1-1-2005

Signal processing techniques for multimedia information security

Arunan Ramalingam
Ryerson University

SIGNAL PROCESSING TECHNIQUES FOR MULTIMEDIA INFORMATION SECURITY

by

Arunan Ramalingam
B.E., Anna University, India, 2001

A thesis
presented to Ryerson University

in partial fulfillment of the
requirements for the degree of
Master of Applied Science
in the Program of
Electrical and Computer Engineering

Toronto, Ontario, Canada, 2005

Author's Declaration

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Author's Signature: _____

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Author's Signature: _____

Abstract

Signal Processing Techniques for Multimedia Information Security

Master of Applied Science
Department of Electrical and Computer Engineering
Ryerson University

The digital representation of multimedia and the Internet allows for the unauthorized duplication, transmission, and wide distribution of copyrighted multimedia content in an effortless manner. Content providers are faced with the challenge of how to protect their electronic content. Fingerprinting and watermarking are two techniques that help identify content that are copied and distributed illegally. This thesis presents a novel algorithm for each of these two content protection techniques.

In fingerprinting, a novel algorithm that model fingerprints using Gaussian mixtures is developed for both for audio and video signals. Simulation studies are used to evaluate the effectiveness of the algorithm in generating fingerprints that show high discrimination among different fingerprints and at the same time invariant to different distortions of the same fingerprint.

In the proposed watermarking scheme, linear chirps are used as watermark messages. The watermark is embedded and detected by spread-spectrum watermarking. At the receiver, a post processing tool represents the retrieved watermark in a time-frequency distribution and uses a line detection algorithm to detect the watermark. The robustness of the watermark is demonstrated by extracting the watermark after different image processing operations performed using a third party evaluation tool called checkmark.

Acknowledgments

I would like to sincerely thank my advisor Dr. Sridhar Krishnan for bringing the problem of fingerprinting and watermarking to my attention, and for his constant support, encouragement, and feedback during the development of the work. In the last two years, he shared his ideas and time with me generously. I gratefully acknowledge his support.

I am indebted to Karthikeyan Umapathy for his critical evaluation of my research at different stages and for his timely research tips and motivation throughout my research.

I would like to thank the members of the Signal Analysis Research Group — Jiming Yang, Lam Le, April Khademi, Danoush Hosseinzadeh for providing a stimulating research environment.

I would like to acknowledge Micronet R&D, Canada and the Department of Electrical and Computer Engineering, Ryerson University for providing financial support during my graduate study.

Lastly, but certainly not least, I would like to thank my parents for their constant support, encouragement during the course of my studies.

Dedication

To my mom and dad for their love, support, and dedication ...

Contents

List of Tables

List of Figures

List of Acronyms

AAC	Advanced Audio Coding
A/D	Analog to Digital
AM	Amplitude Modulation
CD	Compact Disc
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DRM	Digital Rights Management
DVD	Digital Versatile Disk
EDM	Electronic Distribution of Multimedia
EM	Expectation Maximization
FFT	Fast Fourier Transform
FM	Frequency Modulation
FT	Fourier Transform
GMM	Gaussian Mixture Models
HMM	Hidden Markov Models
HRT	Hough-Radon Transform
JND	Just Noticeable Difference
JPEG	Joint Photographic Experts Group
MCLT	Modulated Complex Lapped Transform
MDS	Media Description Schemes
MFCC	Mel-frequency Cepstral Coefficient
MP3	Motion Pictures Experts Group 1 Layer - 3
MPAA	Motion Picture Association of America
MPEG	Motion Pictures Experts Group
OPCA	Oriented Principal Component Analysis
P2P	Peer-to-Peer
PC	Personal Computer
PCM	Pulse Code Modulation
PDF	Probability Density Function
PN	Pseudo-Random
RBF	Radial Basis Function
RE	Renyi Entropy

RIAA	Recording Industry Association of America
RMS	Root Mean Square
SB	Spectral Bandwidth
SBE	Spectral Band Energy
SC	Spectral Centroid
SCF	Spectral Crest Factor
SE	Shannon Entropy
SF	Spectral Flux
SFM	Spectral Flatness Measure
SRF	Spectral Roll-off Frequency
SS	Spread Spectrum
STFT	Short-Time Fourier Transform
TF	Time-Frequency
TTP	Trusted Third Party
VQ	Vector Quantization
WMA	Windows Media Audio
WVD	Wigner-Ville Distribution
XML	Extensible Markup Language
ZCR	Zero Crossing Rate

Chapter 1

Introduction

THE electronic distribution of multimedia (EDM) through the Internet offers many advantages to content sellers as well as consumers. Due to the digital representation, sellers face reduced cost of manufacturing, transportation, storage, and display. They can reach a larger number of users when compared to distributing by compact disc (CD) or Digital Versatile Disk (DVD). Consumers, apart from the reduced cost, enjoy numerous benefits. They can have access to large collection of multimedia files, and can purchase and enjoy multimedia content instantly. Obtaining music online also enables consumers to have more control over what they listen to. They can buy individual singles rather than the whole album. By ordering these songs together, consumers can create their own listening experiences and bypassing the context in which artists envisaged their work would be listened to when purchased.

Despite these potential advantages, both the music and the movie industries are reluctant to distribute multimedia content through Internet. One of the reasons is that these industries are afraid of change. But these industries will eventually accept that EDM will be a significant distribution channel in the future. So the main obstacle in the implementation of EDM is piracy. While there are many advantages associated with digital media and digital media distribution, clear disadvantages are present. Prior to digital technologies, content was created, displayed, and stored in analog means. The advent of personal video recorder in the 1980s presented an opportunity to view video at home; but also provided an oppor-

tunity to make an illegal copy. However, when a copy is made from a recorded content, the new copy is inferior in quality to the original one. Any further copies made from that copy are very much reduced in quality to be of any commercial use. This discouraged people from copying and prevented piracy efforts from reaching alarming proportions. The new digital technologies represent multimedia content in digital format (1s and 0s). These bits can be efficiently stored in an optical or magnetic media. Since digital recording is a process where by each bit in the source stream is read and copied to the new medium, an exact replica of the content is obtained. The digital copies can be created with low cost equipment such as a CD recorder.

Though the digital representation helped to make identical copies easily, the full resolution multimedia files comprise large amounts of data. Transferring or storing them took a large amount of bandwidth. In the beginning of the 1990s, several compression technologies were developed and new standards such as Motion Picture Experts Group (MPEG) for video and MPEG 1 Layer - 3 (MP3) for audio reduced the size of the multimedia files by an order of magnitude. This coupled with the reduction in the cost of storage media allowed PC users to have thousands of songs or movies stored in their computers. Toward the end of the 1990s, the increasing availability of high-speed Internet provided an easy and cheap way of distributing movies and songs. The development of peer-to-peer (P2P) networks (such as Kazaa [1], BitTorrent [2], eDonkey [3]) to exchange files helped Internet users an easy way to search and exchange multimedia files effortlessly in the Internet. The illegal sharing of multimedia content incurred severe loses to the copyright holders. According to the Recording Industry Association of America (RIAA), the volume of sold audio CDs dropped by 5% in 2001 [4] and by 11% in the first half of 2002 [5]. Motion Picture Association of America (MPAA) estimates that the movie industry annually loses US $3 billion through physical good piracy [6]. This figure does not include Internet piracy. Hence developing digital rights management (DRM) to protect digital multimedia content is a crucial problem for which immediate solutions are needed.

1.1 Digital Rights Management

DRM is a collection of commercial, legal, and technical measures that enable technically enforced licensing of digital information. DRM makes it possible for content distributors to distribute valuable content electronically, without destroying the copyright holders revenue stream. There is no unique definition for DRM. Some of the definitions are [7]

> 'the technologies, tools and processes that protect intellectual property during digital content commerce.'
>
> - The Association of American Publishers

> 'a system of information technology (IT) components and services that strive to distribute and control digital products.'
>
> - Gordon

> 'digital rights management entails the operation of a control system that can monitor, regulate, and price each subsequent use of a computer file that contains media content, such as video, audio, photos, or print.'
>
> - Einhorn

Though DRM can be designed to protect any digital information, in this thesis, the focus is restricted only on the technologies that are designed to protect multimedia content. DRM ensures that access to protected content (such as video or audio) is possible only under the conditions specified by the content owner. Any unauthorized access must be prevented because such access is an opportunity for an unprotected version of the content to be obtained. If unprotected content is obtained, then it can be distributed and used in any manner, bypassing DRM. DRM prevents the creation of unauthorized copies (copy protection) and provides a mechanism by which copies can be detected and traced (content tracking). The following section describes the various components of DRM.

1.2 Components of DRM

A typical DRM for multimedia has the following components as shown in Figure 1.1. The figure shows only the technical aspects of DRM. Other than the technical aspects, DRM includes legislative and regulatory solutions for copy protection and management of digital rights.

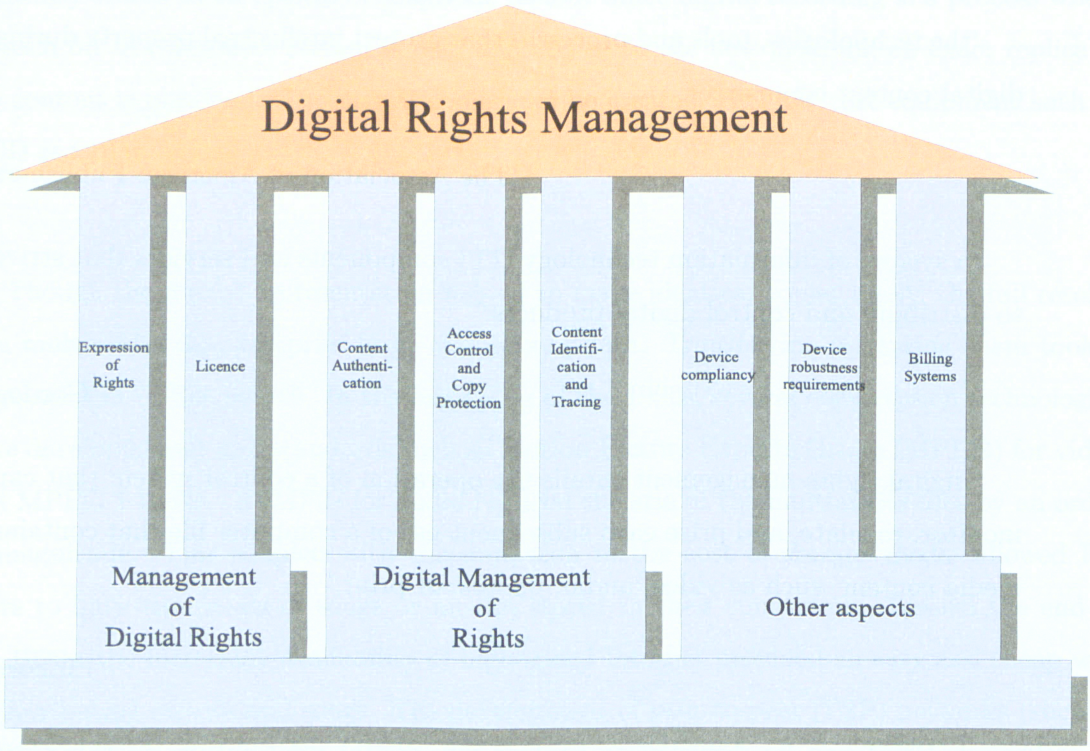

Figure 1.1: Components of DRM

1.2.1 Management of Digital Rights

The management of digital rights include expression of rights and distribution of rights.

- **Expression of rights:** The content producer expresses the conditions under which the content can be accessed using rights expression languages (REL) such as the open digital rights language (ORDL) and the eXtensible rights Markup Language (XrML).

In DRM the usage rules can be adapted to the business models. For example, access can be restricted to selected users, a limited time, or a limited number of accesses. Initial access to the data may even be free (e.g.,the first playback of an audio track), while subsequent access has to be paid for.

- **Licenses:** Licenses are used as a mechanism to distribute rights that are expressed in the rights language. By bringing a license and a digital item together in a device, the device can inspect the right to see what it may do with the digital item.

1.2.2 Digital Management of Rights

The digital management of rights include

- **Content authentication:** DRM should protect the authenticity and integrity of the content. Authenticity is securing the content what it claims to be. Integrity means securing the content from any alteration during distribution to the consumer.

- **Access control and copy protection:** DRM controls who has access to the content and how the content has been used. DRM should prevent the illegal copying of the content once it has been decrypted. Depending on the usage rules, no/one/several/unlimited copies of the multimedia data are allowed, with or without the right to produce copies of the copies. DRM enforces those copy restrictions using sophisticated technology such as watermarking.

- **Identification and tracing:** In DRM, the authorized users have access to play the content. The playback medium is analog and it is possible for the users to make copies from the analog output. Thus, analog copies in general can hardly be prevented. But it is possible to identify and trace back analog and digital copies of distributed media. This can be done by individual digital watermarking (traitor-tracing) of the distributed data.

1.2.3 Other Aspects

- **Device compliancy:** DRM rely on device compliancy to function properly. Device compliancy requires that devices that implement (part of) DRM functionality function according to the rule imposed by DRM. This means that devices do not access digital items in case of absence of a license and also that they do not carry out operations on digital items that are not allowed by the associated rights (e.g., copying or sending content in the clear over unprotected links).

- **Device robustness requirements:** Since devices manipulate licenses, rights, and keys, the manipulation and storage of these items needs to take place in a secure environment. As a result, DRM imposes hardware and software tamper resistance requirements on devices.

- **Billing systems:** The business models for media distribution usually involves monetary transactions. Therefore, DRM should contain mechanisms to perform those transactions. Billing systems should be able to handle different pricing models such as pay per use, monthly subscription.

1.3 Technologies for Digital Management of Rights

The technologies used for the digital management of rights include encryption, fingerprinting and watermarking and are described in the following paragraphs. The technologies used in a multimedia DRM are shown in Figure 1.2.

1.3.1 Encryption

In DRM, encryption can be used to package the content securely and force all accesses to the protected content through the rules enforced by DRM. If the content is not packaged securely, the content could be easily copied. Encryption scrambles the content and renders the content unintelligible unless a decryption key is known. Once the content has been encrypted, the encrypted content may be transmitted over a distribution network or recorded onto a media.

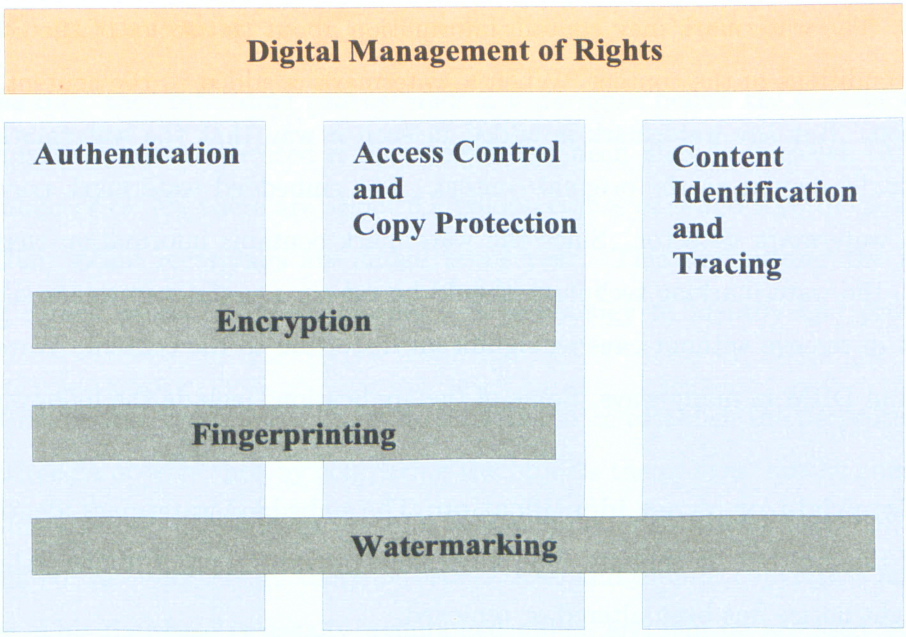

Figure 1.2: DRM technologies

The encrypted content cannot be decoded or displayed without the decryption key. When the content is to be decoded and displayed, the decryption key is provided to the decoder only after DRM has verified that the conditions for accessing to the content are satisfied. In this way, DRM can secure the distribution of content and ensure that access to the content is consistent with the usage rights. Other examples of using encryption in DRM include device authentication and the secure exchange of keys and authorization information.

1.3.2 Watermarking

Encryption packages the content securely and provides restricted access to the content. However, once an authorized user has decrypted the content, it does not provide any protection to the decrypted content. Encryption does not prevent an authorized user from making and distributing illegal copies. Watermarking and fingerprinting are two technologies that can provide protection to the data after it has been decrypted.

A watermark is a signal that is embedded in the content to produce a watermarked

content. The watermark may contain information about the owner of the content and the access conditions of the content. When a watermark is added to the content, it introduces distortion. But the watermark is added in such a way that the watermarked content is perceptually similar to the original content. The embedded watermark may be extracted using a watermark detector. Since the watermark contains information that protects the content, the watermarking technique should be *robust*, *i.e.* the watermark signal should be difficult to remove without causing significant distortion to the content. Watermarking can be used in DRM in many ways. Some of the applications include the following [8].

- **Copyright or owner identification:** The embedded watermark identifies the owner of the multimedia content. The watermark provides a proof of ownership if the copyright notice has been altered or removed.

- **Copy protection:** The watermark encodes the number of times the multimedia content may be (legally) copied. A compliant device checks the watermark and determines whether creating an additional copy is allowed. Each time a copy is made, the watermarked content is modified to decrement the count of allowable copies.

- **Access control:** The watermark encodes the usage and access rights that are granted by the content owner. Compliant devices detect the watermark and complies with the encoded usage restrictions.

- **Content tracking:** In many applications the copyrighted content may be distributed to many users. Some of the users may copy the content and distribute. To identify the particular user who pirates content, the watermark encodes the identification of the user or recipient of the video. This implies that each user obtains a unique or personalized copy of the video. If a copy of the video is found in a suspicious location (such as being shared by a peer-to-peer program), the embedded watermark can identify the source of the suspected copies.

1.3.3 Fingerprinting

In watermarking, the embedding process adds a watermark before the content is released. But watermarking cannot be used if the content has been already released. According to Venkatachalam *et al.* [9], there are about 0.5 trillion copies of sound recordings in existence and 20 billion sound recordings are added every year. This underscores the importance of securing legacy content. Fingerprinting is a technology to identify and protect legacy content.

In multimedia fingerprinting [1], the main objective is to establish the perceptual equality of two multimedia objects: not by comparing the objects themselves, but by comparing the associated fingerprints. The fingerprints of a large number of multimedia objects, along with their associated meta-data (e.g., name of artist, title, and album, copyright) are stored in a database. This database is usually maintained online and can be accessed by recording devices.

Fingerprinting can be used in DRM in many ways. Some of the applications include the following.

- **Copyright or owner identification:** When an audio or video content is played in a compliant device, the device can extract a fingerprint from the content and compare with the database and retrieve the owner and the copyright information of the content.
- **Copy protection:** As in copyright identification, a compliant device checks the fingerprint in the database and determines whether creating an additional copy is allowed.
- **Filtering:** In P2P networks, the exchanged multimedia files can be monitored. The fingerprints of the shared files are extracted and identified by comparing with the database. If the exchanged content is a copyrighted one, then the exchange could be blocked.

[1]The term "fingerprinting" has also been employed as a special case of watermarking in which each legal copy of a multimedia content is watermarked uniquely. However, the same term has been used to name techniques that associate a multimedia signal to a much shorter numeric sequence (the "fingerprint") and use this sequence to identify the content. In this thesis, the term fingerprinting refers to the latter meaning.

1.4 Organization of the Thesis

The main objective of this thesis is to propose novel algorithms to identify and protect multimedia content from unauthorized use. The remainder of the thesis is organized as shown in Figure 1.3.

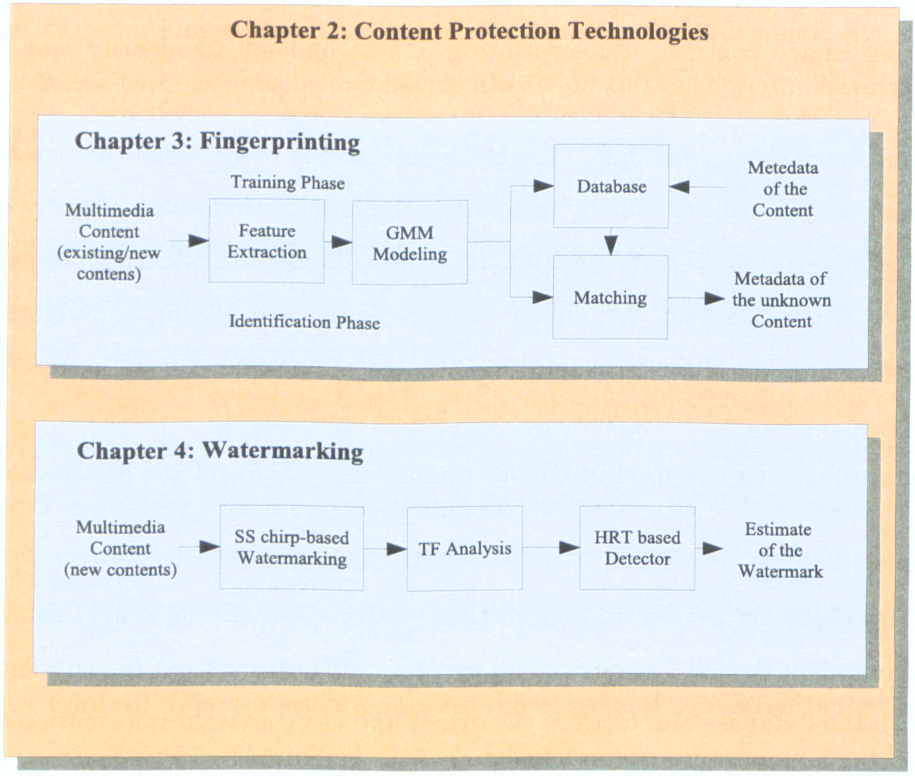

Figure 1.3: Organization of the thesis

Chapter 2 describes the main content protection technologies such as encryption, fingerprinting and watermarking. Encryption provides security to data during transmission and storage. Once the content is decrypted, content can be protected by fingerprinting and watermarking. These two technologies are described in detail along with their applications.

