ROMA TRE
UNIVERSITÀ DEGLI STUDI

Scuola Dottorale di Ingegneria
Sezione di Ingegneria dell’Elettronica Biomedica,
dell’Elettromagnetismo e delle Telecomunicazioni

MULTIMEDIA COMMUNICATIONS OVER TETRA 2 NETWORKS

Michela Cancellaro

Advisor: *Prof. Alessandro Neri*
Roma Tre University
Department of Applied Electronics

A dissertation submitted to

Roma Tre University
Department of Applied Electronics

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Rome, March 8th, 2010

Abstract

Comunicazioni multimediali su reti TETRA 2

Il sistema TETRA è lo standard digitale di radiocomunicazione ad accesso collettivo per la trasmissione sicura di voce e dati ad alta qualità; viene utilizzato per servizi di emergenza, di pubblica sicurezza, per trasporti pubblici e privati o per enti che hanno la necessità di comunicare all'interno di un'area definita. Attualmente lo standard prevede la possibilità di integrare tecnologie quali ad esempio la tecnologia WAP e la tecnologia multislot che permettono agli utenti di accedere ai servizi di Internet e di Intranet attraverso l'uso dei propri terminali. Innumerevoli sono le applicazioni possibili come ad esempio la possibilità di consultare database, documenti, mappe, immagini di persone o di scene in caso di emergenza, recuperare dati medici ecc.

Tuttavia il bit rate offerto dalla tecnologia TETRA resta limitato a 28.8 Kbit/sec nel caso in cui la trasmissione dei dati utilizzi tutti i 4 slot del canale ed in assenza di protezione. Questo e la sempre crescente richiesta di servizi da parte dei fruitori del sistema, hanno portato negli ultimi anni alla definizione di una nuova release: il TETRA2. Sebbene il TETRA Enhanced Data Service (TEDS), il nuovo servizio di trasmissione dati ad altà velocità, supporterà servizi multimediali a banda larga ad una velocità massima di 500 Kbit/sec grazie all'uso di modulazioni adattative, resta necessario definire dei sistemi di trasmissione dati che raggiungano un compromesso ottimale tra la ridondanza aggiunta per proteggere i dati, l'efficienza della

compressione e la complessità computazionale degli algoritmi di codifica, i quali verrano effettuati da terminali a bassa potenza. Infatti, se da un lato i 500 Kbit/sec sono nominali in condizioni ottimali, dall'altro lato questo bit rate non risulta adeguato a garantire una qualità sufficiente per i servizi di nuova generazione.

In questa tesi vengono progettati degli algoritmi innovativi ed efficienti di codifica di immagini e di video, che garantiscono al tempo stesso l'affidabilità e la sicurezza dei dati trasmessi. In particolare sono analizzate tecniche di codifica distribuita, tecniche di codifica a descrittori multipli e tecniche congiunte di cifratura e marchiatura per rendere sicuri i dati.

Fino ad oggi nella compressione dell'informazione si sono sfruttate tecniche dove la conoscenza della statistica della sorgente veniva sfruttata esclusivamente in fase di codifica, in modo tale da inviare al decoder solo l'evoluzione tra simboli o gruppi di simboli della sorgente. Questo ha portato a sistemi di trasmissione caratterizzati da un'alta complessità necessaria per eliminare il più possibile la ridondanza del segnale da inviare. Al contrario i ricevitori sono stati progettati seguendo tecnologie relativamente semplici, ma efficienti nella ricostruzione del segnale. La codifica distribuita è un approccio innovativo e completamente diverso. E' il decodificatore, infatti, che deve sfruttare la conoscenza sulla sorgente per decomprimere il segnale, svolgendo la maggior parte dei calcoli. La codifica distribuita si basa sui risultati ottenuti da Slepian-Wolf e da Wyner-Ziv. Ambedue gli studi furono effettuati e poi pubblicati negli anni '70, ma solo negli ultimi anni sono stati ripresi ed estesi ad applicazioni pratiche. In particolare, gli algoritmi di codifica distribuita basati sul teorema di Slepian-Wolf sono di tipo lossless (senza perdita), mentre quelli basati sul teorema di Wyner-Ziv sono di tipo lossy (con perdite). Questo nuovo approccio alla comunicazione è di grande interesse in scenari come TETRA in cui la compressione dei video viene effettuata da terminali dove la memoria e la capacità di calcolo sono limitate, soprattutto per limitare le dimensioni e i costi del prodotto. In questo modo le operazioni più complesse le effettua il dispositivo base che viene condiviso tra molti utenti risparmiando sui costi.

Nel lavoro di tesi diversi metodi di codifica stereo distribuita sono stati confrontati in base alla qualità della *side information*, che rappresenta il nodo cruciale del paradigma della codifica distribuita: le prestazioni in termini di *rate distortion* del metodo di DVC proposto nella tesi sono migliori rispetto ai metodi esistenti in letteratura [1] e al metodo di codifica convenzionale H.264/AVC per i bit rate minori di 400Kbit/sec che sarannno disponibili con l'avvento del TEDS. In particolare il metodo proposto presenta i seguenti benefici:

- possibilità di utilizzare dei Group of Pictures (GOP) di lunghezza variabile mentre i metodi proposti in letteratura analizzano solitamente GOP di lunghezza 2. L'utilizzo di GOP più grandi riduce la complessità del codificatore dal momento che deve essere codificato tramite codifica convenzionale (alta complessità computazionale) 1 frame ogni 9 (se ad esempio GOP=9) piuttosto che 1 frame ogni 2 (GOP=2).

- utilizzo della combinazione della trasformata Tree Structured Haar (TSH) con la trasformata Discreta Coseno (DCT). La decomposizione TSH è una trasformata wavelet generalizzata in cui la decomposizione in sottobande può essere adattata in base al contenuto del frame. In caso di codifica con perdita, è possibile scartare le sottobande dei dettagli e ricostruire il frame solo a partire dalla basse frequenze senza un'eccessiva perdita in qualità del frame ricostruito. Ovviamente la trasmissione della sottobanda LL riduce considerevolmente la quantità di dati da trasmettere. Al contrario, la trasfomrata DCT viene usata per garantire una maggiore robustezza dell'algoritmo sul canale TETRA. Dal momento che la sola sottobanda delle basse frequenze è trasmessa, la complessità del codificatore è ulteriormente semplificata.

- La *side information* sfrutta sia la correlazione temporale che quella spaziale tra le viste.

Al fine di valutare la qualità delle sequenze stereoscopiche decodificate, sono stati eseguiti degli esperimenti soggettivi informali ispirati allo standard [2] e i risultati sono stati confrontati con le prestazioni ottenute valutando il Peak Signal-to-Noise Ratio (PSNR) ed il Video Quality Metric (VQM).

Un diverso approccio è stato proposto per uno scenario che solitamente vede l'utilizzo sul campo di diversi terminali che trasmettono la medesima scena o le stesse immagini ad un centro di controllo; l'uso di tecniche di codifica video a descrittori multipli risulta particolarmente appropriato per limitare gli effetti dovuti alla perdita di pacchetti nel canale TETRA. Solitamente indicata MDC, la tecnica a descrittori multipli si basa sull'idea di codificare i dati relativi ad una sorgente in due (o più) flussi, detti descrittori, in modo che ciascuno fornisca da solo una qualità non ottima ma accettabile del segnale ricostruito, e che, all'aumentare del numero di descrizioni ricevute, vi sia un incremento della qualità fino al raggiungimento della qualità del segnale originale. L'obiettivo è rendere il sistema di trasmissione più robusto rispetto alle perdite di dati, tipiche della rete TETRA. Nel caso più comune i descrittori sono due, hanno stessa importanza e stesso bit rate, vengono impacchettati individualmente e inviati sullo stesso canale o su canali fisicamente separati; a meno che non siano persi simultaneamente, può quindi essere sempre ottenuta da uno di essi una rappresentazione con una qualità di base accettabile.

Il metodo proposto, basato su un sistema M-JPEG2000, crea due descrizioni bilanciate di un video codificando singolarmente ogni frame senza l'utilizzo di tecniche di stima e compensazione del movimento. L'obiettivo è ottenere la più alta qualità possibile garantendo contemporaneamente il minimo overhead rispetto ad un sistema di codifica a singola descrizione (SDC) M-JPEG2000. A questo scopo:

- Una tecnica di marchiatura nel dominio wavelet è usata per ridurre la ridondanza del sistema; in particolare, la sottobanda delle basse frequenza del secondo livello di decomposizione è nascosta nelle restanti sottobande dei dettagli e non viene dunque trasmessa. Dal momento che la sottobanda delle approssimazioni concentra l'energia del segnale ed è costituita dai valori assoluti più grandi della trasformata, l'uso della marchiatura porta ad una considerevole riduzione della

quantità di dati da inviare.

- Il rate-distortion rate viene adattato al contenuto del frame grazie all'utilizzo della trasformata intera reversibile TSH che permette di variare la grandezza delle sottobande.

Infine, il terzo aspetto affrontato nella tesi riguarda la trasmissione sicura dei dati multimediali quali audio, immagini o video. Premesso che nel sistema TETRA tutte le comunicazioni sono cifrate e che risula cruciale la definizione degli algoritmi di scambio delle chiavi di cifratura che garantiscono un processo sicuro di autenticazione e autorizzazione dei terminali e dunque degli utenti, si è pensato di incrementare il livello di sicurezza dei dati utilizzando tecniche di marchiatura. Diversi sono i metodi proposti in letteratura che uniscono marchiatura e cifratura in modo non commutativo. Il problema tuttavia è che i dati cifrati hanno bisogno di essere protetti anche dopo essere stati decifrati: infatti, quando un dato cifrato viene decifrato da un utente autorizzato, l'informazione non è più protetta e può essere facilmente modificata o rubata. Dunque, invece di marchiare il documento e poi cifrarlo, la ricerca scientifica sta analizzando la possibilità di effettuare le due operazioni in modo simultaneo ma disgiunto in modo da poter ad esempio rilevare la presenza del marchio e quindi autenticare l'immagine sia che sia cifrata sia che sia stata decifrata. In questo lavoro, le tecniche di marchiatura sono state combinate con le tecniche di cifratura in modo commutativo. Inoltre sono stati utilizzati due domini di trasformazione parametrici, quali il dominio TSH e quello di Fibonacci Haar (FHT), per fornire un ulteriore livello di sicurezza rispettivamente per la protezioni di immagini in scala di girigi e a colori.

Parole chiave: TETRA2, codifica video distribuita, codifica video a descrittori multipli, cifratura, marchiatura, sicurezza, trasformata Tree-Structured Haar, sequenze video stereoscopiche.

Abstract

Multimedia communications over TETRA 2 networks

Video coding is required wherever digital video communication, storage, processing, acquisition, and reproduction occur [3]; consequently, video compression becomes an absolute requirement for the growth and success of the low bandwidth transmission and storage of digital video signals.

Several international standards have recently been adopted for video compression, each serving a different type of application; among them, M-JPEG, M-JPEG2000, H.261, MPEG2, H.264 provide very good quality of the data up to several megabits per seconds; however, below 64 kbit/sec these algorithms lead to annoying artifacts or require additional processing, resulting in low quality and long point-to-point delay. Therefore, novel coding schemes have to be designed to obtain an acceptable data quality at low bit rates.

The main issue in video compression over a limited bandwidth is the ability to find a good balance between the amount of introduced redundancy, i.e. the error resilience, and the coding efficiency while preserving an acceptable visual quality of the decoded video.

All the above considerations become crucial when dealing with data transmission over TErrestrial Trunked RAdio (TETRA) networks: this the only ETSI standard for digital trunked land mobile radio designed for the Professional Mobile Radio (PMR) market comprising Public Safety, Government, Military, Transportation, Utilities, Industrial and Military user organizations as well as Public Access Mobile Radio

(PAMR) operators.

TETRA technology provides a *trunked* and *direct* mobile-to-mobile radio capability with a range of facilities including voice and data, where data transfer includes Packet Switched Data Services and Circuit Switched Data Services single/multislot modes at a maximum data rate of 28.8 Kbit/s when all the four slots of a channel are used without protection. Moreover the reliability of the communication between emergency teams is guaranteed through error correction codes, protocols with error detection and re-transmission capability while the data is protected through several encryption algorithm.

In the last decade, a new release of the technology is being approaching to achieve an improved data rate efficiency and to respond to the increasing user demand for new services, enhanced mobility, improved ad-hoc functionality and international interoperability; among the new features that are included in TETRA release 2, TETRA Enhanced Data Service (TEDS) is a new TETRA High Speed Data (HSD) service using different RF channel bandwidths and data rates for flexible use of PMR frequency bands. It will support wideband multimedia services whilst ensuring reliable link performance over heavily time-frequency selective fading mobile channels, to beyond 500 kbit/s [4].

