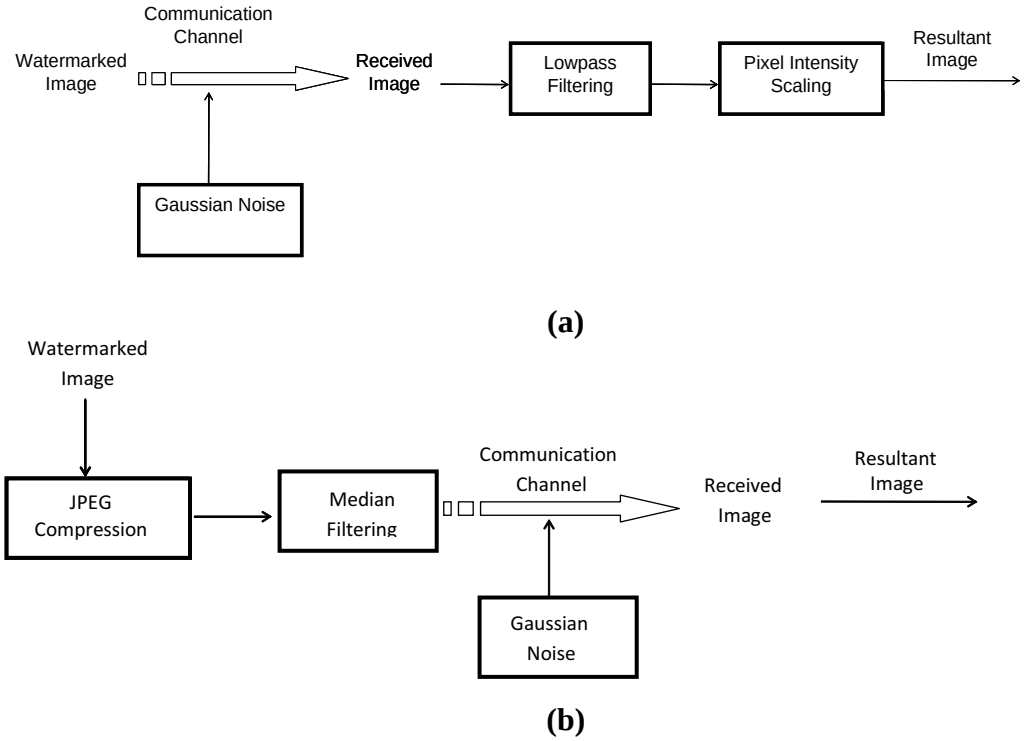


Fig.3.4 Attack sequences on the watermarked image representing (a) Attack Scenario A, and (b) Attack Scenario B.

3.8.1. Improving Visual Quality by Exploiting HVS

Figure 3.5 shows the watermark structuring ability of the proposed scheme. It demonstrates that almost invisible distortion is introduced by embedding a high energy watermark as a result of using intelligent selection of the right frequency band and strength of embedding. The difference image between the original and the watermarked image depict that the proposed scheme is able to learn the content of the cover work, as mostly, high strength embedding is performed in high textured areas. On the other hand,

the severity of the distortion a watermarked image may suffer after a series of attacks are presented is also shown in figure 3.5.

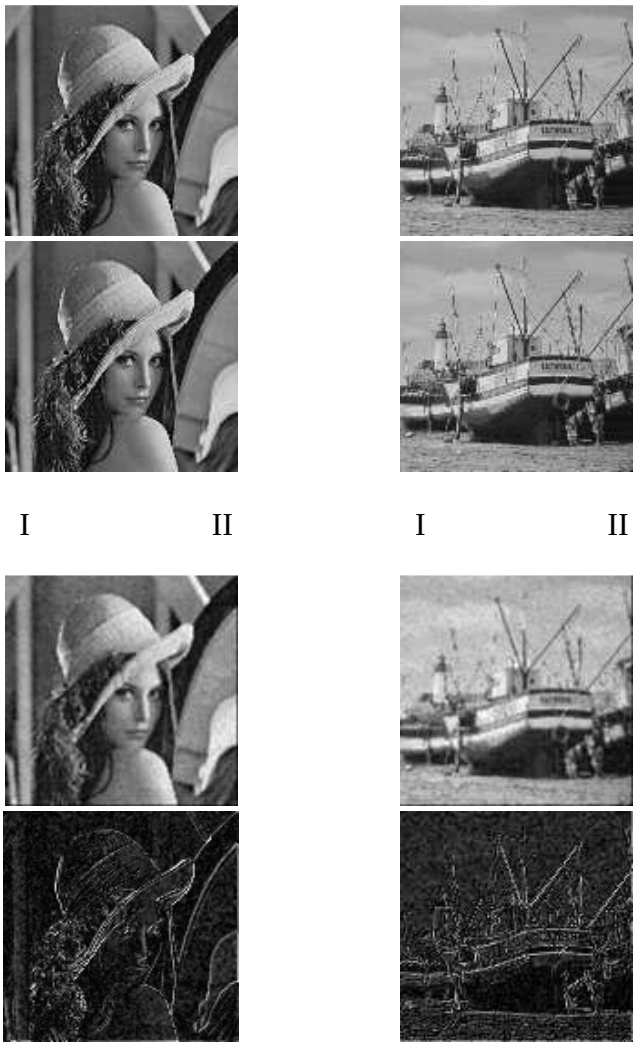


Fig.3.5 Watermark Structuring: I.original image, II. watermarked image, III. attack sequence B applied on watermarked image, and IV. scaled difference image of I and II for (a) Lena image and (b) Boat image.

Figure 3.6 illustrates the distribution of watermark strength in frequency domain for Lena image and demonstrates the ability of the genetic VTF to select suitable frequency bands in view of the hostile series of attacks. In addition, the strength of alteration performed to a selected DCT coefficient for watermark embedding, increases as we move from low to high frequency in zigzag scan order. This attribute is in line with the sensitivity of a human eye.

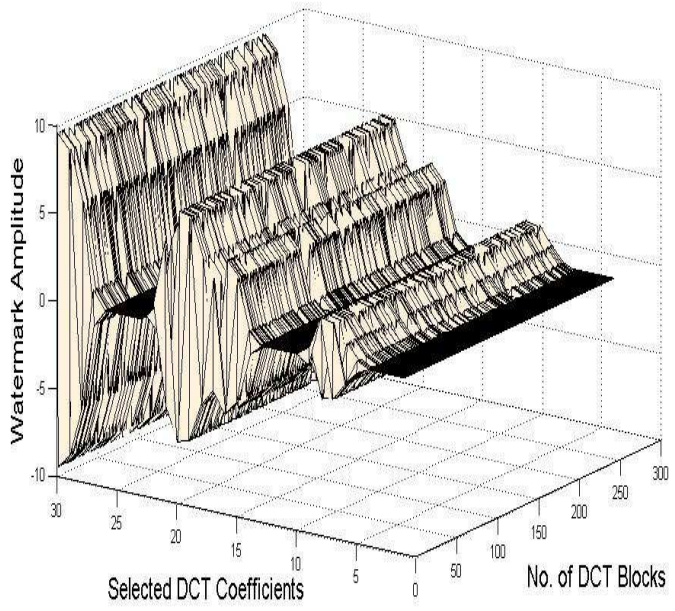


Fig.3.6 Watermark strength distribution in frequency domain for Lena image

Table 3.2 shows average performance of different objective measures on ten training images for the attack sequences A and B respectively. In case of both the attack sequences, it is noticeable that the proposed scheme outperforms the conventional ones in terms of visual tuning by achieving improvement in *SNR*, *wPSNR* and Mean Squared Error (*MSE*) as compared to other approaches. Note that the *SSIM* values are kept the same for all the schemes by using an appropriate scaling factor, since our primary objective is to achieve a considerable improvement in robustness.

Table 3.2 Performance comparison in terms of average robustness and imperceptibility of the proposed technique (SP-VTF; based on strength and position optimization) with that of Watson's VTF, and the VTF obtained by implementing [16](S-VTF; based on strength selection only). Note that each value corresponds to the average value of 10 training images.

Attack Scenario	Watermarking Scheme	MSS	Mean Squared Error	Signal to Noise Ratio	Document to Watermark Ratio	wPSNR	SSIM	$Fitness_{imp}$	BCR
Attack Scenario A	SP-VTF (BCH)	5.5076	5.4861	34.6201	34.6184	44.6251	0.9839	0.9770	1.0
	SP-VTF	11.4636	5.3550	34.7115	34.7099	44.6626	0.9829	0.9770	0.9750
	S-VTF	16.1460	6.7888	33.1260	33.5110	43.7830	0.9817	0.9667	0.8825
	WPM	18.1569	8.4799	32.7520	32.7493	44.0671	0.9838	0.9709	0.8375
Attack Scenario B	SP-VTF (BCH)	25.1470	25.0648	28.3311	28.3242	38.7891	0.9520	0.8986	1.0
	SP-VTF	47.7742	22.3002	28.5645	28.5582	39.2461	0.9510	0.9010	0.9937
	S-VTF	60.4053	28.30639	28.8711	28.8633	38.4536	0.95187	0.8938	0.8219
	WPM	60.3381	28.1774	27.6236	27.6157	38.9333	0.9519	0.89913	0.70625

If we have a closer inspection of all the results presented in table 3.2, we observe that for every technique, the imperceptibility related fitness measure is almost the same. Whereas, the last column depicting the *BCR* values of the training data shows sizable improvement. Both for the attack scenario A and B, *BCR* values are 1 when the proposed technique is applied in conjunction with BCH coding. It is also noticeable that high strength embedding does not always imply a high degree of robustness. The *MSS* values corresponding to the proposed technique and other techniques verify this aspect.

Table 3.3 demonstrates the comparison of the IARW scheme with other intelligent schemes proposed in [17] and [18] in terms of correlation values between the

original and the watermarked images for attack scenarios A and B respectively. These correlation values serve as objective test for imperceptibility. The correlation, C, between two images X and X' is given by:

$$C = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{i,j} - \bar{X})(X'_{i,j} - \bar{X}')}{\sqrt{\left(\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{i,j} - \bar{X})^2\right)\left(\sum_{i=1}^{m}\sum_{j=1}^{n}(X'_{i,j} - \bar{X}')^2\right)}} \tag{3.18}$$

Where, $\bar{X}$ and $\bar{X}'$ are the mean values of X and X' respectively of size $m \times n$. The values of C ranges from -1 to 1. In the table, these values are averaged for the training images and the three hundred test images. It can be observed that the IARW scheme performs better in terms of preserving the perceptual quality after watermark embedding. The scheme presented by Shieh et al. [17] employs the optimal location selection, whereas the one proposed by Aslantas et al. [18] employs the optimum strength through intelligent search. In contrast, our proposed scheme employs both the optimum location and the optimum strength, which thus yields improvement in the results in table 3.3. For attack scenario A Shieh et al.'s technique gives and average value of 0.9989 for training data, which is superior to the value given by the proposed technique (0.9976). But, the proposed technique outperforms Shieh et al.'s technique on testing data by quiet a margin. With a correlation value of 0.9946 on 300 test images from everyday life, compared to Shieh et al.'s 0.9741, the proposed watermarking system could be very useful for practical applications. The margin of enhancement depicts the usefulness and generalization ability of the proposed technique for real world applications.

Table 3.3 Performance comparison in terms of correlation values between the original image and the watermarked image for the proposed technique (SP-VTF) with that of Aslantas et al.' [18] and Shieh et al.'s [17] techniques.

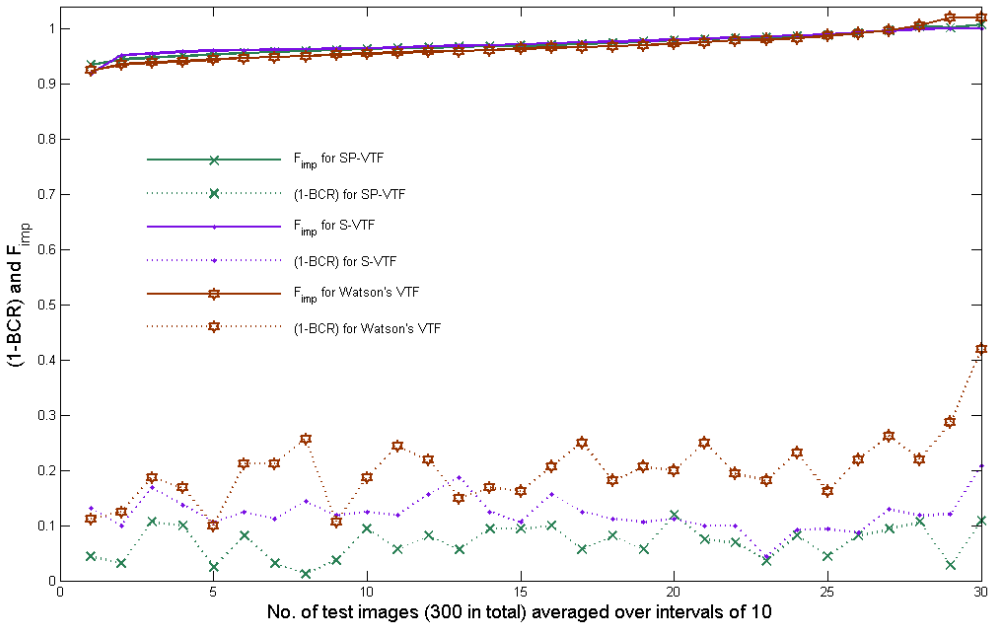
Attack Scenario	Watermarking Scheme / Images	SP-VTF (BCH)	Aslantas et al.'s [18]	Shieh et al.'s [17]
Attack Scenario A	Training	0.9976	0.9862	0.9989
	Testing	0.9946	0.9503	0.9741
Attack Scenario B	Training	0.9990	0.9877	0.9967
	Testing	0.9980	0.9786	0.9833

3.8.2. Robustness comparison

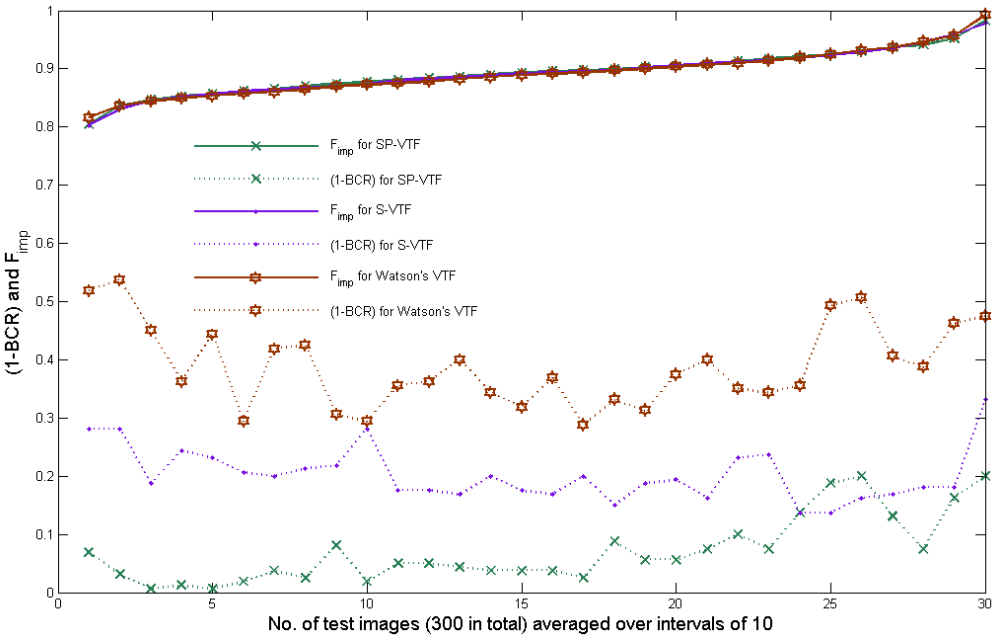
From the values of BCR in table 3.2, it is evident that IARW technique demonstrates sizable improvement in robustness in the training data, besides lower comparative values of MSS and DWR. This shows that high strength embedding and high DWR are not indispensable for high robustness. Rather, selection of appropriate frequency bands for embedding is also vital in achieving superior robustness, as in our proposed scheme. In addition, employment of error correction along with suitable strength and position selection, (SP-VTF(BCH) in table 3.2) enhances robustness further as explained in section 3.8.3. Figure 3.7 shows comparison of the proposed scheme with that of the existing approaches on 300 test images. We first sort the test images according to their watermark imperceptibility measures and then, compare their robustness measures. Keeping imperceptibility almost the same, our proposed scheme outperforms the existing schemes in terms of low bit in-correct ratio (1-BCR) and thus, offers high resistance against a sequence of attacks.