Chapter 3 proposes a novel technique to design fingerprints that are compact, provide discrimination among different multimedia clips, and at the same time, invariant to the distorted versions of the same clips. The fingerprints are generated by modeling audio

or video features using Gaussian mixture models (GMM). Simulation results are used to demonstrate that the fingerprints are invariant to a wide range of distortions and provide very good discrimination among different fingerprints.

Chapter 4 proposes a novel robust image watermarking algorithm that embeds multiple watermark bits in the host image. Linear chirps are embedded as watermark messages, where the slopes of the chirp on the time-frequency (TF) plane represent watermark messages, such that each slope corresponds to a unique message. The Hough-Radon transform (HRT) is a widely used tool to detect directional elements that satisfy a parametric constraint in images. Since linear chirps are localized as a straight line in the TF plane, the HRT is used to detect the watermark messages at the receiver. The HRT robustly estimates the slope of the chirps in the presence of any attacks. Simulation results show the improved performance of the scheme to a wider variety of distortions such as compression, resampling, and filtering.

Chapter 5 summarizes the important results in the proposed schemes and points out directions for future research.

Chapter 2

Content Protection Technologies

THE introduction briefly described about the technologies used for the DRM namely encryption, fingerprinting and watermarking. These are essentially content protection technologies. While encryption protects the content during transmission from the producer to the consumer, fingerprinting and watermarking protects the content once the content reaches the consumer. Encryption is a mature area and the interested reader is referred to the texts [10] and [11]. The remainder of the thesis will focus only on fingerprinting and watermarking.

From a signal processing perspective, both fingerprinting and watermarking share similar steps as shown in Figure 2.1. The objective in both these techniques is to identify a multimedia signal from a distorted version of the original signal. In fingerprinting, features are extracted from the original signal and its distorted versions and they are compared to identify the signal. In watermarking, features are embedded to the original signal. Then, features are extracted from its distorted versions and compared with the inserted features to identify the original signal. The remainder of the chapter gives a description of these technologies and their applications in detail.

2.1 Fingerprinting

The primary objective of fingerprinting is to establish a mechanism to evaluate the perceptual similarity of two multimedia signals by comparing the digests of the signals called

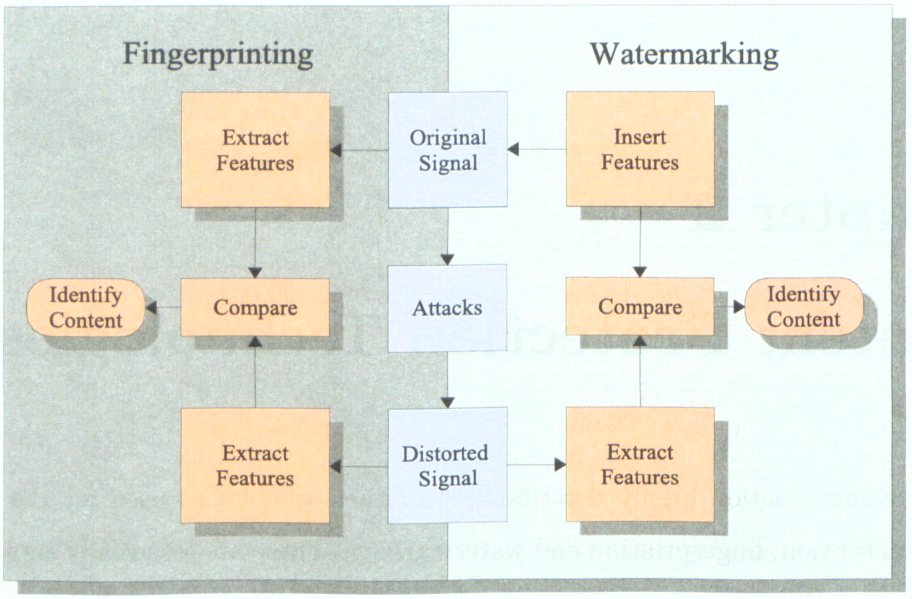

Figure 2.1: Fingerprinting and Watermarking

fingerprints. The fingerprint extraction is done in such a way that each multimedia object produces an unique fingerprint. The advantages of using fingerprints for perceptual comparison instead of the entire signals are:

1. Reduced memory, storage, and processing requirements due to relatively small size of the fingerprint.

2. Efficient comparison: since the fingerprints hold only the perceptually relevant components and any distortion or noise will be removed.

3. Faster searching due to smaller database size.

The overall functionality of fingerprinting is shown in Figure 2.2. From the figure, two operating modes can be identified in a fingerprinting system.

1. **Populating the database:** In this stage, the collection of multimedia signals to be recognized is presented to the system. The system processes the multimedia signals by extracting unique fingerprints based on the characteristics of the content. This

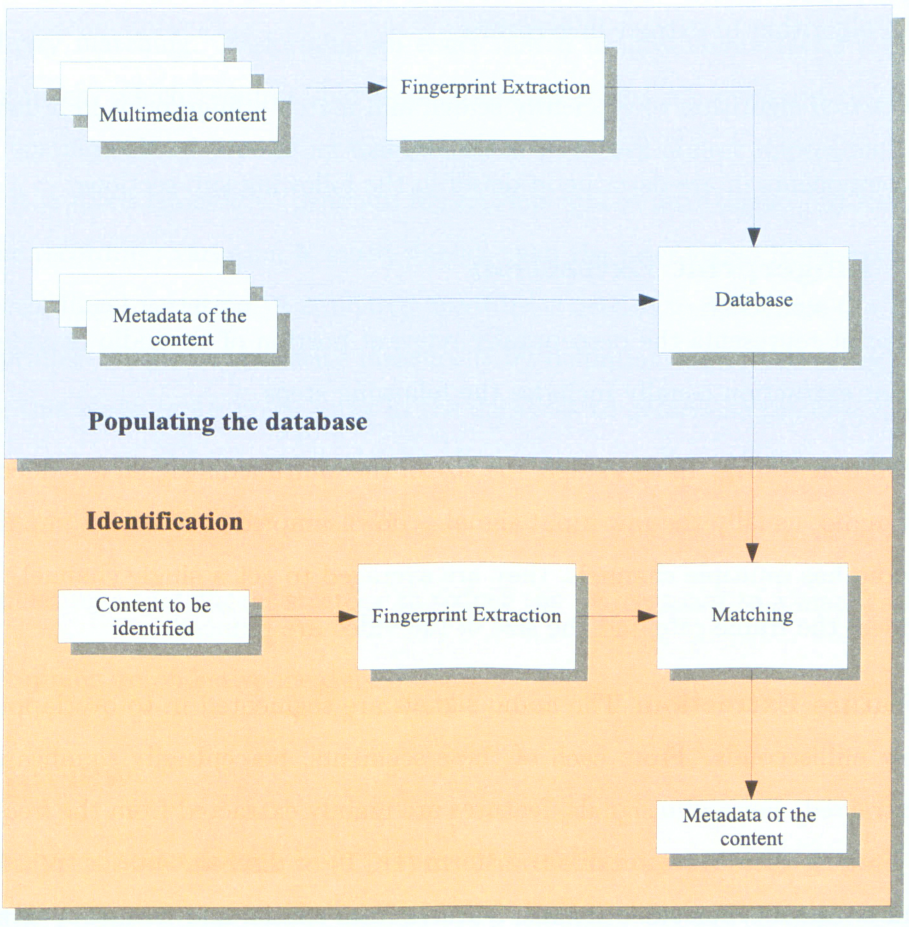

Figure 2.2: Overview of fingerprinting

compact and unique representation is stored in a database along with a tag or other metadata such as artist name, album, and copyright information of each recording.

2. **Identification:** The signal that has to be identified is presented to the system. The signal is processed to extract the fingerprint. The fingerprint is then compared to the fingerprints of the database. If a match is found, the tag or metadata associated with the content is obtained from the database. A confidence of the match can also be provided.

It can be observed that a fingerprinting system consists of two main components:

- An algorithm to extract fingerprints.
- A search algorithm to efficiently search and match a fingerprint in a database.

These components in are describe in detail in the following sub sections.

2.1.1 Fingerprint Extraction

A fingerprint represents the perceptually relevant portion of an audio or video signal. The fingerprint extraction usually includes the following steps.

1. **Preprocessing:** In this stage, the size of the multimedia signal is reduced. In the case of audio, usually the raw input signal is downsampled to a lower sampling rate. If the audio has multiple channels, they are averaged to get a single channel. In the case of video, the frame rate and the size of the video are reduced.

2. **Feature Extraction:** The audio signals are segmented in to overlapping sections of few milliseconds. From each of these segments, perceptually significant features are extracted. For audio signals, features are mainly extracted from the frequency domain by taking the discrete cosine transform (DCT) or discrete Fourier transform (DFT) of the segments. For video signals, features from spatial and frequency domain are used.

3. **Post-processing:** In addition to the absolute features, the derivatives of the features can be added to better characterize the temporal variations of the signal. The features are also mean subtracted and normalized.

4. **Fingerprint Modeling:** In order to capture the underlying statistical variations of the audio or video segment, some authors model the segment using codebooks generated by vector quantization (VQ), hidden Markov models (HMM) or GMM.

2.1.2 Fingerprint Matching

In fingerprint matching, a good search scheme is designed to retrieve the best match from a large database of several millions in a short time. This can be achieved using either

exact or fuzzy matching. Performing an exact match usually means using a direct table lookup approach, which requires that fingerprints generated are invariant to compression and other common signal processing manipulations — a task almost impossible to achieve. Therefore it is more realistic to generate fingerprints whose intra-signal (different variants of the same recording) variation is much smaller than the inter-signal (different variants of different recordings) variation. A similarity measure is needed to determine the closeness of the fingerprints. Since the similarity measure is by definition inexact or fuzzy, it requires computing this measure for every entry in the database to determine the best match. This process is not practical in large databases. Hence a fingerprint matching scheme should address the following two important issues:

1. Formulating an intelligent strategy to reduce the search space to a manageable size.

2. Determining an objective measure of match.

2.2 Review

A large number of schemes have been proposed for audio fingerprinting. A good review of audio fingerprinting schemes can be found in [12]. In [13], Allamanche *et al.* use loudness, spectral flatness measure (SFM) and spectral crest factor (SCF) as features and model the fingerprints using code books generated by VQ. For each of the music items in the database, the associated code book is generated. In the identification process, the features vectors of the test item is subsequently approximated by all stored codebooks using some standard distance metric. The test item is assigned to the music item which yields the smallest accumulated approximation error.

Haitsma *et al.* [14] use energies in the 33 logarithmic spaced frequency bands as features. The fingerprints are modeled as the sign of the difference in adjacent frequency band energy differences between successive time frames. The fingerprint database is a series of binary vectors. To test an unknown audio, its fingerprint is compared with the database and the music item which gives a bit error rate below a certain threshold is chosen.

Cano *et al.* [15] and Batlle *et al.* [16] use Mel frequency cepstral coefficients (MFCC) as features and model fingerprints as a sequence of HMM. Initially, an alphabet of sounds that best describe the music is extracted in an offline process. These audio units are modeled with HMM. During the training phase, the set of songs to be identified are decomposed into these audio units ending up with a database of HMM sequences representing the original songs. By approximate string matching, the song sequences that best resembles the sequence of an unlabeled audio is obtained.

Burges *et al.* [17] argue that the commonly used audio features are heuristic and hence may not be optimal. They use oriented principal component analysis (OPCA) on the modulated complex lapped transform (MCLT) coefficients of the signal to extract optimal features. By taking in to account of some predefined distortions, they use OPCA to reduce dimension and choose features that are invariant to distortions. Sukittanon and Atlas [18] argue that the spectral features are inadequate when there is frequency distortion. They extract features from the joint acoustic and modulation frequency representation of the audio signal. The music identification is based on the cross-entropy between the test item and the items in the database.

In [19], Lu proposes features based on the wavelet modulus of the continuous wavelet transform. Mapelli and Lancini [20] use energies in a single frame and a whole block of frames as features and model the fingerprint as the sign of energy difference between the frame and block energies. In a recent work, Venkatachalam *et al.* [9] use the mean and root mean square (RMS) power of the energies in different frequency bands of successive time frames. To search a fingerprint in a database, they use only a few components of the fingerprint to get a reduced set of elements and later use all the components to select an item from the reduced set.

The literature in video fingerprinting is limited when compared to audio fingerprinting. Oostveen *et al.* [21] introduced the concept of video fingerprinting as a tool for video identification. They use spatial and temporal derivatives as features and quantize them to one-bit to generate a fingerprint. This binary structure allows them to use a fast retrieval of finger-

print from the database. However, for severe distortions such as scaling or shifting, the bit errors in the fingerprints increase and the efficiency of the database search fails. In [22], Joly *et al.* propose a scheme based on image local descriptors. The fingerprint extraction includes a key-frame detection, an interest point detection these key-frames, and the computation of local differential descriptors around each interest point.

Lancini *et al.* [23] calculate the variance of the luminance values of the frames. The fingerprint vector is the minima of the variance values in 16×16 blocks of the image. They show that the scheme is robust to MPEG and DivX compression. Mohan [24] matches videos based on the similarity of temporal activity instead of matching key frames in a video. He uses the ordinal measure [25] of the frames as features. Indyk *et al.* [26] use a fingerprint based on the shot boundaries of the video and show that the fingerprints are robust to many types of attacks. Hampapur *et al.* [27] developed a frame matching algorithm and compared its robustness to resolution changes in MPEG1 encoding.

The proposed fingerprinting scheme in this thesis uses GMM as a modeling tool. The algorithm is described in detail in chapter 3.

2.3 Watermarking

A watermark is a signature or message which is imperceptibly added to the host-signal in order to convey some hidden information. The watermark is added to the host data to be watermarked such that the watermark is unobtrusive and secure in the watermarked data but can partly or fully be recovered from the watermarked data later on if the correct cryptographically secure key needed for recovery is used. To be effective, watermark should satisfy the following requirements [28]:

- Unobstructive – that is perceptually imperceptible, when embedded in the host signal.
- Discreet – undetectable to prevent unauthorized removal.
- Robust – depending upon the type of watermarks, watermarks should remain intact in the host signal when subjected to intentional removal attacks and common signal

processing manipulations.

- Easily extractable – authorized watermark extraction must be easy and reliable to prove ownership.

The overall watermarking process is shown in Figure 2.3. Watermarking has two main processes: watermark embedding and watermark extraction.

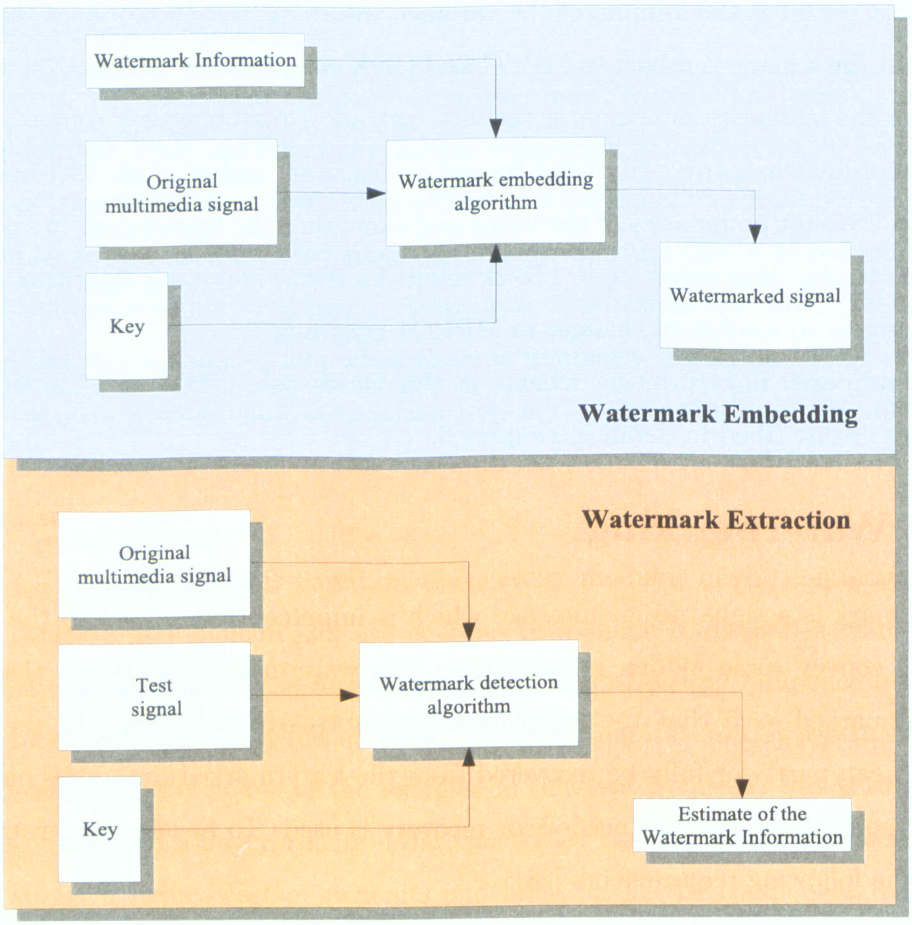

Figure 2.3: Overview of watermarking

2.3.1 Watermark Embedding

The embedding process involves

1. Design of the watermark $\mathbf{W}$ to be added to the host signal. Typically, the watermark depends on a key $\mathbf{K}$, watermark information $\mathbf{I}$ and host signal $\mathbf{X}$

$$\mathbf{W} = \mathrm{f}_0\left(\mathbf{K}, \mathbf{I}, \mathbf{X}\right). \tag{2.1}$$

2. Design of the embedding method itself that embeds the watermark signal into the host signal to get the watermarked data $\mathbf{Y}$

$$\mathbf{Y} = \mathrm{f}_1\left(\mathbf{W}, \mathbf{X}\right). \tag{2.2}$$

The embedding process introduces small changes in the host signal. To ensure imperceptibility of the modification caused by watermark embedding, perceptual models are used. These models provide the bound for amount of distortions tolerable by the host signal without showing any perceptible difference. As a consequence of the required imperceptibility, the individual samples (e.g., pixels or transform coefficients) that are used for watermark embedding can only be modified by an amount relatively small to their average amplitude.

To ensure robustness despite the small allowed changes, the watermark signal is added to the perceptually significant regions of the signal. Hence any attempt to remove the watermark signal severely affects the quality of the signal. Also the watermark information is usually redundantly distributed over many samples (e.g., pixels) of the host data. This helps in the recovery of the watermark from a small fraction of the watermarked data.

The security of the system comes from the uncertainty of the key. Without having access to this key information the watermark cannot be detected or removed.

2.3.2 Watermark Extraction

The watermark extraction process involves

1. Design of the corresponding extraction method that recovers the watermark information $\mathbf{I}'$ from the watermarked data $\mathbf{Y}$ using the key and with or without the help of

the original $\mathbf{X}$.

$$\mathbf{I}' = g(\mathbf{K}, \mathbf{Y}).$$

or

$$\mathbf{I}' = g(\mathbf{K}, \mathbf{Y}, \mathbf{X}). \tag{2.3}$$

The generic watermark extraction process is depicted in Figure 2.3. At the detector the watermarked data, the secret or public key, and, depending on the method, the original data and the original watermark are the inputs. The output of the watermark recovery process is either the recovered watermark or some kind of confidence measure indicating how likely it is for the given watermark at the input to be present in the data under inspection.

2.4 Review

In recent years, numerous watermarking techniques have been proposed. In general, the invisible watermarking schemes can be classified in to fragile watermarking and robust watermarking. Fragile watermarks [29], [30], [31], [32], [33] are added to the host data such that any unwanted manipulation of the watermarked signal alters the watermarked message and provides information about the tampering process. In robust watermarking, the watermarks are designed and embedded in such that it cannot be removed by any incidental signal processing manipulations or intentional watermark removal attacks.

Among these robust watermarking schemes, those requiring both the original data and the secret keys for the watermark bit decoding are called private watermark schemes [34], [35], [36], [37], [38]. Those requiring the secret keys but not the original data are called public or blind watermark schemes [39], [40], [41], [42], [43], [44], [45]. Usually, private watermark schemes are superior in robustness to attacks. However, private schemes are not feasible in situations such as watermark detection in DVD players, because the original data is not available. Blind watermark schemes, on the other hand, detect the watermarks without the original data and are feasible in those situations. The trade-off is that the blind schemes

are usually less robust and have relatively higher false alarm rate compared with the private schemes.

The watermarking schemes can also be classified by the number of watermark bits that can be embedded. Most of the schemes embed single bit [34], [46]. But multiple bit watermarking schemes also exist [47], [48]. In the former case, a single watermark sequence is embedded and the receiver just detects the presence of watermark. The receiver checks for a watermark from the set of all possible watermarks. Hence as the size of the watermark set increases the detection becomes impractical. In the second case, the detector should check for the presence of the watermark, and if the watermark is present, the detector should decode the message correctly.

In a recent work, Erkucuk [49] proposed an audio watermarking algorithm by embedding chirp signals as watermark and detecting the watermark by using TF analysis followed by a HRT detector. The algorithm is capable of embedding multiple watermark bits and is very robust to several attacks. The tests done by the author show that the watermark is extractable from all the ten attacks used in the experiments. The high robustness and the capability of carrying multiple bit watermark messages is the motivation to extend the scheme to images. The proposed image watermarking algorithm is described in chapter 4.