The first part of this work focuses on how video signals should be efficiently coded and transmitted over the error-prone, low bit rate TETRA2 channels to achieve the best possible quality. Although the advent of TEDS will improve the data rate up to about 500 Kbit/sec, it is very challenging to find the optimal trade off among the redundancy added to protect the data, the efficiency of the compression algorithm and the computational complexity of the encoding procedure carried out by low power TETRA2 devices, while guaranteeing a high level of data security. The aim of this work is to provide a performance analysis of video transmission over TETRA2 channel, comparing H.264/AVC coding efficiency with new advanced video coding strategies, named Distributed Video coding (DVC) and Multiple Description Coding (MDC) that can be adopted in TETRA2 system depending on the scenario and the application. In particular, Distributed Video Coding [5][6] (DVC) theory states that, for a given distortion, it is theoretically possible to separately encode and jointly decode two or more statistically dependent sources at the same rate obtained when the same sources are joint encoded and decoded. In other terms, while standard video coders exploit the statistical dependencies of the source signal in order to remove spatial and temporal redundancies, in DVC each video frame is encoded independently knowing that some side information will be available at the decoder in order to remove transmission errors and improve the video quality. This approach considerably reduces the complexity of the video encoder by shifting all the complex interframe processing tasks to the decoder. This property can be very interesting for power/processing limited systems such as TETRA2 devices that have to compress and send video to a fixed base station in a power-efficient way.

The proposed DVC method for stereo sequences has been evaluated in presence of channel errors and has resulted to give better rate distortion performance when compared to conventional video coding standard H.264/AVC and to state of the art method [1] at bit rates lower than 400 Kbit/sec.

On the other side, MDC based schemes has been studied to limit the effect of packet errors: TETRA copes with this problem by applying specific codes as Shortened Reed Muller codes, cyclic codes and Rate Compatible Punctured Convolutional codes depending on the scenario/application. MDC approach can be advantageous when different devices transmit the same source message over particularly noisy and unreliable channels: the message is split and encoded into two (or more) complementary descriptions, which are independently transmitted to the receiver by using separate channels or paths; the more descriptors are received, the higher the reconstruction quality. The advantage of this technique, compared with progressive coding, is that the receiver can get a useful image more quickly in case of packet losses.

The proposes MDC method, inspired to M-JPEG2000 coder scheme, achieves the highest possible quality, while preserving the minimum overhead by exploiting the content adaptivity property of the TSH transform and a wavelet based data hiding

technique. The evaluation of the perceptual impact versus the overall bit rate has demonstrated that the use of a data hiding technique to insert a variable-size average subband in the TSH transform domain allows to sensibly reduce the bit rate needed to transfer the multimedia signal with a low perceptual distortion.

In the last part of this thesis, novel approaches to secure digital images to be transmitted over TETRA2 networks are studied. According to ITU-T, Rec. X.800 [7] and IETF RFC 2828 [8], the security of data is pursued by assuring, among others, the following services: authentication, to verify the identity claimed by or for any system entity; data confidentiality, to protect data against unauthorized disclosure; data integrity, to verify that data have not been changed, destroyed, or lost in an authorized or accidental manner. To partially satisfy these constraints all the data transmitted over TETRA channel are encrypted. However, encrypted data need an additional level of protection in order to keep control on them after the decryption phase. The systems proposed in this thesis address the security requirements in digital image transmission by combining watermarking and ciphering schemes with a commutative approach to avoid unauthorized use of the confidential information, thus providing different levels of security [9]. In this way, it is possible to cipher a watermarked image without interfering with the embedded signal or to watermark an encrypted image still allowing a perfect deciphering. Both operations are performed on key dependent transform domains, the TSH and the Fibonacci Haar domain, as the key dependence increases itself the security of the overall system.

Key words: TETRA2, Distributed Video Coding, Multiple Description Coding, Encryption, Watermarking, Security, Tree-Structured Haar transform, Stereo video sequences.

Acknoledgements

The work presented in this thesis has been carried out at the University of ”Roma TRE”, Rome, Italy.
First and foremost I wish to express my gratitude to my supervisor, Professor Alessandro Neri, for his support and guidance throughout these years. I am deeply thankful to Dr. Marco Carli for the numerous fruitful discussions that have been an important contribution to this work.
I am grateful to Professor Karen Egiazarian from Tampere University of Technology, Finland for his valuable contribution in this thesis.
A special thanks goes to Professor Patrizio Campisi for his support and advices.
The support provided by the ”Istituto Superiore delle Comunicazioni e delle Tecnologie dell'Informazione” ISCOM is gratefully acknowledged.
I want to thank my family, my friends and all my colleagues at the University of ”Roma TRE” for the nice atmosphere I was surrounded during these years.

Rome, March 2010
Michela Cancellaro

Ai miei genitori

Contents

List of Figures

List of Tables

List of Abbreviations

AES	Advanced Encryption Standard
AGA	Air-Ground-Air
AIE	Air Interface Encryption
AVC	Advanced Video Coding
AMR	Adaptive Multiple Rate
AVO	Audio Visual Object
BCH	Bose Chaudhuri Hochquenghem
BPP	Bit Per Pixel
CRC	Cyclic Redundancy Check
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DMO	Direct Mode Operation
DPSK	Differential Phase Shift Keying
DQPSK	Differential Quadrature Phase Shift Keying
DPV	Discontinuity Point Vector

DSC	Distributed Source Coding
DVC	Distributed Video Coding
DWT	Discrete Wavelet Transform
ECC	Error Correcting Codes
E2EE	End To End Encryption
ETSI	European Telecommunications Standards Institute
FEC	Forward Error Correction
FHT	Fibonacci-Haar Transform
GOP	Group Of Pictures
GPRS	General Packet Radio Service
GSM	Groupe Spécial Mobile - Global System for Mobile Communications
HSD	High Speed Data
HVS	Human Vision System
IETF	Internet Engineering Task Force
IP	Internet Protocol
ISDN	Integrated Services Digital Network
ISO	International Organization for Standardization
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
JPEG	Joint Photographic Experts Group

JSCC	Joint Source - Channel Coding
JVT	Joint Video Team
LSB	Least Significant Bit
MAE	Mean Absolute Error
MAP	Maximum A Posteriori
MCP	Motion Compensated Prediction
MDC	Multiple Description Coding
MELP	Mixed Excitation Liner Predictive
M-JPEG	Motion - Joint Photographic Experts Group
MMI	Man-Machine Interface
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MS	Mobile Station
MSB	Most Significant Bit
MSC	Mobile switching center
MSE	Mean Square Error
NATO	North Atlantic Treaty Organization
PABX	Private Automatic Branch eXchange
PAMR	Public Access Mobile Radio
PC-RSC	Parallel Concatenation of Recursive Convolutional Codes

PDF	Probability Density Function
PDO	Packet Data Optimized
PEI	Peripheral Equipment Interface
PMR	Private Mobile Radio
PSK	Phase Shift Keying
PSNR	Peak Signal-to-Noise Ratio
PSTN	Public Switched Telephone Network
QAM	Quadrature Amplitude Modulation
QIM	Quantization Index Modulation
QoS	Quality Of Service
RD	Rate Distortion
RF	Radio Frequency
RGB	Red Green Blue
RS	Reed Solomon
SDC	Single Description Coding
SVD	Singular Value Decomposition
SwMI	Switching and Management Infrastructure
TCP	Transmission Control Protocol
TEDS	TETRA Enhanced Data Service
TETRA	TErrestrial Trunked RAdio

TSH	Tree-Structured Haar Transform
TMO	Trunked Mode Operation
UMTS	Universal Mobile Telecommunication System
VLC	Variable Length Coding
VQM	Video Quality Metric
WAP	Wireless Application Protocol
WPSNR	Weighted Peak Signal-to-Noise Ratio
YUV	Full-color video signal format: Y (luminance), U and V (chrominance)
YCbCr	Non absolute color space: Y (luminance or luma), Cb (blue) and Cr (red) chroma components

Chapter 1

Introduction

1.1 Motivation

The possibility to access data in real time and multiple format (audio, video, images, text, etc,...) represents a key factor in most communication systems. However, if the digital technology generates an enormous amount of data, factors like the limited storage capacity of the physical devices, the variable data transfer rates of the electronic devices and the bandwidth constraints of some telecommunication standard, limit the delivery and the fruition of the data itself. Based on these considerations, the scientific community is focusing the attention on the development of efficient compression algorithms which can produce acceptable quality of the data at very high compression ratios: a powerful compression algorithm at low bit rate is recognized as an 'enabling technology' for the implementation of many advanced digital applications.

In particular, video coding at low bit rate is becoming increasingly important due to the emergence of new applications such as video-conferencing, video-CD playback, video on demand, remote banking, telehealth, security monitoring, interactive games and other value-added services. An efficient compression scheme alone, however, may not be the complete solution in some cases like image/video database browsing and multipoint video distribution over heterogeneous networks. There is also a growing need for features as video *scalability*, which allows selective transmission of different substreams (in terms of data packets) of compressed video to different parties, depending on their respective needs. In this manner, each party can have the best possible quality session, independently of other party's constraints.

Several international standards have recently been adopted for video compression, each serving a different type of application; among them, M-JPEG, M-JPEG2000, H.261, MPEG2, H.264 provide very good quality of the data up to several megabits per seconds; however, below 64 kbit/sec these algorithms lead to annoying artifacts or require additional processing, resulting in low quality and long point-to-point delay. Therefore, novel coding schemes have to be designed to obtain an acceptable data quality at low bit rates.

The main issue in data compression is the ability to reduce the amount of data required to represent a given quantity of information, that is the removal of redundant data. Given that *data* are the means by which *information* is conveyed, video compression schemes aim to exploit both the spatial and temporal redundancies of the video sequence to find the best coding efficiency.

On the other hand, the removal of the redundancy, typical of the compression step in classical communication frameworks, results in error-prone data and, since communication channels are normally noisy, the quality of the received video can be severely degraded. For this reason, channel coding is typically applied to protect the data and allow the decoder to correct the occurred errors. The cost of the use of channel coding is the additional redundancy introduced into the bitstream. Therefore, an important aspect when dealing with video coding algorithms over a limited bandwidth is the need to find a good balance between the amount of introduced redundancy, i.e. the error resilience, and the coding efficiency while preserving an acceptable visual quality of the decoded video.

All the above considerations become crucial when dealing with data transmission over TErrestrial Trunked RAdio (TETRA) networks: this the only ETSI standard for digital trunked land mobile radio designed for the Professional Mobile Radio (PMR) market comprising Public Safety, Government, Military, Transportation, Utilities, Industrial and Military user organizations as well as Public Access Mobile Radio (PAMR) operators.

TETRA technology provides a *trunked* and *direct* mobile-to-mobile radio capability with a range of facilities including voice and data, where data transfer is defined for User Status messages and Short Data Service, including Packet Switched Data Services (based on IP packet access, TCP/IP or X.25 protocol, depending on application) and Circuit Switched Data Services single/multislot modes at a maximum data rate of 28.8 Kbit/s when all the four slots of a channel are used without protection. Moreover the reliability of the communication between emergency teams is guaranteed through error correction codes, protocols with error detection and re-transmission capability while the data is protected through *Air Interface Encryption* between the terminal and base station, *End-to-End Encryption* between the terminals end-to-end to preserve the data confidentiality.

The limited bandwidth offered by TETRA makes it very challenging to find the optimal trade off among the redundancy added to protect the data, the efficiency of the compression algorithm and the computational complexity of the encoding procedure carried out by low power TETRA devices, while guaranteeing a high level of data security. Hence, a new release of the technology is being approaching to achieve an improved data rate efficiency and to respond to the increasing user demand for new services, enhanced mobility, improved ad-hoc functionality and international interoperability; among the new features that are included in TETRA release 2, TETRA Enhanced Data Service (TEDS) is a new TETRA High Speed Data (HSD) service using different RF channel bandwidths and data rates for flexible use of PMR frequency bands. It will support wideband multimedia services whilst ensuring reliable link performance over heavily time-frequency selective fading mobile channels, to beyond 500 kbit/s [4].

The first part of this work focuses on how video signals should be efficiently coded and transmitted over the error-prone, low bit rate TETRA2 channels to achieve the best possible quality.

With the advent of TEDS, an image or a video sequence will be captured and transmitted to and from portable terminals, cameras in surveillance and environmental tracking systems, laptop computers connected through the SwMi or other devices connected through other wireless technologies; typically the devices rely on a battery with a limited power supply. Thus, the efficient use of bandwidth and energy becomes highly important in the deployment of TETRA2 video applications.

The aim of this work is to provide a performance analysis of video transmission over TETRA2 channel, comparing H.264/AVC coding efficiency with new advanced video coding strategies, named Distributed Video coding (DVC) and Multiple Description Coding (MDC) that can be adopted in TETRA2 system depending on the scenario and the application. In particular, Distributed Video Coding [5][6] (DVC) theory states that, for a given distortion, it is theoretically possible to separately encode and jointly decode two or more statistically dependent sources at the same rate obtained when the same sources are joint encoded and decoded. In other terms, while standard video coders exploit the statistical dependencies of the source signal in order to remove spatial and temporal redundancies, in DVC each video frame is encoded independently knowing that some side information will be available at the decoder in order to remove transmission errors and improve the video quality. This approach considerably reduces the complexity of the video encoder by shifting all the complex interframe processing tasks to the decoder. This property can be very interesting for power/processing limited systems such as TETRA2 devices that have to compress and send video to a fixed base station in a power-efficient way.

On the contrary, MDC is to be adopted when different devices transmit the same source message over particularly noisy and unreliable channels: the message is split and encoded into two (or more) complementary descriptions, which are independently transmitted to the receiver by using separate channels or paths; the more descriptors are received, the higher the reconstruction quality. The advantage of this technique, compared with progressive coding, is that the receiver can get a useful image more quickly in case of packet losses.