Figure 3.8(a) and (b) shows the comparison of the proposed technique with other intelligent techniques presented in [17] and [18] for both attack scenarios A and B respectively. These correlation values are between the original watermark and the extracted watermark and are averaged over intervals of ten in case of test images. Note that during the training phase of [17] and [18], we have used a sequence of attacks instead of single attacks. However, the substantial improvement in results is due to the intelligent dual search for location and strength, incorporated in our proposed scheme.

The only drawback with the proposed scheme that we have observed is the high computational time required to reach an optimal solution using GP. In our case, with Pentium Core 2 Duo machine (2.4 GHz and 2 Gb RAM) and Generations, G=90, and Population size, P= 120, it takes about 2 hours to develop an optimal solution. However, as recently proposed [61], parallel implementation of GP is becoming highly effective. As compared to conventional GP, the simulation time with parallel implementation could almost be reduced to 1/P. This certainly, raises the chances of using GP for online watermarking applications, such as protecting the transmission of medical data from mobile medical units to main hospitals or print-to-web technology.

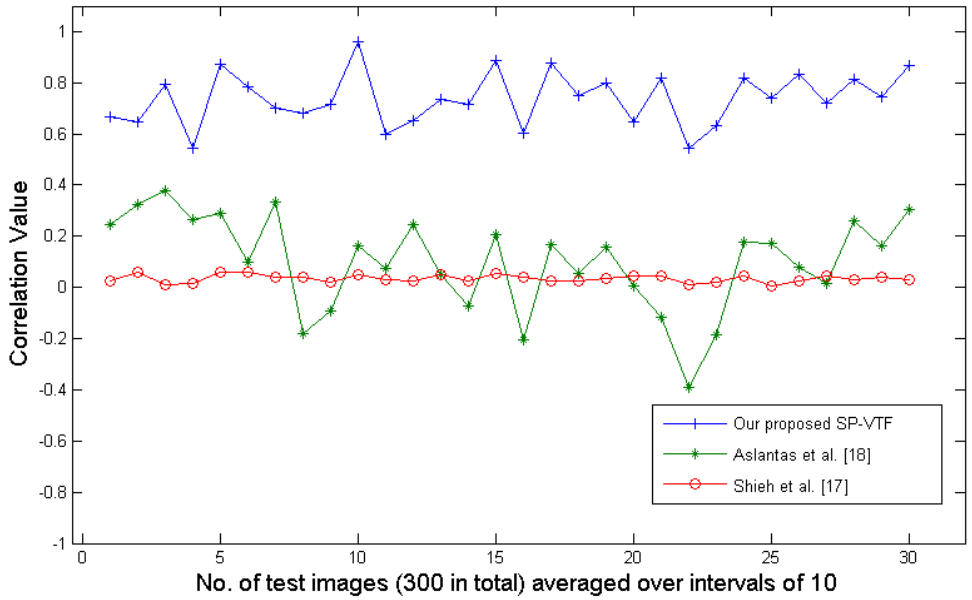


(a)

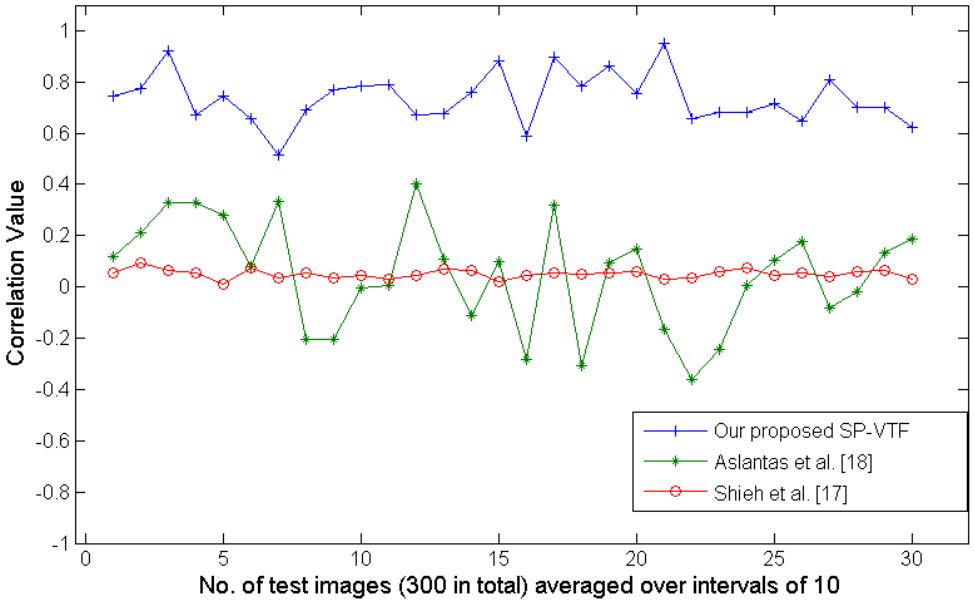


(b)

Fig.3.7 (1-BCR) and Fitness due to imperceptibility comparison of the proposed technique (SP-VTF; based on strength and position optimization) with that of Watson's VTF, and the VTF obtained by implementing [16] (S-VTF; based on strength selection only) for (a) attack sequence A, and (b) attack sequence B.



(a)



(b)

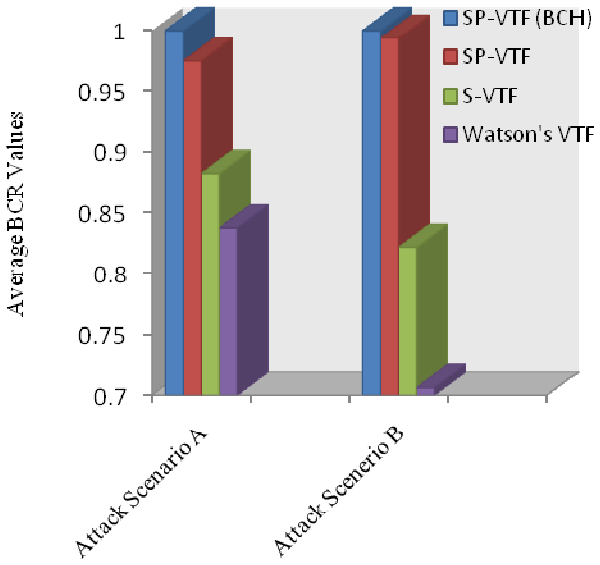
Fig.3.8 Performance comparison of the proposed technique in terms of correlation values between the original watermark and the extracted watermark, with that of Aslantas et al. [18] and Shieh et al. [17] for (a) attack sequence A, and (b) attack sequence B.

GP has another advantage of being considered as the most prospective machine learning approach capable of receiving pixel intensities directly from camera and performing some pattern recognition tasks with the output being valuable attributes [62]. This attribute of GP makes our proposed approach very effective in case of secure digital camera and wireless applications.

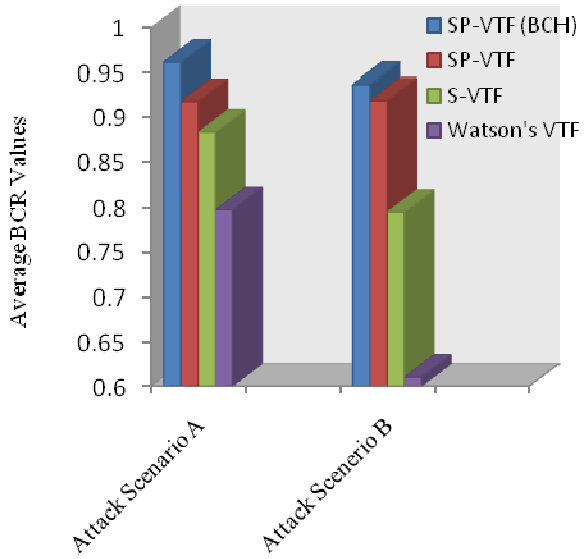
3.8.3. GP based visual tuning in conjunction with BCH coding

One of the main advantages of IARW scheme is its flexibility to incorporate different error correction strategies in order to further enhance robustness. For this purpose, we use BCH coding as explained in section 3.2.1. Experimental results demonstrate that the employment of error correction in conjunction with intelligent strength selection and frequency band selection during GP simulation, improves robustness as illustrated in figure 3.9.

Comparison of the average BCR values for both the training data (figure 3.9 (a)) and the testing data (figure 3.9 (b)) exhibit the strength of our proposed watermarking system in providing high robustness with no compromise on imperceptibility.



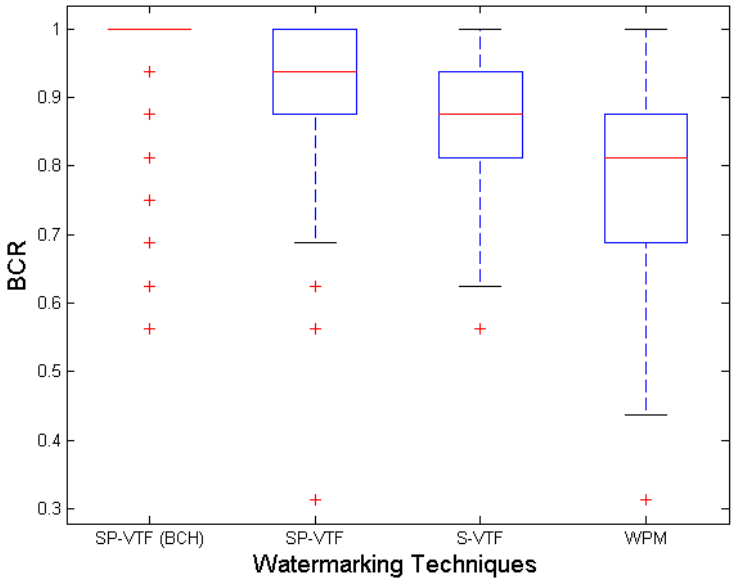
(a)



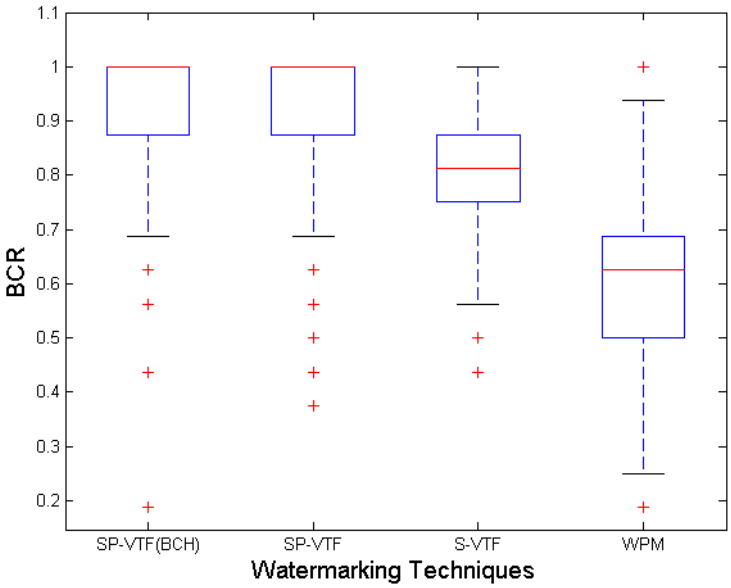
(b)

Fig.3.9 Robustness comparison of GP based visual tuning in conjunction with BCH coding for (a) training data (10 images), and (b) testing data (300 images). Note that there is sizable improvement in BCR values after employing BCH coding in our proposed visual tuning approach.

The box-and-whisker diagrams in figure 3.10 (a) and (b), corresponding to attack scenarios A and B respectively, shows the performance of the proposed watermarking system in terms of its generalization. Spread of the data comprised of 300 images and is shown for different perceptual shaping functions in context of BCR values. For attack scenario A, the inter quartile range for SP-VTF (BCH) is equal to 1 with some scattered values or outliers. Most of the outliers lie within the inter quartile range of SP-VTF, S-VTF and WPM. Similar behavior is observed for attack scenario B, where the inter quartile range for both SP-VTF (BCH) and SP-VTF includes the maximum admissible value of 1. It is observable that approximately 75% of the data lies within the BCR range of [0.88, 1], with median lying at the value of 1. These values confer a sizable improvement in the robustness of a large data set as compared to S-VTF and WPM, and hence favor the generalization ability of our proposed technique. The improvement after employing BCH coding can be observed in figure 3.10(a), where this range is shifted to the value of 1 for approximately all of the data.



(a)



(b)

Fig.3.10 Box-and-whisker diagrams corresponding to different VTFs for (a) attack scenario A, and (b) attack scenario B.

3.8.4. Robustness-imperceptibility versus message size

Figure 3.11 presents a comparison of watermark robustness and imperceptibility with respect to the message size using the best evolved VTF for attack scenario A. It is to be noted that the proposed technique produces the same watermark imperceptibility regardless of the message size. This is because the evolved VTF has learnt the embedding positions and the watermark strength to embed in those positions in view of HVS. On the other hand, robustness decreases with a sufficient increase in message size. This is due to the fact that an increase in the message bits decreases the amount of repetition, and thus, degradation in decoding and error correction performance; which leads to low BCR

values. Nevertheless, the robustness produced is adequately higher when compared to conventional methods using similar message sizes (section 3.8.2 and 3.8.3).

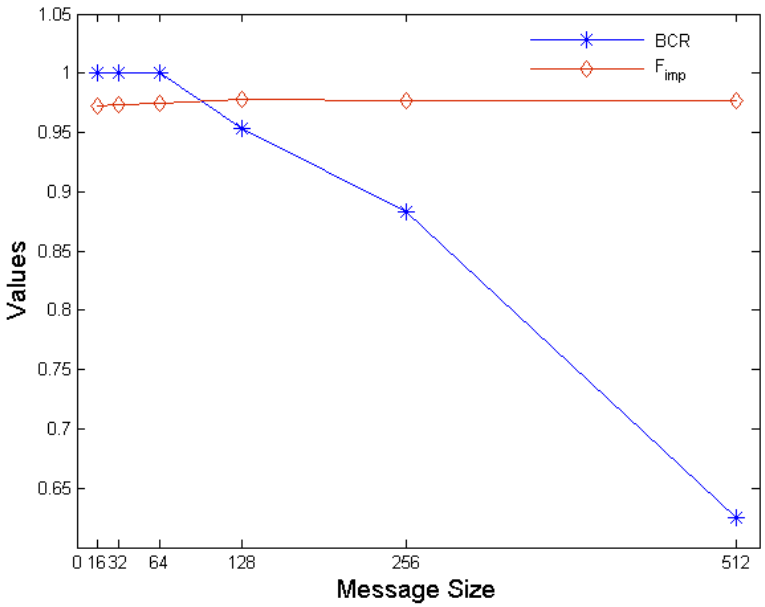


Fig.3.11 Watermark robustness and imperceptibility plot with respect to message size for attack scenario A

3.9. Summary

This chapter discusses IARW scheme capable of adaptively structuring a watermark in DCT domain. The developed VTFs are able to learn and exploit the frequency content of a given image. It has shown substantial improvement in robustness, even in applications, where series of attacks are anticipated. Applying an error correction strategy during GP simulation, such as BCH coding, further improves robustness. It is easy to implement and is applicable to all watermarking schemes that exploit human visual system. In addition, if the developed VTFs are not made public, the proposed approach becomes more secure towards unauthorized decoding. The proposed approach

thus has prospective applications in context aware networks, print-to-web technology, and medical image watermarking, where high robustness at low power embedding is extremely desirable.

4. Intelligent Embedding and Extraction: A New Paradigm in Watermark Encoding and Decoding Structures

4

"Only the most foolish of mice would hide in a cat's ear,

but only the wisest of cats would think to look there."

Andrew Mercer

4.1. Introduction

The need and advantages of intelligent watermark embedding in view of HVS and a number of attacks, as expected in real world applications, have been discussed in chapter 3. Most of the decoding techniques proposed in literature [16-18, 28], make use of fixed watermark extraction strategies which are not adaptive in dynamic applications of watermarking, where different attacks are expected at different instances. In addition, most of the existing watermark extraction strategies [26, 27] do not consider the existence of attacks during the training phase and thus are not adaptive. Similarly, watermarking approaches [28, 29] that do not exploit Computational Intelligence, generally use simple TD, and thus, are also not adaptive as far as the attack on the watermark is concerned. These approaches neither consider the alterations that may incur to the features nor exploit the individual frequency bands; rather treat all the frequency bands collectively. Consequently, these approaches use a single feature for extraction of a message bit. A

more sophisticated approach is to intelligently embed and decode hidden messages in view of HVS as well as a sequence of attacks. This chapter describes such a technique in detail and discusses our proposed Attack Adaptive Watermarking (AAW) scheme, whereby, we implement CI based embedding as well as decoding of the watermark.