2.5 Applications

Chapter 1 briefly mentioned how fingerprinting and watermarking can be useful in DRM. This section explains in detail how these technologies are used in DRM. Also the applications other than DRM are also presented [50].

2.5.1 Digital Rights Management

Proof of Ownership

Content creators are often concerned about the possibility that their work is to be appropriated by another person. Let us consider this example. Let artist A produces and distributes the content. Another artist B copies the content of artist A and distributes as his own. Now

there should be a way for artist A to prove that she is indeed the original author of the content. This problem can be solved by using both fingerprinting and watermarking with the help of a trusted third party (TTP). A TTP is a usually a government agency that acts as a repository of content.

In a fingerprinting based solution, the original artist A registers only the fingerprint of her work with the TTP instead of the whole work. The TTP extracts an unique fingerprint for each content and stores the fingerprint in a database along with the owner of the content. Now the TTP can extract the fingerprint from the copy released by artist B and check with the database to get the original owner artist A.

Watermarking can also be used to solve this problem. The TTP provides a unique key or *signature* to each content owner. The content owner embeds the signature as watermark message in the content. The presence of the signature in the content uniquely identifies the owner of the content.

Access Control and Copy Protection

When a content producer distributes a content to an user, she specifies a set of rules for accessing the content. The rules may specify whether the content can be copied and the number of times it can be copied. These rules are obeyed by compliant devices such as CD players, computers used at the consumer end. But there should be a way to bind these rules to the content.

A robust watermark can be used to specify the rules and can be embedded in to the content. The rules are associated to the content irrespective the format of the content. Since the watermark is robust, unauthorized removal is not possible. The compliant device reads the usage rules and complies with it.

In fingerprinting, addition of usage rules to the content is not possible. But the compliant device can extract of the fingerprint of the content and connect to an online database and retrieve the usage rules of the particular content. This may sound unrealistic at the present time, but with the increasing tendency of electronic devices other than computers to connect

the Internet, it should not be a problem in the near future.

Filtering

P2P networks are excellent channels for music and movie piracy. During a court battle with recording industry, Napster, a music sharing service, was forced to filter copyrighted music files from shared.

A fingerprinting system can be used to filter the copyrighted content. When an user shares a music file with another, the system extracts a fingerprint from the shared content and compares with a database to identify the content shared. Once the content is identified, the copyright information of the shared music is retrieved. If the content is copyrighted the transfer is prevented or the user is asked to pay a fee for the content. Contents that are copyrighted are allowed to be shared freely.

In a watermark based system, all copyrighted contents have a watermark. Before a content is shared, the system checks for the presence of watermark. If a watermark is present, then the content is considered to be copyrighted. Contents without a watermark are allowed to be shared.

Content Tracking

Content tracking is the identification of a particular copy of copyrighted content. To illustrate its importance, let us consider the following example. Suppose a video owner contracts the services of various mastering and distribution companies to create and distribute the video on media. Unscrupulous companies or employees may conspire to leak illicit copies to pirates. So the owner wants to identify the companies that distribute the copies illegally.

Watermarking can be used to identify a particular copy. The owner of the content embeds a different watermark into the copies she provides to each mastering company. If illegal copies bearing a specific company's watermark are found before the official release of the video, the video owner can extract the watermark from the illegal copy and identify the mastering company that distributed the video illegally. Then the owner may sue the company or choose not to deal with that company in the future.

Fingerprinting is a not suitable for this application since the fingerprints of all the copies will be the same.

2.5.2 Broadcast Monitoring

A broadcast monitoring system automatically monitors the number of times a commercial, song or film is broadcasted by radio or TV stations. There are many organizations interested in broadcast monitoring. Advertisers want to ensure that they receive all of the air time they purchase from broadcasters. Performers want to ensure that they get the royalties from advertising firms. Also, broadcast monitoring helps a company to monitor the number of times and the channels in which its rival's commercials are broadcasted.

In a fingerprint based broadcast monitoring, the channels are monitored continuously and fingerprints are generated once a few seconds. The fingerprints are then compared with the database, and if a match is found then the time, channel of the commercial are recorded. Here the fingerprint generation and the matching with the database should be done in real-time which is quite challenging if the database is large.

In a watermarking based broadcast monitoring, each commercial to be monitored is embedded with a unique watermark. The channel is continuously monitored for the watermark. Here the robustness requirements of the watermarks are less strict than that of the copyright applications. However, the watermark should be able to distortions that are common in broadcasting such as A/D conversion, reshaping, compression, and cropping.

2.5.3 Connected Audio

Connected audio is a general term for consumer applications where music is somehow connected to additional and supporting information. Audio fingerprinting can provide a universal linking system for audio content. A popular example is to obtain the name and title of an audio broadcast by recording a small portion of it using a mobile phone and sending it to a service provider who would identify the audio clip and provide you the details such as name, artist and where to buy the album.

From a technical point of view, this application is challenging since the audio signal in this case is severely degraded due to processing by radio stations, AM/FM transmission, A/D conversion, speech coding and finally the transmission over the mobile network. The fingerprints should be robust to these distortions.

2.5.4 Content-related Services

Content information is defined as information about the content that is relevant to the user or necessary for the intended application. Some of the examples include [50]:

- Content information describing an audio excerpt: rhythmic, timbral, melodic or harmonic description.
- Meta-data describing a musical work, how it was composed and how it was recorded. For example: composer, year of composition, performer, date of performance, studio recording/live performance.
- Other information concerning a musical work, such as album cover image, album price or artist biography.

These information can be usually associated with a file as a header. But once the format of the content is changed such as encoding to MP3, the encoder does not necessarily attach these information with the content. Hence it is advantageous to tie these information with the content. These information can be embedded in to the content as a watermark. But the information that can be contained in a watermark is usually limited. So fingerprinting is a better solution for this application. Some of the applications of fingerprinting for providing content-related services are

1. **Automatic Music Library Organization:** Nowadays, PC users have a large collection of songs. These songs are usually obtained from different sources such as ripping from a CD, downloading from the Internet. Hence the meta-data associated with the music files are often inconsistent, incomplete or even incorrect. Audio fingerprinting

can make the library consistent and allow easy organization based on, for example, artist or album.

2. **Media Description Schemes (MDS):** The MPEG-7 standard proposes a MDS for multimedia content based on the Extensible Markup Language (XML) metalanguage providing for easy data interchange between different equipments. Some fields of an MDS can be automatically extracted from audio recordings, with greater or lesser success. This extraction is performed by means of signal processing techniques and psychoacoustic analysis. Nevertheless, many features, which are important for certain applications cannot be automatically extracted, such as the title, author, performer or edition year of a musical work. This information is usually stored externally to the audio data, either in a database or inserted in the structure of the audio file or transmitted frame, depending on the communication channel or storage support. Fingerprinting can then be used to identify a recording and retrieve the corresponding MDS, regardless of support type, file format or any other particularity of the audio data.

3. **Quality Assessment:** The Music2Share [51] project proposes an architecture for copyright-compliant music sharing based in peer-to-peer protocols, cryptographic algorithms, watermarking and perceptual fingerprinting based music identification. In Music2Share project, fingerprints are not only used to identify the music content but also to assess the perceptual quality of the content. A user, for example, should pay less for a song which is MP3 encoded at 32 kbps than for a song encoded at 192 kpbs.

2.6 Advantages and Disadvantages

The various applications of fingerprinting and watermarking are described in the previous section. For many applications, both fingerprinting and watermarking provide solutions. This section presents the relative advantages of fingerprinting and watermarking. Figure 2.4 summarizes the merits and demerits of these two technologies [50].

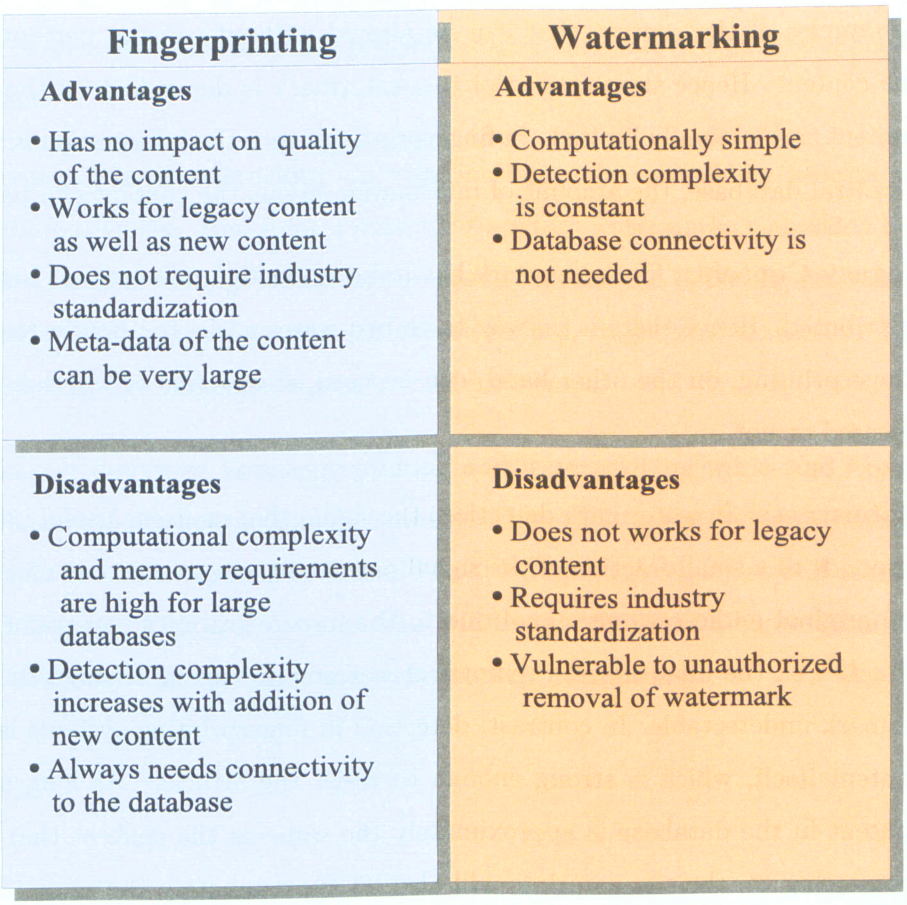

Figure 2.4: Advantages and disadvantages of fingerprinting and watermarking

1. **Modification of Original Signal:** In watermarking, the original signal is modified to embed the watermark information. The power of the watermark is inversely related to the imperceptibility of the watermark. But the detection performance of the watermark increases by increasing the power of the watermark. Hence there is a trade-off between imperceptibility, detection performance, and power of the watermark. In fingerprinting, there is no such trade-off: the system monitors the content, constructs a description of it, and searches for a matching description in its database.

2. **Information Capacity:** The amount of information that can be embedded by watermarking is dependent on the *perceptual capacity* of the host content *i.e.* the number

of samples of the content that can be altered without causing perceptual change in the content. Hence the capacity of the watermark is dependent on the nature of the content and is usually limited. In fingerprinting, since the information is retrieved from a central database, the amount of information about the content can be very large.

3. **Legacy Content:** The watermark has to be embedded to a content before it is being distributed. Hence there is no way to embed watermarks to already released content. Fingerprinting, on the other hand, can be used irrespective of whether the content is released or not.

4. **Robustness:** In watermark detection, the signal that contains useful information corresponds to a small fraction of the signal power as the watermark is much weaker than the original audio content signal due to the imperceptibility constraint. In addition, *attacks i.e.,* the unauthorized removal of watermark, can be strong and make the watermark undetectable. In contrast, detection in fingerprinting systems is based on the content itself, which is strong enough to resist the attacks. As long as the original content in the database is approximately the same as the content that the system is monitoring to, their fingerprints will also be approximately the same. The similarity of the fingerprint depends on the fingerprint extraction procedure and therefore the robustness of the system will also depend on it. Most fingerprinting systems use a robust fingerprint extraction procedure and hence they are inherently more robust.

5. **Requirement of a fingerprint database:** Fingerprinting needs a database containing the fingerprints and the corresponding meta-data. As the number of items in the database increases, memory requirements and computational costs also grow; thus, the complexity of the detection process increases with the size of the database. In contrast, no database is required for detection in a watermarking system, as all the information associated with the content is contained in the watermark itself. The detector checks for the presence of a watermark and, if one is found, it extracts the data contained therein. Hence watermarking requires no update as content is added, and

the complexity of the detection process remains constant irrespective of the number of content.

6. **Industry Standardization:** In watermarking, the embedding process is done by the content owner. Hence for a detector to detect watermarks embedded by different owners, there should be a standard to specify a common detecting process. However, in fingerprinting, the detector can choose its own fingerprint extraction method and still identify the content released by content owners.

This chapter described how fingerprinting and watermarking works and explained how they can be used to protect multimedia content. Then their applications are discussed along with their relative advantages. DRM should be able to protect legacy as well as new content. Hence both fingerprinting and watermarking will be used in DRM. Chapters 3 and 4 present novel algorithms for fingerprinting and watermarking respectively.

Chapter 3

Fingerprinting

FINGERPRINTING is a tool to evaluate the perceptual similarity of two multimedia signals not by comparing the multimedia signals directly but by comparing their fingerprints or signatures. A fingerprint is a content-based compact representation of a multimedia signal. These fingerprints are used to identify a piece of audio or video uniquely based on the content instead of the meta-data. A well designed fingerprint should be able to identify multimedia files even if they are severely distorted by perceptual coding or common signal processing operations.

A typical framework of a fingerprinting system is shown in Figure 3.1. A fingerprinting system basically consists of two parts: fingerprint extraction and a matching algorithm. The fingerprint extraction derives a set of relevant characteristics of a multimedia object in a concise and robust form. Given a fingerprint derived from a recording, the matching algorithm searches a database of fingerprints to find the best match.

One of the critical issues in fingerprinting is the design of fingerprints. The fingerprints should satisfy the following requirements:

1. Discrimination power over a large number of other fingerprints.

2. Invariance to distortions.

3. Computational simplicity.

4. Smaller size.

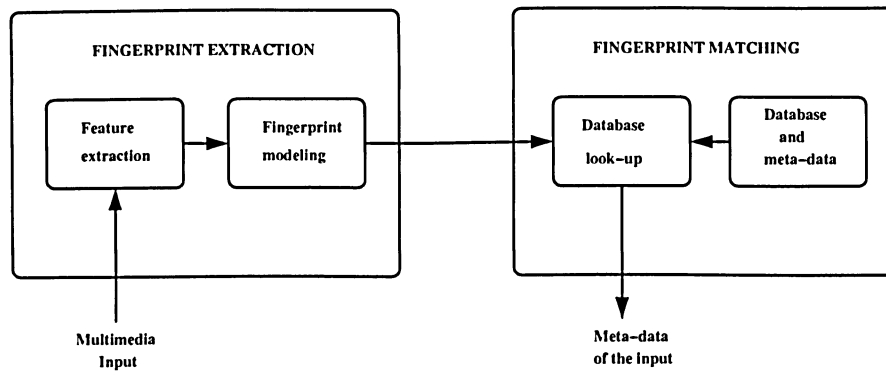

Figure 3.1: Fingerprinting framework

In this work, fingerprints are designed addressing the above issues by modeling an audio or video clip by GMM [52], [53]. The audio fingerprints are generated using various spectral and cepstral features. For video fingerprinting, the luminance component of the pixels within frames and across frames are used as features and the entire video clip is modeled as a single entity instead of sequence of frames. In the works described in section 2.2, fingerprints are generated by using modeling tools such as VQ, HMM, or by just quantizing the features. In this work, GMM are used to model the fingerprints. The motivation for using GMM are:

- GMM outperforms many modeling schemes such as VQ, HMM, radial basis function (RBF), for text-independent speaker identification [54]. Since most of the fingerprinting techniques are adapted from speech identification research, it is instructive to use GMM for fingerprinting.

- GMM reduce the dimensionality of the input greatly. Instead of storing the array of feature vectors, only the GMM parameters have to be stored in the database.

- Fingerprints modeled using GMM are robust to various distortions due to the approximation of the feature space. Also, if the multimedia clips are known to undergo some known distortion, then these distortions can be used in the training to make the fingerprints robust to those distortions.

3.1 Audio Fingerprinting

The overview of the proposed audio fingerprinting scheme is shown in Figure 3.2. First the incoming audio clip is preprocessed and features are extracted from them. Then using these features, the audio clip is modeled using Gaussian mixtures. In the training phase, the mixture models of all the audio clips are stored in the database along with the meta-data information. In the identification phase, features from an unknown audio clip are used to evaluate the likelihood of all the models in the database. Then the model that is most likely to generate features is identified as the correct audio clip.

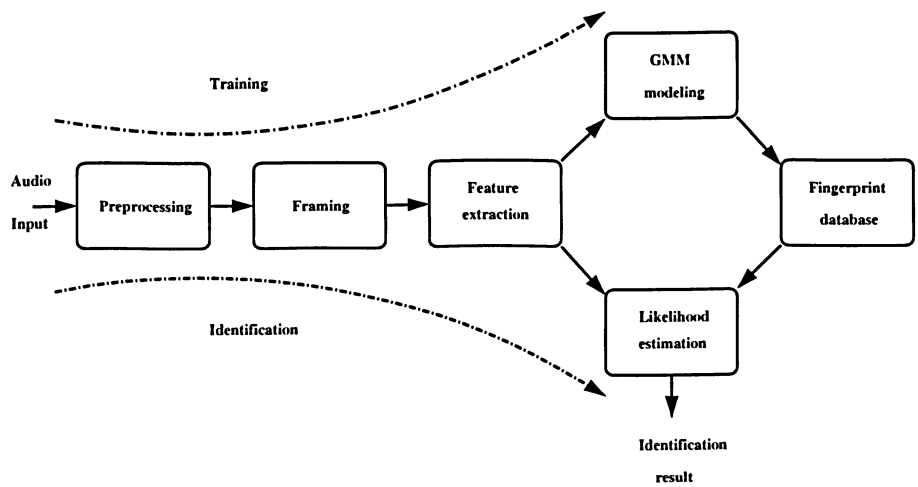

Figure 3.2: Proposed fingerprinting system

3.1.1 Preprocessing

In the first step, the audio signal is converted to a standard format (16 bit, PCM). If the audio signal is stereo, it is converted to mono by averaging the right and left channels. Then the signal is downsampled to 11025 Hz and the dynamic range is normalized to [-1, 1]. Audio signals are highly non-stationary signals. However, it can be assumed that the spectral characteristics of the signals does not change over an interval of few milli seconds. Hence the audio signals are divided in to frames of length 23 ms. To increase robustness to

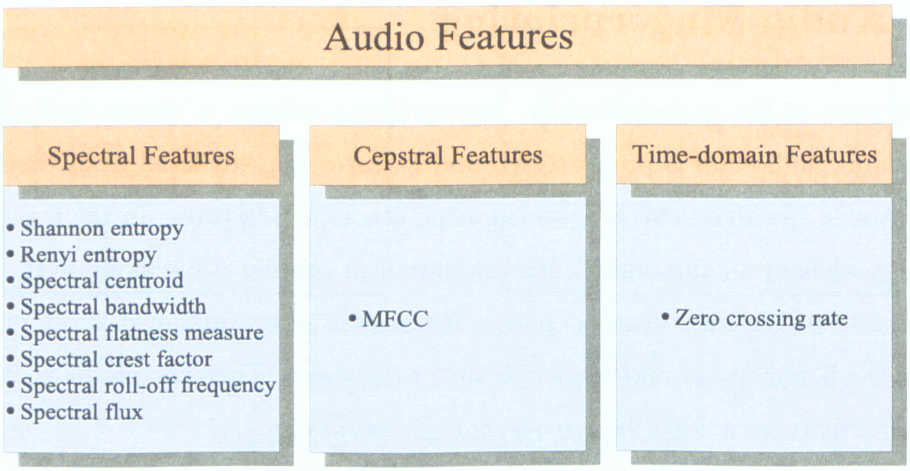

Figure 3.3: Audio features

shifting, the adjacent frames have an overlap of 50%. To reduce discontinuity at the edges, a Hamming window is applied to the frames.

3.1.2 Feature Extraction

In a typical audio fingerprinting application such as in DRM, the initial fingerprint generation and the fingerprint matching are done in a server with large computing and memory resources. In the identification phase, the fingerprint generation is usually done in a client machine such as a personal computer or a mobile device with limited resources. Hence features used to generate fingerprints must be simple to compute. The features used in the fingerprinting algorithm are shown in Figure 3.3. Though it is possible to extract features from the TF domain and wavelet domain, these features are not considered as they are computationally complex for real-time multimedia processing tasks.

Spectral Features

Let $s_i(n)$ represents the i^{th} frame of the signal with $n = 1, \ldots, N$. Let $F_i = f_i(u), u \in (0, M)$, be the Fourier transform of the i^{th} frame, where M is the index of the highest frequency band. To increase the robustness of the fingerprint, features are not extracted on the whole spectrum but on non-overlapping logarithmically spaced bands. Let l_b and u_b be the lower

and upper edges of the frequency band b. The lower and upper edge frequencies of the bands are given in Table 3.1. The definition of the features considered in this work are given below. Some of these features have been used successfully in audio fingerprinting [13] and music classification [55].