In the last part of this thesis, novel approaches to secure digital images to be transmitted over TETRA2 networks are studied. According to ITU-T, Rec. X.800 [7] and IETF RFC 2828 [8], the security of data is pursued by assuring, among others, the following services: authentication, to verify the identity claimed by or for any system entity; data confidentiality, to protect data against unauthorized disclosure; data integrity, to verify that data have not been changed, destroyed, or lost in an authorized or accidental manner. To satisfy these constraints several methods have been proposed in literature, such as watermarking and cryptography. On one side, watermarking techniques are suitable for copyright protection: the cover is the object of communication and the goal is the protection of its ownership. On the contrary, cryptography scrambles the messages so that they cannot be understood: the aim is to make the information not intelligible to any unauthorized entity who might intercept them. In this case, the data content is kept secret and the security of the methods lays in the secret key involved in the process. Actually, encrypted data need an additional level of protection in order to keep control on them after the decryption phase. In fact, when the ciphered data is deciphered by the authorized user, it is unprotected and it can be easily modified, tampered, or stolen. Moreover, one may want or need to check the presence of a watermark without deciphering the data, thus increasing the security of the system. Hence, the scientific community started focusing on the possibility to combine encryption algorithms and data hiding schemes in order to provide different levels of security [9].

The systems proposed in this thesis address the security requirements in digital image transmission by combining watermarking and ciphering schemes with a commutative approach to avoid unauthorized use of the confidential information.

1.2 Overview of the thesis

The thesis consists of five parts: each of them describes the designed solution that can be adopted over the TETRA 2 channel to provide reliable, cost-effective, and secure delivery of digital data.

Chapter 2 describes the architecture and the technical features of the communication system TETRA. The attention is focused on TETRA release 2 and on the new advanced services and improvements in data transmission that will be offered.

In Chapter 3 standard video coding schemes are analyzed towards the definition of new advanced video coding techniques that are specifically suitable for low bit rate channels according to TETRA2 channel requirements. The Tree-Structured Haar (TSH) transform domain is presented: the techniques outlined in the thesis exploit the content-adaptive peculiarity of this generalized Haar wavelet domain.

In Chapter 4 novel DVC methods for stereo sequences are proposed for TETRA2 low bit-rate communication channels in terms of side information generation, which differentiates them from the monoview case. More specifically, different stereo prediction techniques for side information generation are described and compared in terms of prediction quality, complexity and compression efficiency. The effectiveness of the DVC approach is also evaluated by means of subjective tests.

Chapter 5 focuses on Multiple Description Coding Technique approach in the error-prone TETRA2 channel. A TSH based method is presented: the goal is to achieve the highest possible quality, while preserving the minimum overhead through the use of a wavelet data hiding technique.

Chapter 6 deals with security over TETRA2 channel and presents two commutative watermarking and ciphering schemes for gray scale and color digital images to protect the data to be transmitted. Basically, the commutative property of the proposed method allows to cipher a watermarked image without interfering with the embedded signal or to watermark an encrypted image still allowing a perfect deciphering. Both operations are performed on key dependent transform domains, the TSH and the Fibonacci Haar domain, as the key dependence increases itself the security of the overall system.

Chapter 2

Low bit-rate communication system: TETRA

This chapter describes the architecture and the technical features of the communication system TETRA. In particular, the attention is focused on TETRA release 2 and on the new advanced services and data transmission improvements that will be offered.

2.1 Overview of the system

The mobile communications requirements of the public safety and disaster recovery (PSDR) organizations are continuously evolving. Currently, the information exchanged between the control center and the dispatched on-site units and directly between on-site emergency teams such as police, ambulances, hospital and control headquarters, law enforcement, fire, search and rescue agents, government administration and so forth, are characterized by complex context (voice, data and real time video services, highly mobile end terminals, etc.), very high level quality of service (QoS), end-to-end security, fast and reliable data transmission.

Nowadays, TETRA [12] [13] specifications are worldwide adopted by most communication systems targeting security and emergency that need to realize a mission critical communications environment as the one depicted in Figure 2.1. It is the only ETSI standard for digital trunked land mobile radio designed for the Professional Mobile Radio (PMR) market comprising Public Safety, Government, Military, Transportation, Utilities, Industrial and Military user organizations as well as Public Access Mobile Radio (PAMR) operators.

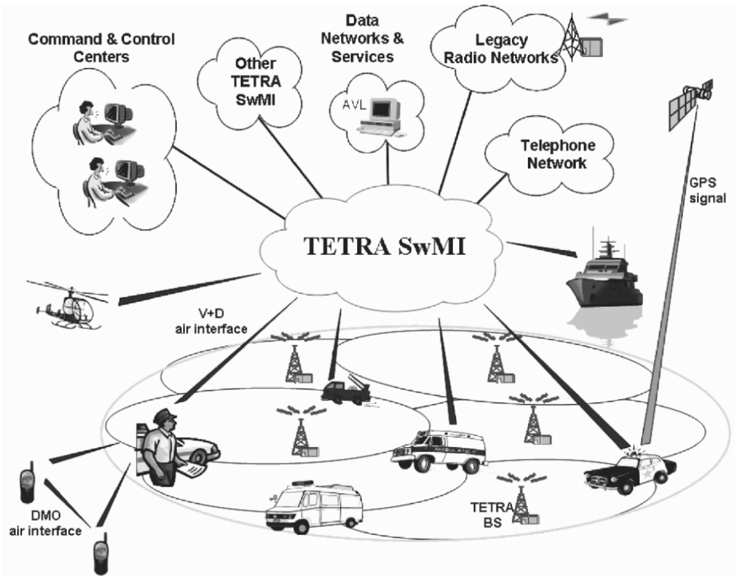


Figure 2.1: General TETRA environment.

In particular the main objectives of the projects can be summarized as:

- to define a digital Professional Mobile Radio (PMR) standard to satisfy the current and future needs of especially Public Safety and other PMR users across Europe;
- to unify the fragmented PMR markets into one common market;
- harmonized use of the frequency spectrum;
- to fulfill European authority co-operation requirements (Schengen Treaty) to enable European Integration-crossborder operation using common frequency spectrum;

- to guarantee an open standard supported by many manufacturers, which provides the customers with a wider choice.

TETRA networks are expected to satisfy specific organizational needs such as, serving an airport where the network will provide coverage to just a small area, or to facilitate regional emergency services, where the network coverage will be extended to a wide area. Therefore, TETRA networks facilitate a wide range of connections to external TETRA and non-TETRA networks that can be accessed from the mobile terminals. These networks can be public or private telephone networks, different types of data networks and command or control systems (see Figure 2.2). The connectivity to different networks combined with bandwidth-on-demand makes TETRA a superior platform for data application development.

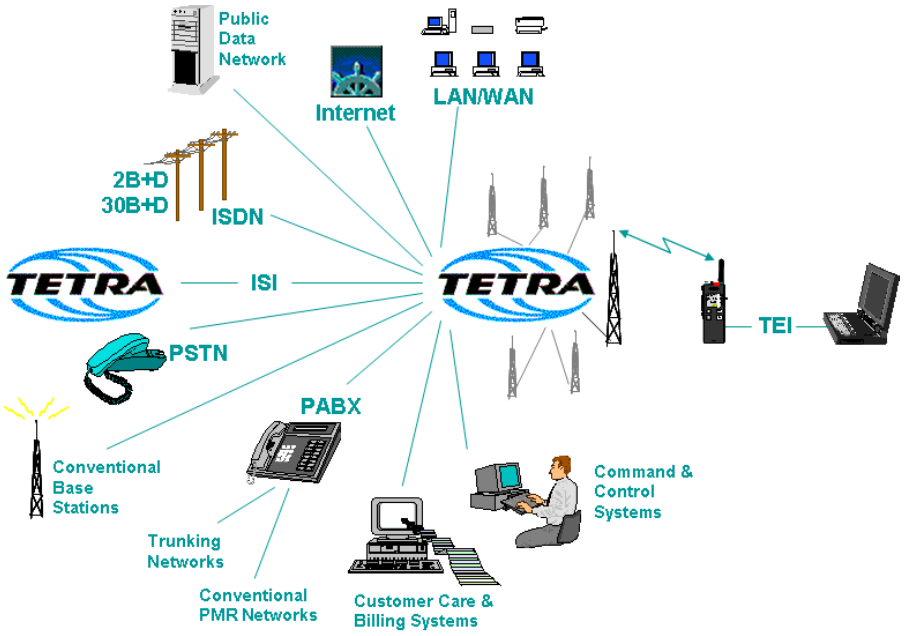


Figure 2.2: TETRA interconnection.

All major system manufacturers, user organizations, network operators, regulators, test houses and application software developers formed the TETRA MoU (Memorandum of Understanding) Association in December 1994, a joint effort to support and promote fast and consistent implementation of TETRA systems in the member countries. To ensure a wide open market, the TETRA MoU strives for the maximum interoperability of equipment from different manufacturers.

2.1.1 Services

All TETRA services are characterized by mission-critical performance figures such as fast call setup (typically, between 300-500 msec), fast message transmission, priority-based call handling, advanced encryption and authentication, etc. TETRA standard classifies the offered services in two categories depending on the point of access [14]:

- Teleservices provide complete communication capability between end users, including all terminal functions. In TETRA standards, teleservices cover all voice communications services:
  - Individual Call connects one user of the network with another user, comparable to a public telephone system.
  - Group Call connects one user with a group of users; groups do not have to be fixed but can be formed dynamically.
  - Broadcast Call is sent out of a control center to inform all users.
  - Emergency Calls are handled by the TETRA Mobile Switching Centers (MSC) with a high priority to enable a quick connection to a dispatcher or a group.
  - Direct Mode Operation (DMO): Two users connect themselves directly in simplex mode without using a TETRA network.
  - Open Channel is a service that behaves very similar to an analogue two-way radio channel in which every participant can talk or listen free.
  - Call Include allows adding additional users to a group call that is already established and running.

- Bearer services provide communication capability between terminal network interfaces, excluding the functions of the terminal. Hence, they define the data transfer for User Status messages and Short Data Service, including Packet Switched Data Services (based on IP packet access, TCP/IP or X.25 protocol, depending on application, with a maximum data rate of 28.8 kbit/s packet) and Circuit Switched Data Services single/multislot modes. In detail, circuit mode allows to transmit data at variable data rate depending on the applied protection as illustrated in Table 2.1.

Table 2.1: Circuit mode data rate.

	Circuit mode data (Kbit/sec)			
Error protection	1 time slot	2 time slots	3 time slots	4 time slots
high	2.4	4.8	7.2	9.6
normal	4.8	9.6	14.4	19.2
none	7.2	14.4	21.6	28.8

There are also supplementary services for very flexible system applications:

- Call Authorized by Dispatcher: dispatcher verifies call request before allowing call to proceed.

- Area Selection: defined areas of operation for users, redefined on a call by call basis.

- Access Priority: Radio Unit uplink access prioritization during congested periods.

- Priority Call: access to network resources can be prioritized.

- Late Entry: latecomers may join a call in progress.

- Pre-emptive Priority Call: this call has the highest uplink priority and highest priority access to network resources. If system is busy the lowest priority communication will be dropped lo allow this call to continue.

- Discrete Listening: authorized control room or user may monitor a communication without being identified.

- Ambience Listening: dispatcher may turn on the transmitter of a device without any indication being provided to the user. Can be used in critical situations to listen what is happening in the car.

- Dynamic Group Number Assignment: allows the dispatcher to program new group numbers into the devices over the air. Can also be used to group participants in an ongoing call.

ETSI has specified three modes of operation, identified in terms of services they support (see Figure 2.3):

- Trunked Mode Operation TMO or Voice plus Data (V + D): this mode is optimized for simultaneous transfer of voice and data.

- Direct Mode Operation DMO: in this mode, intercommunication of mobile stations is independent on the network, i.e. without mediation of base stations. In this way it is guaranteed the operation outside of the coverage of a TETRA Infrastructure, in poor signal strength areas or when there are capacity limitations on a network. This mode of operation supports circuit mode speech and data and short data services over one time slot.

- TETRA PDO (Packet Data Optimized): this is a specific version, optimized for packet data transmission via radio channels over the equivalent of four time slots in V + D system.

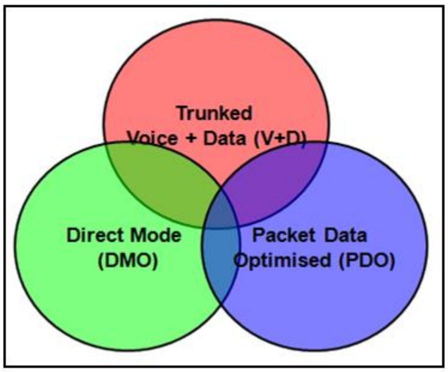


Figure 2.3: TETRA modes of operations.

2.2 Architecture

TETRA V + D specification states no constraints on the form of the radio network architecture. As shown in Figure 2.4, the infrastructure is only defined in terms of [15]:

- Trunked mode air interface: this interface operates in trunking mode, i.e. it allocates and releases the available radio resources dynamically and on demand basis so that the available radio spectrum can be efficiently shared across many different groups of users, or even across many different PSDR agencies.

- Direct mode air interface: guarantees communication between terminals also beyond network coverage in event of Base Station failure.

- Peripheral Equipment Interface (PEI): this interface defines the connection of the radio terminal to an external device, and supports data transmission between applications resident in the device and the connected TETRA radio. The PEI also allows the control of the radio from the external device and/or application.