Figure 4.1 shows the basic architecture of the proposed AAW scheme. It comprises of two main modules which constitute the encoding and decoding sides. At the encoding side, an optimal Perceptual Shaping Function (PSF) is evolved both in view of HVS and a sequence of attacks. Once optimal genetic PSF is developed during the training phase, the GP training phase is stopped. The training phase of the machine learning based decoding starts at the decoding side and the evolved genetic PSF is then used for feature extraction. We have selected the ASSW scheme [16, 28] (section 2.3-2.5), as a baseline approach. First we discuss the performance measures related to both imperceptibility and robustness. We elaborate the proposed genetic based watermark embedding in section 4.2, and then describe the machine learning based adaptive decoding scheme in section 4.3.

4.1.1 Performance Assessment

For comparing performance of watermarking systems, a set of general objective measures are defined. In this regard, robustness is generally measured in terms of BCR [8, 9], bit error rate and false alarm probability. Whereas, imperceptibility of a watermark is usually measured in terms of SSIM [39], wPSNR [38], and WDR [40].

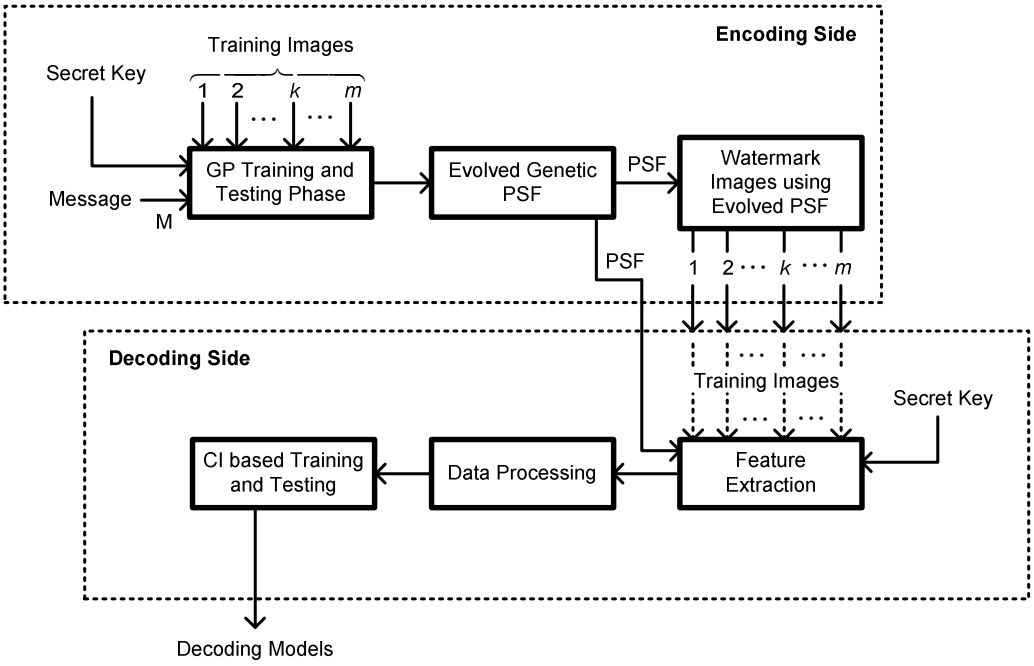


Fig.4.1 Basic Flowchart of the proposed AAW scheme

For performance evaluation of each individual of a GP population, a fitness function is defined based on the above stated set of objective measures for watermark robustness and imperceptibility. The fitness evaluation is based on how well are the imperceptibility and BCR values. The numerical rating of the fitness function is used to rank each individual of the population. The greater the fitness is, the better is the individual performance.

$$Fitness = W_1 \cdot Fitness_{imp} + W_2 \cdot BCR_{attack} \tag{4.1}$$

where,

$$Fitness_{imp} = w_1 * SSIM_{E.S} + w_2(wPSNR / 46.0) \tag{4.2}$$

represent watermark imperceptibility measures. W_1, W_2, w_1, and w_2 represent the corresponding weights of different measures used in the fitness. $SSIM_{E.S}$ denotes the

structural similarity index measure of the marked image at a certain level of estimated robustness. Details on *SSIM* and *wPSNR* are discussed in section 3.5. BCR_{attack} , is the second objective of the fitness function and represents BCR after the sequence of attacks are carried out in sequence. It is also defined in detail in section 3.5.

4.2. GP based Watermark Embedding Phase

The training phase is based on GP search mechanism, which uses a population of functions to represent a generation and enhances, or evolves, the solution generation by generation. To exploit the solution space; representing different possible forms of dependencies of the watermark shaping on the characteristics of HVS, we consider PSF as a candidate function and the characteristics of HVS as independent variables. Before representing a candidate function with a GP tree, one needs to define suitable *GP function set, GP terminal set* and the *fitness* criteria according to the optimization problem. The GP function set comprises of simple functions including four binary floating arithmetic operators (+, -, *, and protected division), *LOG, EXP, SIN, COS, MAX and MIN*. The GP terminal set consists of the current value of Watson's PSF-based perceptual mask, DC and AC DCT coefficients of 8×8 block, and the coefficient indices *(i,j)* of the current DCT coefficients as variable terminals. Random constants in the range [-1, 1] are used as constant terminals.

4.2.1. Selection of optimal strength of Alteration:

Let A denote the information pertaining to the distortions caused by the set of conceivable attacks. The functional dependency of the PSF on the characteristics of HVS and a cascade of conceivable attacks can be represented as follows:

$$\alpha_{GeneticPSF}(k_1,k_2) = f\left(X_{0,0},\ X(i,j),\ \alpha(k_1,k_2),\ i,j,A \right) \tag{4.3}$$

where, $X_{0,0}$ is the DC DCT coefficient, $X(i,j)$ is the AC DCT coefficient of the current block. They represent the luminance sensitivity and contrast masking characteristics of HVS respectively, whereas, frequency sensitivity of HVS is incorporated in $\alpha(k_1,k_2)$. Whereas, i and j represent the indices of the current AC coefficient corresponding to the selected bandpass DCT coefficients N_c.

The information regarding the position of a coefficient inside a block DCT is thus, provided through the indices for exploitation by the GP search mechanism. This helps the GP search mechanism to find appropriate frequency bands as per the new attacks mounted on the watermarking system. On the other hand, the dependence of an evolved PSF on A is not explicit. Rather, it is learned during the GP training phase, which repeatedly feedbacks the performance regarding the set of conceivable attacks.

4.2.2. Selection of Suitable frequency band:

We use GP to search for appropriate strength of alteration as well as suitable location selection in different DCT coefficients. To help GP make decisions of whether to perform or not perform embedding, we use the concept of wrapper. By doing so, GP learns appropriate positions in the 8×8 DCT block for optimum tradeoff in subsequent

generations. Position of coefficients inside 8×8 block actually represent different frequency bands. Thus, the proposed GP based embedding phase incorporates searching of suitable frequency bands and allowable alterations in view of both the anticipated attacks and HVS. Using the concept of wrapper, equation 4.3 is modified as follows:

$$\hat{\alpha}_{GeneticPSF}(k_1,k_2)=\begin{cases}\alpha_{GeneticPSF}(k_1,k_2) & if\ \alpha_{GeneticPSF}(k_1,k_2)\geq 0\\ 0 & otherwise\end{cases} \tag{4.4}$$

To incorporate generalization and unbiased behavior in our training phase, we score and check the performance of every individual on a set of training images. The average of these scores is then taken as the final score of a particular individual. Let $S=\{X_n=(i_n,j_n);\ n=1,\ldots\ldots,N\}$ represent the training set with N being the number of training images. We consider a function space $\mathcal{F}$ consisting of PSFs in the form of equation 4.3 to represent the decision space. An initial random population $\mathcal{P}_1$ of size z is created such that $\mathcal{P}_1\subset\mathcal{F}$. We represent an individual from this population as β_t, where $t=1,\ldots\ldots,z$. Algorithm 4.1 presents the steps required for a generation of genetic watermark embedding, which is explained and justified in what follows.

Algorithm 4.1: Training algorithm of a GP generation for the intelligent embedding phase of the proposed AAW scheme.

FOR each $\beta_t: t=1,\ldots\ldots,z$, *and* $\beta_t\in\mathcal{F}$

FOR each image $\mathrm{X}_n: n=1,\ldots\ldots,N$ *and* $\mathrm{X}_n\in S$

Step 1. Compute 8×8 block DCT

Step 2. Compute perceptual mask $\alpha_{t,n}$ (equation 4.3 and 4.4)

Step 3. Embed message M using $\alpha_{t,n}$ (equation 2.9 and 2.10)

Step 4. Compute MSS using equation 2.1

Step 5. Compute $Fitness_{imp}$ (equation 4.2)

Step 6. Perform the selected sequence of attacks

Step 7. Perform decoding to retrieve the decoded message M'.

Step 8. Compute BCR_{attack} (equation 2.2)

Step 9. Calculate fitness based on the following rule

IF $BCR_{attack} \geq T_1$ *and* $C_1 \leq MSS \leq C_2$

$Fitness_{t,n} = W_1 \cdot Fitness_{imp} + W_2 \cdot BCR$

ELSE

$Fitness_{t,n} = BCR_{attack}$

END

END

Step 10. Compute average fitness score for β_t using $Fitness_t = \sum_{n=1}^{N} Fitness_{t,n} \Big/ N$

END

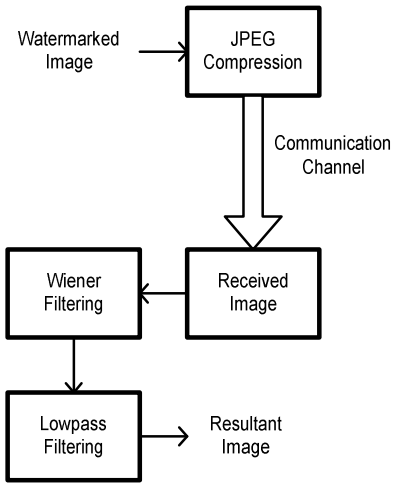
Every individual β_t from a generation is scored against each image X_n in the training set S. Steps 1 to 9 in algorithm 4.1 represent the scoring procedure for an image from S. Initially, 8×8 block DCT transformation is applied on X_n. Operating β_t on all of the DCT coefficients, we obtain a perceptual mask $\alpha_{t,n}$ for X_n. A secret message M is embedded using spread spectrum based watermarking technique and the perceptual mask ($\hat{\boldsymbol{\alpha}}_{GeneticPSF}$). To estimate robustness during GP simulation, we compute *MSS* which characterizes the watermark power. As the image is watermarked, the imperceptibility component of the fitness measure is computed before operating a sequence of attacks on the watermarked image.

4.2.3. Attack Sequences

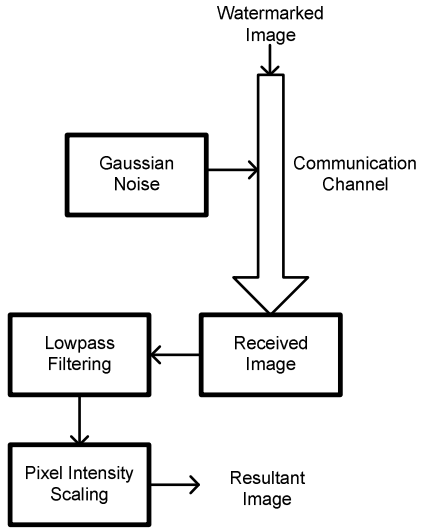
The sequence of attacks actually represent the conceivable attacks which may be legitimate or illegitimate and are mostly application dependent as discussed in Chapter 2. In the present work, we have considered three different attack sequences comprising of intentional and/or unintentional attacks. In the first set of attacks, the watermarked image is reduced before transmission to reduce its size. An adversary, after having access to the

watermarked image, attempts to destroy the watermark through an estimation attack, such as Weiner filtering attack followed by lowpass filtering based on his assumption that the watermark would be most probably residing in the high frequency region. This set of attack is presented in figure 4.2(a).

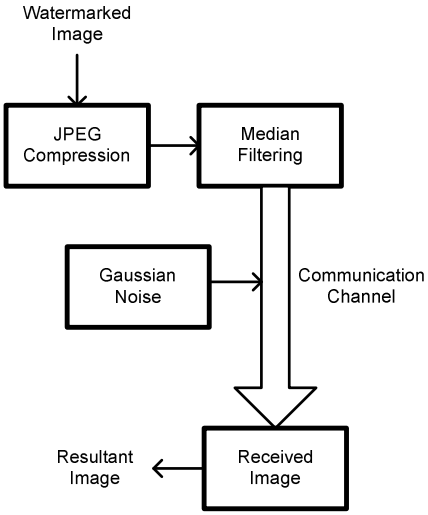
In the second case, a set of normal image processing operations are considered that comes in the category of unintentional attacks. A watermarked image is transmitted through a channel characterized by Gaussian noise. In order to remove noise, the received image is processed with lowpass filtering. Furthermore, the intensity of the resultant image is scaled to 0-255, representing a type of value-metric attack. Figure 4.2(b) depicts this set of attacks. In the case of third attack set, a watermarked image is first compressed to reduce its size. Later, the owner noticed small impulse noise in the produced work and performed median filtering to reduce it, and transmitted it via a channel characterized by Gaussian noise. On the receiver side, the received watermarked image has suffered through a sequence of unintentional attacks. This attack scenario is presented in figure 4.2(c).



(a)



(b)



(c)

Fig.4.2 Sequences of attacks on the watermarked image representing (a) attack scenario A, (b) attack scenario B and (c) attack scenario C

After the set of conceivable attacks are carried out, the distorted image is presented to the decoder. At the GP training phase, we assume that the decoder is fixed and does not modify in accordance to the attacks. The same PSF, as used in the embedding process, is used to obtain the perceptual mask for the attacked image.

Sufficient statistics for the maximum likelihood based decoder are computed using the perceptual mask to perform decoding of the embedded message. The decoded message M' is then used to calculate *BCR* after attack using equation 2.2, which is the second objective of the fitness measure.

4.2.4. Overall Fitness Computation

The fitness score, $Fitness_{t,n}$, corresponding to X_n and β_t is calculated based upon the fitness rule presented in step 9 of algorithm 4.1. If β_t has made a good tradeoff between robustness and imperceptibility, it is given bonus fitness. Otherwise, its fitness equals the value of BCR_{attack}. The bonus fitness is the amount of imperceptibility of the watermark when it was embedded. We are interested in those PSF that lie under some level of acceptable robustness. Therefore, T_1 is the lower bound of BCR_{attack}, while C_1 and C_2 are the upper and lower bounds of *MSS* respectively. These bounds are set empirically. Bounding of *MSS* value is necessary. If it is very low, low watermark power and consequently faint attack resistance is expected. On the other hand, if *MSS* is very high, it may violate the imperceptibility requirement. With each GP generation, those PSFs that make a good tradeoff between robustness and imperceptibility, emerge and conceptually replace the rest of the population.

After computing the fitness of β_t for X_n, steps 1 to 9 in algorithm 4.1 are repeated for the rest of the elements in S. The final fitness corresponding to β_t is the average fitness value of all the images in S, as illustrated in step 10 of algorithm 4.1. We now

present the overall training algorithm for the embedding side of the proposed system which evolves the best PSF.

Let $\mathcal{G} = \mathcal{P}_s : [\ \mathcal{P}_s = \beta_t : \{\ \beta_t \in \mathcal{F} \wedge t = 1,\ldots\ldots,z\ \} \wedge s = 1,\ldots\ldots,V\ ]$ represent the set of generations composed of population $\mathcal{P}_s$ of size z, with V being the total number of generations and $t = 1,\ldots\ldots,z$. Algorithm 4.2 presents the overall training algorithm of our GP based intelligent embedding scheme. It starts by creating an initial population of PSFs. At every subsequent generation, the population of PSFs is converged towards optimum with the help of genetic operators. Once the termination criterion is satisfied, the best evolved PSF is saved and the simulation is stopped. The saved PSF is then used in the testing phase and adaptive watermark decoding as discussed in the section following this.

Algorithm 4.2: Overall training algorithm for the Genetic based watermark embedding phase of the proposed AAW scheme.

Step 1. Create *z* randomly generated solutions to form the first generation P_1.

FOR each $\mathcal{P}_s : s = 1,\ldots\ldots,V$ and $\mathcal{P}_s \subset \mathcal{F}$

Step 1.1. Compute fitness of all the individuals in P_s following the procedure presented in algorithm 4.1.

Step 1.2. Evaluate fitness using the fitness function.

Step 1.3. Check the following condition:

IF termination criteria is satisfied, go to step 2.

ELSE

Create new generation using genetic operators and their corresponding probabilities.

Step 1.3.1. **Crossover:** Generate an offspring population Q_t as follows:

a. Choose two solutions τ_1 and τ_2 from P_s based on the fitness values.

b. Using a crossover operator, generate offspring and add them to Q_t.