Table 3.1: Frequency bands

Band b	Lower edge (Hz)	Upper edge (Hz)
1	0	62.5
2	62.5	125
3	125	250
4	250	500
5	500	1000
6	1000	2000
7	2000	5512

1. Shannon Entropy (SE): The Shannon entropy of a signal is a measure of its spectral distribution of the signal. Shannon entropy is defined as

$$SE_{i,b} = \sum_{u=l_b}^{u_b} |f_i(u)| \log_2 |f_i(u)| . \tag{3.1}$$

2. Renyi Entropy (RE): The Renyi entropy of a signal is also a measure of its spectral distribution. Renyi entropy is defined as

$$RE_{i,b} = \frac{1}{1-r} \log_2 \left(\sum_{u=l_b}^{u_b} |f_i(u)|^r \right) . \tag{3.2}$$

3. Spectral Centroid (SC): The spectral centroid is the center of gravity of the magnitude spectrum of the STFT and is a measure of spectral shape and "brightness" of the

spectrum. SC is defined as

$$SC_{i,b} = \frac{\sum_{u=l_b}^{u_b} u.\,|f_i(u)|^2}{\sum_{u=l_b}^{u_b} |f_i(u)|^2}. \tag{3.3}$$

4. Spectral Bandwidth (SB): The spectral bandwidth is measured as the weighted average of the distances between the spectral components and the spectral centroid. SB is defined as

$$SB_{i,b} = \frac{\sum_{u=l_b}^{u_b} (u - SC_{i,b})^2\,.\,|f_i(u)|^2}{\sum_{u=l_b}^{u_b} |f_i(u)|^2}. \tag{3.4}$$

5. Spectral Band Energy (SBE): The spectral band energy is the energy in the frequency bands normalized by the energy in the whole spectrum. SBE is defined as

$$SBE_{i,b} = \frac{\sum_{u=l_b}^{u_b} |f_i(u)|^2}{\sum_{u=0}^{M} |f_i(u)|^2}. \tag{3.5}$$

6. Spectral Flatness Measure (SFM): The spectral flatness measure quantifies the flatness of the spectrum and distinguishes between noise-like and tone-like signal. SFM is defined as

$$SFM_{i,b} = \frac{\left[\prod_{u=l_b}^{u_b} |f_i(u)|^2\right]^{\frac{1}{u_b - l_b + 1}}}{\frac{1}{u_b - l_b + 1} \sum_{u=l_b}^{u_b} |f_i(u)|^2}. \tag{3.6}$$

7. Spectral Crest Factor (SCF): The spectral crest factor is also a measure of the tonality of the signal. SCF is defined as

$$SCF_{i,b} = \frac{\max\left(|f_i(u)|^2\right)}{\frac{1}{u_b - l_b + 1} \sum_{u=l_b}^{u_b} |f_i(u)|^2}. \tag{3.7}$$

8. Spectral roll-off Frequency (SRF): The spectral roll-off frequency is defined as

$$SRF_i = max\left(h \middle| \sum_{u=0}^{h} f_i(u) < TH. \sum_{u=0}^{M} f_i(u)\right), \tag{3.8}$$

where TH is a threshold between 0 and 1. A threshold value of 0.8 is used in this work.

9. Spectral Flux (SF): The spectral flux is defined as

$$SF_i = \sum_{u=0}^{M} |||f_{i+1}(u)| - |f_i(u)|||. \tag{3.9}$$

Among the spectral features, spectral band energy, spectral flatness measure, spectral crest factor have been already used for audio fingerprinting. In this thesis, in addition to these features novel spectral features such as Shannon entropy, Renyi entropy, spectral centroid, bandwidth, spectral roll-off frequency, spectral flux are considered in the fingerprinting scheme.

Cepstral Features

Cepstral features are also extracted from the frequency domain but are based on the log-amplitude of the spectrum. In this work, only MFCC are considered. MFCC are perceptually motivated features based on the STFT. After taking the log-amplitude of the magnitude spectrum, the Fourier transform coefficients are grouped and smoothed according to the perceptually motivated Mel-frequency scaling. Finally, in order to decorrelate the resulting feature vectors, a DCT is performed. In this work, 13 coefficients are used since this parameterization has been shown to be quite effective for speech recognition and speaker identification applications [56].

Time Domain Features

In time domain, only the zero crossing rate (ZCR) is considered. ZCR is a correlate of the spectral centroid. It is defined as the number of time-domain zero-crossings within the processing frame.

$$ZCR_i = \frac{f_s}{N}\left(\sum_{i=1}^{N-1} |sign\left(s_i(n)\right) - sign\left(s_i(n-1)\right)|\right), \tag{3.10}$$

where N is the number of samples in the frame i and f_s is the sampling frequency.

Let $\mathbf{X_i}$ be the set of features extracted for the frame i. $\mathbf{X_i}$ can be any one of the features described above. In order to better characterize the temporal variations of the signal, the first derivatives of the above features

$$\delta_i = \delta_i - \delta_{i-1}, \tag{3.11}$$

are also included in the feature matrix. In an audio clip, successive frames are related in time. To include this time dependency, a time vector is added to the feature matrix. This time vector is taken as an incremental counter. Thus, the feature matrix of the entire audio clip can be described as

$$\mathcal{F}'_{\mathcal{M}} = \begin{bmatrix} \mathbf{X_1}, \delta_1, t_1 \\ \mathbf{X_2}, \delta_2, t_2 \\ \vdots \\ \mathbf{X_N}, \delta_N, t_N \end{bmatrix}, \tag{3.12}$$

where N is the number of frames in the audio clip. Finally the feature matrix $\mathcal{F}'_{\mathcal{M}}$ is mean subtracted and component wise variance normalized to get a normalized feature matrix $\mathcal{F}_{\mathcal{M}}$.

3.1.3 Fingerprint Modeling

The feature matrix generated can be used as a fingerprint. The feature matrix is usually large. Hence it increases the size of the database and the computational complexity during fingerprint matching. The fingerprint size can be reduced by exploiting the redundancies

across the audio frames. A modeling tool can be used to generate a concise form of a fingerprint. This work uses GMM for modeling which has been successfully used in audio classification [55] and content based retrieval [57]. Here GMM is used to model an audio fingerprint as a probability density function (PDF), using a weighted combination of Gaussian component PDFs (mixtures). GMM are commonly used to model multi-modal, or other non-Gaussian distributions [58]. A GMM is just a weighted average of several Gaussian PDFs, called the component PDFs. The density function of a random variable $X \in R^d$ with a mixture of k Gaussians is defined as:

$$f(x|\theta) = \sum_{j=1}^{k} \alpha_j \frac{1}{\sqrt{(2\pi)^d |\Phi_j|}} \exp\left\{-\frac{1}{2}(x-\mu_j)^T \Phi_j^{-1}(x-\mu_j)\right\}, \tag{3.13}$$

with parameter set $\theta = \{\alpha_j, \mu_j, \Phi_j\}_{j=1}^{k}$ having the following parameters:

- weight $\alpha_j > 0, \sum_{j=1}^{k} \alpha_j = 1$
- mean $\mu_j \in R^d$, and
- covariance matrix Φ_j, a $d \times d$ positive definite matrix, where d is the dimension of the feature matrix.

Given a set of feature vectors $x_1, x_2 \ldots, x_n$, the maximum likelihood estimation of θ is:

$$\theta_{ML} = \arg\max_{\theta} L(\theta|x_1, x_2, \ldots, x_n) \tag{3.14}$$

$$= \arg\max_{\theta} \sum_{i=1}^{n} \log f(x_i|\theta) \tag{3.15}$$

The Expectation Maximization (EM) algorithm [59] is an iterative method to obtain θ_{ML}. Given the current estimation of the parameter set θ, each iteration of the EM algorithm reestimates the parameter set according to the following two steps:

1. Expectation step:

$$w_{ij} = \frac{\alpha_j f(x_i|\mu_j, \Phi_j)}{\sum_{l=1}^{k} \alpha_l f(x_i|\mu_l, \Phi_l)} \tag{3.16}$$

$$j = 1, \ldots, k \tag{3.17}$$

$$i = 1, \ldots, n. \tag{3.18}$$

The term w_{ij} is the posterior probability that the feature vector x_i was sampled from the j^{th} component of the mixture distribution.

2. Maximization step:

$$\alpha'_j = \frac{1}{n} \sum_{i=1}^{n} w_{ij} \tag{3.19}$$

$$\mu'_j = \frac{\sum_{i=1}^{n} w_{ij} x_i}{\sum_{i=1}^{n} w_{ij}} \tag{3.20}$$

$$\Sigma'_j = \frac{\sum_{i=1}^{n} w_{ij} (x_i - \mu'_j)(x_i - \mu'_j)^T}{\sum_{i=1}^{n} w_{ij}} \tag{3.21}$$

The first step in applying the EM algorithm is to initialize the mixture model parameters. The $k-means$ algorithm [58] is utilized to initialize the values of the parameter set θ. The covariance matrix in the GMM can be full or diagonal. As the feature vectors in this work have reasonably uncorrelated components, computationally convenient diagonal covariance matrices can be used. To reduce the computational complexity, diagonal covariances are used for each Gaussian mixture. The parameter set θ is updated repeatedly until the log-likelihood is increased by less than a predefined threshold from one iteration to the next to get the maximum likelihood parameters θ_{ML}. During the training phase, using the feature matrix $\mathcal{F}_{\mathcal{M}}$, an audio item is modeled as a GMM and the parameter set θ is stored in the database along with the meta-data of the audio.

3.1.4 Fingerprint Matching

The fingerprint matching is based on the log-likelihood value given by Equation (3.15). The features from the unknown audio are used to evaluate the log-likelihood value of the all the models present in the database. If we know that the unknown audio is present in the database, then we can choose the model that gives the highest log-likelihood. If that information is not available, then there should be a criteria to reject the audio clips that are not in the database. An audio clip can be rejected if the log-likelihood is below a certain threshold. The choice of threshold affects the false positive and false negative values.

3.2 Results and Discussion

A database containing 250 five-second audio clips chosen from the categories of rock, pop, country, classical, and jazz is used for simulation experiments. The audio clips are chosen from random portions of songs from CDs. As described in Section 3.1, the audio clips are modeled as mixture of Gaussian distributions.

3.2.1 Robustness to Undistorted Audio

In order to describe the robustness of the scheme to undistorted audio, the log-likelihood of each of the 250 clips are evaluated with all the 250 models in the database. The audio clips used in testing are shifted in time by quarter of the frame width so that all the corresponding frames in the training and testing phase have only 50% of the samples in common. In this case there are 250 target/target log-likelihood values and 62250 target/non-target log-likelihood values. If the log-likelihood values of these two classes are far apart, then using a simple threshold, these two classes can be separated easily. In Figure 3.4, the solid line shows the log likelihood values for target/target case sorted in increasing order. The dotted line shows the highest 250 log likelihood values for target/non-target case sorted in decreasing order. From the figure, it is clear that the two categories can be separated with an appropriate threshold.

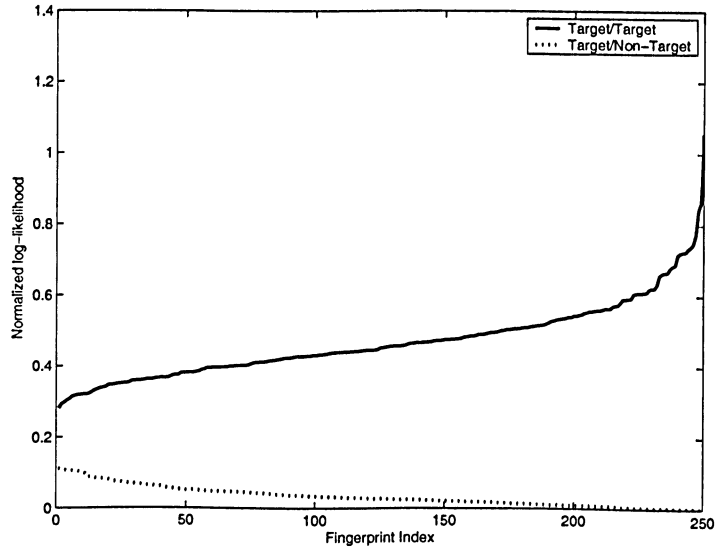

Figure 3.4: Log-likelihood values for target/target and target/non-target cases

3.2.2 Robustness to Distortions

In the previous test, only the undistorted audio clips are used. However, in practical cases, the test audio clips usually undergo some kind of distortion. The following distorted versions are used in our tests.

I. Compression – 1) MP3 at 32 kbps, 2) Advanced Audio Coding (AAC) at 32 kbps, 3) Windows Media Audio (WMA) at 32 kbps, 4) Real encoding at 32 kbps.

II. Amplitude distortion – 1) 3 : 1 Compression above 30 dB, 2) 3 : 1 Expander below 10 dB, 3) 3 : 1 compression below 10 dB, 4) Limiter at 9 dB, 5) 'Super-loud' amplitude distortion, 6) Noise gate at 20 dB, 7) De-esser, 8) Nonlinear amplitude distortion.

III. Frequency distortion – 1) Nonlinear bass distortion, 2) Midrange frequency boost, 3) Notch Filter, 750 - 1800 Hz, 4) Notch Filter 430 - 3400 Hz, 5) Telephone bandpass, 135 - 3700 Hz, 6) Bass cut, 7) Bass boost.

IV. Change in pitch – 1) Lower pitch 2 - 6 %, 2) Raise pitch 2 - 6 %.

V. Change in speed – 1) Linear speed increase 2 - 6%, 2) Linear speed decrease 2 - 6%.

VI. Resampling at 8 kHz

VII. Echo addition

Table 3.2: Mean recognition (in %) rate for distortions. **T1:** Without using distorted versions in training, **T2:** Using some distorted versions in training

	T1	T2
Shannon entropy	91.8	98.9
Renyi entropy	94.4	96.7
MFCC	87.8	97.7
Zero crossing rate	94.0	97.2
Spectral centroid	99.0	99.9
Spectral bandwidth	88.6	98.3
Spectral band energy	87.6	96.8
Spectral flatness measure	95.9	97.0
Spectral crest factor	96.6	99.1
Spectral roll-off frequency	83.7	84.8

Table 3.2 shows the percentage of clips correctly identified for the distortions described above. The results show that fingerprints using some features such as MFCC, spectral bandwidth, spectral band energy are not robust to the distortions. To increase the robustness of the fingerprints, in addition to the original audio, some distorted versions of the audio are also used in training. The following distorted versions are used in our training: 1) Undistorted audio, 2) 3 : 1 Compression above 30 dB, 3) Nonlinear amplitude distortion, 4) Nonlinear bass distortion, 5) Midrange frequency boost, 6) Notch Filter, 750 - 1800 Hz, 7) Notch Filter 430 - 3400 Hz, 8) Raise Pitch 1%, 9) Lower Pitch 1%. As a first test, the log-likelihood of the test clips are evaluated for all the models in the database. Then the model that gives the highest log-likelihood is taken as the correct match. Thus, this test assumes the presence of the test clip in the database. Table 3.3 shows the percentage of clips that are correctly identified for different features for distortions used in training as well as for

distortions not used in training. The values under the columns 'Train' and 'Test' represent the mean recognition rate for distortions used in training and distortions not used in training respectively. The column 'Mean' gives the mean recogntion rate for all the distortions. The results show that it is not necessary to train the model for all possible distortions. By training the model to some representative distortions, robustness can be obtained to a wide variety of distortions.

Table 3.3: Mean recognition rate (in %) for distortions

	Train	Test	Mean
Shannon entropy	99.2	98.6	98.8
Renyi entropy	99.4	98.7	98.9
MFCC	98.7	96.2	96.9
Zero crossing rate	97.7	95.5	96.2
Spectral centroid	99.4	98.8	99.0
Spectral bandwidth	99.4	98.8	99.0
Spectral band energy	99.4	98.5	98.8
Spectral flatness measure	98.8	98.2	98.4
Spectral crest factor	99.4	98.7	98.9
Spectral roll-off frequency	85.0	89.6	88.3

3.2.3 False Positive Analysis

In the previous section, it was assumed that the test clip is present in the database. Hence the model that gives the highest log-likelihood value is identified as the correct match. However, it is possible that the test clip may not be in the database. So there should be a criteria to reject the audio clips that are not in the database. A suitable threshold for log-likelihood can be used to vary the false positive and false negative rates.

Model Order

The false positive rates and the corresponding identification rates is dependent on the number of Gaussian mixtures used in the modeling. For example, Figures 3.5 and 3.6 show the iden-

tification rate curves for spectral centroid and spectral flatness measure. The identification rate curves for spectral centroid essentially converges for model order greater than 8. Similar results were obtained for Shannon entropy, Renyi entropy, MFCC, spectral bandwidth, spectral band energy and spectral crest factor. However, for some features like spectral flatness measure, zero crossing rate, spectral rolloff and spectral flux the curves does not converge. A higher model order is not used for these features because of the increased computational complexity and fingerprint size.

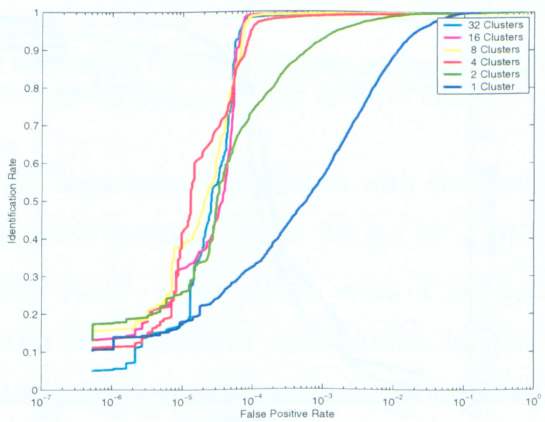

Figure 3.5: Identification rate (in %) vs. false positive rate for different cluster size using spectral centroid as feature.

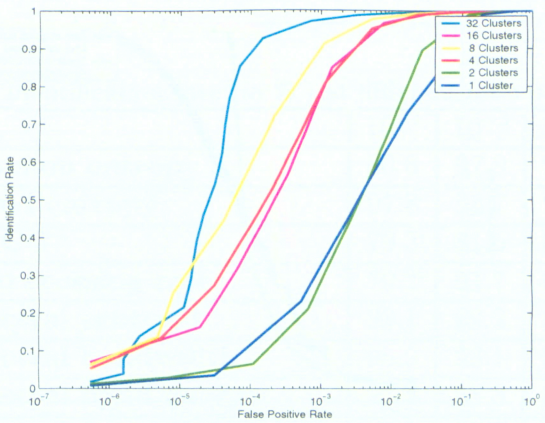

Figure 3.6: Identification rate (in %) vs. false positive rate for different cluster size using spectral flatness measure as feature.

Identification Rate vs. False Positive Rate

The false positive and the corresponding identification rates are shown in Figures 3.7, 3.8 and 3.9. The percentage of audio clips correctly identified at different false positive rates are shown in Table 3.4. Among the different features used, spectral centroid gives the highest identification rate of 99.1% with a false positive rate of 10^{-4}. MFCC gives an identification rate of 79.3% at a false positive rate of 10^{-4}. Spectral roll-off frequency and spectral flux are the two features with very low identification rates.

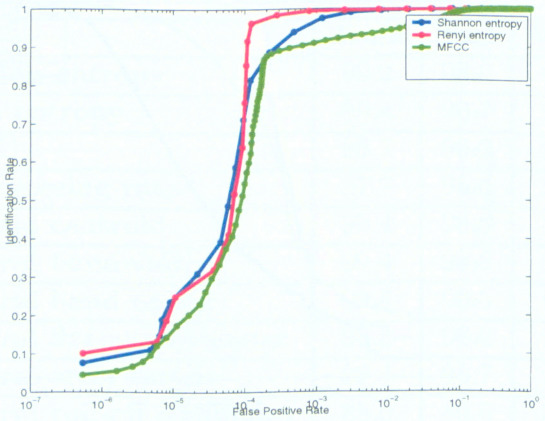

Figure 3.7: Identification rates (in %) at different false positive rates for Shannon entropy, Renyi entropy, and MFCC

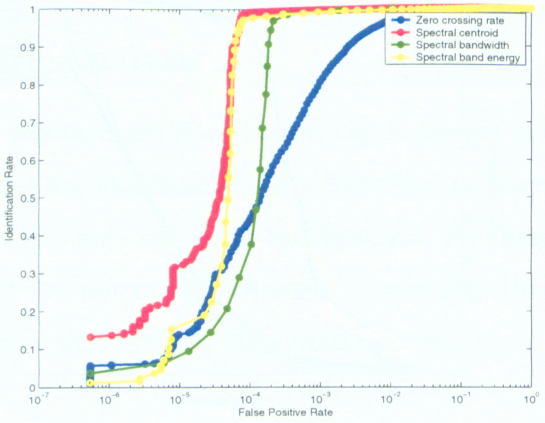

Figure 3.8: Identification rates (in %) at different false positive rates for zero crossing rate, spectral centroid, spectral bandwidth, and spectral band energy

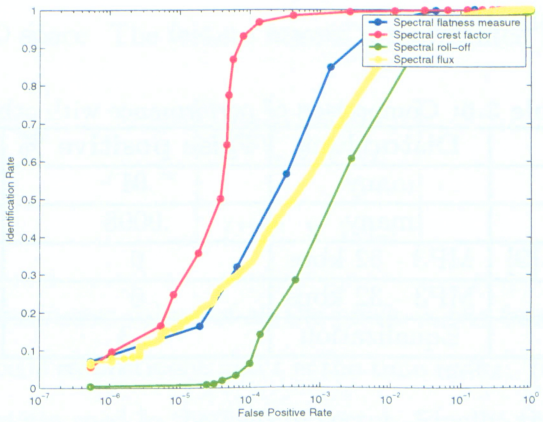

Figure 3.9: Identification rates at (in %) different false positive rates for spectral flatness measure, spectral crest factor, spectral roll-off frequency, and spectral flux

The performance of the algorithm is compared with the results obtained by other authors in Table 3.5. While some authors such as Burges *et al.* [17] tested the robustness of the algorithms using a wide range of distortions, most of the authors have tested only for few distortions such as MP3 compression or equalization. In spite of using a wide range of distortions, the proposed algorithm is comparable in performance to many of the schemes except that of Burges [17]. Burges *et al.* [17] derive optimal features using oriented principal component analysis (OPCA) which is computationally intensive and hence may not be

Table 3.4: Identification rate (in %) at different false positive rates

	10^{-4}	10^{-3}	10^{-2}
Shannon entropy	80.8	97.9	99.1
Renyi entropy	97.8	98.7	99.3
MFCC	79.3	81.6	85.3
Zero crossing rate	79.2	91.7	97.0
Spectral centroid	99.1	99.2	99.7
Spectral bandwidth	95.5	98.1	98.9
Spectral band energy	97.1	98.3	98.9
Spectral flatness measure	85.1	97.0	98.7
Spectral crest factor	94.4	97.9	98.9
Spectral roll-off frequency	26.6	53.1	88.3
Spectral flux	59.1	89.9	97.0

suitable for some applications.