- Man-Machine Interface (MMI): since it was recognized that a comprehensive MMI standard would inhibit innovation in important Human Factor areas, MMI standardization is limited mainly to keypad presentation on radio terminals.

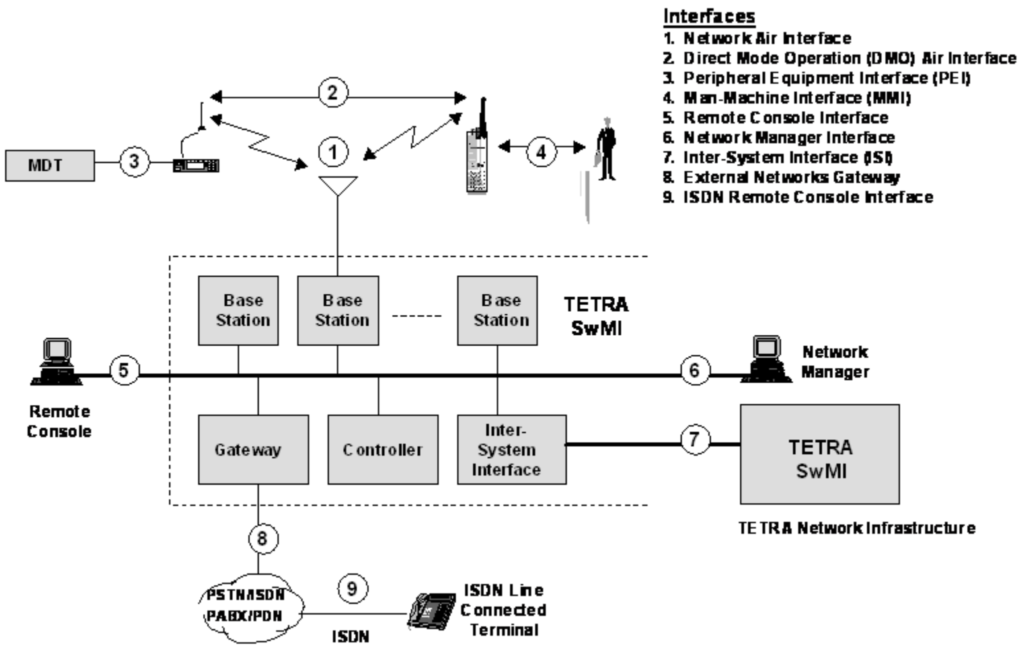

Interfaces
1. Network Air Interface
2. Direct Mode Operation (DMO) Air Interface
3. Peripheral Equipment Interface (PEI)
4. Man-Machine Interface (MMI)
5. Remote Console Interface
6. Network Manager Interface
7. Inter-System Interface (ISI)
8. External Networks Gateway
9. ISDN Remote Console Interface
MDT
1
2
3
4
Base Station
Base Station
Base Station
TETRA SwMI
5
Remote Console
6
Network Manager
Gateway
Controller
Inter-System Interface
7
TETRA SwMI
TETRA Network Infrastructure
8
PSTN/ISDN PABX/PDN
9
ISDN
ISDN Line Connected Terminal

Figure 2.4: TETRA architecture.

- Remote Dispatcher Interface: this was originally intended to allow connection to remote wire line dispatcher consoles located in most control rooms. Unfortunately, work on this interface was dropped because of the complexity to provide a universal interface without degrading performance. Hence, remote dispatcher consoles can be connected to TETRA networks, but the type of interface will be manufacturer specific.

- Network Management Interface: like the local dispatcher interface, it was recognized that a common network management interface was impractical. It was later decided that this interface should be a recommended interface rather than a standard interface.

- Inter-System Interface: this allows infrastructures supplied by different TETRA manufacturers to inter-operate with each other allowing interoperability between two or more TETRA networks covering information transfer using circuit mode. A parallel standard interface, called IPI, is available in TETRA for IP packet data connection between two or more TETRA networks.

- Gateway to PSTN/ISDN/PABX: it enables TETRA to interface with the PSTN, the ISDN directly and/or with PABXs indirectly as required by user organizations.

- Remote Line Connected Terminal: it was intended to cover the signalling protocol required to support an ISDN line connected terminal, but this standardization was not completed due to the poor interest in this facility.

Finally, the Switching and Management Infrastructure (SwMI) is the core component of the system and comprises the necessary networking, switching, management and service provision elements. It also provides digital narrowband radio services to a wide geographical area by means of a plurality of TETRA Base Stations (BS), deployed in strategic locations according to the overall radio coverage, traffic and availability requirements. Even though some ETSI Project TETRA members felt that a standard base station to switch interface would be useful (as provided in GSM), it was decided to allow TETRA infrastructure manufacturers the flexibility in design while guaranteeing the interoperability with other networks.

2.3 Technical Features

The frequency allocation for public safety market is in the range 380-400 MHz and 2 by 5 MHz is established with a duplex spacing of 10 MHz (380-385 MHz for the uplink and $390 - 395$ MHz for the downlink). For commercial users, spectrum is allowed in one or more of the following bands: 410 - 430 MHz, 450 - 470 MHz, 806 - 870 MHz, 870 - 890 / 915 - 933 MHz. The channel spacing is 25 kHz and 4 voice/data communications are provided in 2x25 kHz frequency band (current DMO version allows for one communication in 25 kHz, providing spectrum efficiency which is halved compared to the V+D one; future standardization activity will let direct mode provide same spectrum efficiency as in trunked mode). Frequency spectrum is shown in Figure 2.5.

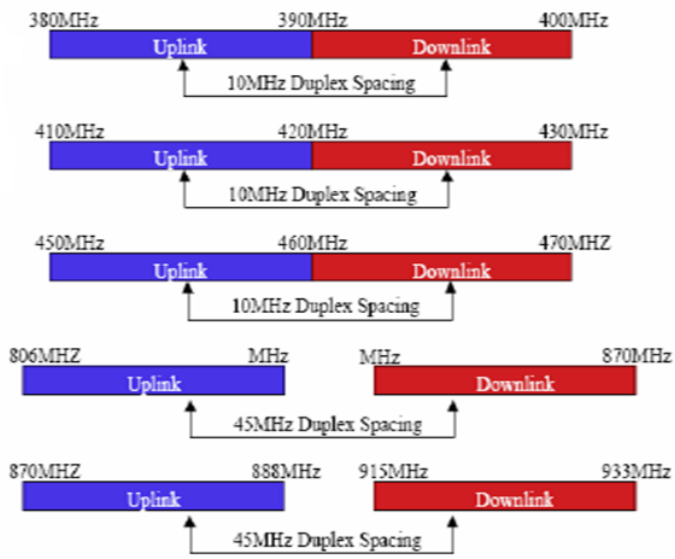


Figure 2.5: TETRA frequencies.

TETRA employees trunking technology to increase network capacity and/or RF spectrum efficiency. An automatic and dynamic assignment of a small number of communication channels shared amongst a relatively large number of users leads to more radio users per RF channel for a given Grade of Service (GoS). This mode of operation also allows the effective and economical sharing of the usage of the network between several organizations without compromising the security and privacy.

TETRA radio access protocol is based on a four slots per carrier Time Division Multiple Access TDMA arrangements (see Figure 2.6). This technology has offered the optimum solution to balance the cost of equipment with that of supporting the services and facilities required by user organizations for a medium to high capacity network, providing single site local RF coverage and/or multiple site wide area RF coverage.

One TDMA frame lasts 57.67 ms. Frames are organized in multiframes and hyperframes as shown in Figure 2.7.

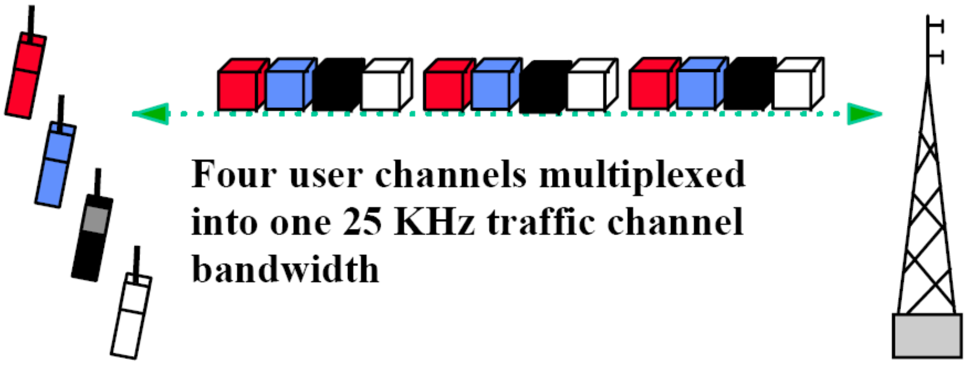

Four user channels multiplexed into one 25 KHz traffic channel bandwidth

Figure 2.6: TDMA four slots channel.

The modulation chosen for TETRA is $\pi/4$-DQPSK that has an alphabet of 4 symbols; it transmits two bits per symbol period giving a gross data rate of 36kbps. After TDMA oveheads (guard and ramp periods) are removed, this results in 4 TDMA channels, each with a gross data rate of 7.2kbps. One of four allowed phase changes $\Delta\varphi \in -\pi/4, \pi/4, -3\pi/4, 3\pi/4$ is transmitted in each symbol interval.
The advantages of choosing this modulation scheme are:

- Two bits are transmitted in each symbol, giving a bandwidth efficiency of up to 2 bits/second/Hz.
- Since the information is transmitted as phase changes, no estimate of the absolute phase of the carrier is required and some very simple demodulation schemes can be used (but only where no channel equaliser is needed).
- No transitions pass through zero amplitude, which is helpful for the design of a linearised RF power amplifier.

RF spectrum efficiency is a combination of three main factors: the occupied bandwidth per communication channel, the frequency re-use factor determined by the Carrier to Interference protection ratio C/I and the trunking technology.
Figure 2.8 shows that the TDMA technology used in TETRA provides 4 independent communications channels in a 25 kHz RF bandwidth Channel, making it twice

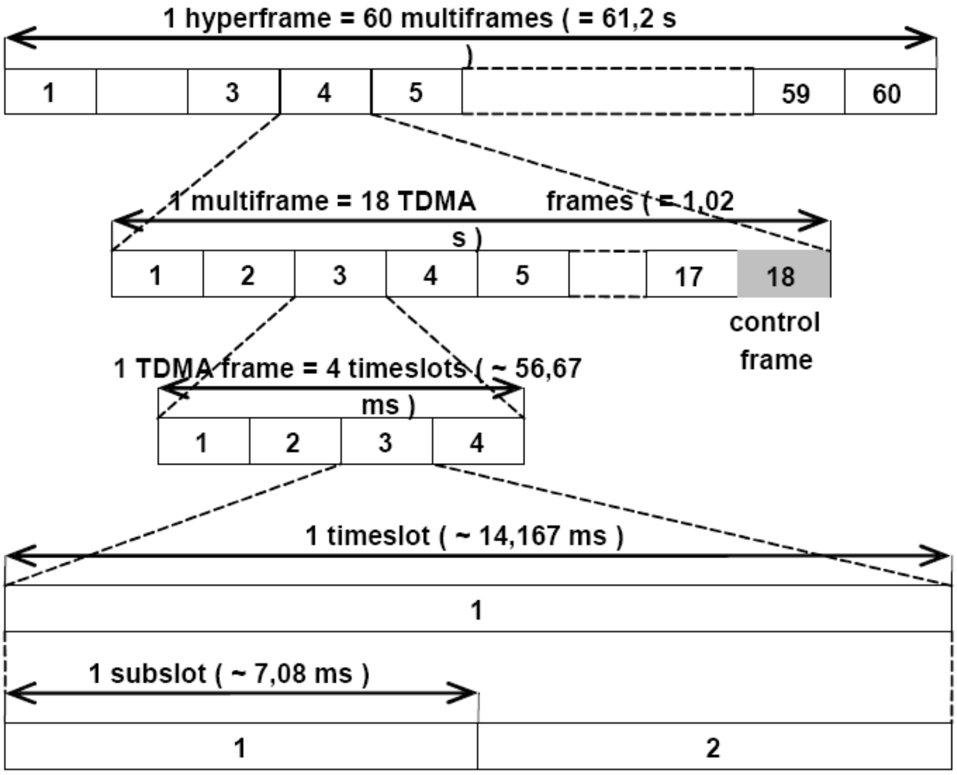

1 hyperframe = 60 multiframes (= 61,2 s)
1
3
4
5
59
60
1 multiframe = 18 TDMA frames (= 1,02 s)
1
2
3
4
5
17
18
control frame
1 TDMA frame = 4 timeslots (~ 56,67 ms)
1
2
3
4
1 timeslot (~ 14,167 ms)
1
1 subslot (~ 7,08 ms)
1
2

Figure 2.7: TDMA frame.

efficient in occupied bandwidth compared with traditional 12.5 kHz RF bandwidth FDMA channel [16]. Although FDMA technologies tend to have a better C/I performance than TDMA, the overall spectrum efficiency advantage lies with TETRA, especially for medium to high capacity networks.

2.4 TETRA Release2

TETRA technology is widely accepted (especially in Europe) and is considered as one of the most mature and prominent technologies for the Public Safety and Disaster Recovery (PSDR) market as well as for the Public Access Mobile Radio (PMR/PAMR) markets.
In 2006 figures, some 40% of countries around the world had adopted this standard.