Step 1.3.2. **Mutation:** Mutate each solution $\tau \in Q_t$ with a predefined mutation rate.

Step 1.3.3. Select *V* solutions from Q_t based on their fitness and assign them to P_{s+1}.Delete the low ranked individuals.

END IF

END

Step 2. Save the best evolved Genetic PSF, and terminate the procedure.

4.3. Computational Intelligence based Adaptive Watermark Decoding

Due to the range and diversity of attacks that could be mounted on a watermarking system, there is no such watermark decoding scheme that can perform well under all hostile attacks. However, we know that watermarking applications are increasing and becoming complex and that a specific set of attacks might be expected for a particular application at a particular instance of time. Therefore, with the emergent need of sophisticated watermarking applications, we require a decoding scheme that can adapt in view of a specific application. Normally, regular signal processing, channel noise, and JPEG compression are the most common attacks, however, geometric as well as security oriented attacks are also becoming important. When a watermarked image is attacked, the blind extraction of the hidden information becomes complicated. For instance, in case of tele-surgery, extraction of valuable embedded information about the identity of the patient, hospital, medical instruments, etc is challenging. This is because the watermarked image may have been intentionally and/or unintentionally processed, and JPEG compressed, before being communicated through medical data networks linking hospitals. Medical data networks, such as connecting hospitals of different islands to mainland hospital, are now widely used in countries such as Japan.

However, fixed decoding schemes, like simple TD model, can accurately classify bits, when distribution of the features in the decoding space does not overlap. This is because using a threshold, only linear bifurcation could be possible with the values on the left of the threshold considered as -1 and those on the right as +1. Nonetheless, in case of an attack on the watermarked image, the distributions of the features of a decoding model overlap. As a result, linear bifurcation is not possible and thus a simple *TD* model is unable to decode the message bits efficiently. To tackle such message decoding problems in watermarking, we assume that a non-separable message in lower dimensional space might be separable if it is mapped to higher dimensional space. In our proposed AAW scheme, this mapping to higher dimensional space is performed by the hidden layers in case of ANN and the kernel functions in case of SVM. Our decoding scheme comprises of the following two modules; (i) Dataset Generation, and (ii) CI based Decoding Modules.

4.3.1. Generating Attacked Watermarked Images

With the purpose of analyzing the performance of the proposed adaptive *CI* based extraction of watermark, we have generated a dataset of 800 bits by considering five different images, each of size 128×128 as shown in algorithm 4.3. Next, in each image, a message of size 16 bits is embedded. The whole process is repeated 10 times by changing the secret key used to generate the spread spectrum sequence. The anticipated attack sequence is operated on each image, which corrupts the embedded message as well. Similarly, in order to compare the different approaches in a harsher environment, the

watermark message size is increased, while the rest of the parameters are kept the same. This generates another dataset of 1600 bits.

Algorithm 4.3: Generation of the attacked watermarked image database for the CI based decoding phase of the proposed AAW scheme.

Step 0. Encode the message using an error correction technique and expand it to form a vector b.

FOR each image $X_n : n = 1, \ldots\ldots, N$

Step 1. Compute 8×8 block DCT of X_n

Step 2. Compute perceptual mask α_n corresponding to X_n

Step 3. Do the following repetitions:

FOR each image secret key $Q_m : m = 1, \ldots\ldots, M$

Step 3.1. Generate the spread spectrum sequence S_m

Step 3.2. Compute the watermark using $W_m^n = \alpha_n \cdot S_m \cdot \mathrm{b}$

Step 3.3. Perform watermark embedding using $Y_m^n = X_n + W_m^n$

Step 3.4. Perform inverse 8×8 block DCT

Step 3.5. Perform the selected sequence of attacks on the watermarked image.

END

END

The embedded message is blindly extracted by modeling the coefficients of each frequency band of an 8x8 block DCT domain and applying maximum likelihood estimation. In our proposed scheme, we consider the sufficient statistics for the watermark decoding as feature for the hidden bit classification. A statistic $T(Z)$ is sufficient for an underlying parameter θ, precisely if the conditional probability distribution of the data Z, given the statistic $T(Z)$, is independent of the parameter θ. Nonetheless, as opposed to [28] that utilize a single feature for the *TD* model, we do not compute sufficient statistics collectively across all the frequency bands; rather we compute it individually across each frequency band. The reason is that we assume the frequency channels are independent, but not identical. This assumption is maintained by the fact that a single attack has different effects on the different frequency channels.

Besides, this allows us to keep the sufficient statistics across each frequency band as a separate feature itself. Accordingly, corresponding to a single bit, the number of features is equal to the number of selected frequency bands. Considering all the selected frequency channels to be alike, the sufficient statistics r_i corresponding to each embedded bit is computed using equation 4.6. We modify this model by assuming that each frequency band is distorted differently by an attack. Hence, the sufficient statistics corresponding to a single bit are computed separately for each frequency band:

$$r_i = \sum_j r_i^j \quad j = 1, 2, ...J_{\max} \tag{4.5}$$

where J_{max} is the maximum number of selected frequency bands, and r_i^j is defined as:

$$r_i{}^j \quad \sum_{\mathbf{k} \in Q_i^j} \frac{\left| Y\left[\mathbf{k}\right] + \alpha\left[\mathbf{k}\right] s\left[\mathbf{k}\right] \right|^{c\left[\mathbf{k}\right]} - \left| Y\left[\mathbf{k}\right] - \alpha\left[\mathbf{k}\right] s\left[\mathbf{k}\right] \right|^{c\left[\mathbf{k}\right]}}{\sigma\left[\mathbf{k}\right]^{c\left[\mathbf{k}\right]}} \tag{4.6}$$

where Q_i^j is defined as the sample vector of all DCT coefficients in different 8×8 blocks that correspond to a single bit i and the *jth* frequency band. [**k**] represents 2-d vector indices. Algorithm 4.4 presents the procedure for feature extraction in the CI based decoding phase of the proposed watermarking system.

Algorithm 4.4: Feature extraction in the CI based decoding phase of the proposed AAW scheme:
FOR each image $X_n : n = 1, \ldots\ldots, N$

Step 1. Compute 8×8 block DCT of X_n

Step 2. Compute perceptual mask α_n corresponding to X_n

Step 3. Do the following repetitions:

FOR each image secret key $Q_m : m = 1, \ldots\ldots, M$

Step 3.1. Generate the spread spectrum sequence S_m

Step 3.2. Estimate C and σ for each frequency band

Step 3.3. Compute feature r_i corresponding to each bit using equation 4.6.

END

END

4.3.2. Adaptive Decoding

For bipolar signal, the estimated bit $\hat{b_i}$ in TD model is computed using equation 2.17. On the contrary, we treat the sufficient statistics as features for the CI based decoding scheme. With the aim of making these features linearly separable, they are first mapped to a higher dimensional space. Consider n training pairs (r_i, q_i), where $r_i \in R^N$ and $q_i \in [-1,1]$, then, in case of SVM, we have the nonlinear mapping:

$$f(r) = \sum_{i=1}^{N_S} \alpha_i q_i K(r_i, r) + b = \sum_{i=1}^{N_S} \alpha_i q_i \Phi(r_i) \cdot \Phi(r_v) + b \qquad (4.7)$$

where $\Phi(r)$ is nonlinear mapping function and b is bias. The attractiveness of this approach is its ability to generate an optimal hyper-plane in view of a new attack by learning the distortion in the features. We use three different kernels; Linear, Radial basis function (Rbf), and Polynomial:

- $K(x_i, x_j) = x_i^T . x_j$ (Linear kernel with parameter C)
- $K(x_i, x_j) = \exp(-\gamma \left\| x_i - x_j \right\|^2)$ (RBF with kernel parameters γ, C)
- $K(x_i, x_j) = [\gamma < x_i, x_j > + r\ ]^d$ (Polynomial kernel with parameters γ, r, d and C)

Similarly, r_i are provided as features to the multi-layered neural Network. Back-propagation learning algorithm is employed during training phase. This algorithm computes the error, e, for output neuron, m, as:

$$e_m(t) = z_m(t) - p_m(t) \qquad (4.8)$$

where z_m and p_m are the actual and target output for neuron m for iteration t. To minimize the average error, the weight vector and the bias are modified according to

Levenberg-Marquardt Algorithm. Further details of SVM and ANN based decoding strategies can be found in [30].

4.4. Security Enhancement

Albeit, we are mainly concerned with making a suitable tradeoff (between robustness and imperceptibility) at the embedding phase and improving robustness only at the decoding phase, we would also discuss the security related aspects. Due to the inherent spreading and anti-jamming characteristics of spread spectrum sequences, the secret key used to generate the spread spectrum sequence is of prime importance. In order to reduce the chances of security leakage, a second key may be used for randomly permuting the elements of the matrix of the selected DCT coefficients before embedding. This helps in introducing uncertainty about the correspondence of a codeword element and the selected DCT coefficients. In this fashion, it is less likely for an attacker to know which coefficients are altered matching to a codeword element.

Besides keys, there is another important factor of our proposed scheme that boosts the security of the system. Although, both the evolved genetic PSF and the trained decoding models could be made public or private depending on the application, however, let us assume the conservative scenario that they are made publically available. Even then, due to the nonlinearities introduced both at embedding and decoding phases, it is very difficult for an attacker to illicitly gain information about the secrecy of the system by analyzing the embedding or detection space. Especially, at the decoding phase, both SVM and ANN transform the input space to high dimensional space through kernel functions and hidden layers respectively. In this case, an adversary can hardly study the

effect of random change in keys on the decoder behavior. On the other hand, if the genetic PSF and trained decoding models are not made public, it is almost impossible for an adversary to perform unauthorized embedding and/or decoding.

4.5. Potential Applications

Nowadays, there is a constant competition between a watermark embedder and an adversary. Additionally, due to the rapid expansion of new, dynamic, and complex watermarking applications, fixed watermarking strategies may become obsolete. Few potential watermarking applications, where dynamic and adaptive watermarking strategies are highly desirable are; print-to-web technology [63], fingerprinting (or transaction tracking)[64], broadcast monitoring [65, 66], securing data sent through context aware medical networks [15, 58], secure digital camera [56], data transmission through wireless networks, agent based online auction systems etc. Such watermarking applications have to be guarded against both intentional and unintentional attacks that may change with respect to time. For example, in case of context aware medical networks, the images may be compressed and transmitted at different rates at one instant of time, but this may rapidly change depending on the network congestion and/or the requirement of the physician. In cases, where the type of attack is the same, but only its intensity changes with time, the retraining of only the CI based decoding model is recommended. There may be no need of the retraining of the GP based embedding phase. However, in applications, where the type of attack also changes dynamically, then both embedding and decoding models should be retrained for coping with the distortions introduced by the dynamic attacks.

4.6. Results and Discussion

In the foremost discussion of this section, we analyze the performance of the first phase of our proposed approach; making effective tradeoff between imperceptibility and robustness using intelligent GP based embedding. Later on, we discuss the performance of the second phase of our proposed approach; increasing bit extraction performance using CI based decoding strategies.

4.6.1. Tradeoff between Imperceptibility and Robustness

Figure 4.3(a) and (i) shows the original *Trees* and *Lena* images respectively. Whereas, figure 4.3(e) and (m) shows the watermarked *Trees* and *Lena* images respectively. It can be examined that a human eye can hardly notice the difference between the original and watermarked images. Figure 4.3(b)-(d), (f)-(h), (j)- (l), and (n)-(p) demonstrate the severity of distortions that a set of attacks can bring into any watermarked medium. Resisting such distortions has always been a challenge in the watermarking community, and our proposed technique provides robustness against all such legitimate and illegitimate manipulations. Usually, high strength embedding means high distortion of the original image and thus low imperceptibility. Therefore, we show that even though keeping low watermark strength and consequently, a fair amount of imperceptibility, our proposed AAW approach is able to retrieve the embedded message from the attacked watermarked image. In watermarking, imperceptibility, robustness, and capacity are three important, but contradicting properties. In this current discussion regarding imperceptibility versus robustness tradeoff, we have kept the capacity of the system constant.

a) Original *Trees* Image | b) Gaussian noise attack on (e) | c) (Gaussian + lowpass filtering) attack on (e) | d) (Gaussian + lowpass filtering + pixel intensity scaling) attack on (e)

e) Watermarked *Trees* image | f) Scaled difference image between (e) and (b) | g) Scaled difference image between (e) and (c) | h) Scaled difference image between (e) and (d)

i) Original *Lena* image | j) JPEG compression attack on (m) | k) (JPEG + median filtering) attack on (m) | l) (JPEG + median filtering + Gaussian noise) attack on (m)

m) Watermarked *Lena* image | n) Scaled difference image between (m) and (j) | o) Scaled difference image between (m) and (k) | p) Scaled difference image between (m) and (l)

Fig.4.3 Severity of distortions caused by the sequence of attacks on watermarked Trees and Lena images

High imperceptibility lays down a limit on the capability of the decoding models, as robustness usually requires high power embedding. In this regard, figure 4.4 shows the distribution of the watermark strength for *Trees* image and attack scenario B across both frequency and number of blocks. This watermark strength distribution has been generated using the GP based embedding phase of the proposed AAW scheme. It can be straightforwardly observed that inside each block, the amplitude of the watermark strength increases as we move towards high frequency. This attribute of the developed genetic PSF for watermark structuring is consistent with HVS modeling, according to which high watermark strength should be used for high frequencies to reduce the resultant distortion. There is also a variation in amplitude from block to block. In figure 4.4, if we pick a selected DCT coefficient and vary the block number, then we can recognize the watermark strength variation across a single frequency band. Each of these selected frequency bands have dissimilar variation in watermark strength and thus should be dealt with separately.

This analysis also validates the effectiveness of the GP based intelligent selection of the frequency bands, as some frequency bands may be very resistive against certain types of attacks. The features corresponding to these frequency bands may thus not be altered heavily by the attacks under question. On the other hand, this analysis also supports our idea of exploiting the frequency bands individually for learning their corresponding distortion using SVM and ANN based approaches.

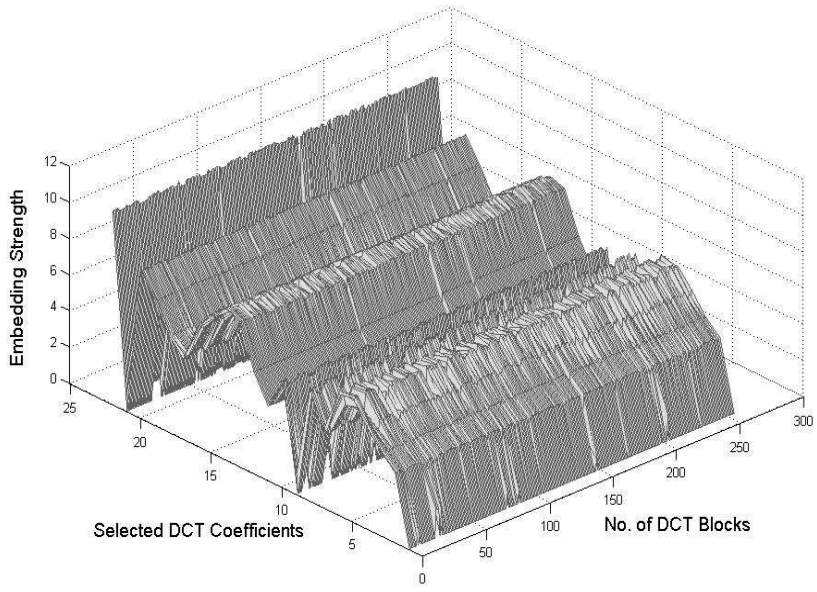


Fig.4.4 Watermark Strength distribution of Trees image for Attack Scenario B

The statistics of the different imperceptibility and robustness measures for the Watson's PSF and our GP based AAW scheme against the three sequences of attacks are given in table 4.1-4.3. It can be observed that there is quite a margin of improvement for the GP based embedding in case of all the three attack scenarios. Keeping the SSIM values the same, the mean and the standard deviation of both the wPSNR and BCR are better for genetic PSF, as compared to that of Watson's PSF for different message sizes and against all attack scenarios.

Figure 4.5 shows watermark strength distribution for attack scenario C. It demonstrates the ability of the GP based embedding strategy to change both the amplitude and frequency bands selection as per the new attack. Table 4.4 provides the details of the GP parameter selection using Matlab based GPLab toolbox [54].