Table 3.5: Comparison of performance with other schemes

Author	Distortions	False positive %	Identification rate %
Proposed	many	.01	99.1
Burges [17]	many	.0008	99.2
Venkatachalam [9]	MP3 - 32 kbps	0	96.64
Cano [15]	MP3 - 32 kbps	0	92.5
Sukittanon [18]	Equalization	2.3	98.3

3.3 Video Fingerprinting

In the previous section, audio fingerprints are modeled using Gaussian mixtures. The fingerprints have good discrimination and are invariant to distortions. This motivated us to extend the scheme to video fingerprinting. This section describes a video fingerprinting algorithm that uses spatio-temporal features and model them using GMM. The robustness of the fingerprints is evaluated using MPEG compression and other video manipulations.

3.3.1 Feature Extraction

In audio, features extracted from the frequency domain are good in representing the perceptual characteristics of the audio. However, in video it is less clear which domain is suitable. To reduce the computational complexity, features are extracted from the spatio-temporal domain. The motivation for choosing spatio-temporal domain is to not only capture the features in the individual frames but also to capture the temporal activity present during the duration of the clip.

For each pixel, the luminance value is calculated. This is the first dimension of the feature vector. In order to include spatial information, the (x, y) position of the pixel is appended to the feature vector. The time feature t is added next. The time descriptor is taken as an incremental counter. Following the feature extraction stage, each pixel is represented with a 4D feature vector, and the image sequence as a whole is represented by a collection of

feature vectors in the 4D space. The feature matrix is represented as

$$\mathcal{F}'_{\mathcal{M}} = \begin{bmatrix} L^i, X, Y, t_1 \\ L^{i+n}, X, Y, t_2 \\ \vdots \\ L^{i+kn}, X, Y, t_N \end{bmatrix}, \tag{3.22}$$

where L^i is a vector of luminance values of all the pixels in the frame i, and X, Y are the x and y positions of the pixels respectively, and t is the time index. To reduce the complexity, one frame every n frames are used in the feature vector. Finally, the feature vector is mean subtracted and component wise variance normalized to get a normalized feature matrix $\mathcal{F}_{\mathcal{M}}$.

3.3.2 Fingerprint Modeling

In [60], Greenspan *et al.* propose a statistical framework for modeling video content into coherent space-time segments within the video frames and across frames. They term such segments "video-regions." Unsupervised clustering, via Gaussian mixture modeling, enables the extraction of space-time clusters, or "blobs," in the representation space. Space and time are treated uniformly and the video is modeled as a single entity, as opposed to a sequence of separate frames. Following their approach, the video fingerprints are modeled based on the features described in the previous section. During the training phase, the GMM parameters of all video clips are extracted and stored in the database along with the meta-data of the video clips.

3.3.3 Results and Discussion

A database containing 100 six-second video clips with a resolution of 288 lines and 352 pixels per line are used in our tests. The frame rate of the video clips is 15 frames per second. But only 5 frames per second are used in the feature matrix given by Equation (3.22). During fingerprint extraction, each video frame is resized to 144×176, and are modeled as described in the previous sections. The tests included the following processing:

1. MPEG1 compression at a bit rate of 1 Mbps

2. Median filtering using 3×3 adjacent pixels

3. Vertical shifting with gray band at the bottom

4. Spatial resizing

5. Spatial Gaussian noise addition with different variance values

The results are shown in Table 3.6 and Figures 3.10 - 3.12. The results show that the proposed technique is very robust to perceptual coding and median filtering. For bit rates above 1 Mbps, 100 % recognition rate is obtained for MPEG compression. The results indicate that the algorithm is robust to processing that are done on a local basis such as compression and filtering. However, for global geometric operations such as scaling and resizing, the algorithm is not very robust.

Table 3.6: Recognition rate (in %) for MPEG compression and median filtering.

Original	MPEG 1	Median filtering
100	98	97

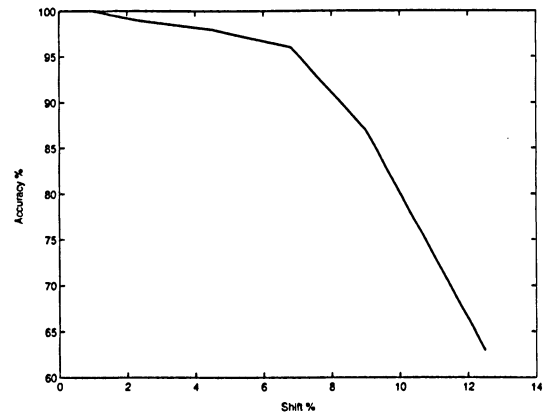

Figure 3.10: Recognition rate (in %) for vertical shifting.

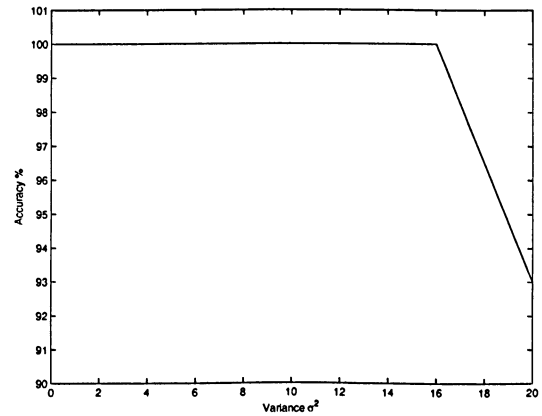

Figure 3.11: Recognition rate (in %) for additive white Gaussian noise.

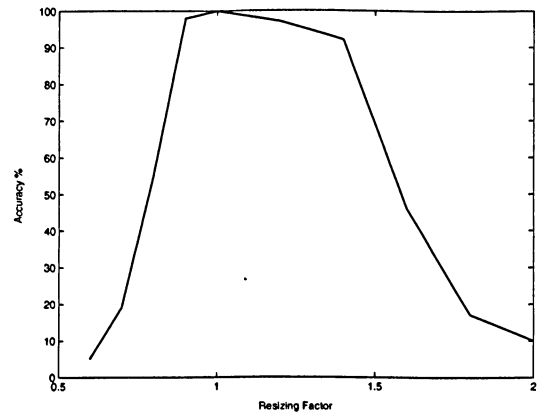

Figure 3.12: Recognition rate (in %) for spatial resizing.

The performance of the algorithm is comparable to other works such as [22] and [21]. The algorithms is quite robust to MPEG compression when compared to [21]. To resizing operations the algorithm performs better than [22]. However, the algorithm is less robust to shifting operations than [22].

This chapter discussed a novel fingerprinting technique that extracts features and models them using GMM. By modeling the fingerprints by GMM, the size of the fingerprint is reduced. GMM also provides as approximation of the feature space over the length of the multimedia clip providing invariance to various distortions. The technique is applied to

both audio and video fingerprinting. Spectral features are used in audio fingerprinting and spatio-temporal features are used in video fingerprinting. The simulation results show that the algorithm is robust to several distortions. The next chapter describes a novel algorithm for image watermarking.

Chapter 4

Watermarking

The huge success of the Internet allows for the transmission, wide distribution, and access of electronic data in an effortless manner. Content providers are faced with the challenge of how to protect their electronic data. One of the possible solutions in the area of copyright protection is data watermark, which is added to multimedia content by embedding an imperceptible and statistically undetectable signature. Thereby, multimedia data creators and distributors are able to prove ownership of intellectual property rights without forbidding other individuals to copy the multimedia content itself. The embedded watermark should be satisfy the following requirements [28]:

- Unobstructive – that is perceptually imperceptible, when embedded in the host signal.
- Discreet – undetectable to prevent unauthorized removal.
- Robust – depending upon the type of watermarks, watermarks should remain intact in the host signal when subjected to intentional removal attacks and common signal processing manipulations.
- Easily extractable – authorized watermark extraction must be easy and reliable to prove ownership.

In this chapter, a novel robust image watermarking scheme [61] is developed addressing the above issues and the performance of the proposed scheme is evaluated using a third party evaluation tool, the checkmark [62].

4.1 Proposed Scheme

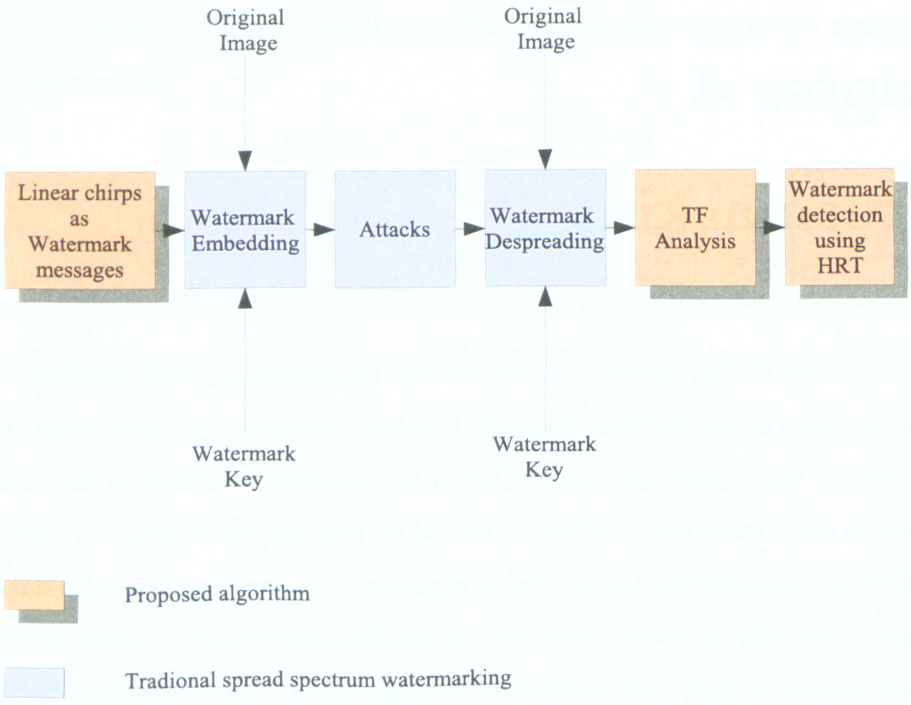

Figure 4.1: Overview of the proposed scheme.

The proposed novel robust watermarking scheme is capable of embedding multiple bits in images. Figure 4.1 shows an overview of the proposed scheme. The motivation for the proposed image watermarking algorithm is to embed, detect, and reliably extract watermark messages from a large alphabet even in the presence of bit errors caused by image or channel manipulations. To achieve this, linear frequency modulation signals (chirps) are embedded as watermark messages [63]. Different chirp rates, i.e., slopes on the TF plane, represent watermark messages such that each slope corresponds to a different message. As a result of any incidental or intentional image manipulations, some message bits of the embedded chirp, extracted by the detector may be in error potentially resulting in the detection of the wrong watermark message. Hence the objective is to detect and extract the slope of the chirp accurately in the presence of noise or discontinuities that could arise as a result

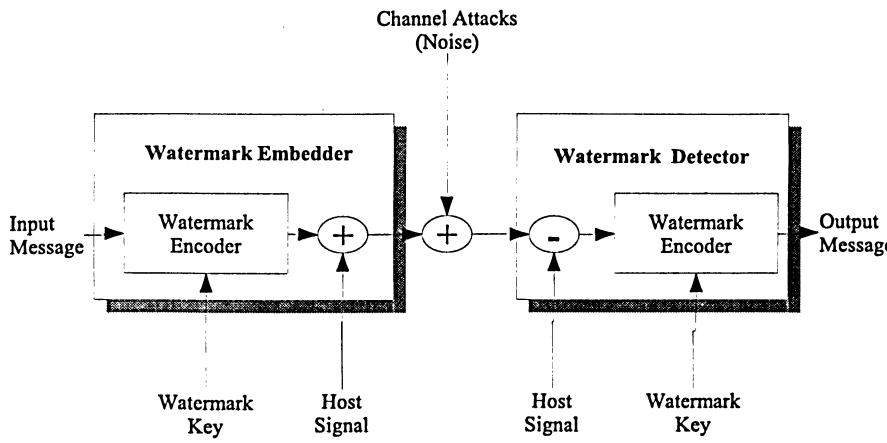

Figure 4.2: Watermarking system

of bit errors introduced by image manipulations. TF representations combine time-domain and frequency-domain analysis and hence for non-stationary signals they yield a potentially more revealing picture of temporal localization of signals spectral components. Chirps are non-stationary signals and they can be best detected in a TF plane. At the receiver, the HRT is used to detect the chirp, which appears as a straight line in the image of the TF plane. The individual blocks in the proposed algorithm are described in the following subsections.

4.1.1 Spread Spectrum Watermarking

Watermarking can be thought of a communication problem; communicating the watermark is analogous to transmission of information through a communication channel as shown in Figure 4.2. The watermark embedder and watermark detector are similar to the transmitter and receiver in a communication system. The channel in the communication system is the original content. After the watermark is embedded in to the content, the watermarked content is processed in some way, and the effect of processing can be modeled as the addition of noise. Hence solutions to communication can be applied to watermarking. Secure communications systems are important for military communications. In a secure communication system, first the unauthorized reading of the message by an adversary should be prevented. Second the adversary should not be able to find out whether there is a message sent through

the channel. Finally the signal jamming, i.e., the intentional jamming of communication between two or more people by the adversary, should be prevented. To address these issues, spread spectrum (SS) communications were developed in the early 1940s and used in the World War II. Watermarking has also the same requirements. The transmission of information through the original content should be invisible to the users and unauthorized removal of watermark should be prevented. Hence SS techniques are ideal solutions for watermarking. The next subsection describes the traditional SS communication techniques.

Discrete Spread Spectrum Communication

The discrete SS communication model developed here is based on approach followed by Flikkema [64]. Let us consider a simple communication system in two forms. In the first case, the data is transmitted as it is and in the second case data is transmitted by spreading with a pseudo-random (PN) sequence. First let us consider the non-spread digital communication model. Let the message sequence be $m_k \in \{\pm 1\}$. For keeping our discussion simple, message bits are assumed to be equiprobable. The received sequence is transmitted through the channel and exposed to additive noise. It can be expressed as

$$r_k = Em_k + n_k, \tag{4.1}$$

where E is the energy of the transmitted pulse and n_k is the additive white noise added to the signal. Let the noise n_k be a zero mean signal with a auto-covariance

$$\mathrm{E}[n_k n_{k+l}] = \sigma^2 \delta(l), \tag{4.2}$$

where E[.] is the expectation operator. The optimal receiver in this case is a level detector:

$$\begin{aligned} r_k \quad &> \quad 0; \quad \hat{m}_k = 1 \\ r_k \quad &<= \quad 0; \quad \hat{m}_k = -1, \end{aligned} \tag{4.3}$$

$\hat{m}_k$ is the estimate of the transmitted bit. r_k is a random variable with mean Em_k and variance σ^2. In this case, the probability of bit error is a function of energy of the pulse and

the variance of the noise [64]:

$$P_b = Q\left(\frac{E}{\sigma}\right), \tag{4.4}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{y^2}{2\sigma^2}} dy$.

Now let us consider a spread spectrum discrete system. In this case, instead of transmitting one bit for a duration of T, N bits $b_n, n = 0, \quad \ldots N-1$ and $b_n \in \pm 1$ are transmitted in the same duration T. b_n is obtained by spreading a message bit m_k using a spreading sequence c_n.

$$b_n = E_c m_k c_n, \quad n = 0, \ldots N-1 \tag{4.5}$$

where the energy of the bit is $E_c = E/N$ since each bit of b_n is transmitted only for a duration of T/N. The ideal spreading sequence has a mean of

$$\mathrm{E}[c_n] = \frac{1}{N} \sum_{n=0}^{N-1} c_n \approx 0, \tag{4.6}$$

and a discrete time autocorrelation of

$$\mathrm{E}[c_n c_{n+l}] = \frac{1}{N} \sum_{n=0}^{N-1} c_n c_{n+l} \tag{4.7}$$

$$\approx \begin{cases} 1 & \text{if } l = 0 \\ 0 & \text{if } l \neq 0. \end{cases} \tag{4.8}$$

These two conditions are ideal but can be closely approximated in practice using the techniques described in [65]. For example, in maximal length sequence, the number of +1 bits differs from the number of bits -1 bits. Hence the mean of the sequence is not exactly zero.

The received sequence r_n for the k^{th} transmitted bit with additive white noise j_n is given by

$$r_n = E_c m_k c_n + j_n. \quad n = 0, \ldots N-1 \tag{4.9}$$

The optimal receiver in this case correlates the spreading sequence with the received sequence and uses a level detector on the correlated value to estimate the transmitted message bit

m_k. The decision variable is

$$y = \sum_{n=0}^{N-1} (E_c m_k c_n + j_n)\, c_n, \tag{4.10}$$

$$= NE_c m_k + \sum_{n=0}^{N-1} j_n c_n, \tag{4.11}$$

and the output of the decision device

$$\hat{m}_k = \begin{cases} 1 & \text{if } y > 0 \\ -1 & \text{if } y \leq 0. \end{cases} \tag{4.12}$$

The expected value of the decision variable y in Equation 4.11 is zero as shown in the following equation.

$$\mathrm{E}[y] = \mathrm{E}[NE_c m_k] + \mathrm{E}\left[\sum_{n=0}^{N-1} j_n c_n\right] \tag{4.13}$$

$$= NE_c m_k + 0 \tag{4.14}$$

$$= Em_k. \tag{4.15}$$

The variance of the decision variable is

$$\mathrm{V}[y] = \mathrm{V}[NE_c m_k] + \mathrm{V}\left[\sum_{n=0}^{N-1} j_n c_n\right] \tag{4.16}$$

$$= 0 + N\frac{\sigma^2}{N} \tag{4.17}$$

$$= \sigma^2, \tag{4.18}$$

where V[.] is the variance operator. Again the probability of bit error depends only on the energy E and the variance σ^2, and is same as in Equation 4.4. This means that the performance of the non-spread spectrum and spread-spectrum communication systems are same in the presence of additive white noise in the channel. However, in the military communications the main concern is the jamming of the signal by the adversary. In a SS system, the spreading sequence c_n is a broad band signal and its spectrum is like a white noise. Since the message signal is spread by the spreading sequence, the transmitted signal also has a spectrum like a white noise. For an adversary, who does not know the secret spreading

sequence, the transmitted signal looks like a white noise. To jam the transmitted signal, the adversary can use a high-power narrow band signal, since using a high-power broad-band signal is practically not possible. SS systems have an inherent property of suppressing a narrow band interference [64].

Interference Suppression

Let us consider an interferer in the channel: an unknown constant I is added to the received signal. It can be shown that the decision variable for the non-spread spectrum system has an expected value of $N(E_c m_k + I)$ which will render the system unusable for sufficiently large $|I|$. Now for the SS system, the received signal is

$$r_n = E_c m_k c_n + i_n + j_n, \quad n = 0, \ldots, N-1 \tag{4.19}$$

with $i_n = I$. The decision variable at the receiver is

$$\begin{aligned} y &= \sum_{n=0}^{N-1} (E_c m_k c_n + i_n + w_n)\, j_n && (4.20)\\ &= N E_c m_k + I \sum_{n=0}^{N-1} c_n + \sum_{n=0}^{N-1} j_n c_n && (4.21)\\ &\approx N E_c m_k + 0 + \sum_{n=0}^{N-1} j_n c_n. && (4.22) \end{aligned}$$

The decision variable has a mean of Em_k and a variance of σ^2 and the interference is suppressed by despreading operation. In the above discussion, it was assumed that I is a constant signal. However, the mean and the variance of the decision variable will be close to the above mentioned values if the signal I is a narrow band signal. In watermarking, most of the attacks (the unauthorized removal of watermarks) can be modeled as a narrow band interference. Hence SS techniques are ideally suited for watermarking.

4.1.2 Time Frequency Analysis

The watermark detector has to detect and estimate the slope of the chirp. To detect the chirp, first the chirp is represented in a TF plane. This section explains why Fourier transform is

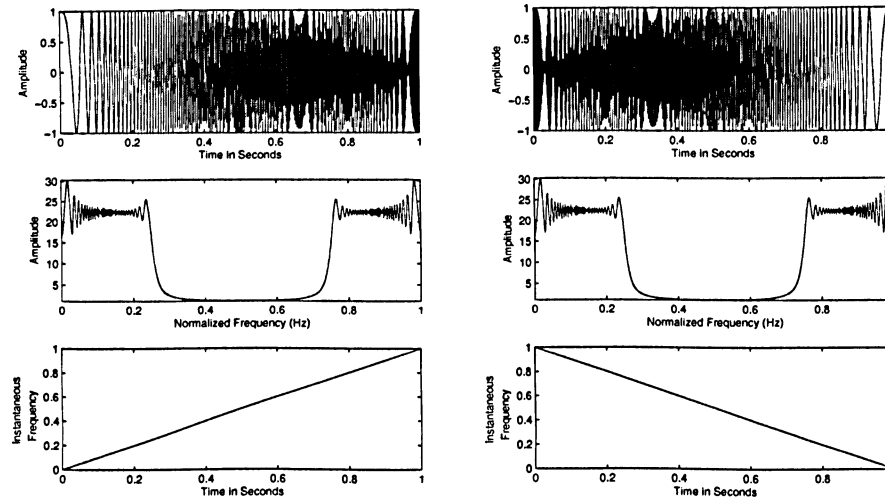

Figure 4.3: Top: Time domain representation of two chirps. **Middle:** Fourier transform **Bottom:** TF representation

insufficient for representing a chirp and explain the ways of representing a chirp by different TF distributions.