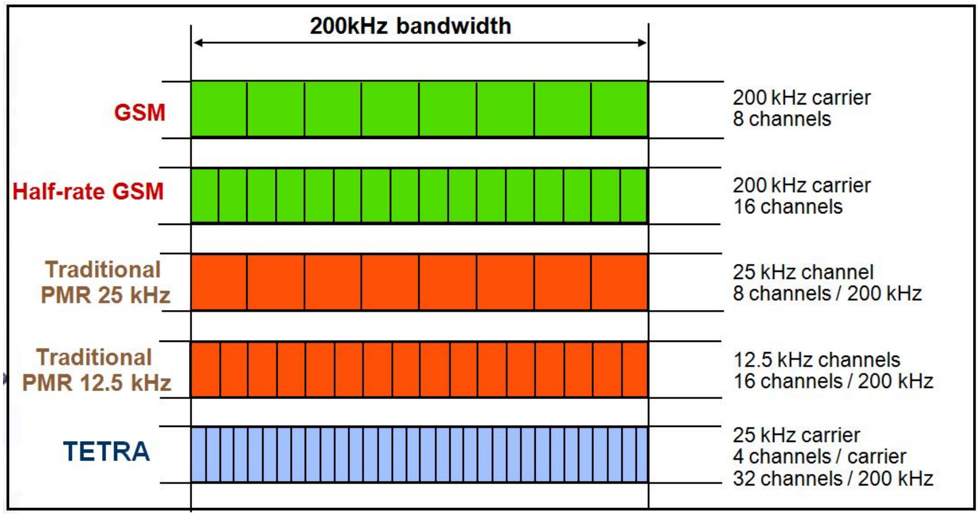


Figure 2.8: RF spectrum efficiency.

Almost half of the usage in the PMR sector is attributed to Public Safety organizations whilst a quarter has been taken up by the transportation industry. Utilities and military applications amount to 6% each. In addition, the PAMR sector's usage of TETRA, currently at (3%), is on the rise.

TETRA specifications are constantly being evolved by ETSI and new features are being introduced to satisfy increasing user demand for new services, such as remote patient monitoring, 2-way real-time video, 3-D positioning and GIS, mobile robots or enhanced telemetry, enhanced mobility, improved ad-hoc functionality and international interoperability as well as gleaning the benefits of new technology.

In 1999 interest groups, comprising both users and manufacturers within Technical Committee (TC) TETRA and the TETRA Association, identified the areas to be enhanced, resulting in the following services and facilities being standardized at the end of 2005 as part of TETRA Release 2:

- Trunked Mode Operation (TMO) Range Extension: uplink and downlink bursts, as well as guard times were modified to operate beyond the 58 km range limit (a function of TETRA's TDMA structure). The TMO range of TETRA is extended up to 83 km for Air-Ground-Air (AGA) applications. Note that DMO has no TDMA structure range limitation as synchronization takes place in DMO at the start of each transmission).

- Further enhancements to the air interface to provide increased benefits in terms of spectrum efficiency, subscriber capacity, system performance, quality of service, size and cost of terminals, etc.

- Production and/or adoption of standards to provide improved interworking and roaming between TETRA and public mobile networks, such as GSM, GPRS and UMTS.

- Evolution of the TETRA SIM, with the aim of convergence with the universal SIM (USIM), to meet the needs for TETRA-specific services while gaining the benefits of interworking and roaming with public mobile networks, such as GSM, GPRS and UMTS.

- Adaptive Multiple Rate (AMR) Voice Codec: The AMR codec, operating in the 4.75 kbits/s only mode, has been chosen for possible future applications in TETRA but the definition of the Air Interface Standard to accommodate the AMR codec is still in progress.

- Mixed Excitation Liner Predictive, enhanced (MELPe) Voice Codec: The STANAG 4591 (MELPe codec) has been standardized by NATO for its own military communication applications because of its low bit rate (2400 bit/s), immunity to high background noise and acceptable voice quality performance. Main potential benefits have been identified in interworking with government systems, improved RF Coverage using spare bits available for extra FEC and simultaneous V+D using spare bits available for data. However, the way the MELPe codec needs to be implemented in TETRA increases ”end to end” voice delay, which needs to be balanced against its possible benefits.

- TETRA Enhanced Data Service (TEDS): TEDS is a new TETRA High Speed Data (HSD) service using different RF channel bandwidths and data rates for flexible use of PMR frequency bands. TEDS is fully compatibility with TETRA Release 1 and allows for ease of migration. In order to support wideband multimedia services whilst ensuring reliable link performance over heavily time-frequency selective fading mobile channels, a number of up-to-date technological choices have been made including [4]:

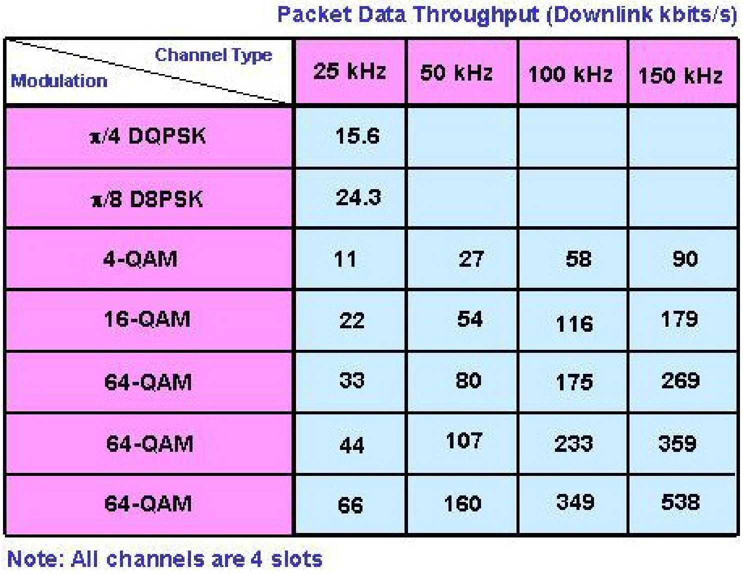
- adoption of Multi-Carrier (MC) filterbank-based signalling to achieve robust performance even in frequency- selective fading channels, for a total number of (2.7 kHz spaced) subcarriers ranging from 8 (25 kHz channel) to 48 (150 kHz channel);
- spectral-efficient multilevel modulation schemes, i.e.:
    - pi/4 DQPSK (for common TETRA V+D and TEDS control channel);
    - pi/8 D8PSK (for early migration requiring modest increase in speed);
    - 4 QAM (for efficient links at edge of coverage);
    - 16 QAM (for moderate speeds);
    - 64 QAM (for high speeds).
- powerful turbo-code for payload channel encoding with rates 1/2 and 2/3 and 1 (uncoded case);
- separate channel encoding for short "header" blocks to exceed the payload performance and enable reliable slot decoding and network operations;
- flexibility of selecting the required data throughput from a wide range extending to beyond 500 kbit/s;
- link adaptation techniques to improve the system performance (e.g., overall message throughput), based on choosing adaptively the modulation level, the coding rate and possibly the RF channel bandwidth according to the varying channel propagation conditions; in detail the channel bandwidth are 25 KHz, 50 kHz, 100 kHz and 150 kHz. Even though TEDS is capable of providing High Speed Data in 150 kHz RF channels, the current limitation

caused by insufficient RF spectrum to support the growth of TETRA will probably limit early deployments to 50 kHz RF channel assignments only.

– sectored antennas as a means of extending the TEDS high speed channel coverage to that of the TETRA 1 control channel without a need for additional base station sites.

Three classes of data have been defined, i.e., a real-time class for live audio and video transmissions, a telemetry class for applications with intermittent transmissions of small volumes of data, and a background class for file transfer and Internet browsing applications. For each data class, the TEDS protocols allow negotiation of Quality of Service (QoS) attributes, such as throughput, delay, precedence and reliability.

Figure 2.9 shows the different RF channel bandwidths and data rates supported in TEDS.



Packet Data Throughput (Downlink kbits/s)

Channel Type Modulation	25 kHz	50 kHz	100 kHz	150 kHz
π/4 DQPSK	15.6			
π/8 D8PSK	24.3			
4-QAM	11	27	58	90
16-QAM	22	54	116	179
64-QAM	33	80	175	269
64-QAM	44	107	233	359
64-QAM	66	160	349	538

Note: All channels are 4 slots

Figure 2.9: TEDS data rate.

Chapter 3

Advanced video coding

Video coding is required wherever digital video communication, storage, processing, acquisition, and reproduction occur [3]; consequently, video compression becomes an absolute requirement for the growth and success of the low bandwidth transmission and storage of digital video signals.

In this chapter standard video coding schemes are analyzed towards the definition of new advanced video coding techniques that are specifically suitable for low bit rate channels according to TETRA2 channel requirements. The Tree-Structured Haar (TSH) transform domain is finally presented: Distributed Video Coding (DVC) and Multiple Description Coding (MDC) techniques outlined in next chapters exploit the content-adaptive peculiarity of this domain.

3.1 Requirements in video coding design

Traditional video-based compression seeks to minimize information redundancy of the source either in spatial and temporal domain. In detail, *intra-frame coding* considers the spatial redundancy in a single image or frame; however, in most cases, consecutive frames differ only slightly, raising to a high temporal redundancy of visual information. *Inter-frame coding* exploits the redundancy existing among two or more consecutive frames to achieve more efficient coding. Hence, to gain high compression

ratios, intra and inter frame coding have to be combined.
There are two main families of video compression standard: the H.26X family and the Moving Picture Expert Groups (MPEG) one. MPEG standards have been developed by the ISO and are primarily aimed at motion picture storage and communications. H.26X compression schemes have been proposed by the ITU to address videoconferencing applications. The evolution of compression standards generated by MPEG and H.26X are very closely related and many of the techniques adopted by MPEG's latest compression standard borrow from recent developments in H.26X's latest release, and vice versa.
These standards are application-oriented and address a wide range of issues as bitrate, complexity, picture quality, and error resilience. Specifically, the design of a video compression system that achieves the optimal trade off between bit rate and quality, should consider the following items:

- Video characteristics: parameters such as the dynamic range, source statistics, pixel resolution, and noise content can affect the performance of the compression system.

- Transmission/bit rate requirements: a lossless compression may be needed for high storage capacity or if the data is to be transmitted in a high rate channel; conversely, extremely low bit rate requirements may dictate compression systems that trade-off image quality for a large compression ratio. In case the transmission bandwidth exceeds the compressed video bandwidth, then progressive coding allows for transmission and reconstruction of each resolution independently from low to high resolution. Moreover it is important to be aware of channel errors that normally affect system performance and the quality of the reconstructed video.

- Compression system - characteristics and performance: the characteristics of the application, like multimedia telephony, multimedia messaging service (MMS), video on demand or streaming, will dictate the suitability of the video compression algorithm for particular system implementations. For example, videoconferencing demands that each participant in the interactive video session must have the same video encoding and decoding capabilities, and that the system performance requirements must be met by both the encoder and decoder; on the other extreme, television broadcast video has significantly greater performance requirements at the transmitter because it has the responsibility of providing real-time high-quality compressed video that meets the transmission channel capacity.

- Rate-distortion requirements: the video encoder must be able to provide the bit rate(s) and video fidelity (or range of video fidelity) required by the application. Otherwise, any aspect of the system may not meet specifications.
- Standards requirements: video encoder compatibility with existing and future standards is an important consideration if the digital video system is required to inter-operate with existing and/or future systems.

3.2 Video coding standards

The earliest efforts at video compression were based on methods developed for image compression: in particular, Motion JPEG (MJPEG) applies the Joint Photographic Experts Group (JPEG) image compression standard to video exploiting the spatial redundancy of each frame. Due to its poor compression ratio performances, it is largely used in applications such as video editing, enhanced VCR functionality, and high-quality video applications in the motion picture industry.

A video compression standard that exploits both spatial and temporal redundancies was firstly proposed in MPEG-1 framework in 1991, followed by MPEG-2 in 1994. These standards use a motion compensation technique to extract temporal redundancy and increase the compression ratio for the transmission of digital video information for multimedia and television formats (1.5 Mbps - 100 Mbps).

A dramatic change in approach, emphasizing content-based hierarchical audio-visual object (AVO) representation and composition, was used in the development of MPEG-4. This standard targets bit rates between 5 and 64 kbit/s for mobile and public switched telephone network (PSNT) applications and up to 4 Mbit/s for TV and film applications; it supports audio and video as well as synthetic and animated images, text, graphics, texture, and speech synthesis, provides improved video compression efficiency, content-based interactivity such as audio-video object-based access, universal access, including increased robustness in error prone scenarios/the ability to add and drop audio-video objects/object resolution scalability. Despite the novel approach and initial excitement surrounding the release of the MPEG-4 standard, MPEG-4 profiles that focus on the use of AVOs for content-based video compression have had very little impact on the commercial world due to two reasons. Firstly, detection of AVOs requires the use of efficient segmentation and tracking techniques and secondly, the subsequent release of the H.264 video compression standard has been demonstrated to provide superior video compression compared with MPEG-4 without the use of AVOs.

The greatest impact of MPEG-4 on practical video communication systems has focused on a couple of specific profiles: MPEG-4 Part 2-Simple Profile/ Advanced Simple Profile (SP/ASP) has been adopted by various commercial video codecs and MPEG-4 Part 10-Advanced Video Coding (AVC). It is worthy noticing that the latter has been developed by simply including the H.264 video compression standard as a new part in MPEG-4.