4.6.2. Performance Comparison against a Series of Attacks

In this section, using different series of attacks, we compare the robustness performance of the proposed AAW scheme with that of the existing schemes, ASSW [28] Machine Learning Decoding (MLD) [30], and Genetic Watermark Shaping Scheme (GWSS) [16]. For this purpose, we consider two cases. In the first case, we compare watermark extraction performance using a watermark message size of 16. We then further make the conditions harsh by increasing the message size to 32 and keeping all the rest of parameters the same. This results in high capacity of the system but less available DCT coefficients corresponding to a single bit, which thus reduce the number of samples for training because of the repetition coding.

Table 4.1 Imperceptibility and BCR based performance statistics for 50 training images for attack scenario A

Message Size	Watermark Embedding Method	Type of measure	Lowest	Highest	Average	Standard Deviation
16 Bits	Watson's PSF	SSIM	0.9372	0.9803	0.9591	0.0139
		wPSNR	37.773	41.745	39.0191	1.468
		BCR	0.6250	1	0.8612	0.1012
	Genetic PSF	SSIM	0.9362	0.9793	0.9581	0.0139
		wPSNR	38.8380	42.026	40.0379	1.0844
		BCR	0.7500	1	0.9075	0.0830
32 Bits	Watson's PSF	SSIM	0.9371	0.9802	0.9591	0.0138
		wPSNR	37.7770	41.7480	39.0193	1.4689
		BCR	0.4688	1	0.7656	0.0991
	Genetic PSF	SSIM	0.9361	0.9792	0.9581	0.0138
		wPSNR	38.8380	42.0260	40.0379	1.0844
		BCR	0.6875	1	0.8263	0.0804

Table 4.2 Imperceptibility and BCR based performance statistics for 50 training images for attack scenario B

Message Size	Watermark Embedding Method	Type of measure	Lowest	Highest	Average	Standard Deviation
16 Bits	Watson's PSF	SSIM	0.9625	0.9865	0.9746	0.0077
		wPSNR	40.0020	43.4630	41.1668	1.2342
		BCR	0.8125	1	0.9237	0.0695
	Genetic PSF	SSIM	0.9615	0.9855	0.9736	0.0077
		wPSNR	41.174	43.6620	42.0835	0.8307
		BCR	0.8750	1	0.9762	0.0397
32 Bits	Watson's PSF	SSIM	0.9624	0.9865	0.9746	0.0077
		wPSNR	40.0120	43.4670	41.1688	1.2333
		BCR	0.6563	0.9688	0.8481	0.0717
	Genetic PSF	SSIM	0.9615	0.9855	0.9736	0.0077
		wPSNR	41.1740	43.6620	42.0835	0.8307
		BCR	0.7500	1	0.9281	0.0510

Table 4.3 Imperceptibility and BCR based performance statistics for 50 training images for attack scenario C

Message Size	Watermark Embedding Method	Type of measure	Lowest	Highest	Average	Standard Deviation
16 Bits	Watson's PSF	SSIM	0.9734	0.9921	0.9830	0.0059
		wPSNR	41.8410	45.8400	43.0707	1.4033
		BCR	0.4375	1	0.6787	0.1300
	Genetic PSF	SSIM	0.9725	0.9911	0.9820	0.0058
		wPSNR	42.8710	45.7030	43.9412	0.9379
		BCR	0.6250	1	0.9087	0.0921
32 Bits	Watson's PSF	SSIM	0.9734	0.9920	0.9830	0.0059
		wPSNR	41.8480	45.8250	43.0700	1.4043
		BCR	0.3438	0.8438	0.6212	0.1080
	Genetic PSF	SSIM	0.9724	0.9911	0.9820	0.0058
		wPSNR	42.8710	45.7030	43.9412	0.9379
		BCR	0.4688	1	0.8500	0.0934

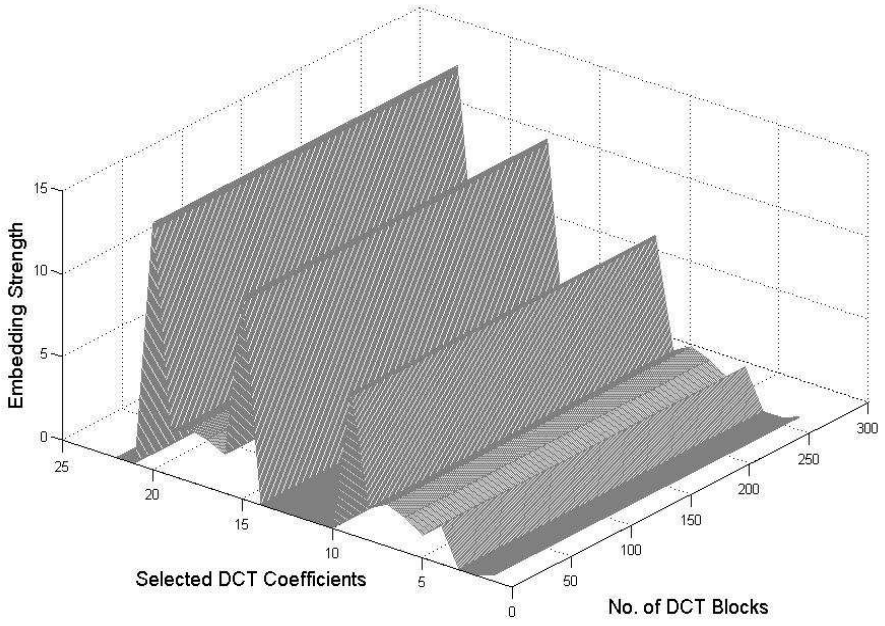


Fig.4.5 Watermark Strength distribution of Trees image for Attack Scenario C

Figure 4.6 shows the overlapping distribution of the decoding features due to the sequence of attacks; (a), (c), (e) using Watson's PSF as used in ASSW and (b), (d), (f) using genetic PSF as employed in the proposed AAW scheme. It can be observed that there is relatively less overlapping of the features in case of GP based embedding as compared to that of Watson's PSF. This makes it easier for the subsequent CI based classification strategies to classify the hidden bits. The parameter selection details for ANN are provided in table 4.5.

Table 4.4 GP Parameters Settings

Objective:	**To perform intelligent watermark structuring**
Function Set:	+, -, *, protected division, *SIN*, *COS*, *LOG*, *MAX* and *MIN*
Terminal Set:	Constants: *random constants in range* of [-1, 1] Variables : $X_{0,0} / 1024$, $\lvert X(i,j)\rvert$, $\alpha(k_1, k_2)$, i, j
Fitness :	$\text{Fitness} = (W_1 * \text{SSIM} + W_2 * \text{wPSNR} / 46) + \text{BCR}_{\text{attack}}$
Selection:	Generational
Population Size:	160
Initial max.Tree Depth	8
Initial population:	Ramped half and half
Operator prob. type	Variable
Sampling	Tournament
Expected no. of offspring	rank89
Survival mechanism	Keep best
Real max level	30
Termination:	Generation 40

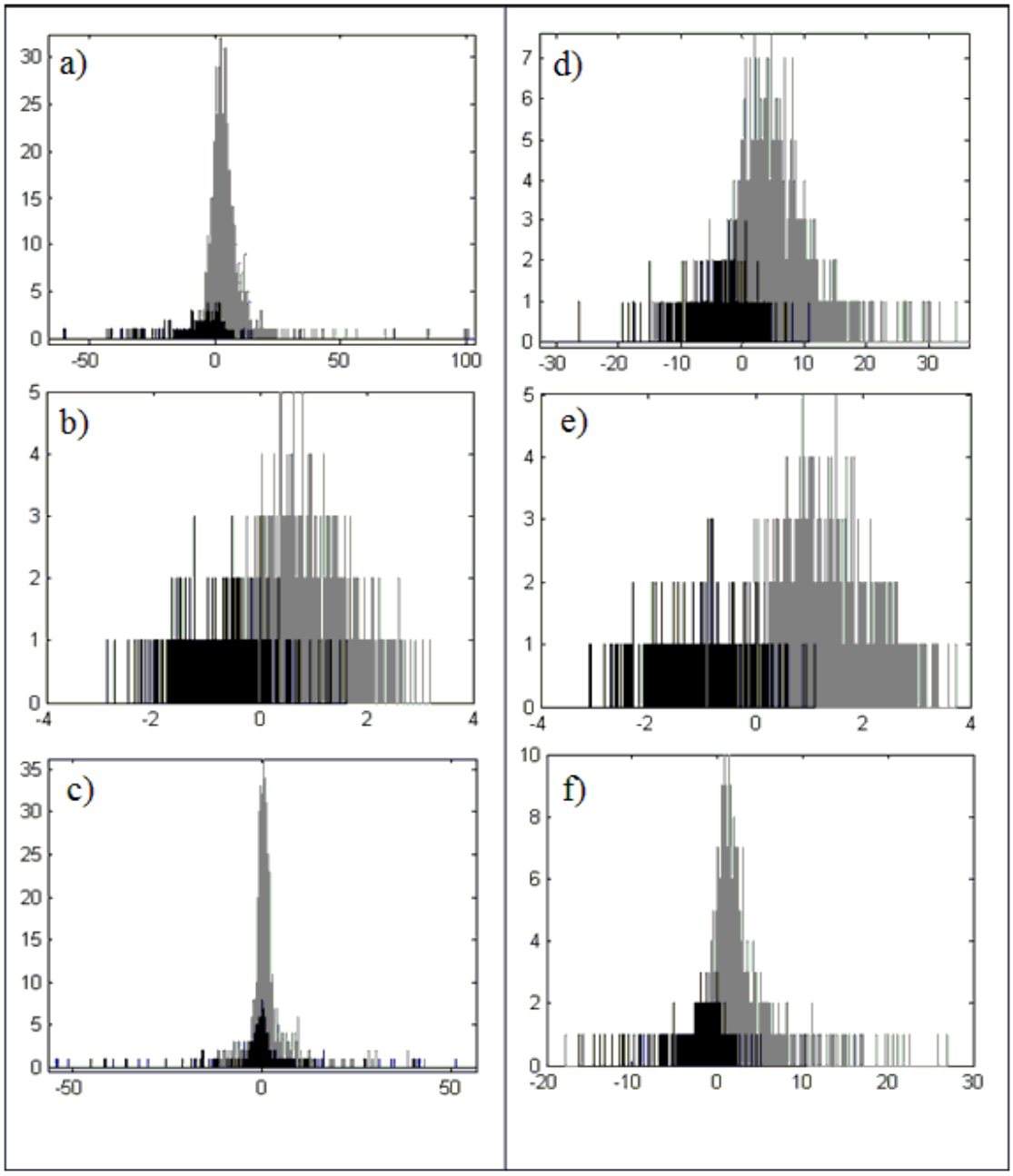


Fig.4.6 Distribution of the decoding features using Watson's PSF (a), (c), (e) and genetic PSF (b), (d), (f) for watermark structuring against the three attack scenarios A, B, and C respectively

Table 4.5 Parameter selection of ANN based decoding strategy

Features	22
Data Size	1600
Epochs	50
Hidden layers	3 [8,4,2]
Activation function of Hidden Layer	'tansig'
Activation function of Output Layer	'pure linear'
Training Algorithm	'Levenberg-Marquardt'

Performance comparison using watermark message size of 16

Table 4.6 shows the performance comparison of the different approaches in terms of BCR. It can be observed that the performance of the proposed AAW scheme is superior to MLD, which uses Watson's PSF, an all the three attack scenarios. For example, considering attack scenario B and using Poly SVM, the performance of the proposed AAW is 1 as compared to that of 0.9975 for MLD. Almost the same behavior is shown by the ANN based classification in both strategies. On the other hand, the performance of ASSW, which is based on simple TD, lags behind. This fact shows that it is far better to intelligently employ CI approaches at the watermark decoding stage, so as to learn the effect of distortions on the embedding feature space introduced by the series of attacks.

The details of parameter selection; performed using grid search, for the three SVM models are provided in tables 4.6 and 4.7. Similarly, temporal comparisons are also presented in tables 4.6 and 4.7 for the different decoding strategies.

Table 4.6 BCR based comparison of the different approaches. Note: message size=16 bits, features=22, and total bits=800.

Attack Sequence	Types of Attacks	**Watermark Decoding strategies**			**Watermark Shaping Strategies**			
		Models	Parameters		Using Watson PSF		Using GP based Embedding	
			C	γ	Time (sec.)	BCR	Time (sec.)	BCR
Scenario A	JPEG compression, Wiener attack, and Lowpass filtering.	Linear SVM	2	-	2.09	0.9587	0.23	0.9850
		Poly SVM	2	2	3.82	**1**	0.48	**1**
		RBF SVM	2	2	0.89	**1**	0.406	**1**
		ANN	-	-	10.218	**1**	3.53	**1**
		TD	-	-	0.01	0.8612	0.01	0.9075
Scenario B	Gaussian noise addition., Lowpass filtering., and Intensity Scaling	Linear SVM	2	-	0.194	0.9413	0.062	0.9875
		Poly SVM	2	2	0.407	0.9975	0.156	**1**
		RBF SVM	2	2	0.218	0.985	0.328	0.9988
		ANN	-	-	11.75	0.9938	2.76	**1**
		TD	-	-	0.09	0.9237	0.01	0.9762
Scenario C	JPEG compression, Median filtering., and Gaussian Noise addition	Linear SVM	2	-	0.91	0.7725	0.156	0.9487
		Poly SVM	2	2	61.4	0.999	50.4	**1**
		RBF SVM	2	2	0.419	0.9163	0.422	0.9925
		ANN	-	-	16.87	0.8962	15.235	0.9925
		TD	-	-	**0.05**	0.6787	0.01	0.9087

Performance comparison using watermark message size of 32

Table 4.7 present the BCR performance, when the watermark message size is increased to 32. As expected, both MLD and TD are not able to cope with these harsh conditions that may be regularly faced in real world applications. However, the performance of the proposed AAW scheme is impressive due its use of two different CI approaches in conjunction to thwart the distortions introduced by the series of attacks.

In attack scenario A, the BCR of the proposed AAW scheme (using simple threshold at decoding stage) is 0.83, while that of MLD is 0.76. However, using ANN for decoding, BCR for AAW scheme is 0.99 and that for MLD is 0.98. This fact clearly demonstrates the effectiveness of intelligent watermark extraction strategies. In case of attack scenarios B and C; the achieved BCR of the proposed AAW scheme using ANN for extraction, is 0.99 and 0.97 respectively. While that of MLD using ANN for extraction is 0.96 and 0.90 respectively. The bold values in tables 4.6 and 4.7 indicate the highest BCR being achieved for a given series of attacks. The relatively low performance of SVM as compared to that of ANN for 32 bit watermark may largely be attributed to the parameter selection of SVM. We have used grid search for SVM parameter selection, and therefore the performance of SVM may increase further with the refinement of grid search.

In summary, for all the attack scenarios, almost all the decoding strategies show an improvement when used with conjunction with GP based embedding. Temporal cost is also reduced largely because of the implicit feature selection provided by the frequency band selection mechanism of the GP based embedding.

Both bit extraction performance and reduced training time are important factors, when CI based strategies are used for resisting the complex and dynamic attacks on a watermarking system. In future, we intend to use more advanced CI techniques, which can further increase the robustness and/or offer less training time for coping with the dynamic applications of watermarking.

Table 4.7 BCR based comparison of the different approaches. Note: message size=32 bits, features=22, and total bits=1600.

Attack Sequence	Types of Attacks	Watermark Decoding strategies			Watermark Shaping Strategies			
		Models	Parameters		Using Watson PSF		Using GP based Embedding	
			C	γ	Time (sec.)	BCR	Time (sec.)	BCR
Scenario A	JPEG compression, Wiener attack, and Lowpass filtering.	Linear SVM	2	-	26.859	0.92	1.672	0.96
		Poly SVM	2	2	222.4	0.984	273.05	**1**
		RBF SVM	2	2	3.359	**1**	2.82	**1**
		ANN	-	-	14.06	0.98	13.28	0.99
		TD	-	-	0.1	0.76	0.1	0.83
Scenario B	Gaussian noise addition, Lowpass filtering, and Intensity Scaling	Linear SVM	2	-	0.41	0.90	0.36	0.96
		Poly SVM	2	2	2.781	0.96	0.844	**0.99**
		RBF SVM	2	2	0.75	0.93	0.625	0.97
		ANN	-	-	12.907	0.96	15.56	**0.99**
		TD	-	-	0.1	0.84	0.1	0.93
Scenario C	JPEG compression, Median filtering, and Gaussian Noise addition	Linear SVM	1	-	298.7	0.8125	15.56	0.92
		Poly SVM	1	1	5224.2	0.88	4923.6	0.93
		RBF SVM	2	2	1.33	0.85	0.578	0.95
		ANN	-	-	13.28	0.90	13.73	**0.97**
		TD	-	-	0.1	0.62	0.1	0.85

Exemplary for watermarking small images

In recent years, widespread of portable devices and e-business, such as online catalogues, has increased the use of small sized images more than ever before. These images require less watermark bits in order to maintain visual quality. Whereas, on the contrary, no compromise can be made on the robustness as well. Our proposed system is very applicable in such applications as demonstrated in table 4.8. The results achieved demonstrates the effectiveness of intelligent encoding and decoding structures. Structures which are able to learn the hidden dependencies in HVS, and separability in the decoding

feature space, which were otherwise neglected.