Fourier transform is a widely used technique to analyze the energy content of the signal at different frequencies. Fourier transform assumes that the signal to be stationary and evaluates the frequency content of the whole signal. To demonstrate the need for TF analysis, consider the representation of two chirps as shown in Figure 4.3. Chirps are non-stationary signals whose frequency varies with time. In Figure 4.3, two chirps are shown; one chirp whose frequency increases with time, and another one, whose frequency decreases with time. The magnitude spectrum of the two chirps are same. The inclusion of phase spectrum of the two chirps can help us differentiate the chirps, but with the magnitude spectrum alone, it is not possible to differentiate the chirps. But the TF representation clearly shows how the frequency of the chirps changes over time. TF representations show the change in frequency content of the signals over time.

Short Time Fourier Transform

The simplest TF representation is the STFT. In STFT, a sliding window divides the signal in to overlapping time segments and evaluates the Fourier transform in each segment. It is assumed that the signal is stationary for the period of the window. The discrete STFT of the signal is given by

$$X(n,\omega) = \sum_{\tau=-\infty}^{\infty} x(n+\tau)w(\tau)e^{-j\omega\tau}, \tag{4.23}$$

where $w(n)$ is the window function. $X(n,\omega)$ is the two dimensional TF representation with discrete time variable n and frequency variable ω. Usually the squared modulus of the STFT, the spectrogram is used for representation of signals and is defined as

$$S(n,\omega) = |X(n,\omega)|^2. \tag{4.24}$$

In STFT, the size of the window determines the frequency resolution. If the length of the window is long, the frequency resolution will be good but any changes in the frequency content of the signal within the period of window will not be captured. On the other hand, a small window provides a good time resolution at the cost of frequency resolution. The fixed time and frequency resolution is an important drawback of the STFT. Figure 4.4 shows the representation of chirp using a spectrogram. Chirps are highly non-stationary signal and its frequency changes with every time instant. The spectrogram exhibits poor frequency resolution due to the fixed window size. Though it is possible to identify the presence of a chirp in a spectrogram, it is not possible to accurately estimate the slope of the chirp from the spectrogram. Hence TF representations with higher frequency and time resolution are needed for accurate representation of chirp signals.

Wigner-Ville Distribution

The Wigner-Ville Distribution (WVD) has been developed to provide high time as well as frequency resolution. It is a quadratic distribution and is a special case of Cohen's class of

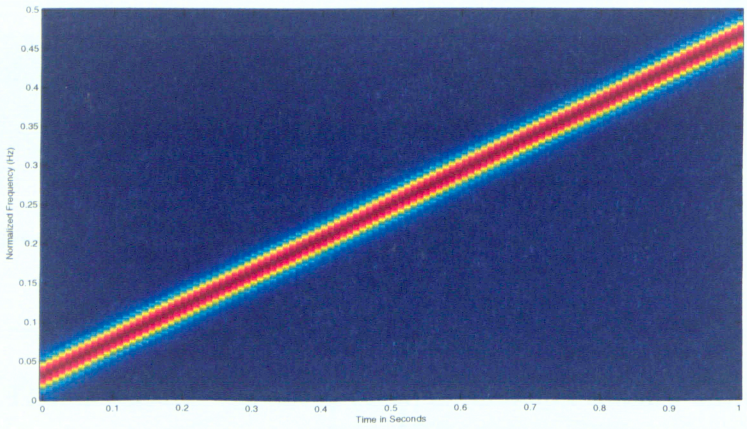

Figure 4.4: Spectrogram representation of the chirp signal

bilinear distribution [66]. The WVD of a signal $x(n)$ is given by

$$W(n,\omega) = \sum_{\tau=-\infty}^{\infty} x(n+\tau)x^*(n-\tau)e^{-j\omega\tau}, \tag{4.25}$$

where $x^*(n)$ is the complex conjugate of the signal. $W(n,\omega)$ can be seen as the Fourier transform of the autocorrelation of $x(n)$. The WVD of a chirp signal is shown in Figure 4.5. From the figure, it can be clearly seen that the resolution of WVD is much better than the spectrogram and the chirp itself is represented by a thin line. The main disadvantage of the WVD is the presence of cross terms for multi-component signals. Cross terms are spurious oscillating components that occur in the TFD. For a multi-component signal with a frequency of ω_1 at time n_1 and ω_2 at time n_2, the WVD contains a cross term $\omega_{12} = (\omega_1 + \omega_2)/2$ at time $n_{12} = (n_1 + n_2)/2$. There are many approaches to reduce the cross terms in a TFD [66]. A common approach to reduce the cross terms is to smooth the WVD. Since the cross terms are highly oscillating in nature, smoothing reduces the cross terms but also affects the resolution of the TFD. But since our objective is estimate the presence of a mono component chirp signal, WVD provides a optimal representation of the chirp signal. The next section describes how to detect and estimate the slope of the chirp from the WVD using the HRT.

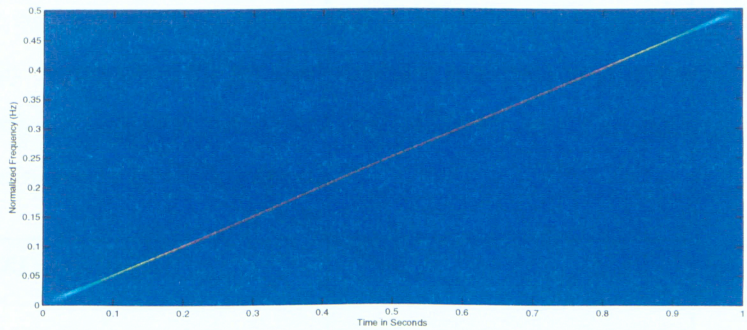

Figure 4.5: Wigner-Ville distribution of the chirp signal

4.1.3 Hough-Radon Transform

The HRT is developed by Rangayyan and Krishnan [67] to detect linear and non-linear frequency modulated signals from the image of the TFD. The Radon Transform (RT) and Hough Transform (HT) are described first and the need for a combined HRT is discussed later.

Radon Transform

The RT is used to identify straight lines in an image or a two-dimensional distribution. It is widely used in computer tomography [68]. The RT integrates a function over lines in the plane, mapping a function of position to a function of the parameters of a straight line. The RT can be expressed as

$$R\{f(x,y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(\rho - (x\cos\theta + y\sin\theta))dxdy, \tag{4.26}$$

where θ, is the angle of the ray path of integration, ρ, is the distance of the ray path from the center of the image, and δ is the Dirac delta function. The equation represents integration of $f(x,y)$ along the line $\rho = x\cos\theta + y\sin\theta$ as shown in Figure 4.6. Here the function $f(x,y)$ represents a TFD $W(t,\omega)$. Basically, the RT adds up the pixel values in the given image or TFD along a straight line in a particular direction and at a specific displacement. The RT may be applied to both binary images and gray level of any intensity distribution.

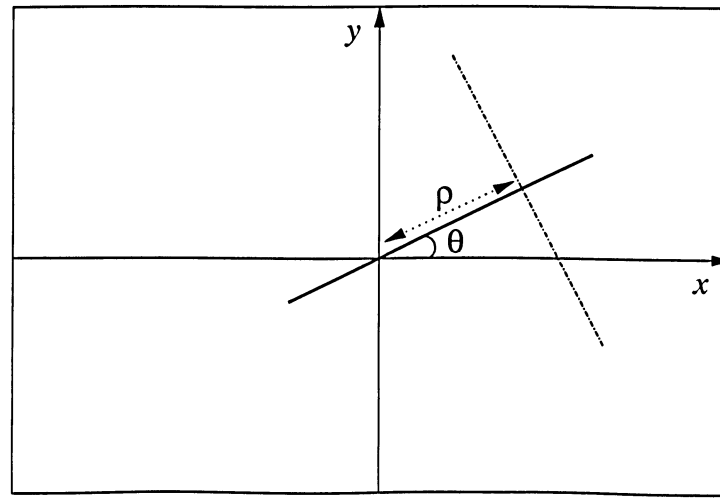

Figure 4.6: Illustration of the Radon transform

Hough Transform

The HT is a pattern recognition tool invented by Paul Hough and patented by IBM [69]. HT is used to identify lines or curves in a image that satisfy a parametric constraint and it is widely used in object detection [70], texture analysis [71] and character recognition[72]. The parametric constraint may be expressed as

$$f(X, \Delta) = 0, \tag{4.27}$$

where X is a point in the space of possible features and Δ is a point in a space of parameters. The parameter space is commonly referred to as the Hough space. Depending on the application, the constraint may be a curve, a line, or a surface. Each point Δ_0 in the parameter space represents a particular constraint that is a particular instance of a curve, line, or a surface. The constraint may be mapped into the Hough space by evaluating

$$\{X : f(X, \Delta_0) = 0\}. \tag{4.28}$$

The set of lines or curves that pass through a given feature point X_0 are given by

$$\{\Delta : f(X_0, \Delta) = 0\}. \tag{4.29}$$

Given a number of feature points that satisfy a constraint specified by parameters Δ_0, the sets generated by Equation 4.29 for each feature will contain the point Δ_0. The feature points that satisfy a particular constraint will all intersect at a common point capital Δ_0 that gives the parameters of the constraint.

Hough Radon Transform

The combined HRT is proposed by [67] is an extension of the HT. Line detection by the HT is performed by quantizing the parameter space and by incrementing the accumulator cells by one value for each pixel on a straight line. HT by itself counts only the number of pixels in a straight line. The HT is a robust method which is insensitive to missing data, and can cope well with huge sets of data. The major drawback with the HT is that it is defined for a binary image and cannot be readily applied to a gray-level image. In RT, the parameter space quantization of HT is not possible.

Given the advantages and the drawbacks of the RT and the HT individually, it is appropriate to combine the RT with the HT to identify complex TF features with varying gray level or intensity. In HRT, instead of incrementing each accumulator cell by one value for each pixel on the straight line, the energy (or gray scale value) of each pixel is added to the accumulator. This method can be applied to gray level images, and can be implemented to identify any feature that satisfies a parametric constraint.

Detection of Chirps

Let us consider the problem of detecting a chirp represented by a TFD $W(n,\omega)$. The TFD is treated as a gray scale image and the chirp is identified by detecting a straight line in the TFD image. The straight line is represented by

$$x\cos\theta + y\sin\theta = \rho. \tag{4.30}$$

The parameter space (ρ,θ), also known as the HT space, is now bounded in $\theta \in [0,\pi]$ and ρ, by the greater of rows and columns (say rows) $\pm rows/\sqrt{2}$. The HT space is divided into accumulator cells and the cell at coordinates (i,j) with accumulator value $A(i,j)$ corresponds

to the partition of the space associated with the parameter coordinates (θ_i, ρ_j). Initially, the cells are set to zero.

1. For every point (n_k, ω_k) in the TF plane, let the parameter θ equal each of the allowed subdivision values on the $\theta-$axis and solve for the corresponding ρ using Equation 4.30. The resulting ρ value is rounded to the nearest allowed value on the ρ-axis. If a particular θ_i value results in the solution ρ_j, add the gray scale intensity of the pixel is added to the accumulator $A(i, j)$.

2. A specific point $W(n_k, \omega_k)$ is represented as a sinusoidal curve in the HT space. All of the sinusoids resulting from a mapping of a line in the TF plane have a common point of intersection in the HT space. Thus, the straight line correspond to a high intensity point in the HT space. A suitable threshold value is chosen for the accumulator cell values and the straight line (chirp) is declared as detected only at least one of the accumulator value is above the threshold. The cell with the highest value gives the parameters of the line.

This section described the components used in the proposed scheme. The following sections describe the proposed algorithm and results obtained for various robustness tests.

4.2 Watermark Embedding

The overview of the embedding scheme is shown in Figure 4.7. The main steps of the proposed algorithm are as follows:

1. Selection of the watermark message.

2. Generation of watermark, which includes spreading the watermark message bits using a watermark key to generate a spread spectrum signal.

3. Finding perceptually significant regions in the images using perceptual models based on the human visual system.

4. Embedding the watermark message.

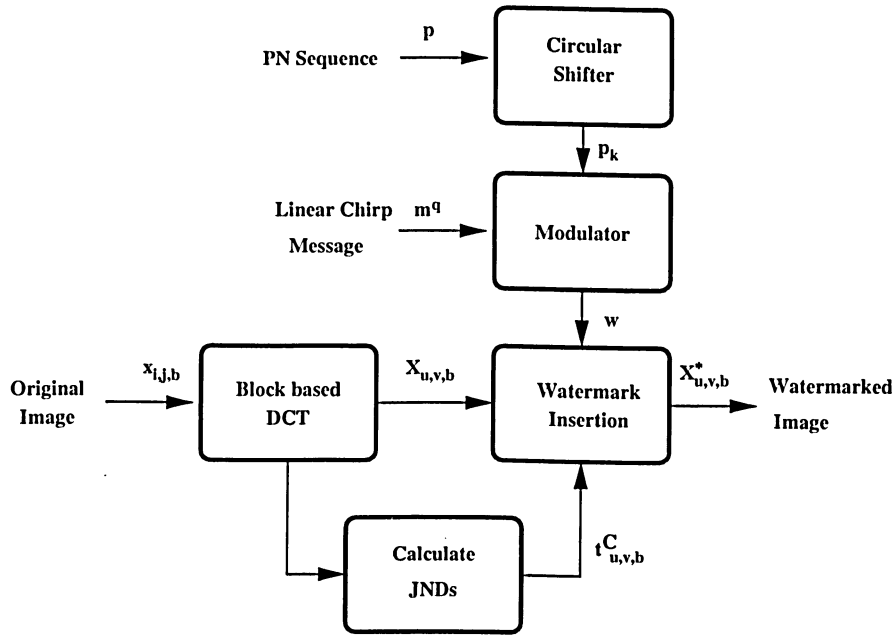

Figure 4.7: Proposed watermark embedding scheme

4.2.1 Selection of the Watermark Message

As described earlier, linear chirps are used as watermark messages. The watermark message is the slope of the normalized chirp $\mathbf{m}$ on a TF plane, with initial and final frequencies f_0 and f_1, respectively. Distinct pairs of initial and final frequencies form different watermark messages, each represented by a unique slope in the TF plane. $\mathbf{m}$ is an N_m - sample long sequence sampled at the sampling frequency f_s, and its n^{th} element m_n is defined as

$$m_n = \sin\left(\pi \frac{f_1 - f_0}{N_m} n^2 + 2\pi f_0 n\right). \tag{4.31}$$

$\mathbf{m}$ takes continuous values in the interval [-1,1], and needs to be quantized. $\mathbf{m}^q$ is the quantized version of $\mathbf{m}$ formed according to the sign of the sample values of $\mathbf{m}$, taking values -1 and 1.

$$\mathbf{m^q} = [m_1^q, m_2^q, \ldots, m_{N_m}^q] \tag{4.32}$$

The WVD of the continuous and quantized versions of the chirps are shown in Figure 4.8. The quantization introduces additional cross terms in the WVD, but it does not affect the accuracy of the slope estimation because of the use of HRT in the receiver.

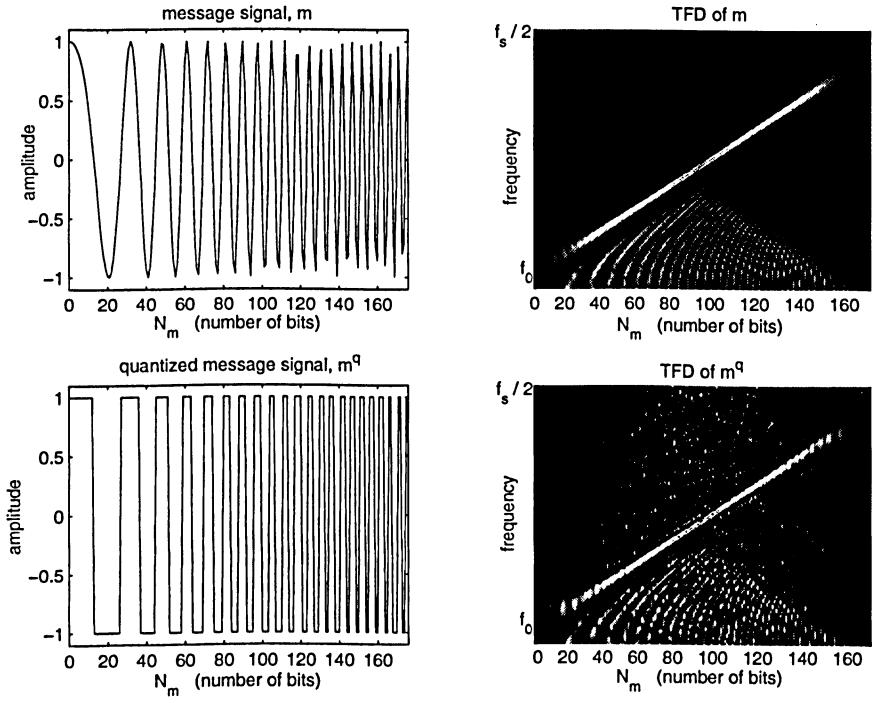

Figure 4.8: Time-domain and TF representation of chirp signals

4.2.2 Generation of Watermark Message

To embed each watermark message bit in $\mathbf{m^q}$ into the image, each bit m_k^q is spread with a cyclic shifted version $\mathbf{p_k}$ of a binary PN sequence with a chip length of N and summed together to generate the wideband noise vector $\mathbf{w}$.

$$\mathbf{w} = \alpha \sum_{k=0}^{M} m_k^q \mathbf{p_k}, \tag{4.33}$$

where M is the number of watermark message bits in $\mathbf{m^q}$ and α is a scaling factor which is chosen appropriately to keep embedded watermark imperceptible. There is always a possibility to make the trade-off between the embedded data size and robustness of the algorithm; as the PN length is decreasing, the algorithm is able to add more bits into the host image but the detection of the hidden bits and resistance to different attacks is decreased. The wideband noise vector $\mathbf{w}$ formed is added to the image in perceptually significant regions to ensure robustness of the watermark against attacks. The length of $\mathbf{w}$ and hence the number of watermark bits that can be embedded depends on the perceptual entropy of the image.

4.2.3 Perceptual Model

To find the perceptually significant regions in the image, models that describe the masking characteristics of the human visual system can be utilized [73]. Among such models, a model based on the *just noticeable difference* (JND) paradigm [74] is used. A set of JNDs is associated with a particular invertible transform T. Given that a multimedia signal is transformed using T, the JNDs provide an upper bound on the extent that each of the coefficients can be perturbed without causing perceptual changes to the signal quality. The set of signal and transform dependent JNDs can be derived using complex analytic models or through experimentation. The JND paradigm is widely used in image compression, and image watermarking applications. The JNDs are used to determine the perceptually significant regions and also to find the perceptual entropy of the image. This work uses one such model based on the DCT, which is discussed in the following paragraphs.

The model used was proposed by Watson [75] that has been applied to JPEG coding. In this method, the original image is decomposed into non-overlapping 8×8 blocks, and the DCT is performed independently for every block of data. Let the original image pixels are

represented as $x_{i,j,b}$, where i and j represent the pixel elements in block b, and $X_{u,v,b}$ denotes the DCT coefficients for the basis function located in the position u, v of the block b. A frequency threshold value is derived for each DCT basis function and in this case result in an 8×8 matrix of $t^F_{u,v}$ threshold values. These threshold values are determined for various viewing conditions by Peterson *et al.* [76]. The visual model used is for a minimum viewing distance of four picture heights and a D65 monitor white point. Watson further refines this model by adding a luminance sensitivity and contrast masking component. Luminance sensitivity threshold is estimated by the formula,

$$t^L_{u,v,b} = t^F_{u,v} \left(\frac{X_{0,0,b}}{X'_{0,0}} \right)^a, \tag{4.34}$$

where $X_{0,0,b}$ is the DC coefficient of the DCT for block b, $X'_{0,0}$ is the DC coefficient corresponding to the mean luminance of the display, and a is a parameter which controls the degree of luminance sensitivity. The authors in [76] suggest to set the value of a to 0.649. Given a DCT coefficient and a corresponding threshold value derived from the viewing conditions and local luminance masking, a contrast masking threshold is derived as

$$t^C_{u,v,b} = \max\left[t^L_{u,v,b}, |X_{u,v,b}|^{w_{u,v}}, (t^L_{u,v,b})^{1-w_{u,v}} \right], \tag{4.35}$$

where $w_{u,v}$ is a number between zero and one, and is empirically derived as 0.7 by the authors in [76]. The contrast masking threshold $t^C_{u,v,b}$ is the calculated JND for the image and represents the extent to which the DCT coefficients can be perturbed without producing any visual artifacts.

4.2.4 Embedding the Watermark

The watermark embedding scheme is based on the model proposed in [38]. The watermark encoder for the DCT scheme is described as

$$X^*_{u,v,b} = \begin{cases} X_{u,v,b} + t^C_{u,v,b} w_{u,v,b}, & \text{if } X_{u,v,b} > t^C_{u,v,b}; \\ X_{u,v,b}, & \text{otherwise} \end{cases} \tag{4.36}$$

where $X_{u,v,b}$ refers to the DCT coefficients, $X^*_{u,v,b}$ refers to the watermarked DCT coefficients, $w_{u,v,b}$ is obtained from the wideband noise vector w, and $t^C_{u,v,b}$ is the computed JND calculated from the visual model described in Equation 4.35.

4.3 Watermark Detection

The received image may be different from the watermarked image due to some intentional or unintentional image processing operations such as lossy compression, shifting and down-sampling. Figure 4.9 shows the block diagram of the described watermark decoding scheme.