Finally, the latest MPEG developed standards are MPEG-7 and MPEG-21: the first [17], formally named ”Multimedia Content Description Interface”, defines audiovisual content storage and retrieval services while the second [18] looks to define the technology needed to support users to exchange, access, consume, trade, and otherwise manipulate Digital Items (defined as the fundamental units of distribution and transaction, that is, content such as web page, picture, movie, etc.) in an efficient, transparent, and interoperable way.

The International Telecommunications Union (ITU) has similarly defined four encoding standards, as H.261, H.262, H.263 and H.264. H.261 is intended for transmission at affordable telecom bit rates and to support full motion video transmission over Integrated Services Digital Network (ISDN). In its definition, H.261 describes the video compression methods that were later adopted by the MPEG standards. H.263 encoder is an extended version of H.261 and was developed for supporting low bandwidth applications on telephony and data networks. H.262 is functionally equivalent to the MPEG-2 encoder.

H.264/AVC, developed by a Joint Video Team (JVT) that was formed as a partnership project between the ISO, IEC, and ITU-T, is the latest developed video compression standard and incorporates a number of transform, quantization, and motion compensation improvements that achieve significantly better rate distortion performance than its predecessors at the cost of greater computational complexity. Then it is universally recognized as a major advance of video technology and has became widely adopted in essentially all digital video applications.

3.3 Hybrid Block-based Motion Compensation video coding

All the above-mentioned video compression standards are based on the Hybrid Block-based Motion Compensation (HBMC) approach and share the same block diagram, as shown in Fig. 3.1.

The system is designed for two modes of operation: the Intraframe mode and the Interframe mode, that operate on image blocks; alternately, a given frame can be simply replicated from the previously decoded frame. The latter is denoted as Skip modes. Temporal encoders normally process macroblocks of size 16x16. The intraframe mode spatially encodes an entire current frame I_k on a periodic basis, such as every 15 frames, to ensure that systematic errors do not continuously propagate. Intraframe mode will also be used to spatially encode a block whenever the interframe

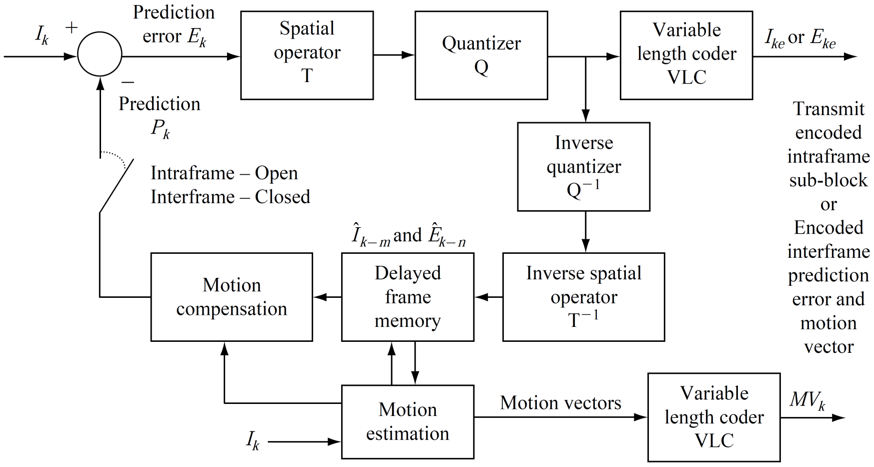

I_k
+
−
Prediction error E_k
Spatial operator T
Quantizer Q
Variable length coder VLC
I_{ke} or E_{ke}
Transmit encoded intraframe sub-block or Encoded interframe prediction error and motion vector
Prediction P_k
Intraframe – Open
Interframe – Closed
Inverse quantizer Q^{-1}
$\hat{I}_{k-m}$ and $\hat{E}_{k-n}$
Motion compensation
Delayed frame memory
Inverse spatial operator T^{-1}
I_k
Motion estimation
Motion vectors
Variable length coder VLC
MV_k

Figure 3.1: Hybrid Block-based Motion Compensation scheme.

encoding mode cannot meet its performance threshold.

The current frame I_k is spatially encoded, quantized and subsequently encoded by the variable length coder (VLC) generating I_{ke}, which is transmitted to the decoder. The receiver decodes I_{ke} producing the reconstructed image $\hat{I}_k$.
More in detail the functional elements of this process are:

- Spatial operator: this element decorrelates the signal data so that the associated energy in the transform domain is distributed into a small number of coefficients, that result to be easier to encode.
  The Discrete Cosine Transform (DCT) is the most widely used spatial operator for image and video coding that provides excellent decorrelation of signal components. DCT is also preferred for a couple of important reasons. The DCT has fast O(n logn) implementations using real calculations. It is very simple to compute because it does not require the use of complex numbers. Moreover the reconstructed input of the Inverse DCT (IDCT) avoids generation of spurious components between the edges of adjacent image blocks, i.e. DCT is not susceptible to discontinuities at the periodic edges of the signal or pixel block

caused by the periodicity in the reconstructed signal.

- Quantizer: this element quantizes the transform coefficients according to the bit rate and distortion specifications. It introduces loss of information and is the primary source of the compression gain. The quantization method generally adopted is scalar and nonuniform. The scalar quantizer simplifies the complexity of the operation compared with vector quantization. The nonuniform quantization interval is tuned according to the distribution of the transform coefficients to minimize the bit rate and the distortion created by the quantization process or, alternatively, it can be adjusted based on the performance of the Human Visual System (HVS).
- VLC: the lossless VLC is used to exploit the "symbolic" redundancies contained in each block of quantized transform coefficients. Entropy encoders are used for information sources that are memoryless (sources in which each value is independently generated), and try to minimize the bit rate by assigning variable length codes for the input values according to the input probability density function (pdf). Predictive coders are suited to information sources that have memory, that is, a source in which each value has a statistical dependency on some number of previous and/or adjacent values. Predictive coders can produce a new source pdf with significantly less statistical variation and entropy than the original. The transformed source can then be fed to a VLC to reduce the bit rate. Entropy and predictive coding are good examples for presenting the basic concepts of statistical coding theory.
  The scanning pattern used to generate the bitstream to be coded should also be chosen with the objective of maximizing the performance of the VLC. The MPEG encoder for instance describes a zigzag scanning pattern that is intended to maximize transform zero coefficient run lengths. Alternatively, the H.261 VLC is designed to encode these run lengths using a variable length Huffman code.

The core of the encoder is the motion-compensated prediction (MCP). The first step is the motion estimation, aiming to find the region in the previously reconstructed frame that best matches each macroblock in the current frame. The offset between the MB and the prediction region is known as the motion vector, that is VLC encoded as MV_k.
The second step in MCP is motion compensation, where the current frame prediction P_k is obtained by applying the motion field to the previously reconstructed frame. The prediction error E_k, known as the displaced frame difference (DFD), is obtained by subtracting the current frame prediction P_k from the current frame I_k. The prediction error is then spatially and VLC encoded to form E_{ke} and is transmitted along with the VLC encoded motion vectors MV_k .
Note that the intraframe encoding mode does not receive any input from the feedback loop.
The decoder can reconstruct the current frame $\hat{I}_k$ using the previously reconstructed frame $\hat{I}_{k-1}$ (stored in the decoder), the current frame motion vectors, and the prediction error. The motion vectors MV_k operate on $\hat{I}_{k-1}$ to generate the current prediction frame P_k . The encoded prediction error E_{ke} is decoded to produce the reconstructed prediction error $\hat{E}_k$. The prediction error is added to the prediction to form the current frame $\hat{I}_k$.

Hence, the feedback loop requires the Inverse Quantizer, the Inverse Spatial Operator, the Delayed Frame Memory, the Motion Estimator and the Motion Compensator:

- Inverse operators: Q^{-1} and T^{-1} are applied to the encoded current frame I_{ke} or the current prediction error E_{ke} to reconstruct and store the encoded frame for the motion estimator and motion compensator to generate the next prediction frame.

- Delayed frame memory: both current and previous frames must be available to the motion estimator and motion compensator to generate a prediction frame.

The number of previous frames stored in memory can vary based on the requirements of the encoding algorithm.

- Motion estimation: this element estimates the rigid body motion between two or more successive frames. The motion estimator operates on all current frame 16x16 image blocks and, for each block, generates the pixel displacement or motion vector whose size is constrained by the search neighborhood.
  The search space size determines the complexity of the motion estimation algorithm; full search methods are generally not used practically. If a search does not meet a minimum Mean Square Error (MSE) or Mean Absolute Error (MAE) threshold criteria, the motion compensator will indicate that the current block is to be spatially encoded using intraframe mode.

- Motion compensation: this element uses the current frame motion estimates MV_k and the previously reconstructed frame $\hat{I}_{k-1}$ to generate the current frame prediction P_k. A threshold criteria is used to decide which blocks will be encoded as prediction error blocks using motion vectors and which blocks will only be spatially encoded.

The model described above does not address some video compression system details such as the bit-stream syntax (which supports different application requirements) or the specifics of the encoding algorithms. These issues are dependent on the video compression system design.

Some technical details of H.264/AVC standard are presented in the following. The main improvements are within error resiliency and compression efficiency through the use of variable block sizes for motion compensation, which enables more localized motion prediction and better exploitation of the temporal redundancies in the sequence during compression. This is an important feature when combined with temporal error concealment techniques that use the motion vectors of neighboring blocks to estimate the motion vector of a lost block.

The accuracy of motion estimation is refined to quarter-pixel resolution for the luma component of the video signal. H.264 also uses an in-loop adaptive deblocking filter for each reference frame to improve its perceptual quality and compression efficiency. The compression performance is also improved by the use of a spatial prediction from neighboring macroblocks for the encoding of intra macroblocks, i.e. only the residual signal is transform coded. However, this form of prediction can lead to error propagation if the spatially neighboring macroblocks are intercoded and use unreliable reference blocks for MC. Hence, in packet lossy channels, only pixels from other intracoded macroblocks can be used for intraspatial prediction.

Unlike in previous video compression standards, H.264 allows for motion compensation of bipredictive (B) coded pictures from previously coded B pictures in the sequence. In addition to greater compression efficiency, this also enables temporal scalability through hierarchical biprediction, which is a useful feature for scalable video coding (SVC). The latter is a novel coding that results in a bit stream that offers progressive refinement of video quality by dynamically adapting the source rates to changing network and channel conditions.

3.4 New video coding models

As illustrated in the previous section, most video encoding standards use block-based motion estimation and compensation to address inter frame redundancy. However, this approach tends to produce poor results at very low bit rates: large prediction errors can be generated in some regions of an image, since motion boundaries do not necessarily coincide with block boundaries. Such errors must be coded and transmitted to avoid serious image degradations but at very low bit rates this cannot be successfully achieved. Hence, visible artifacts, called mosquito effects, appear at object boundaries in reconstructed images. In addition, block coding of prediction errors also rise to blocking artifacts, a line structure in reproduced frames.

Due to the above considerations, alternative video encoding models have been the focus of current research. On the other hand, it is important to consider the great advances in VLSI, ASIC, and microcomputer technology in the last decade: the real-time nature of video communications necessitates the use of specialized high-performance hardware devices and in the near future advances in design and manufacturing technologies will create hardware devices that will allow greater adaptability, interactivity, and interoperability of video applications.

This work mainly focuses on how video signals should be efficiently coded and transmitted over the error-prone, low bit rate TETRA2 channels to achieve the best possible quality. As highlighted in the Introduction, an in-depth analysis of the scenarios and applications is the first step towards the definition of the most suitable coding architecture to be used.

TETRA already provides voice and data transmission, where the data is intended to be photos, fax, vehicle locations, maps etc. The advent of TEDS in TETRA2 will provide quality real-time mobile video transmission and will offer several advantages to PMR users. A common example in the literature is the "Technocop", a police officer who wears a helmet with a visor that can receive video and other types of data like maps, locations and photos [19]. In addition, a camera would be embedded in the helmet so he could send video sequences of the crime scene to other officers. Another example of real time video transmission for PSDR agencies is during an accident where a camera could be used to transmit a video of the accident scene back to the police and ambulance services which could then accurately assess the situation and dispatch the relevant medical equipment and aid [20] [21].

In general, an image or a video sequence will normally be captured and transmitted to and from portable terminals, cameras in surveillance and environmental tracking systems, laptop computers connected through the SwMi or other devices connected through other wireless technologies; typically the devices rely on a battery with a limited power supply. Note that energy in mobile devices is mainly used for (i) computation to run the operating system software and to encode and decode the audio and video signals, (ii) transmission to transmit and receive the RF audio and video signals, (iii) display and driving speakers. Computational energy consumption is especially a concern for video transmission due to motion estimation and compensation, forward and inverse DCT (IDCT) transformations, quantization, and other processing units. Thus, the efficient use of bandwidth and energy becomes highly important in the deployment of TETRA2 video applications. Research has focused on performance analysis when transmitting MPEG-4 coded video over the error-prone TETRA2 channel [21] [22].