Other examples where the proposed system can be very effective include watermarking biometric images such as fingerprints [67] and iris scans [68].

Table 4.9 summarizes different statistics for GP training versus testing performance in temrs of BCR values for the three different attack scenarios. For test dataset, the difference of standard deviation (training - testing) in terms of attack sequence A is relatively higher as compared to sequences B and C. But still it is in close proximity to that of the training data. In case of attack scenario C, the standard deviation for both the training and testing data is almost the same. This demonstrates the precision of the proposed watermarking system and hence its generalization capability. Once the system is trained on a number of training images, it can be suitably applied in the real world applications.

Table 4.8 Decoding performance comparison against attack sequences for small images (size = 128 x128)

Type of Attack	Intensity of Attack	Decoding Model	Time (sec)	BCR
Cascade: Pixel Shift + Median filtering	Shift=1, Window Size =3x 3	Linear SVM	1.28	0.80
		Poly SVM	1.22	1
		RBF SVM	0.9	1
		ANN	4.1	0.97
		TD	0.1	0.575
Cascade: Rotation + jpeg	Rotation=20°, JPEG Quality Factor=70	Linear SVM	0.78	0.67
		Poly SVM	0.93	1
		RBF SVM	0.97	0.98
		ANN	4.4	1
		TD	0.1	0.59

Table 4.9 Statistics for Training versus Testing Performance in Terms of BCR

Attack scenario	Data for	Min	Max	Average	Standard Deviation
A	*Training*	0.875	1	0.9625	0.0559
	Testing	0.7500	1	0.9075	0.083
B	*Training*	0.9375	1	0.9875	0.0282
	Testing	0.8750	1	0.9762	0.0397
C	*Training*	0.8125	1	0.9125	0.0948
	Testing	0.6250	1	0.9087	0.0921

4.7. Summary

In this chapter, we have discussed a novel watermarking system with intelligent techniques employed at both the encoding and decoding stages. In view of HVS, GP is used to embed a high energy watermark in frequency bands that are least affected when presented to a sequence of attacks. This embedding is performed using the evolved appropriate PSFs which take into account the possible attack environment. Similarly, learning capabilities of SVM and ANN are utilized to learn about the induced distortions at the decoding side and estimate the hidden message accordingly. Experimental results on a dataset of test images demonstrate marked improvement in terms of BCR values and visual quality. The proposed system is easy to implement and is practically viable. In addition, security of the system can be further enhanced by employing an overlaying cryptographic layer.

5. Reversible Watermarking using Machine Learning

5

"There was a level on which I believed that;
What had happened remained reversible."
Joan Didion

5.1. Introduction

Traditional watermarking techniques cause irreversible degradation of an image. Although the degradation is perceptually sparse, it may not be admissible in applications like medical, legal or military imagery. For applications such as these, it is desirable to recover the embedded information as well as the sensitive host image. For example, in a database of medical images marked with patient information, it is desired to extract the patient information along with recovery of the original host signal for proper diagnosis. Unlike robust watermarking [16-18], in reversible watermarking the original image is completely restored from the watermarked image, thus, preserving the originality of the cover work.

An efficient reversible watermarking scheme should be able to embed more information with less perceptual distortion, and equally, be able to restore the original cover work content. Watermark capacity and imperceptibility are two contradicting

properties. If one increases, the other decreases and vice versa. Therefore, one needs to make an optimum choice between them.

Many reversible data hiding techniques are proposed in literature. These include Chen and Kao's DCT based zero replacement reversible watermarking [31], Ni et al.'s [32] histogram manipulation based reversible watermarking technique, Xuan et al.'s [33] spread spectrum based watermarking method and Tian's [34] lossless data hiding technique based on difference expansion transform of a pair of pixels. Alattar [35] extended Tian's work and proposed a technique which uses difference expansion of a vector of several pixels to achieve larger capacity. These approaches are simple and efficient but are unable to achieve the optimum tradeoff between capacity and imperceptibility. Xuan et al. [36] proposed a reversible watermarking scheme that uses integer wavelet transform and threshold embedding. However, they have used a fixed threshold for all of the coefficients in different sub-bands of integer wavelet transform. Xuan et al. [37] also proposed a bitplane compression based technique which losslessly compresses one or more middle bitplanes to save space for data embedding. This chapter presents a reversible watermarking technique which uses machine learning based intelligent approach for making efficient watermark capacity-imperceptibility tradeoff. We use GP to exploit the hidden dependencies of the wavelet coefficients in different frequency sub-bands. By selecting the appropriate coefficients for embedding, it is possible to make a good choice between watermark payload and imperceptibility.

5.2. Intelligent Reversible Watermarking

The proposed GP based watermarking system evolves a mathematical function capable of selecting coefficients for LSB embedding. The GP evolved function acts like a map but with a reduced size. The other advantage is that we do not need to push the non-selected coefficient away from the selected ones (as in [36]), and thus, introduce less distortion. The general architecture of our proposed scheme is based on training and testing phases as shown in figure 5.1. For a given image x, a candidate expression is used to embed and extract the watermark by using the watermark embedder and extractor. Details about the watermark embedder and extractor, as shown in figure 5.2, and how an individual expression makes coefficient selection for watermark embedding, are discussed later in sections 5.2.2-5.2.4. For every candidate expression, its fitness is calculated in order to score its performance. The fitness function is based upon watermark payload and imperceptibility measures for assessing the performance of an individual GP expression. The procedure is repeated for the whole population in a generation. After a generation is scored, the best individuals are selected as parents and offspring are created from the parents by probabilistically applying the genetic operators of crossover, mutation and replication on them [50]. The procedure continues generation by generation with scoring, selection and offspring creation in place till a termination condition is reached as explained in Chapter 2. This concludes the training phase. The best GP expression, represented by $\Im$, is saved.

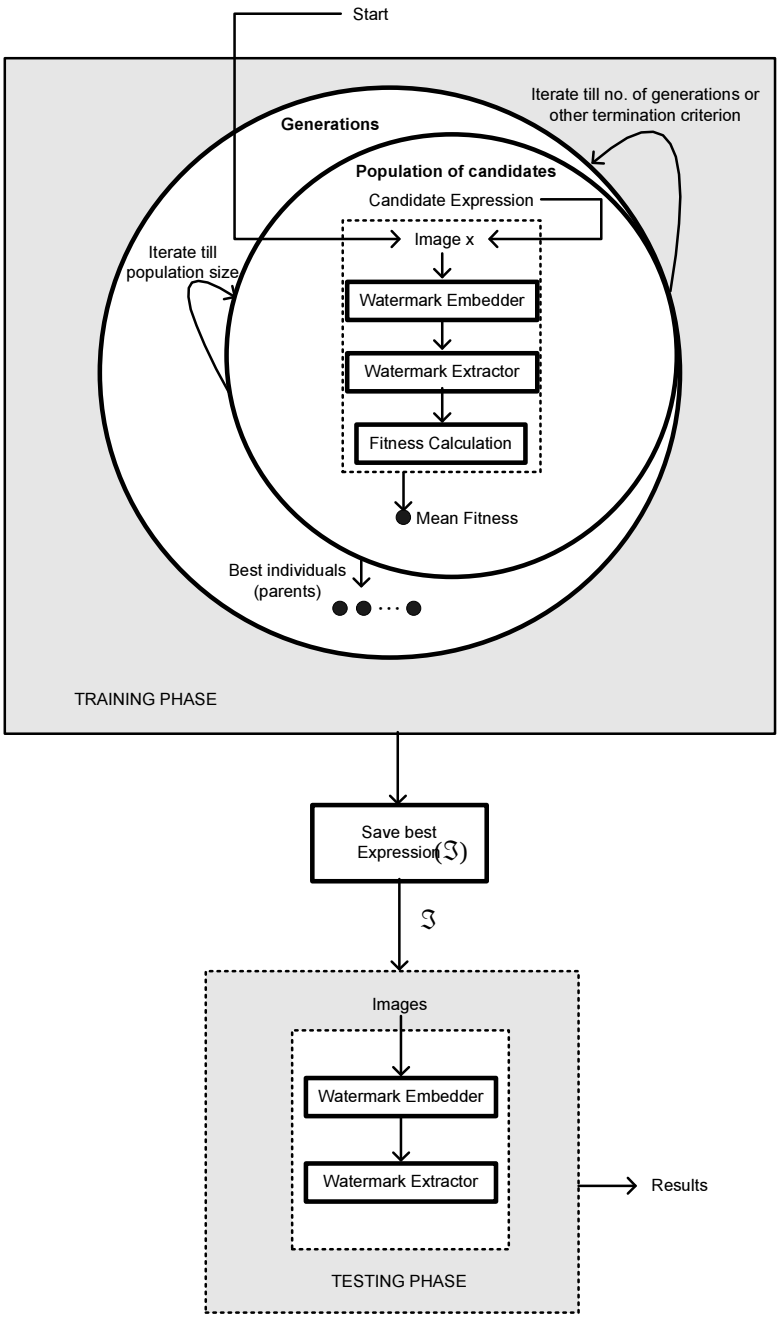


Fig.5.1 General architecture of the proposed reversible watermarking scheme.

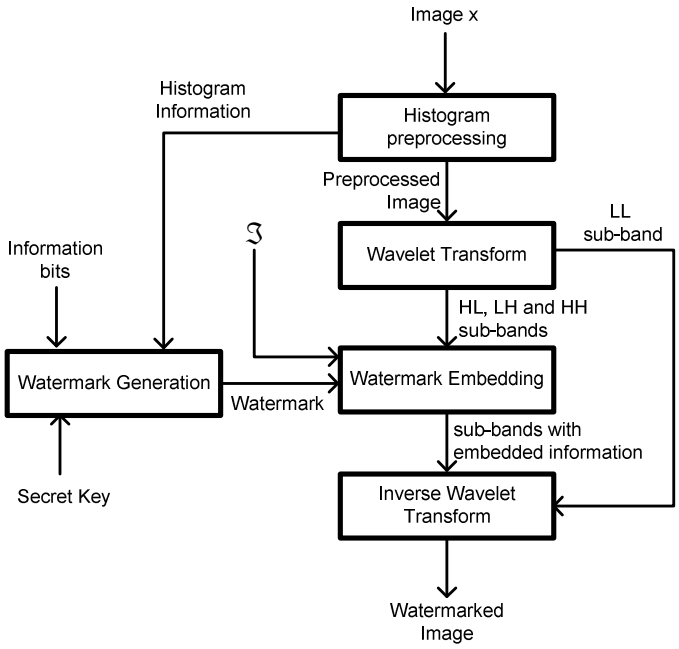
After training, $\Im$ is used in the testing phase or in the given watermarking application for watermark embedding. It may be hard wired in the watermark extractor or

communicated over a secret channel for extraction purposes. It can also be made public if the secret key used in watermark generation is kept secret.

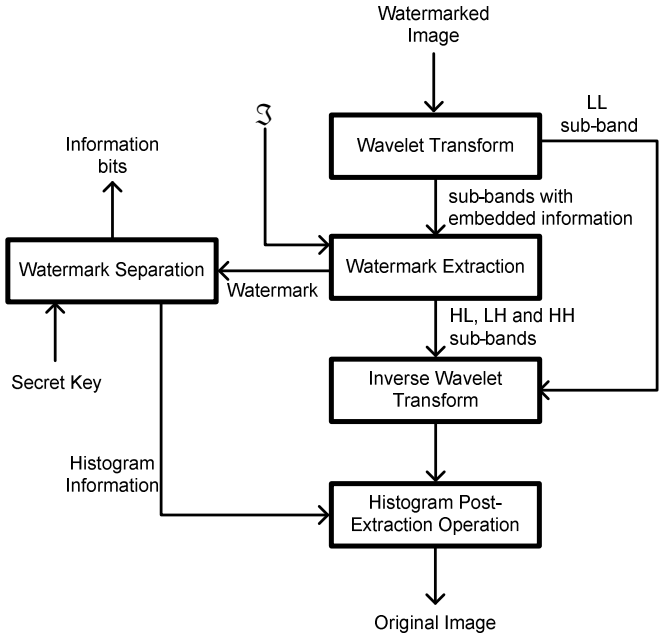
5.2.1. Histogram Pre-processing and Wavelet Decomposition

As is the case with most of the reversible watermarking approaches, some preprocessing needs to be performed to avoid possible overflow/underflow. Therefore, we apply the histogram preprocessing before embedding, and histogram post-extraction operation after extracting the watermark. These operations are well documented in [36]. The histogram post-extraction operation is necessary to extract the original image contents.

Transformation of the image to frequency domain is considered as more beneficial for obtaining a large bias between 0's and 1's such as to create more space for hiding data by bit compression. To avoid round-off error, we use the second generation CDF (Cohen-Daubechies-Fauraue) integer wavelet transform. This wavelet transform maps integer to integer and is used in JPEG2000 as well. The phenomenon of frequency masking suggests that the horizontal (LH), vertical (HL), and diagonal (HH) detail sub-bands are less susceptible to embedding distortions. Therefore, we use these sub-bands for watermark embedding and leave the approximation (LL) sub-band unchanged.



(a)



(b)

Fig.5.2 Block diagrams of (a) watermark embedder and (b) watermark extractor.

5.2.2. Expression representation

GP is a directed random search technique, inspired by biological evolution, for solving optimization problems. In GP, a candidate solution is represented by a data structure such as a tree. For representing a candidate solution with a GP tree, suitable GP *terminal set* and GP *function set* are defined. We use the coefficient amplitude, sub-band, and local 8x8 block information within a sub-band to exploit their hidden dependencies in making a payload–imperceptibility tradeoff. In order to let GP learn about different sub-bands, we give them different numbering (LH=3, HL=4 and HH=5) as shown in figure 5.3. Therefore, in our proposed scheme the GP terminal set comprises of the following:

i. $|X(i,j)|$: Absolute value of the wavelet coefficient $X(i,j)$ in a sub-band.

ii. S_{Label} : Numerical value assigned to different watermarkable sub-bands in order to distinguish them from each other. In our simulation, we have used their values as LH=3, HL=4 and HH=5.

iii. S_{mean} : Average value of the coefficients in a particular sub-band.

iv. B_{number} : Position of an 8x8 block, *B*, within a sub-band, $B_{number} = 1, 2, \ldots, \frac{size(subband)}{size(B)}$.

v. B_{mean} : Average value of coefficients in *B*.

vi. m, n : Row and column indices of coefficients in *B*.

The above parameters for the terminal set helps GP in finding the right expression for selecting suitable coefficients and incorporates a kind of dependence between the neighborhoods. The GP function set comprises of simple functions including four binary floating arithmetic operators (+, -, *, and protected division), *LOG, EXP, SIN, COS, MAX*

and MIN. In order to evaluate the performance of a candidate expression, we use the following fitness function:

$$fitness = (\kappa_1 * bpp) + (\kappa_2 * PSNR / 50) \tag{5.1}$$

The above fitness function is based on watermark payload and imperceptibility objective measures, *bpp* is the bit per pixel ratio and *PSNR* represents peak signal-to-noise ratio of the watermarked image compared to the original image. κ_1 and κ_2 are the corresponding weights. The choice of κ_1 and κ_2 is not easy. As a rule of thumb, for applications requiring more payload, some sacrifice can be made in terms of imperceptibility, i.e. κ_1 should be given more weightage. On the other hand, for those requiring high imperceptibility, κ_2 should be increased. *PSNR* is divided by 50 in order to scale its value, such that equal weightage could be given to both of the objective measures.