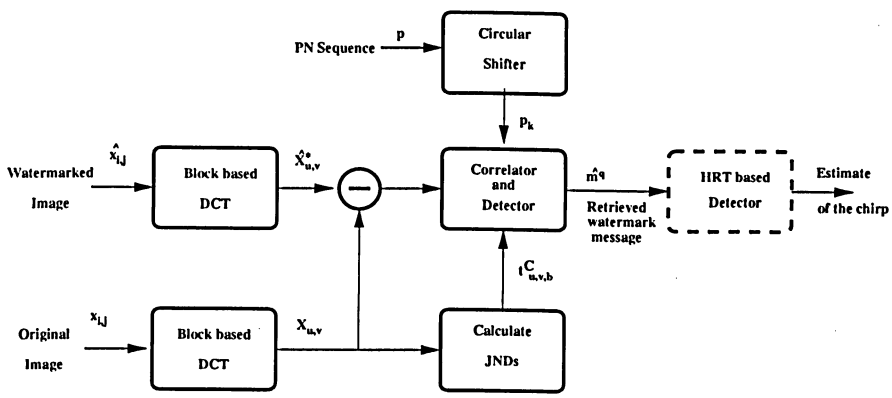

Figure 4.9: Proposed watermark detection scheme

The detection scheme for the DCT based watermarking can be expressed as

$$\hat{w}_{u,v,b} = \frac{X_{u,v,b} - \hat{X}^*_{u,v,b}}{t^C_{u,v,b}} \tag{4.37}$$

$$\hat{\mathbf{w}} = \begin{cases} \hat{w}_{u,v,b}, & \text{if } X_{u,v,b} > t^C_{u,v,b}; \\ 0 & \text{otherwise} \end{cases} \tag{4.38}$$

where $\hat{X}^{*}_{u,v,b}$ are the coefficients of the received watermarked image, and $\hat{\mathbf{w}}$ is the received wideband noise vector. The received wideband noise vector can be expressed as

$$\hat{\mathbf{w}} = \mathbf{w} + \mathbf{n}, \tag{4.39}$$

where $\mathbf{n}$ is the distortion component resulting from hostile image manipulations and is modeled as a zero-mean random vector uncorrelated with the PN sequence. The watermark key, i.e., the appropriately circular shifted PN sequence $\mathbf{p_k}$ is used to despread $\hat{\mathbf{w}}$, and integrate the resulting sequence to generate a test statistic $\langle \hat{\mathbf{w}}, \mathbf{p_k} \rangle$. The sign of the expected value of the statistic depends only on the embedded watermark bit m^q_k. Hence the watermark bits can be estimated using the decision rule:

$$\hat{m}^q_k = \begin{cases} +1, & \text{if } \langle \hat{\mathbf{w}}, \mathbf{p_k} \rangle > 0; \\ -1, & \text{if } \langle \hat{\mathbf{w}}, \mathbf{p_k} \rangle < 0. \end{cases} \tag{4.40}$$

The bit estimation process is repeated until an estimate of all the transmitted bits are obtained. Though it is possible to form an estimate of the chirp sequence from the received bits, as shown in Figure 4.10, the robustness of the detection algorithm is improved by localizing the chirp in a TFD and detecting the chirp using a line detection algorithm.

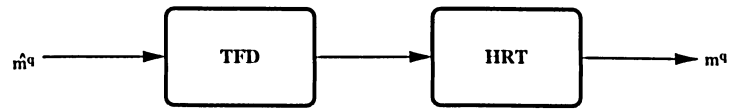

Figure 4.10: Postprocessing of extracted bits

4.4 Post-processing of the Estimated Bits for Watermark Message Extraction

After all watermark bits are extracted, a TFD of the extracted watermark bits in $\{\hat{m}_k, k = 1, \ldots, N_m\}$ is constructed. Each TFD is characterized by a different TF resolution [66]. Let

$$I_T = T(\hat{m}_1, \ldots, \hat{m}_{N_m}) \tag{4.41}$$

be the $K \times N$ image matrix of the TF plane resulting from the TFD of the estimated bits using the TF representation T. K is the number of rows in I_T that corresponds to the number of frequency slots, and N is the number of columns that corresponds to the number of time slots. A directional element detector can be used to detect time-varying energy values. The line detector that can satisfy our needs is a detector that uses the combination of Hough and Radon transforms as proposed in [67]. This detector has been mathematically proven to be an optimal detector as it provides the maximum likelihood identification of a chirp signal [77]. Using the combined HRT, the pixels that form a parametric constraint in a gray level image can be detected. These constraints can be straight lines or curves in the image of the TF plane. Since linear chirps are embedded, the parametric constraint equation corresponding to a line equation expressed in terms of the parameter vector $\Theta = (\rho, \theta)$ is

$$\rho - (n\cos\theta + k\sin\theta) = 0. \tag{4.42}$$

With the above constraint equation the HRT can be expressed as [77]

$$H(\Theta) = \sum_{n=1}^{N}\sum_{k=1}^{K} I(k,n)\,\delta(\rho - n\cos\theta + k\sin\theta). \tag{4.43}$$

In the implementation of the HRT, the transform value at some point (ρ_0, θ_0) in the Hough space contains the total energy in the pixels that satisfy the parameter constraint equation.

Therefore, a HRT based system can be devised to effectively detect directional elements defined by parametric equations: the peak values of the HRT will yield the most likely parameter values. (Figure 4.11 shows the WVD and the Hough space of the linear chirp received with 19% of the bits in error. The prominence of the global maximum in the HRT space provides an indication of the presence of chirp in TFD and thereby leading to successful watermark detection.) There are three main considerations that affect the performance of

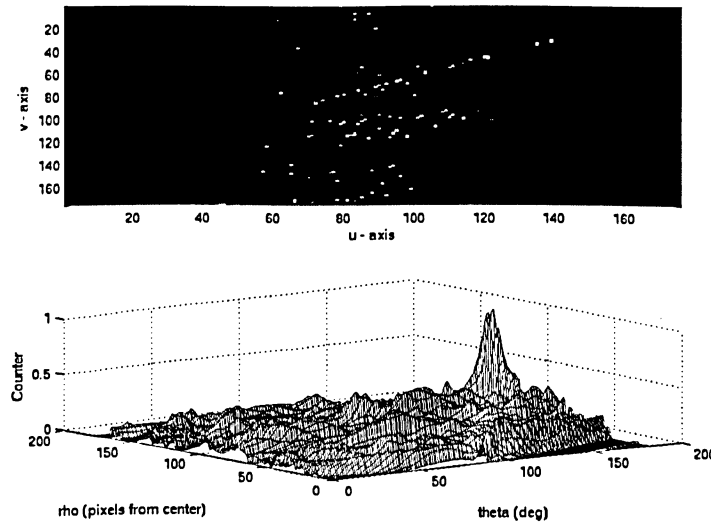

Figure 4.11: Line detection using HRT

the HRT in detecting the chirps:

4.4.1 Selection of the TFD

Since the HRT will operate on I_T, the choice of the TFD method used will have an influence on the detection of the chirp. For example, the STFT may have limited TF resolution, and the WVD may exhibit cross-terms for multi-component signals, which may result in false detection of the watermark. In this case, as mono component signals are detected, the WVD, which provides the highest resolution, can be used.

4.4.2 Quantization of the HRT Parameter Space

Prior to the start of the search for chirp parameters, an appropriate search grid for the HRT parameter space has to be selected. This is equivalent to determining the proper quantization level of the HRT parameters that will allow the potential detection of the parameters of all possible chirps. The approach used by Erkucuk [49] is used to determine the quantization parameters. Let θ_k be the angle occurring from the image of a line with its initial and end points corresponding to the first and the k^{th} frequency slots, respectively. To find the minimum step size for θ, $\Delta\theta_k = \theta_k - \theta_{k-1}$ is evaluated as a function of k such that

$$\Delta\theta_k = \tan^{-1}\left(\frac{k}{N}\right) - \tan^{-1}\left(\frac{k-1}{N}\right), \tag{4.44}$$

where $1 \leq k \leq K$. The minimum value of $\Delta\theta_k$ is achieved at Δ_K as shown in Figure 4.12. Let ρ_k be the distance of the line with initial and end points corresponding to the first and k^{th} frequency slots, measured relative to the center of the image. Then $\Delta\rho_k = \rho_k - \rho_{k-1}$ for each $1 \leq k \leq K$ is calculated as

$$\Delta\rho_k = \left[\left(\frac{K}{2}\cos\theta_k - \frac{N}{2}\sin\theta_k\right) - \left(\frac{K}{2}\cos\theta_{k-1} - \frac{N}{2}\sin\theta_{k-1}\right)\right]. \tag{4.45}$$

The minimum value of $\Delta\rho_k$ is achieved at $\Delta\rho_1$ as shown in Figure 4.12. $\Delta\theta_K$ and $\Delta\rho_1$ are the largest quantization steps needed to evaluate all possible lines in the $K \times N$ image plane.

4.4.3 Threshold Level Selection in the HRT Space

To detect the presence of chirp, the peak value in the Hough space is compared with a threshold. If the peak value is less than the threshold, the detector concludes that there is no watermark. Only if the peak value is above the threshold, the parameters of the chirp are estimated. The choice of the threshold significantly affects the performance of the detector. If a low threshold value is used, there will be a high probability of false detection. The

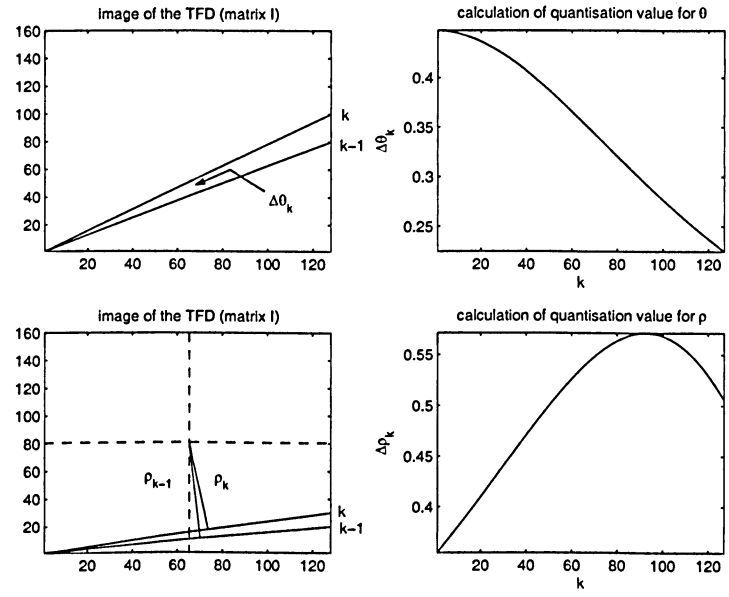

Figure 4.12: Quantization values for θ and ρ

analytical expressions showing the relation between the threshold and the probabilities of correct and false detection are shown in [61].

4.5 Results and Discussion

The proposed scheme is evaluated using the non geometric attacks proposed in the checkmark benchmark scheme [62] for five different images of size 512×512. The test images are shown in Figure 4.13. The sampling frequency f_s of the watermarks equal 1 kHz. Hence the initial and final frequencies, f_0 and f_1 of the linear chirps representing all watermark messages are constrained to [0-500] Hz. These chirps are embedded in to the images for a chip length of N, which depends on the perceptual entropy of the image. It was found experimentally found that for reliable detection of chirp under various image processing attacks, the chip length should be at least 10000 samples. If the image can support more 10000 samples, then multiple chirps can be embedded and the payload can be increased.

In out tests, a single watermark sequence having 32 and 128 message bits is used. For

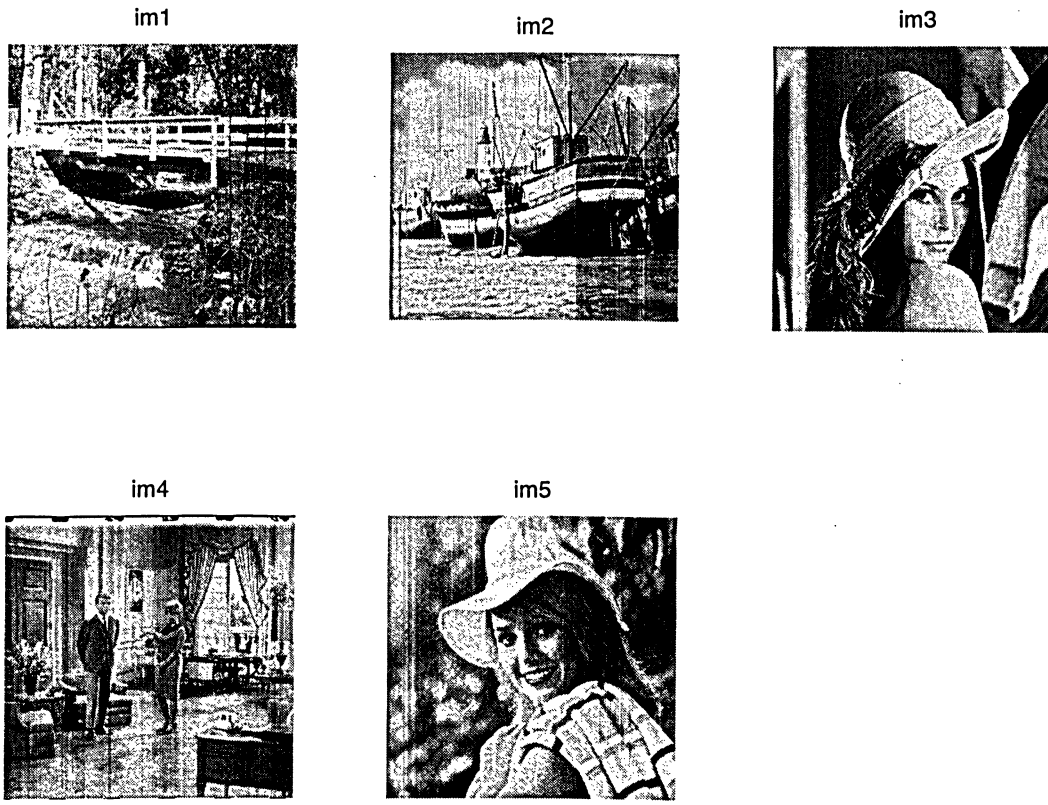

Figure 4.13: Test images.

each test, the watermark embedded image $X^*_{u,v}$ is processed to generate the received signal $\hat{X}^*_{u,v}$ resulting from the image manipulation operations described in that checkmark tests. The watermark extraction algorithm described in Section 4.3 is applied on the received signal $\hat{X}^*_{u,v}$. Then the 128×128 image matrix I_T is formed. Therefore, the total number of possible messages is 128^2. Due to the size of I_T, θ was bounded by $\pm\pi/4$ (i.e., $\theta_{max} = \pi/4$) and ρ by 64 (i.e., $\rho_{max} = 64$). As a result of the HRT, the global maxima found in the HRT spaces exactly corresponded to the frequency parameters of the embedded chirp, thus resulted in extracting the correct watermark messages. In all the tests performed, it was assumed that the PN sequences used in the watermark embedding and extraction processes are synchronized. Table 4.1 and Table 4.2 show the watermark detection results on five watermarked images after performing the attacks specified in the checkmark benchmark.

The numbers in the brackets under category 'Attack' represent the number of attacks in that particular class of attacks. The number of attacks for which the watermark is detected is shown under the 'Images' category. The 'Average' represents the percentage of attacks for which the watermark is detected under each class of attacks.

Table 4.1: Watermark detection results for checkmark benchmark attacks with chirp length 32 bits.

Attacks	Images					
	im1	im2	im3	im4	im5	Average (%)
Remodulation(4)	2	0	4	1	1	40
MAP(6)	6	6	6	6	6	100
Copy(1)	1	1	1	1	1	100
Wavelet(10)	9	9	9	9	10	92
JPEG(12)	12	12	12	12	12	100
ML(7)	4	4	4	3	5	57
Filtering(3)	3	3	3	3	3	100
Resampling(2)	2	2	2	2	2	100
Colour Reduce(2)	2	0	1	2	1	60

Table 4.2: Watermark detection results for checkmark benchmark attacks with chirp length 128 bits

Attacks	Images					
	im1	im2	im3	im4	im5	Average (%)
Remodulation(4)	4	2	4	3	2	75
MAP(6)	6	6	6	6	6	100
Copy(1)	1	1	1	1	1	100
Wavelet(10)	10	9	10	9	10	96
JPEG(12)	12	12	12	12	12	100
ML(7)	6	4	6	3	5	69
Filtering(3)	3	3	3	3	3	100
Resampling(2)	2	2	2	2	2	100
Colour Reduce(2)	2	2	1	2	2	90

Table 4.1 presents the results for chirp length 32 bits and Table 4.2 shows the results

for chirp length 128 bits. As expected, higher detection results are obtained with increasing chirp length. For comparison, the results of the watermarking scheme proposed by Pereira [78] are presented in Table 4.3 which can embed 56 bits in the image. Though the number of bits that can be embedded using our algorithm is lesser than [78], the results are significantly better than [78].

Table 4.3: Watermark detection results for checkmark benchmark attacks for the scheme proposed by Pereira *et al.*

Attacks	Images					
	im1	im2	im3	im4	im5	Average (%)
Remodulation(4)	0	0	0	2	0	10
MAP(6)	1	2	2	5	1	37
Copy(1)	1	1	0	1	0	60
Wavelet(10)	3	6	8	7	7	62
JPEG(12)	12	12	12	12	12	100
ML(7)	0	3	3	3	0	26
Filtering(3)	3	3	3	3	3	100
Resampling(1)	1	1	1	1	1	100
Colour Reduce(2)	0	1	1	2	0	40

The algorithm performs well for wavelet and JPEG compression attacks. For these attacks, the bit error rates are quite low, usually less than 20%. Figure 4.14 shows the successful detection of watermark for an image after wavelet compression. The bit error rate in this case is 17%. But in the case of maximum likelihood (ML) or remodulation attacks, the bit error rates are quite high. For example, the Table 4.4 shows the bit error rates for the image im5 for ML and remodulation attacks. When the bit error rates are higher than 20%, the HRT is unable to detect the chirp.

The performance of the algorithm can be compared with other existing schemes by comparing the number of attacks for which the watermark is detected. The checkmark tests have a total of 230 attacks (46 attacks for each image). The proposed algorithm with a chirp length of 128 bits is able to detect the watermark for 216 attacks giving a 94% detection rate. This is better than the detection rates the watermarking schemes proposed by Cox

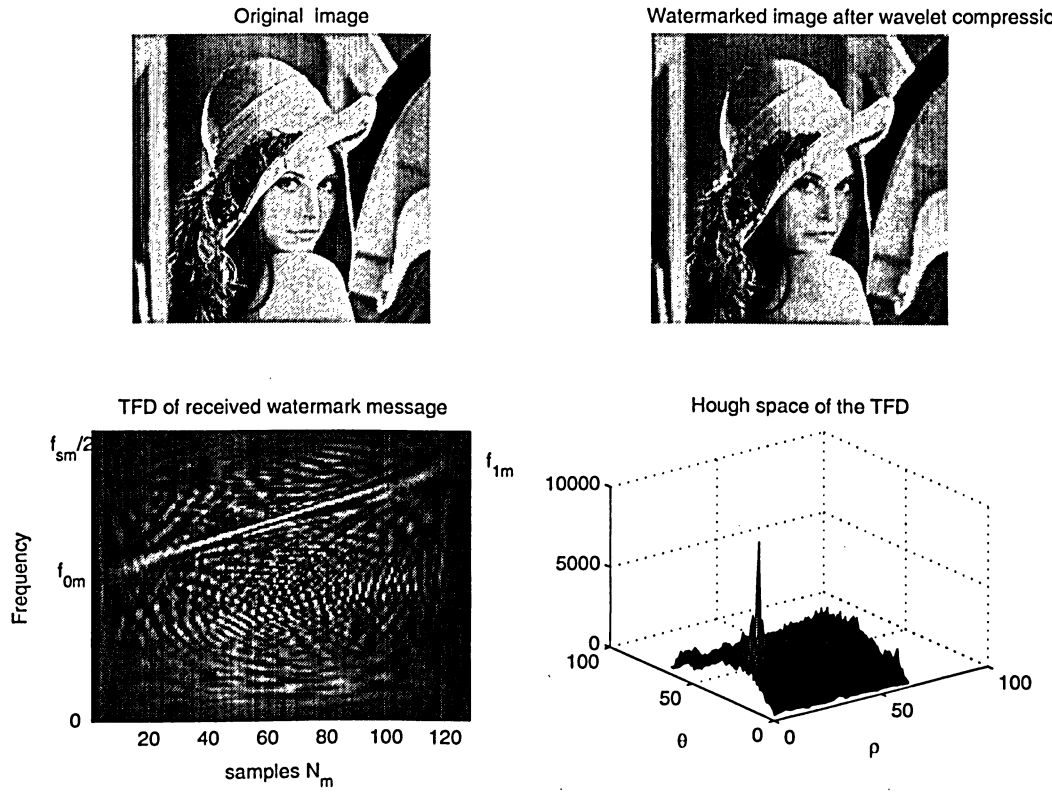

Figure 4.14: Performance of HRT for wavelet compression

[34], Xia [79] whose detection rates are of 90% and 84% respectively. These two schemes can embed a single bit while the proposed scheme can embed multiple bits. The algorithm porposed by Pereira [78] can embed 56 watermark bits and gives a detection result of 61%. However, the algorithm is a blind algorithm which does not use the original image at the receiver.