This thesis aims to provide a performance analysis of video transmission over TETRA2 channel, comparing H.264 coding efficiency with new advanced video coding strategies, Distributed Video coding (DVC) and Multiple Description Coding (MDC) that can be adopted in TETRA2 system depending on the scenario and the application. In particular, a DVC approach results suitable when low power devices encode separately two or more statistically dependent sources at the same rate obtained when the same sources are joint encoded and decoded, thus reducing the amount of data to be transmitted to the base station and simplifying the complexity of the encoder. On the contrary, MDC is to be adopted when different devices transmit the same source message over particularly noisy and unreliable channels: the message is split and encoded into two (or more) complementary descriptions, which are independently transmitted to the receiver by using separate channels or paths; the more descriptors are received, the higher the reconstruction quality. The advantage of this technique, compared with progressive coding, is that the decoder can get a useful image more quickly in case of packet losses.

3.5 Tree Structured Haar Transform

During the last decade, the Discrete Wavelet Transform (DWT) and subband decomposition have gained increased popularity. Recently, there has also been active research applying the DWT to video coding.

The wavelet transform has a number of advantages over other transforms such as DCT that can be exploited for both image compression and watermarking applications[23]:

- **Space-frequency localization**. The wavelet domain provides good space-frequency localization for analyzing image features such as edges or texture areas. These features correspond to the large coefficients in the detail subbands at various resolutions.
- **Multi-resolution representation**. Hierarchical processing is available in a straightforward way and it is important for progressive and scalable transmission.
- **Superior HVS modeling**. Image compression methods can benefit from a good model of the human visual system that allows to adapt the distortion introduced to the masking proprieties of the human eye. The dyadic frequency decomposition of the wavelet transform resembles the signal processing of the HVS and thus permits to excite the different perceptual bands individually.
- **Linear complexity**. The wavelet transform has linear computational complexity, $O(n)$, opposed to $O(nlog(n))$ for the DCT (where n is the length of the signal to be transformed).
- **Adaptivity**. The wavelet transform is flexible enough to adapt to a given images or a particular type of application. The decomposition filters and the decomposition structure can be chosen to reflect the characteristics of the image.

In particular, among the wavelets, the Haar function has been used because it has some proprieties that make it useful in signal processing: (i) its local statistics are relatively constant and easily modeled; (ii) many of its value are close to zero, thus making it an excellent candidate for image compression; (iii) both coarse and fine resolution approximation of the original image can be easily extracted; (iv) there exist a fast algorithm that is computationally more convenient respect to the Fourier transform and to the other types of wavelet.

Due to these reasons, wavelet transform encoding results in excellent bit-rate distortion characteristics, is easily adaptable for progressive transmission and has received special attention due to their inherent feature of full scalability.

This work employees the discrete Tree Structured Haar (TSH) transform, recently developed in [24], which is a generalization of the Discrete Haar transform. To illustrate the TSH transform, a brief description of the classical Haar discrete transform is reported.

Given the discrete interval $I = [1, N]$ with $N = 2^L$, let $\phi^H(t)$ denote the Haar scaling function:

$$\phi^H(t) = \begin{cases} \frac{1}{\sqrt{N}}, \text{ if } t \in [1, N] \\ 0, otherwise \end{cases} \tag{3.1}$$

To construct the representation basis, the interval I is split into two halves $I_0^H = [1, 2^{-1}N]$ and $I_1^H = [2^{-1}N + 1, N]$. Then, it is introduced the (wavelet) function $\psi^H(t)$ defined as follows:

$$\psi^H(t) = \begin{cases} \frac{1}{\sqrt{N}}, \text{ for } t \in I_0^H \\ -\frac{1}{\sqrt{N}}, \text{ for } t \in I_1^H \\ 0, otherwise \end{cases} \tag{3.2}$$

It can be verified that $\phi^H(t)$ and $\psi^H(t)$ are orthogonal and $\|\phi^H(t)\| = 1$ and $\|\psi^H(t)\| = 1$ where $\|\cdot\|$ denotes the L_2 norm. The set $\{\phi^H(t), \psi^H(t)\}$ can be extended by splitting I_0^H into $I_{0,0}^H$ and $I_{0,1}^H$ and I_1^H into $I_{1,0}^H$ and $I_{1,1}^H$ having denoted with subscript 0 and 1 the left and the right halves respectively, thus obtaining:

$$\phi_j^H(t) = \begin{cases} \frac{1}{\sqrt{2^{-1}N}}, \text{ for } t \in I_{j,0}^H \\ \frac{1}{\sqrt{2^{-1}N}}, \text{ for } t \in I_{j,1}^H \\ 0, otherwise \end{cases} \quad \psi_j^H(t) = \begin{cases} \frac{1}{\sqrt{2^{-1}N}}, \text{ for } t \in I_{j,0}^H \\ -\frac{1}{\sqrt{2^{-1}N}}, \text{ for } t \in I_{j,1}^H \\ 0, otherwise \end{cases} \quad j = 0, 1. \tag{3.3}$$

This procedure is iterated $L-1$ times, until each subinterval contains only one integer. More specifically, for a given resolution level k $(1 \leq k < L)$, there are 2^k intervals $I^H_{\alpha_1,\alpha_2,\ldots,\alpha_k} = [a^H_{\alpha_1,\alpha_2,\ldots,\alpha_k}, b^H_{\alpha_1,\alpha_2,\ldots,\alpha_k}]$, with $\alpha_n = \{0,1\}$, with

$$\begin{aligned} a^H_{\alpha_1,\alpha_2,\ldots,\alpha_k} &= \textstyle\sum_{n=1}^{k} \alpha_n 2^{L-n} + 1, \\ b^H_{\alpha_1,\alpha_2,\ldots,\alpha_k} &= \textstyle\sum_{n=1}^{k} \alpha_n 2^{L-n} + 2^{L-k}; \end{aligned} \tag{3.4}$$

the following two basis functions are associated to each interval

$$\phi^H_{\alpha_1,\alpha_2,\ldots,\alpha_k}(t) = \begin{cases} \frac{1}{\sqrt{2^{-k}N}}, \text{ for } t \in I^H_{\alpha_1,\alpha_2,\ldots,\alpha_k} \\ 0, otherwise \end{cases} \tag{3.5}$$

$$\psi^H_{\alpha_1,\alpha_2,\ldots,\alpha_k}(t) = \begin{cases} \frac{1}{\sqrt{2^{-k}N}}, \text{ for } t \in I^H_{\alpha_1,\alpha_2,\ldots,\alpha_k,0} \\ -\frac{1}{\sqrt{2^{-k}N}}, \text{ for } t \in I^H_{\alpha_1,\alpha_2,\ldots,\alpha_k,1} \\ 0, otherwise \end{cases} \tag{3.6}$$

In the TSH discrete transform the interval I is split into two intervals $I_0^{TSH} = [1, \nu_0]$ and $I_1^{TSH} = [\nu_0 + 1, N]$ with $1 \leq \nu_0 < N$; however N does not need to be a power of 2. The basis function $\psi^{TSH}(t)$ is defined as follows:

$$\psi^{TSH}(t) = \begin{cases} \sqrt{\frac{\nu_1}{\nu_0 N}}, \text{ for } t \in I_0^{TSH} \\ -\sqrt{\frac{\nu_0}{\nu_1 N}}, \text{ for } t \in I_1^{TSH} \\ 0, otherwise \end{cases} \tag{3.7}$$

where $\nu_1 = N - \nu_0$. This construction can be iterated by splitting I_0^{TSH} into $I_{0,0}^{TSH}$ and $I_{0,1}^{TSH}$ and I_1^{TSH} into $I_{1,0}^{TSH}$ and $I_{1,1}^{TSH}$ respectively, and so on. In general, given the interval $I^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k} = [a^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k}, b^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k}]$, with $\alpha_j = \{0,1\}$, of length $\nu_{\alpha_1,\alpha_2,\ldots,\alpha_k} = b_{(\alpha_1,\alpha_2,\ldots,\alpha_k)} - a_{(\alpha_1,\alpha_2,\ldots,\alpha_k)} + 1$, if $\nu_{\alpha_1,\alpha_2,\ldots,\alpha_k} > 1$, $I^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ can be partitioned into two intervals $I^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k,0}$ and $I^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k,1}$ of length $\nu_{\alpha_1,\alpha_2,\ldots,\alpha_k,0} \geq 1$ and $\nu_{\alpha_1,\alpha_2,\ldots,\alpha_k,1} =$

$\nu_{\alpha_1,\alpha_2,\ldots,\alpha_k} - \nu_{\alpha_1,\alpha_2,\ldots,\alpha_k,0} \geq 1$. The following recursive relationships hold

$$\begin{aligned} a^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k,0} &= a^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k}, \\ b^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k,0} &= a^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k} + \nu_{\alpha_1,\alpha_2,\ldots,\alpha_k,0} - 1 \end{aligned} \tag{3.8}$$

and

$$\begin{aligned} a^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k,1} &= a^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k} + \nu_{\alpha_1,\alpha_2,\ldots\alpha_k,0}, \\ b^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k,1} &= b^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k} \end{aligned} \tag{3.9}$$

with initial conditions $a^{TSH}_0 = 1$, $b^{TSH}_0 = \nu_0$, $a^{TSH}_1 = \nu_0 + 1$, $b^{TSH}_1 = N$. The functions $\phi^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k}(t)$ and $\psi^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k}(t)$ are defined as follows:

$$\phi^{TSH}_{\alpha_1,\alpha_2,\ldots\alpha_k}(t) = \begin{cases} \frac{1}{\sqrt{\nu_{\alpha_1,\alpha_2,\ldots\alpha_k}}}, \text{ for } t \in I^{TSH}_{\alpha_1,\alpha_2,\ldots\alpha_k} \\ 0, otherwise \end{cases} \tag{3.10}$$

$$\psi^{TSH}_{\alpha_1,\alpha_2,\ldots\alpha_k}(t) = \begin{cases} \sqrt{\frac{\nu_{\alpha_1,\alpha_2,\ldots\alpha_k,1}}{\nu_{\alpha_1,\alpha_2,\ldots\alpha_k,0}\nu_{\alpha_1,\alpha_2,\ldots\alpha_k}}}, \text{ for } t \in I^{TSH}_{\alpha_1,\alpha_2,\ldots\alpha_k,0} \\ -\sqrt{\frac{\nu_{\alpha_1,\alpha_2,\ldots\alpha_k,0}}{\nu_{\alpha_1,\alpha_2,\ldots\alpha_k,1}\nu_{\alpha_1,\alpha_2,\ldots\alpha_k}}}, \text{ for } t \in I^{TSH}_{\alpha_1,\alpha_2,\ldots\alpha_k,1} \\ 0, otherwise \end{cases} \tag{3.11}$$

It has been demonstrated in [24] that the set of TSH functions is a set of orthogonal functions. It is important to remark that each splitting scheme defines a different basis set and, for a given N, each TSH basis is univocally defined by the set $\{a^{TSH}_{\alpha_1,\alpha_2,\ldots,\alpha_k}\}$, named Discontinuity Point Vector (DPV), that defines the splitting scheme. An example of the interval splitting procedure and the corresponding set $\{\psi(t)\}$ is illustrated in Figure 3.2. The DPV randomness allows to obtain a multiresolution analysis in which the details are spread all over the subbands. In particular for a signal of length N, the possible number of DPV for the first order decomposition can be computed as:

$$DPVNumber = (N-1) + \sum_{j=1}^{N-2} \prod_{i=N-1-j}^{N-1} i. \tag{3.12}$$

Let T be the $(N \times N)$ orthogonal matrix, whose rows are constituted by the basis functions. Then, given a column vector $\mathbf{f}$ of size N, its TSH transform $\mathbf{g}$ is

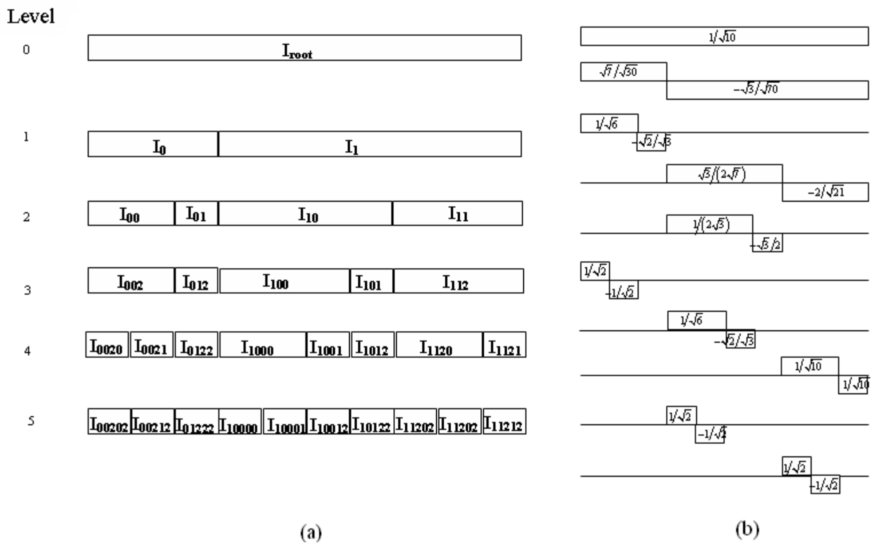


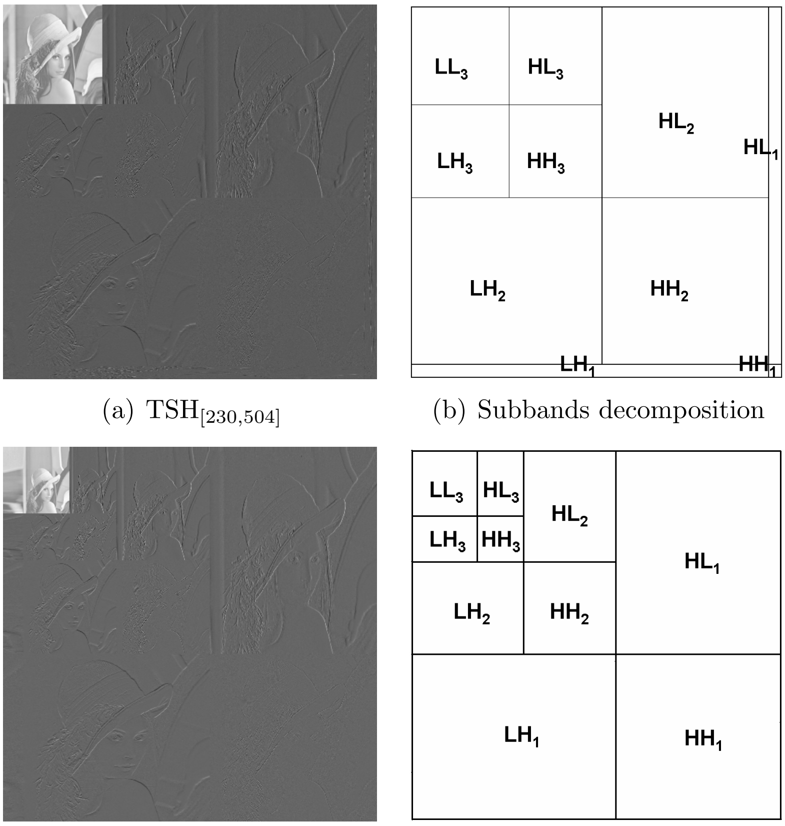
Figure 3.2: Splitting intervals (a) and associated TSH functions (b).