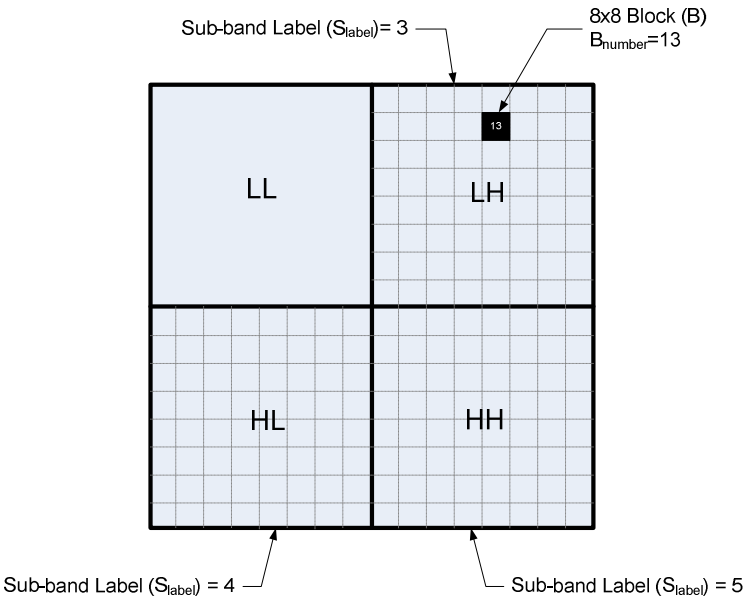


Fig.5.3 Wavelet decomposition showing different sub-bands and 8x8 blocks

5.2.3. Information Embedding

In order to embed information in image x, it is first modified by applying the histogram preprocessing operation to prevent possible overflow/underflow. After the preprocessing operation, CDF(2,2) integer wavelet transformation is applied to generate X as shown in figure 5.2(a). LH, HL, and HH are considered for watermarking. Next, the watermark is generated which constitutes of a header portion and a message portion. In the header portion, bits containing information about the histogram preprocessing operation are stored, so that they can be used for original image retrieval in the post-extraction operation.

Let $\mu(i, j)$ denote the value after applying $\Im$ on a wavelet coefficient $w(i, j)$. Since $\Im$ contains information about the optimum choice, therefore, the value of $\mu(i, j)$ governs which coefficient to select for a watermark bit embedding based on the following rule:

$$w'(i, j) = \begin{cases} 2 \cdot w(i, j) + b, & \text{if } \mu(i, j) \geq 0 \text{ and } \mu'(i, j) \geq 0 \\ w(i, j), & \text{otherwise} \end{cases} \quad (5.2)$$

where, b is the watermark bit to be embedded and $2 \cdot w(i, j) + b$ is equivalent to shifting the binary representation of the coefficient towards left by one bit and replacing the LSB by bit b. $\mu'(i, j)$ is the value computed by applying $\Im$ on $w'(i, j)$. Using the same rule, embedding is performed in all of the selected sub-bands. Different coefficient, block and sub-band values correspond to different $\mu(i, j)$ values. Through GP learning, the best evolved expression, $\Im$, incorporates information about the optimum choice of

coefficients for watermark embedding. After embedding the watermark, inverse integer wavelet transform is applied in order to represent the marked image in spatial domain so that *PSNR* values can be calculated.

5.2.4. Information Extraction

In the watermark extraction stage (figure 5.2(b)), integer wavelet transformation is applied on the watermarked image so that wavelet sub-bands are generated. After applying $\Im$ on a coefficient $w'(i,j)$, if $\mu'(i,j) \geq 0$, the LSB of this coefficient is the watermark bit. Otherwise, we move to the next coefficient since the current coefficient contains no hidden information. Applying this procedure in all of the three watermarkable sub-bands yields the watermark. The original image content is retrieved from LH, HL, and HH sub-bands according to the following restoration rule:

$$w(i,j) = \begin{cases} \lfloor w'(i,j)/2 \rfloor, & \text{if } \mu'(i,j) \geq 0 \\ w'(i,j), & \text{otherwise} \end{cases} \tag{5.3}$$

where $\lfloor \delta \rfloor$ denotes the largest integer value smaller than δ. By applying eq. 5.3 we can restore the original coefficient values. Inverse integer wavelet transformation and histogram post-extraction operation restores the original content of the cover image.

5.3. Results and Discussion

We have used 10 standard gray scale images to analyze the proposed watermarking technique. In order to generate best evolved GP expressions, we have used Matlab based GPLab toolbox [54]. The proposed technique takes about 2 seconds to

watermark Lena image of size 512x512 using Intel Core 2 Duo 2.4 GHz processor. It is to be noted that since only three sub-bands are chosen for watermark embedding, therefore, the maximum payload that can be achieved is 0.75.

In order to generate the best evolved expressions several GP simulations are carried out. The best expression obtained is presented in the prefix notation as follows:

$$\sin(\div(\log|X(i,j)|,\cos(\cos(\cos(0.73593))))) \tag{5.4}$$

Whereas, the second best expression obtained in prefix notation is:

$$\sin(\div(\log|X(i,j)|,\cos(\cos(\cos(\log(S_{mean})))))) \tag{5.5}$$

Figure 5.4 demonstrates the capability of our reversible watermarking scheme in embedding a large amount of information with insignificant perceptual distortion while restoring the original contents. Figure 5.4(a) and (f) shows the original Lena and River images respectively. They are watermarked using the best evolved expression and the results are displayed in figure 5.4(b) and (g). The difference images in figure 5.4(c) and (h) demonstrates imperceptible embedding using the best evolved expression. After extracting the watermark in the extraction stage, original contents of the image are retrieved. The resultant restored images are shown in figure 5.4(d) and (i). In order to demonstrate the ability of the proposed technique to restore the original contents, its difference image from the original image is taken. For Lena and River images, these results are presented in figure 5.4(e) and (j) respectively. The black region shows that the difference between the original image and the restored image is nil. Whereas, moving towards the white region in grayscale represents the areas of the restored image which are

not the same with respect to the original image. It can be observed that the proposed technique restores almost all of the original contents.

(a)Original Lena image | (b)Watermarked Lena image | (c)Difference image between (a) and (b) | (d)Restored Lena image | (e)Difference image between (a) and (d)

(f)Original River image | (g)Watermarked River image | (h)Difference image between (f) and (g) | (i)Restored River image | (j)Difference image between (f) and (i)

Fig. 5.4 Ability of the proposed technique to embed high energy watermark with minimum perceptual distortion while maintaining reversibility.

In figure 5.5, a comparison of the proposed scheme with that of prior works [33, 34, 36, 37] is presented in terms of watermark payload and imperceptibility using Lena image. As can be observed, the proposed scheme achieves high PSNR compared to the existing techniques. It does so by making an intelligent decision of choosing the right coefficient based on the coefficient value, its position, its neighborhood blocks and the sub-band it belongs to. The other reason for this margin of improvement is that we do not need to change the values of the non-selected coefficients as is done in [36], and consequently less distortion is incurred. It can be observed from the figure that the proposed technique achieves a payload of 0.692 bpp at the PSNR value of 39.832 dB. On

the other hand, even with a PSNR value of 41.2 dB, the proposed technique is still able to embed a watermark with payload=0.6. This demonstrates the significant improvement of this work.

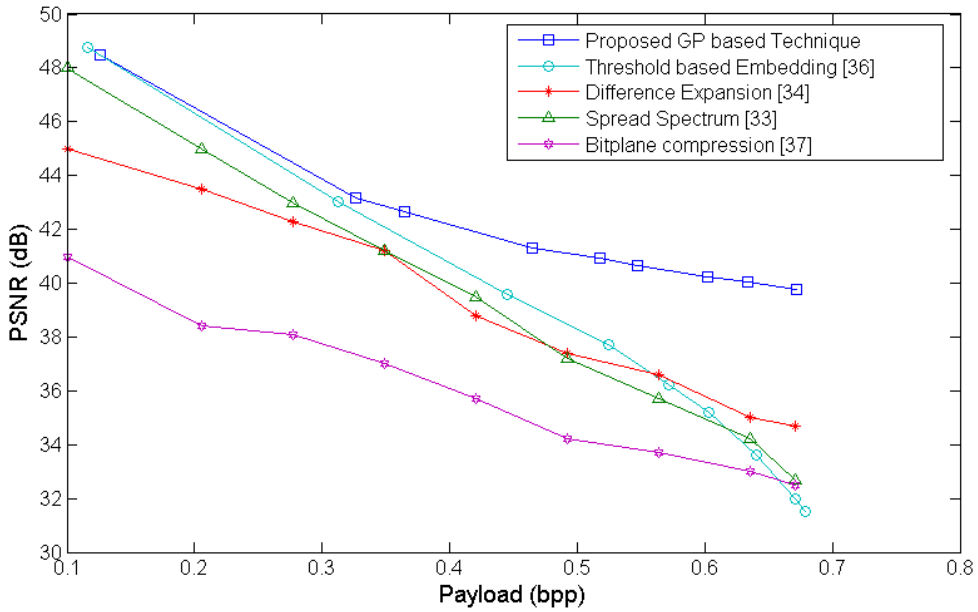


Fig.5.5 Performance comparison with prior techniques in terms of payload and imperceptibility

Table 5.1 presents the performance comparison of the best evolved expression with threshold based embedding [36] (with threshold = 6) for different standard images. Although, the best GP expression is evolved using Lena image alone, high payload and visual quality is achieved using the test images as well. Comparing it to figure 5.5, it can be observed that our technique performs significantly better than the previous techniques for other images as well, and thus, demonstrates the generalization ability of the proposed scheme.

Table 5.1 Performance of the best evolved expression on different images.

Images	Proposed GP based Technique		Threshold based Embedding [36]	
	bpp	PSNR (dB)	bpp	PSNR (dB)
Lena	0.692	39.832	0.603	35.177
River	0.633	36.427	0.541	36.021
Couple	0.639	37.700	0.545	34.855
Boat	0.649	38.560	0.553	35.219
Trees	0.616	38.392	0.518	34.698
Baboon	0.508	38.449	0.354	34.308
Cameraman	0.637	37.149	0.569	36.007
Barbara	0.626	39.112	0.537	35.348
Hill	0.668	36.361	0.552	33.844
House	0.706	39.770	0.632	35.061
Peppers	0.685	35.539	0.596	35.089

5.4. Prospective Applications of the Proposed Scheme.

The proposed scheme is able to produce larger watermark payload with less embedding distortions. In addition, it is able to restore the original image contents. The next generation of digital cameras may include the biological information of a photographer as a watermark in images captured through them. One of the examples includes a recent patent application by cannon [56] which reveals the use of a photographer's iris information for watermarking digital images. In such applications, higher imperceptibility is required with the payload. Since watermark embedding is being performed in the image capturing device, therefore, reversible watermarking is a key technology for retrieving original contents of the image at the time of capturing.

In other applications like electronic patient records and legal imaging, a larger payload is usually required with a higher visual quality. In applications such as these, irreversibility is not always admissible. Mobile context aware medical networks [15, 58] and mobile context aware stories [59] are a recent trend in mobile usage, where image

sizes are small and high payload and imperceptibility is required. The use of a reversible watermarking technique in such scenarios would be highly beneficial.

Besides these applications, the proposed reversible watermarking scheme could be used in all such scenarios where retrieval of the original content is of prime importance along with higher watermark payload and invisibility.

5.5. Summary

In this chapter, an intelligent reversible watermarking technique based on GP and integer wavelet transform is presented. The proposed technique uses GP to evolve optimum expressions that make a delicate balance between watermark capacity and imperceptibility. Furthermore, the reversibility property of the proposed technique makes it extremely useful in applications where irreversible degradation of the original contents after watermark insertion is not admissible. The proposed intelligent coefficient selection scheme can be used in reversible watermarking schemes where a map is used to store information about the embedded coefficient's position. It is robust against the embedding distortion and in future we intend to explore its ability to be robust against a specific attack through its learning mechanism.

6. Conclusions and Future Directions

6

"The hemisphere you deem the sky of your world,
might it be the earth of yet another stratosphere"
Iqbal

This thesis contributes in the field of digital watermarking by applying intelligent techniques in order to achieve optimization between various watermarking constraints. Machine learning is applied in different watermarking scenarios in order to achieve superior results as compared to contemporary techniques. In the first phase of this work, optimum tradeoffs are achieved between watermark robustness and imperceptibility, while keeping the capacity constant. The problem is further analyzed and divided into two distinguished categories namely (i) intelligent watermark embedding and (ii) intelligent watermark embedding as well as extraction. For this purpose, different machine learning techniques are applied which include GP, ANN and SVM. In order to assess and compare the performance of the proposed techniques in phase one, additive spread spectrum based watermarking approach [28] is used as a baseline approach.

In the second phase of this work, we have demonstrated the importance of machine learning in making superior tradeoff between watermark payload and imperceptibility, given the same amount of robustness. For this purpose, we have applied our proposed methodology (intelligent structuring of a watermark) in reversible

watermarking. Primarily due to the fact that reversible watermarking is the next step in digital watermarking and is getting most of the attention from the watermarking community nowadays. Secondly, the need of capacity versus imperceptibility tradeoff can be best taken into account when talking about the reversibility of a system. Therefore, our major concern is to make a capacity versus imperceptibility tradeoff in phase two.

6.1. Intelligent Watermark Tuning

In chapter 3 of this work, we have developed an intelligent technique which genetically evolves appropriate VTFs for watermark embedding in the presence of a series of attacks. These VTFs are intelligent enough to learn the suitable coefficient positions and strength of embedding in different transform domain coefficients. In addition, they are able to shape the watermark in accordance to HVS and a sequence of attacks as expected in real world applications. This is a more realistic approach compared to a single attack, as far as research in digital watermarking is concerned.

Sustaining a series of attacks is a difficult task to achieve because of the fact that each new attack has its own peculiar distortions. The proposed IARW technique presented in chapter 3 counters these problems by introducing the concept of wrapper in GP training phase. Applying an error correction strategy, such as BCH coding, at the embedding stage further improves robustness. IARW technique is easy to implement and is applicable to all watermarking schemes that exploit HVS. In addition, if the developed VTFs are not made public, the proposed approach becomes more secure towards unauthorized decoding. Experimental results demonstrate that intelligent visual tuning of the watermark at the embedding stage outperforms the usage of conventional perceptual

models proposed in literature. This is primarily due to the fact that they are modeled by taking into consideration human perception alone and lack information pertaining to a series of attacks. As a result, they are more inclined towards imperceptibility as compared to robustness, and hence, are unable to deliver a good robustness versus imperceptibility tradeoff. On the other hand, IARW scheme takes into account both human perception as well as hostile attack environment, which gives the margin of improvement.

6.2. Intelligent Embedding and Extraction

In chapter 4 we developed an intelligent AAW system, whereby, intelligent embedding as well as extraction of the watermark is performed using various machine learning approaches. In watermarking community, surviving attacks that are encountered in real world applications is one of the most challenging problems. In order to thwart real world sequence of attacks on the watermarking system, we employ advanced computational intelligence based approaches at two important phases of watermarking. Using GP, we perform intelligent embedding of the watermark by providing flexible and adaptive tradeoff between robustness and imperceptibility. This intelligent embedding not only improves the visual quality of the watermarked image, but also improves its resistance against a series of attacks. These attributes are achieved through the intelligent selection of both appropriate frequency bands and strength of allowable alterations for watermark embedding in the transform domain.

At the watermark extraction phase, SVM and ANNs are applied to adaptively modify the decoding strategy in view of the anticipated sequence of attacks. Conventional approaches [7, 16-18, 23, 25, 28] neither consider the alterations that may

incur to the features nor exploit the individual frequency bands; rather treat all the frequency bands collectively. Consequently, these approaches use a single feature for extraction of a message bit. In case of an attack on the watermarked image, the distributions of the features of a decoding model overlap. As a result, linear bifurcation is not possible and thus a simple TD [28] model is unable to decode the message bits efficiently. We assume that a non-separable message in lower dimensional space might be separable if it is mapped to higher dimensional space. In our proposed AAW scheme, this mapping to higher dimensional space is performed by the hidden layers in case of ANN and the kernel functions in case of SVM. These computational intelligence based models are used to gain knowledge of the distortion that might have incurred varyingly on the different frequency bands due to the attacks. Keeping the same level of imperceptibility, the BCR approaches 1 against a series of attacks. Due to its nonlinear intelligent embedding and decoding, the proposed watermarking approach also enhances the much questioned security of the watermarking system.

The proposed watermarking system can be easily used in medical information handling and sharing, with applications ranging from cooperative working sessions, telediagnosis to telesurgery [55]. The high security risks involved in medical data such as EPRs, and concomitantly, a higher desired visual quality such as in telediagnosis and telesurgery, urges for the use of intelligent watermarking systems to embed the watermark in the most efficient manner. In applications such as these, a high capacity watermark is not always functional to produce high robustness. Rather, exploitation of the hidden dependencies in HVS in respect of the robustness and imperceptibility is a more feasible option.

Training time required for GP based training at the embedding stage, and ANN and SVM based training at the decoding stage is one of the major drawbacks of the proposed technique. Once trained, the proposed watermarking system can be deployed on a chip for broadcast monitoring and device control after generating appropriate VTFs in view of the application environment, and hence the conceivable attacks. E-democracy [57], Mobile context aware medical networks [58] and mobile context aware multimedia stories [59] are some of the recent trends in mobile usage. The use of intelligent watermarking is highly desirable in applications where imperceptibility and robustness requirements constantly change. Furthermore, it is worthwhile in applications where high robustness is required with no compromise on imperceptibility. An appropriate example of such an application could be watermarking a database of faces in a face recognition system [60].