In this chapter, a novel watermarking algorithm for images that uses chirp signals as watermark message is presented. The algorithm uses the traditional SS based watermarking techniques for embedding and detecting the watermark signals. After the watermark signals are extracted, they are post processed: first by representing the extracted watermark message by a TFD and then by detecting the chirp in the image of the TFD by using a line detection tool, the HRT. The robustness of the algorithm is tested using a standard benchmark called

Table 4.4: Bit error rates for ML and Remodulation attacks

Attacks	im5	
	Bit error rates(%)	Detection result after HRT
Remodulation(4)		
dpr1	10.16	1
dpr2	33.59	0
dprcorr1	11.72	1
dprcorr2	35.16	0
ML(7)		
medfilt1	0.78	1
medfilt2	1.2	1
medfilt3	27.18	0
trimmedmean1	3.90	1
trimmedmean2	43.75	0
midpoint1	0.0	1
midpoint2	39.94	0

checkmark and the algorithm is found to be quite robust to many of the attacks.

Chapter 5

Conclusions and Future Research

5.1 Conclusions

THIS thesis proposed novel solutions for two content protection technologies: fingerprinting and watermarking. The Introduction explained why digital data is susceptible to piracy and discussed the need for a DRM system to protect digital content and enable its electronic distribution. Chapter 2 described how encryption alone is not sufficient to protect content and explained two technologies namely fingerprinting and watermarking that offer protection to data after it is decrypted. Chapter 3 proposed a novel fingerprinting scheme that models fingerprints by GMM using spectral features. Chapter 4 proposed a novel SS based watermarking scheme that includes a post processing stage based on TF analysis and HRT detector. The summary of the work and future extension of the work are presented in this chapter.

Fingerprinting

A novel fingerprinting scheme that models audio or video features using Gaussian mixture models is introduced in this thesis. The fingerprints are compact and are found to be invariant to various types of distortions. The fingerprints corresponding to different multimedia clips show excellent discrimination. Under different distortions, the percentage of clips correctly identified at different false positive rates are evaluated. A false positive occurs when a fingerprint of clip matches with the fingerprint of another clip. Based on the simulations

and the results obtained in chapter 3 the following observations could be made:

- GMM help reduce the dimensionality of the inputs significantly. For example, a 5 second audio clip sampled at 44.1 KHz ($5 \times 44,100$ samples/s $= 220,500$ samples) modeled by 16-cluster GMM (16 means + 16 variances + 15 weights = 47 coefficients) results in a reduction of more that 4600 times.

- The invariance to distortions is due to the modeling by GMM. GMM approximates the feature space over the entire duration of the clip. This makes it robust to various distortions. However, this approximation results in reduction in discrimination between fingerprints of different clips and results in false positives.

- For audio fingerprinting, the fingerprints are modeled using various spectral features. The performance of the features can be compared by the identification rate vs. false positive curves. Among the different features used, spectral centroid gives the best identification rates of 99.1%, 99.2%, 99.7% at false positive rates of $10^{-4}, 10^{-3}, 10^{-2}$ respectively. Spectral roll-off and spectral flux performs poorly and their identification rates are 26.6% and 59.1% at a false positive rate of 10^{-4}. The reason for their poor performance is because they are scalar features, while the rest of the spectral features considered are vectors.

- Two novel features Shannon entropy and Renyi entropy are introduced for fingerprinting. Among the two, Renyi entropy performs better than the Shannon entropy. Renyi entropy performs almost as good as spectral centroid with an identification rate of 97.8% at a false positive rate of 10^{-4}.

- The proposed video fingerprinting is very robust to MPEG compression and filtering operations. The fingerprinting scheme is very tolerant to additive white Gaussian noise giving an identification rate of 100% up to a noise variance of 16. But the scheme is not very robust to shifting and resizing operations. When more than 6% of the lines in the video are shifted up, the performance degrades rapidly. Similarly when the video

is resized from 0.9% to 1.2%, the identification rate is above 95% beyond which the performance degrades.

- The identification rate of the proposed audio fingerprinting algorithm is comparable or better than other existing audio fingerprinting schemes. While many schemes have been tested only for a limited number of distortions, the proposed algorithm is robust to wide variety of distortions. However, the false positive rates of the proposed scheme are higher than some of the other schemes. The proposed video fingerprinting algorithm shows a better identification rate for distortions that locally modify the video such as MPEG compression and filtering. For global distortions, the identification rates are comparable to other schemes.

Watermarking

A novel robust watermarking scheme for images that uses linear chirp signals as watermark messages is proposed in this thesis. The watermark messages are embedded to a host image and retrieved from the image using the traditional SS based techniques. After the watermark messages are extracted, they are post processed by representing the received watermark message in a TF plane, and detecting the watermark message using a HRT based detector. The watermarking scheme is capable of carrying multiple bits in the watermark message. The performance of the scheme is evaluated by a benchmark called *checkmark* developed by Pereira [62]. Based on the research in this area and our experimental results, the following deductions could be made:

- Among the checkmark tests performed, the remodulation and maximum likelihood attacks resulted in the maximum bit error rates. A maximum bit error rate of 43.75% is obtained for a maximum likelihood attack. The simulation results show that the HRT applied to a TFD generated by a WVD can successfully detect the watermark and estimate all the bits correctly when the bit error rates are less than 20%.

- The proposed scheme is very robust to JPEG compression, filtering, resampling and

copy attacks and the HRT is able to detect the watermark and estimate the bits accurately for all the attacks in these categories and for all the five test images. For wavelet compression and color reduction, the algorithm is fairly robust, identifying 96% and 90% of attacks correctly.

- The performance of the algorithm depends on the length of the chirp signal embedded. It is intuitive that higher the length of the chirp better the performance is. The overall percentage of attacks for which the watermark is detected is 87% with a chirp length of 32 bits and 94% for a chirp length of 128 bits. Increasing the chirp length further increases the performance. But it may also leave some visual artifacts in the watermarked image.

- The HRT is able suppress the low power cross terms that arise due to the quantization of the chirp signals. The TFD generated by WVD provide the highest resolution for the watermark signal and optimal for the detection of watermarks.

- In the proposed watermarking scheme, it was assumed that the PN sequences used in the receiver and the detector are synchronized. However, attacks like rotation and other geometric transformations removes the synchronization of the PN sequence. This is an outstanding issue and is not discussed here since the objective is to show the watermark detection can be improved by using a TF analysis followed by HRT at the receiver.

- The proposed algorithm has a overall detection rate of 94% for the checkmark tests. This is better than many single bit non-blind watermarking schemes such as Cox [34] and Xia [79] and multiple bit blind watermarking schemes such as Pereira [78].

5.2 Future Research

Fingerprinting

In the fingerprinting scheme proposed in chapter 3, there are potential areas that would improve the performance of the scheme.

- In this work, different spectral features are used for audio fingerprinting. The features used are heuristic and are successfully used in music classification, speech/music discrimination and other applications. However, these features may not be optimal for fingerprinting. One way to extract optimal features is to use OPCA on the DCT coefficients as proposed by Burges *et al.* [17]. Generating fingerprints by modeling these optimal features using GMM might improve the robustness of the system.

- Since the objective in this thesis is to do a preliminary evaluation of the proposed scheme, a relatively small database is used. In a typical fingerprinting application such as broadcast monitoring or mobile music recognition, there will be millions of multimedia clips in the database. It will be interesting to evaluate the robustness of the proposed algorithm with a realistic database.

Watermarking

There are potential areas of research in the proposed watermarking algorithm that could improve its performance and capacity.

- One way to increase the payload of the watermarking algorithm is to use more complex parametric curves such as higher order polynomials as watermark message. This will increase the payload but also increase the complexity in the HRT detector since the dimension of the Hough space increases. But by using faster versions of Hough transform, the complexity can be significantly reduced.

- The empirical limit for the HRT to detect the watermark message is 20% bit error rate. An analytical expression to determine the performance of the HRT at different bit error rates can help understand the error-correcting capability of the HRT.

Appendix A

List of Publications

In this section, the publications resulted from this research work are presented.

Journals

- A. Ramalingam and S. Krishnan, "Robust image watermarking using chirp detection based technique," in press, *IEE Transactions on Vision, Image and Signal Processing,* scheduled for August 2005.

- A. Ramalingam, and S. Krishnan, "Gaussian Mixture Modeling using Short Time Fourier Transform Features for Audio Fingerprinting," submitted to *IEEE Transactions on Information Forensics and Security,* May 2005.

- A. Ramalingam, and S. Krishnan, "Video fingerprinting using space-time and Gaussian mixture models," submitted to *IEEE Transactions on Circuits and Systems - II,* July 2005.

Conferences/Workshops

- A. Ramalingam and S. Krishnan, "Gaussian Mixture Modeling using Short Time Fourier Transform Features for Audio Fingerprinting," in *Proceedings of the International Conference on Multimedia and Expo*, Amsterdam, July 2005.

- A. Ramalingam and S. Krishnan, "Video fingerprinting using space-time and Gaussian mixture models" in *Proceedings of the ninth Canadian Workshop on Information Technology,* Montreal, pp. 288 - 291, June 2005.

- A. Ramalingam and S. Krishnan, "A Novel Robust Image Watermarking Using a Chirp Based Technique," in *Canadian Conference on Electrical and Computer Engineering,* Niagara Falls, vol. 4, pp. 1889 - 1892, May 2004.

- A. Ramalingam and S. Krishnan, "Multimedia Fingerprinting," in Micronet Annual workshop, Ottawa, May 2005.

- S. Esmaili, A. Ramalingam and S. Krishnan, "Watermarking and Retrieval of Multimedia Data Using Advanced DSP Techniques," in Micronet Annual workshop, Aylmer, 2004. **(Best Student Paper Award under systems category)**

Bibliography

[1] "Kazaa software, [online] Available: http://www.kazaa.com."

[2] "Bittorrent, [online] Available: http://www.bittorrent.com."

[3] "edonkey and overnet, [online] Available: http://www.edonkey2000.com."

[4] "IFPI, world sales 2001, [online]
Available: http://www.ifpi.org/site-content/statistics/worldsales.html."

[5] "Market data pages of the web site, [online] Available: http://www.riaa.org."

[6] J. Valenti, "Piracy threatens to destroy movie industry and U.S. economy, in testimony before the senate foreign relations committee, February, 12, 2002."

[7] M. Fetscherin, "Present state and emerging scenarios of digital rights management systems," *The International Journal on Media Management*, vol. 4, no. 3, pp. 164–171, 2002.

[8] E. T. Lin, A. M. Eskicioglu, R. L. Lagendijk, and E. J. Delp, "Advances in digital video content protection," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 171–183, January 2005.

[9] V. Venkatachalam, L. Cazzanti, N. Dhillon, and M. Wells, "Automatic identification of sound recordings," *IEEE Signal Processing Magazine*, vol. 21, no. 2, pp. 92–99, March 2004.

[10] B. Schneier, *Applied Cryptography.* New York: Wiley, 1997.

[11] A. Menez, P. van Oorschot, and S. Vanstone, *Handbook of Applied Cryptography.* Baca Raton, FL: CRC Press, 1997.

[12] P. Cano, E. Batle, T. Kalker, and J. Haitsma, "A review of algorithms for audio fingerprinting," in *IEEE Workshop on Multimedia Signal Processing*, December 2002, pp. 169–173.

[13] E. Allamanche, B. Frba, J. Herre, T. Kastner, O.Hellmuth, and M. Cremer, "Content-based identification of audio material using MPEG-7 low level description," in *Proceeding of the International Symposium on Music Information Retrieval (ISMIR)*, Indiana, USA, October 2002.

[14] J. Haitsma and T. Kalker, "A highly robust audio fingerprinting system," in *Proceedings of the 3rd Int. Symposium on Music Information Retrieval*, Oct 2002, pp. 144–148.

[15] P. Cano, E. Batlle, H. Mayer, and H. Neuschmied, "Robust sound modeling for song detection in broadcast audio," in *Proceedings of 112th AES Convention*, Munich, Germany, 2002.

[16] E. Batlle, J. Masip, and E. Guaus, "Automatic song identification in noisy broadcast audio," in *Proceedings of the SIP*, August 2002.

[17] C. Burges, J. Platt, and S. Jana, "Distortion discriminant analysis for audio fingerprinting," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 3, pp. 165–174, May 2003.

[18] S. Sukittanon and L. Atlas, "Modulation frequency features for audio fingerprinting," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, May 2002, pp. 1773–1776.

[19] C. S. Lu, "Audio fingerprinting based on analyzing time-frequency localization of signals," in *IEEE Workshop on Multimedia Signal Processing*, Dec 2002, pp. 174–177.

[20] F. Mapelli and R. Lancini, "Audio hashing technique for automatic song identification," in *Proceedings of the International Conference on Information Technology: Research and Education*, August 2003, pp. 84–88.

[21] J. Oostveen, T. Kalker, and J. Haitsma, "Feature extraction and a database strategy for video fingerprinting," in *Proceedings of the 5th International Conference on Recent Advances in Visual Information Systems.* Springer-Verlag, 2002, pp. 117–128.

[22] A. Joly, C. Frlicot, and O. Buisson, "Robust content-based video copy identification in a large reference database," in *Lecture Notes in Computer Science*, vol. 2728, January 2003, pp. 414–424.

[23] R. Lancini, F. Mapelli, and A. Mucedero, "Automatic identification of compressed video," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, May 2004, pp. 445–448.

[24] R. Mohan, "Video sequence matching," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, May 1998, pp. 3697–3700.

[25] D. Bhat and S. Nayar, "Ordinal measures for image correspondence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 4, no. 20, pp. 415–423, April 1998.

[26] P. Indyk, G. Iyengar, and N. Shivakumar, "Finding pirate video sequences on the internet," Technical Report, Stanford University, Tech. Rep., 1999.

[27] A. Hampapur and R. Bolle, "Comparison of distance measures for video copy detection," in *IEEE International Conference on Multimedia and Expo*, vol. I, Aug 2001, pp. 737–740.

[28] B. M. Macq and J. Quisquater, "Cryptology for digital TV broadcasting," *Proceedings of the IEEE*, vol. 83, pp. 944–957, June 1995.

[29] S. Walton, "Image authentication for a slippery new age," *Dr. Dobb's Journal*, vol. 20, pp. 18–26, 'April 1995.

[30] L. Xie and G. R. Arce, "A blind wavelet based digital signature for image authentication," *Proceedings of the EUSIPCO-98 - Signal Processing IX : Theories and Applications*, vol. 1, pp. 21–24, 1998.

[31] R. B. Wolfgang and E. Delp, "A watermark for digital images," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, 1996, pp. 219–222.

[32] M. Yeung and F. Mintzer, "An invisible watermarking technique for image verification," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 2, 1997, pp. 680–683.

[33] D. Kundur and D. Hatzinakos, "Toward a telltale watermarking technique for tamper-proofing," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 2, no. 409–413, 1998.

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

[35] C. Hsu and J. L. Wu, "Hidden signatures in images," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, 1996, pp. 223–226.

[36] B. Toa and B. Dickinson, "Adaptive watermarking in the DCT domain," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, vol. 4, 1997, pp. 2985–2988.

[37] X. Xia, C. Bontcelet, and G. Arce, "Wavelet transform based watermark for digital images," *Optics Express*, vol. 3, pp. 497–508, December 1998.

[38] C. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 525–539, May 1998.

[39] P. H. W. Wong, O. C. Au, and J. C.Wong, "Image watermarking using spread spectrum technique in log-2-spatio domain," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 1, May 2000, pp. 224–227.

[40] C. S. Lu and H. Y. M. Liao, "Multipurpose watermarking for image authentication and protection," *IEEE Transactions on Image Processing*, vol. 10, pp. 1579–1592, October 2001.

[41] J. R. Hernandez, F. P. Gonzalez, J. M. Rodriguez, and G. Nieto, "Performance analysis of a 2-D-multi pulse amplitude modulation scheme for data hiding and watermarking of still images," *IEEE Journal of Selected Areas in Communication*, vol. 16, pp. 510–524, May 1998.

[42] 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*, vol. 9, pp. 55–68, January 2000.

[43] G. C. Langelaar and R. Lagendijk, "Optimal differential energy watermarking of dct encoded images and video," *IEEE Transactions Image Processing*, vol. 10, pp. 148–158, January 2001.

[44] P.W.Wong and N. Memon, "Secret and public key image watermarking schemes for image authentication and ownership verification," *IEEE Transactions Image Processing*, vol. 10, pp. 1593–1601, October 2001.

[45] Y. J. Zhang, T. Chen, and J. Li, "Embedding watermarks into both dc and ac components of dct," *Proceedings of SPIE Security and Watermarking of Multimedia Contents III*, pp. 424–435, January 2001.

[46] A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "Threshold selection for correlation-based watermark detection," in *Proceedings of COST 254 Workshop on Intelligent Communications*, Laquila, Italy, June 1998, pp. 67–72.

[47] M. Kutter, F. Jordan, and F. Bossen, "Digital signatures of colour images using amplitude modulation," in *Proceedings of SPIE - International Society of Optical Engineering*, vol. 3022, 1997, pp. 518–526.

[48] S. Pereira, J. Oruanaidh, F. Deguillaume, G. Csurka, and T. Pun, "Template based recovery of fourier based watermarks using log-polar and log-log maps," in *Proceedings of International Conference on Multimedia computing and systems*, June 1999.

[49] S. Erkucuk, "Time-frequency analysis of spread spectrum based communication and audio watermarking systems," Master's thesis, Ryerson University, 2003.

[50] L. Gomes, P. Cano, E. Gmez, M. Bonnet, and E. E. Batlle, "Audio watermarking and fingerprinting: For which applications?" *Journal of New Music Research*, vol. 32, no. 1, 2003.

[51] T. Kalker, D. Epema, P. Hartel, R. Lagendijk, and M. V. Steen, "Music2share - copyright-compliant music sharing in P2P systems," *Proceedings of the IEEE*, vol. 92, no. 6, pp. 961–970, June 2004.

[52] A. Ramalingam and S. Krishnan, "Gaussian mixture modeling using short time fourier transform features for audio fingerprinting," in *Proceedings of the International Conference on Multimedia and Expo*, Amsterdam, The Netherlands, July 2005.

[53] ——, "Video fingerprinting using space-time and Gaussian mixture models," in *Proceedings of the ninth Canadian Workshop on Information Technology*, June 2005, pp. 288–291.

[54] D. A. Reynolds and R. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72–83, January 1995.

[55] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, July 2002.

[56] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition.* Englewood Cliffs, NJ,: Prentice-Hall, 1993.

[57] D. Pye, "Content-based methods for the management of digital music," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, 2000, pp. 24–27.

[58] R. Duda and P. Hart, *Pattern Classification and Scene Analysis.* John Wiley and Sons, 1973.

[59] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1997.

[60] H. Greenspan, J. Goldberger, and A. Mayer, "Probabilistic space-time video modeling via piecewise GMM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 3, pp. 384–396, March 2004.

[61] A. Ramalingam and S. Krishnan, "Robust image watermarking using chirp detection based technique," *IEE Transactions on Vision, Image and Signal Processing*, August 2005.

[62] S. Pereira, S. Voloshynovskiy, M. Madueno, S. Marchand-Maillet, and T. Pun, "Second generation benchmarking and application oriented evaluation," in *Information Hiding Workshop III*, Pittsburgh, PA, USA, April 2001.

[63] S. Erkucuk, S. Krishnan, and M. Zeytinoglu, "Robust audio watermarking using a chirp based technique," in *IEEE International Conference on Multimedia and Expo*, vol. 2, 2002, pp. 513–616.

[64] P. Flikkema, "Spread-spectrum techniques for wireless communication," *IEEE Signal Processing Magazine*, vol. 14, pp. 26–36, May 1997.

[65] R. Peterson, R. Ziemer, and D. Borth, *Introduction to Spread Spectrum Communication.* New Jersey, NY: Prentice-Hall, 1995.

[66] L. Cohen, "Time-frequency distributions – a review," *Proceedings of the IEEE*, vol. 77, pp. 941–981, 1989.

[67] R. Rangayyan and S. Krishnan, "Feature identification in the time-frequency plane by using the hough-radon transform," *IEEE Transactions on Pattern Recognition*, vol. 34, pp. 1147–1158, 2001.

[68] G. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography.* New York, NY: Academic, 1980.

[69] P. Hough, "Methods and means for recognizing complex patterns," *U.S. Patent no. 3069654*, 1962.

[70] R. Duda and P. Hart, "Use of hough transform to detect lines and curves in pictures," *ACM Communications*, vol. 15, no. 1, pp. 11–15, June 1972.

[71] G. Eichmann, "Topologically invariant texture description," *Computer Vision Graphics Image Processing*, vol. 41, pp. 267–281, 1987.

[72] K. A. M. Kushnir and K. Matsumoto, "Recognition of hand printed hebrew characters using features selected in the hough transform space," *Pattern Recognition*, vol. 18, pp. 103–114, June 1985.

[73] M. Sanders and E. Mccormick, *Human Factors in Engineering and Design*, 7th ed. New York: McGraw-Hill, 1993.

[74] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based models of human perception," *Proceedings of the IEEE*, vol. 81, pp. 1385–1422, October 1993.

[75] A. Watson, "DCT quantization matrices visually optimized for individual images," in *Proceedings of the SPIE Conference on Human Vision, Visual Processing, and Digital Display*, vol. 1913, February 1993, pp. 202–216.

[76] H. Peterson, A. Ahumada, and A. Watson, "Improved detection model for dct coefficient quantization," in *Proceedings of the SPIE Conference on Human Vision, Visual Processing, and Digital Display*, vol. 1913, February 1993, pp. 191–201.

[77] S. Krishnan, "Adaptive signal processing techniques for analysis of knee joint vibroarthrographic signals," Ph.D. dissertation, University of Calgary, June 1999.

[78] S. Pereira, S. Voloshynovskiy, and T. Pun, "Optimal transform domain watermark embedding via linear programming," *Signal Processing, Special Issue: Information Theoretic Issues in Digital Watermarking*, May 2001.

[79] X. Xia, C. G. Boncelet, and G. R. Arce, "A multiresolution watermark for digital images," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, Santa Barbara, California, USA, October 1997, p. 548.