6.3. Reversible Watermarking using Machine Learning

In the second phase of this work we developed an intelligent reversible watermarking system based on GP and integer wavelet transform. The proposed technique uses GP to evolve optimum expressions which make a delicate balance between watermark capacity and imperceptibility. During GP training phase, several dependencies are taken into account for selecting the appropriate coefficient that would yield high imperceptibility. These include the absolute value of coefficient in question, its sub band mean, block number, block mean, and sub band label etc. During evolutionary phase, GP is able to guide the search space towards the optimum choice between the number of coefficients selected for embedding and imperceptibility. Furthermore, the

reversibility property of the proposed technique makes it extremely useful in applications where irreversible degradation of the original contents after watermark insertion is not admissible. As an example, the next generation of digital cameras is taking the photographer's copyright to a level further by acquiring and embedding biological information of a photographer in a digital image. One of the examples include a recent patent application by cannon [56] which reveals the use of a photographer's iris information for watermarking digital images. In applications like these, a high imperceptibility is required since the watermark is being embedded at the image capturing stage. Because the watermark is being embedded at the source, the only technology that can deliver, later on, the exact content of the scene at the time of capturing is reversible watermarking. Our proposed watermarking technique may be implemented in the camera hardware and a pre-stored user profile based on iris information may be utilized to watermark batch files after each photographic session.

6.4. Future Directions

Nowadays, there is a constant competition between a watermark embedder and an adversary. Additionally, due to the rapid expansion of new, dynamic, and complex watermarking applications, fixed watermarking strategies may become obsolete. Few potential watermarking applications, where dynamic and adaptive watermarking strategies are highly desirable are; print-to-web technology [63], transaction tracking [64, 67], broadcast monitoring [65, 66], securing data sent through context aware medical networks [15, 58], secure digital camera [56], data transmission through wireless networks, agent based online auction systems etc. Such watermarking applications have

to be guarded against both intentional and unintentional attacks that may change with respect to time. For example, in case of context aware medical networks, the images may be compressed and transmitted at different rates at one instant of time, but this may rapidly change depending on the network congestion and/or the requirement of the physician. In cases, where the type of attack is the same, but only its intensity changes with time, the retraining of only the machine learning based decoding model is recommended. There may be no need of the retraining of the GP based embedding phase. However, in applications, where the type of attack also changes dynamically, then both embedding and decoding models should be retrained for coping with the distortions introduced by the dynamic attacks.

The only drawback of the proposed system is that it is not suitable for geometric attacks, which constitute an important class of attacks. In future, we aim to develop appropriate VTFs that counter geometric attacks as well. For this purpose, statistical moments can be utilized [69, 70]. A more realistic approach would be to take into account the concepts of game theory and data classification [71, 72].

As far as intelligent reversible watermarking is concerned, in future we intend to explore its ability to be robust against a specific attack through its learning mechanism. Reversible watermarking is also susceptible to different attacks in real watermarking applications. We intend to incorporate single attack information in its learning mechanism first and later enhance it to a series of attacks.

Another important aspect that could be taken into account is the security of the watermark itself. For this purpose, different encryption strategies, such as RSA encryption, can be employed on the watermark before embedding. But the drawback of

an encryption strategy is the increase in size of the encrypted watermark. This issue can be resolved in future by adding another layer of lossless compression after watermark encryption. Concluding, there are many aspects which still need to be investigated in future; covering them all is beyond the scope of this thesis.

References

[1] I. J. Cox, M. L. Miller, and J.A. Bloom, *Digital Watermarking and fundamentals*. 2002, San Francisco: Morgan Kaufmann.

[2] Herodotus, *The Histories*. 1996, London: Penguin Books, Translated by Aubrey de Selincourt.

[3] I. J. Cox and M. L. Miller, *The first 50 years of electronic watermarking.* EURASIP Journal on Applied Signal Processing, 2002. **2002**(2): p. 126 - 132

[4] Hembrooke, E.F., *Identification of sound and like signals*, U.S. Patent, Editor. 1961: USA.

[5] K. Tanaka, Y.Nakamura, and K. Matsui,. *Embedding secret information into a dithered multi-level image*. in *IEEE Military Communications Conference*. 1990.

[6] A. Tirkel, G. Rankin, R. van Schyndel, W. Ho, N. Mee, and C. Osborne,. *Electronic water mark*. in *DICTA*. 1993.

[7] A. Piva, M. Barni, F. Bartolini, and V. Cappellini,. *DCT-based watermark recovering without resorting to the uncorrupted original image*. in *International Conference on Image Processing*. 1997. Washington, USA.

[8] M. Barni and F. Bartolini, *Watermarking systems engineering: Enabling digital assets security and other application*. 2004, New York: Marcel Dekker.

[9] H. C. Huang, C. J. Lakhmi, and J.S. Pan, *Intelligent Watermarking Techniques*. 2004: World Scientific Publishing Company Inc.

[10] B. Chen, G. W. Wornell, *Quantization index modulation: A class of provably good methods for digital watermarking and information embedding.* IEEE Trans. Inf. Theory, 2001. **47**(4): p. 1423-1443.

[11] A. B. Watson. *Visual optimization of DCT quantization matrices for individual images*. in *Proceedings of AIAA Computing in Aerospace 9*. 1993. San Diego, CA.

[12] J. A. Solomon, A. B. Watson, and A. J. Ahumada, *Visibility of DCT basis functions: Effects of contrast masking*. in *Proceedings of Data Compression Conference*. 1994. Snowbird, UT.

[13] C. J. V. D. B. Lambrecht and J.E. Farrell. *Perceptual Quality Metric for digitally coded color images*. in *Proceedings of the European Signal Processing Conference*. 1996. Trieste, Italy.

[14] A. B. Watson, G. Y. Yang, J. A. Solomon, J. Villasenor, *Visibility of Wavelet Quantization noise.* IEEE Transactions on Image Processing, 1997. **6**(8): p. 1164-1175.

[15] C. Doukas and I. Maglogiannis, *Adaptive transmission of medical image and video using scalable coding and context-aware wireless medical networks.* EURASIP Journal on Wireless Communications and Networking, 2008, doi: 10.1155/2008/428397.

[16] A. Khan and A. M. Mirza, *Genetic perceptual shaping: Utilizing cover image and conceivable attack information using Genetic Programming.* Information Fusion, 2007. **8**(4): p. 354-365.

[17] C. S. Shieh, et al., *Genetic watermarking based on transform-domain techniques.* Pattern Recognition, 2004. **37**(3): p. 555-565.

[18] V. Aslantas, A. L. Dogan, and S. Ozturk. *DWT-SVD based image watermarking using particle swarm optimizer*. in *Proc. IEEE International Conference on Multimedia and Ex*. 2008. Hannover, Germany.

[19] P. Laskov, C. Gehl, and S. Kruger, *Incremental support vector learning: Analysis, implementation and applications.* Journal of Machine Learning Research, 2006. **7**: p. 1909–1936.

[20] W. C. Chu, *DCT-based image watermarking using sub sampling.* IEEE Transactions on Multimedia, 2003. **5**(1): p. 34–38.

[21] J. J. K. O'Ruanaidh, W. J. Dowling, and F.M. Boland. *Phase watermarking of digital image*. in *Proceedings of the IEEE International Conference on Image Processing 3*. 1996. Lausanne, Switzerland.

[22] M. Barni, F. Bartolini, and A. Piva, *Improved wavelet-based watermarking through pixel-wise masking.* IEEE Transactions on Image Processing, 2001. **10**(5): p. 783–791.

[23] H. C. Huang, F. H. Wang, and J.S. Pan, *Efficient and robust watermarking algorithm with vector quantization.* Electronic Letters, 2001. **37**(13): p. 826–828.

[24] N. Nikolaidis and I. Pitas, *Robust image watermarking in the spatial domain.* Signal Processing, 1998. **66**(3): p. 385–403.

[25] C. I. Podilchuk and W. J. Zeng, *Image-adaptive watermarking using visual models.* IEEE Journal of Selected Areas in Communication, 1998. **16**(4): p. 525–539.

[26] Y. Fu, R. Shen, and H. Lu, *Optimal watermark detection based on support vector machines.* Lecture Notes in Computer Science, 2004. **3137**: p. 552-557.

[27] S. Bounkong, B. Toch, D. Saad, and D. Lowe, *ICA for watermarking digital images.* Journal of Machine Learning Research, 2003. **4**: p. 1471-1498.

[28] J. R. Hernandez, M. Amado, and F. Perez-Gonzalez, *DCT domain watermarking techniques for still images: Detector performance analysis and a new structure.* IEEE Transactions on Image Processing, 2000. **9**(1): p. 55–68.

[29] M. Barni, F.B., A. D. Rosa, and A. Piva,, *A new decoder for the optimum recovery of non-additive watermarks.* IEEE Transactions on Image Processing, 2001. **10**(5): p. 755–766.

[30] A. Khan, et al., *Machine learning based adaptive watermark decoding in view of an anticipated attack.* Pattern Recognition, 2008. **41**(8): p. 2594-2610.

[31] C. C. Chen and D.S. Kao, *DCT-based zero replacement reversible image watermarking approach.* International Journal of Innovative Computing, Information and Control, 2008. **4**(11): p. 3027-3036.

[32] Z. Ni, et al. *Reversible data hiding*. in *Proceedings IEEE International Symposium on Circuits and Systems*. 2003. Bangkok, Thailand.

[33] G. Xuan, Y. Q. Shi, and Z. Ni. *Lossless data hiding using integer wavelet transform and spread spectrum*. in *IEEE International Workshop on Multimedia Signal Processing*. 2004. Siena, Italy.

[34] J.Tian, *Reversible data embedding using a difference expansion.* IEEE Trans. Circuits and Systems for Video Technology, 2003. **13**(8): p. 890-896.

[35] Alattar, A.M., *Reversible watermarking using the difference expansion of a generalized integer transform.* IEEE Trans. Image Process., 2004. **13**(8): p. 1147-1156.

[36] G. Xuan, et al. *Lossless data hiding using integer wavelet transform and threshold embedding technique*. in *IEEE international Conference on Multimedia and Expo*. 2005.

[37] G. Xuan, et al., *Distortionless data hiding based on integer wavelet transform.* IEE Electronics Letters, 2002: p. 1646-1648.

[38] S. Voloshynovskiy, et al., *A stochastic approach to content adaptive digital image watermarking.* Lecture Notes in Computer Science, 1999. **1768**: p. 211-236.

[39] Z. Wang, A. C. Bovik, and H.R. Sheikh, *Image quality assessment: From error visibility to structure similarity.* IEEE Transactions on Image Processing, 2004. **13**(4): p. 600-612.

[40] F. I. Koprulu, *Application to low-density parity-check codes to watermark channels*. 2001, Bogaziei University: Turkey.

[41] A. J. Ahumada and H.A. Peterson. *Luminance-model-based DCT quantization for color image compression*. in *Human Vision, Visual Processing, and Digital Display III*. 1992. San Jose, CA, USA

[42] A. B. Watson, et al. *Visual thresholds for wavelet quantization error*. in *Human Vision and Electronic Imaging, SPIE*. 1996.

[43] M. Kutter and S. Winkler, *A vision-based masking model for spread-spectrum image watermarking.* IEEE Transactions on Image Processing, 2002. **11**(1): p. 16-25.

[44] A. M. Alattar and O.M. Alattar. *Watermarking electronic text documents containing justified paragraphs and irregular line spacing*. in *Security, stenography, and watermarking of multimedia contents VI* 2004. San Jose CA.

[45] S. G. Miaou, T. S. Lee, and C.M. Chen, *BCH coded watermarks for error-prone transmission of MPEG video.* Lecture Notes in Computer Science, 2001. **2195**: p. 654-661.

[46] A. Bastug and B. Sankur, *Improving the payload of watermarking channels via LDPC coding.* IEEE Signal Processing Letters, 2004. **11**(2): p. 90-92.

[47] Todorov, T., *Improving the Watermarking Process With Usage of Block Error-Correcting Codes*. 2005, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences.

[48] S. Voloshynovskiy, et al., *Attack modeling: towards a second generation watermarking benchmark.* Signal Processing,, 2001. **81**(6): p. 1177-1214.

[49] S. Pereira, et al., *Second Generation Benchmarking and Application Oriented Evaluation* Lecture Notes in Computer Science, 2001. **2137**: p. 340-353.

[50] W. Benzhaf, et al., *Genetic Programming An Introduction: On the Automatic Evolution of Computer Programs and Its Applications*. 1998, San Francisco, California: Morgan Kaufmanns Publishers, Inc.

[51] J. R. Koza, M.A.K., M. J. Streeter, M. Mydlowec, and J. Yu, G. Lanza,, *Genetic Programming IV: Routine Human-Competitive Machine Intelligence*. 2005, New York: Springer Science + Business Media Inc.

[52] Sklar, B., *Digital Communications: Fundamentals and Applications*. second ed. 2001: Prentice-Hall Inc.

[53] S. Voloshynovskiy, et al. *A stochastic approach to content adaptive digital image watermarking*. in *Proc. International Workshop on Information Hiding*. 1999. Dresden, Germany.

[54] Silva, S. *GPLAB toolbox and user's manual version 2.1, http://gplab.sourceforge.net/ download.html*, accessed on October 2006.

[55] G. Coatrieux, et al. *A Review of Image Watermarking Applications in Healthcare*. in *Proc. IEEE Int. Conference on Engineering in Medicine and Biology*. 2006. New York.

[56] G. Morikowa and G. Tokura, *Picture taking apparatus and method of controlling same*, in *application no. US 2008/0025574A1*, US Patent and Trademark Office, Editor. 2008: USA.

[57] G. Cabri, L. Ferrari, and L. Leonardi, *A role-based mobile-agent approach to support e-democracy.* Applied Soft Computing, 2005. **6**: p. 85–99.

[58] C. N. Doukas, I. Maglogiannis, and T. Pliakas. *Advanced Medical Video Services through Context-Aware Medical Networks.* in *Proc. The 29th Annual International Conference of the IEEE EMBS Cité Internationale*. 2007. Lyon, France.

[59] L. Doyle, G. Davenport, and D.O. Mahony. *Mobile Context Aware Stories*. in *Proc. IEEE conference on Multimedia and Expo*. 2002. Lausanne, Switzerland.

[60] J. K. Sing, et al., *Face recognition using point symmetry distance-based RBF network.* Applied Soft Computing 2007. **7**: p. 58–70.

[61] Laboratory, A.N., MPICH2, *http://www-unix.mcs.anl.gov/mpi/mpich2/*. 2007.

[62] M. Zhang, V.B. Ciesielski, and P. Andreae, *A Domain-Independent Window Approach to Multiclass Object Detection Using Genetic Programming.* EURASIP Journal on Applied Signal Processing, 2003. **8**: p. 841–859.

[63] Liu, Z., *Print-to-Web Linking Technology Using a Mobile Phone Camera and its Applications in Japan*. Studies in Computational Intelligence. Vol. 58. 2007: Springer Berlin / Heidelberg.

[64] D. Boneh and J. Shaw, *Collusion-secure fingerprinting for digital data.* Lecture Notes in Computer Science, 1995. **963**: p. 452–465.

[65] H. C. Papadopoulos and C. E. W. Sundberg. *Simultaneous broadcasting of analog fm and digital audio signals by means of precanceling techniques*. in *IEEE Int. Conf. on Communications, ICC 98*. 1998.

[66] Sundberg, B.C.a.C.-E., *Digital audio broadcasting in the FM band by means of contiguous band insertion and precanceling techniques.* IEEE Trans. on Communications, 2000. **48**(10): p. 1634–1637.

[67] H. Fronthaler, K.K., J. Bigun, J. Fierrez, F. A. Fernandez, J. O. Garcia, and J. G. Rodriguez,, *Fingerprint image-quality estimation and its application to multi-algorithm verification.* IEEE Transactions on Information Forensics and Security, 2008. **3**(2): p. 331-338.

[68] C. Belcher and Y. Du, *A selective feature information approach for iris image-quality measure.* IEEE Transactions on Information Forensics and Security, 2008. **3**(3): p. 572-577.

[69] L. Parameswaran and K. Anbumani, *A Content Based Image Watermarking Scheme Resilient to Geometric Attacks.* International Journal of Computer Science. **2**(2): p. 118-123.

[70] M. Alghoniemy and A. H. Tewfik. *Geometric distortion through Image Normalization*. in *Proceedings of International Conference on Multimedia Expo*. 2000.

[71] A. Majid, A. Khan, and A.M. Mirza, *Combination of support vector machines using genetic programming*. International Journal of Hybrid Intelligent Systems 3 (2) (2006) 109-125.

[72] A. Majid, *Optimization and combination of classifiers using Genetic Programming*, Faculty of Computer Science, PhD thesis, GIK institute, Pakistan, 2006.