$$\mathbf{g} = T\mathbf{f}. \tag{3.13}$$

A detailed study of matrix representation of TSH transform of a general signal can be found in [25]. Here, the TSH transform is applied to two dimensional signals. In this case different DPVs lead to different subband decompositions of the image (see in Figure 3.3 examples of 3^{rd}-level decomposition depending on two different DPVs).

3.5.1 Fast TSH algorithm

For the TSH transform, an algorithm of linear complexity ($2(N-1)$ additions and $3N-2$ multiplications), based on the transform matrix factorization, has been developed [26]. In the proposed approach a step-by-step decomposition has been used [27]: for each level j of the tree a $\mu_j \times \mu_j$ matrix P_j is generated, where μ_j is the number of nodes of the considered level. This matrix is composed by two sub-matrices: the first one U_j has dimension $\mu_{j-1} \times \mu_j$ where μ_{j-1} is the number of the nodes in the previous level and represents the ”low-pass” operator that allows to obtain the approximation of the signal as in the classical Haar case; the second one



(a) TSH$_{[230,504]}$ (b) Subbands decomposition

(c) TSH$_{[27]}$ (d) Subbands decomposition

Figure 3.3: A 3$^{\mathrm{rd}}$-level TSH transform of the image *Lena*: (a)-(b) with DPV=[230,504]; (c)-(d) with DPV=[27].

V_j has dimension $(\mu_j - \mu_{j-1}) \times \mu_j$ where $(\mu_j - \mu_{j-1}) = \sigma_j$ is the number of splitting nodes in the previous level and represents the "high-pass" operator that allows to obtain the detail of the signal.

In detail the matrices U_j and V_j that form the transform matrix $P_j = \begin{pmatrix} U_j \\ V_j \end{pmatrix}$ for the j-th level of the tree are built as follows:

- Initialize the $(\mu_j \times \mu_j)$ matrix $\widetilde{U}_j = 0$.

  For $i = 1, \ldots, \mu_j$

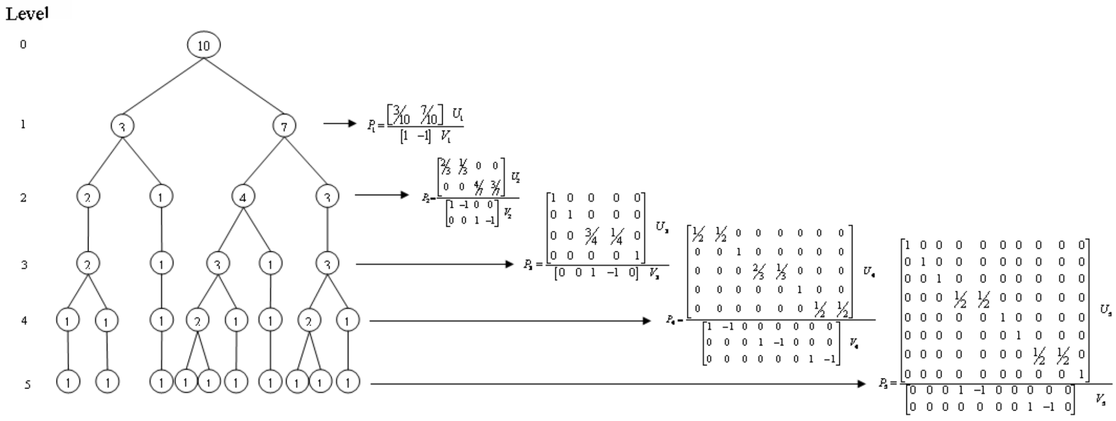


Figure 3.4: Definition of transform matrices for each level of a tree in step-by-step TSH algorithm.

if $c(j,i) =$ left child, then

$$\widetilde{U}_j(i,i) = \tfrac{\nu(c(j,i))}{\nu(c(j,i))+\nu(c(j,i+1))},$$
$$\widetilde{U}_j(i,i+1) = \tfrac{\nu(c(j,i+1))}{\nu(c(j,i+1))+\nu(c(j,i))}$$

if $c(j,i) =$ right child, then $\widetilde{U}_j(i,i) = 0$

if $c(j,i) =$ only child,then $\widetilde{U}_j(i,i) = 1$.

U_j is obtained from $\widetilde{U}_j$ by deleting all zero rows.

- Initialize the $(\mu_j \times \mu_j)$ matrix $\widetilde{V}_j = 0$.

  For $i = 1, \dots, \mu_j$

  if $c(j,i) =$ left child, then $\widetilde{V}_j(i,i) = 1$ and $\widetilde{V}_j(i,i+1) = -1$

  if $c(j,i) =$ right child, then $\widetilde{V}_j(i,i) = 0$

  if $c(j,i) =$ only child,then $\widetilde{U}_j(i,i) = 0$.

  V_j is obtained from $\widetilde{V}_j$ by deleting all zero rows.

An example of transform matrices is presented in Figure 3.4. The step by step decomposition operates so that it is possible to transform a signal until a desired level of decomposition according to the scheme proposed in Figure 3.5 that refers to a second level transform. As in one dimensional case the value of the root of the

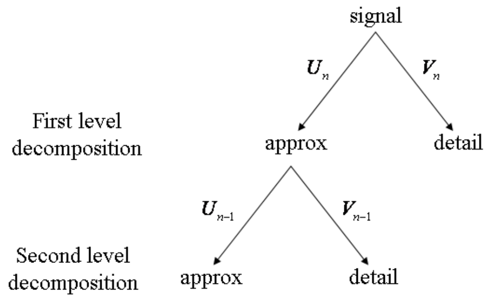


Figure 3.5: Second level step-by-step TSH decomposition procedure.

tree corresponds to the length of the input vector to be processed, in two-dimensional case, the value of the dimension of the image represents the root of the tree and transformation by rows is followed by a transformation by its columns, similar to the classical case.

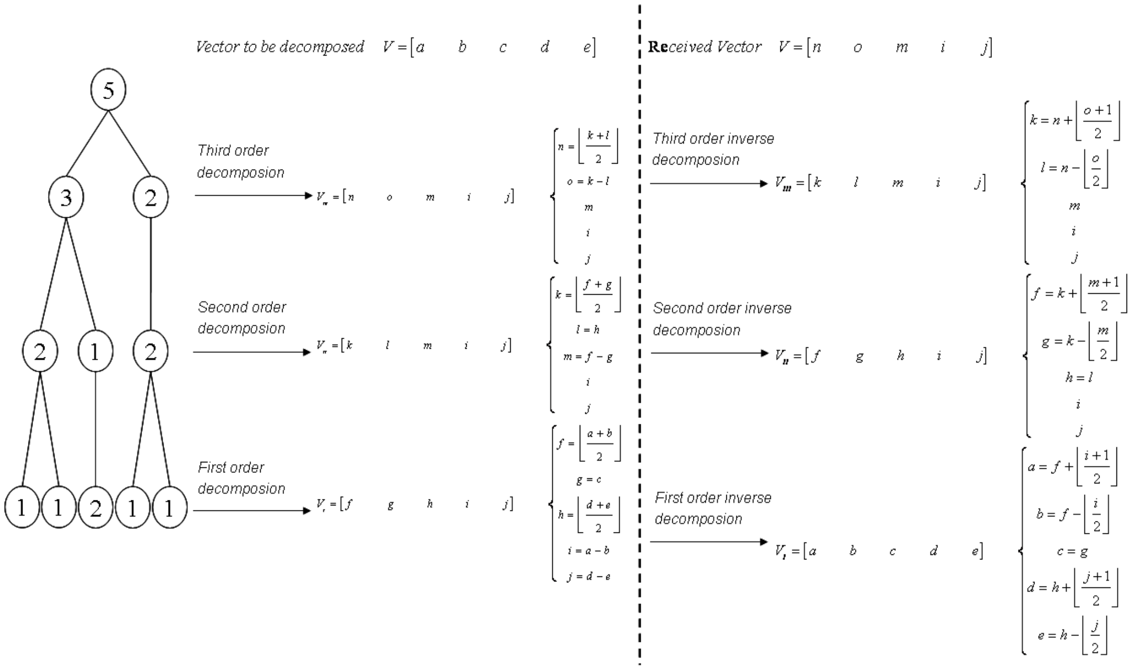

5
3
2
2
1
2
1
1
2
1
1
Vector to be decomposed $V=[a \quad b \quad c \quad d \quad e]$
Third order decomposion
$V_{m}=[n \quad o \quad m \quad i \quad j]$
$n=\left\lfloor\frac{k+l}{2}\right\rfloor$, $o=k-l$, m, i, j
Second order decomposion
$V_{n}=[k \quad l \quad m \quad i \quad j]$
$k=\left\lfloor\frac{f+g}{2}\right\rfloor$, $l=h$, $m=f-g$, i, j
First order decomposion
$V_{t}=[f \quad g \quad h \quad i \quad j]$
$f=\left\lfloor\frac{a+b}{2}\right\rfloor$, $g=c$, $h=\left\lfloor\frac{d+e}{2}\right\rfloor$, $i=a-b$, $j=d-e$
Received Vector $V=[n \quad o \quad m \quad i \quad j]$
Third order inverse decomposion
$V_m=[k \quad l \quad m \quad i \quad j]$
$k=n+\left\lfloor\frac{o+1}{2}\right\rfloor$, $l=n-\left\lfloor\frac{o}{2}\right\rfloor$, m, i, j
Second order inverse decomposion
$V_n=[f \quad g \quad h \quad i \quad j]$
$f=k+\left\lfloor\frac{m+1}{2}\right\rfloor$, $g=k-\left\lfloor\frac{m}{2}\right\rfloor$, $h=l$, i, j
First order inverse decomposion
$V_t=[a \quad b \quad c \quad d \quad e]$
$a=f+\left\lfloor\frac{i+1}{2}\right\rfloor$, $b=f-\left\lfloor\frac{i}{2}\right\rfloor$, $c=g$, $d=h+\left\lfloor\frac{j+1}{2}\right\rfloor$, $e=h-\left\lfloor\frac{j}{2}\right\rfloor$

Figure 3.6: One dimensional integer TSH.

3.5.2 Integer Tree Structured Haar Transform

An integer TSH transform can be performed if the decomposition matrices described in the previous section are factorized [28]. Figure 3.6 shows an example of TSH complete decomposition of an input vector of length 5 reported. After the binary tree generation, depending on a secret DPV, the tree nodes are analyzed level by level to decompose the signal.

Chapter 4

Distributed Stereo Video coding

The use of stereo systems to deliver information is increasing everyday for entertainment, medical applications, remote control, scene analysis, object detection and tracking, and so on [29]. In particular, existing video-surveillance scenarios, typical of a TETRA system, use stereo cameras as simple and cost effective solutions for scene acquisition, where several views of a scene can be combined to produce a complete 3D representation [30]. This scheme originates accurate and detailed information of the acquired video and help the operators in object recognition and data processing.

These applications can benefit from the extra information given by the stereo architecture in resolving occlusions, improving motion estimation, improving compression rate, and so on. The main drawback in using stereo sequences is the increased amount of data to be transmitted. As in multiple camera coding system, the stereo information can be more efficiently coded if the correlation among the sequences is exploited. Two approaches are available: in the first one the two sequences are jointly coded to achieve maximum compression rate. In the second framework, each sequence is coded by an independent encoder which relies on a joint decoding system. A close approach has been proposed by Wyner Ziv [5] with the source coding with side information paradigm. This strategy has been adopted by many authors for the design of high compression rate inter-frame video coding schemes.

Based on simple preliminary approach [31], in this chapter novel DVC methods for stereo sequences are proposed for TETRA2 low bit-rate communication channels in terms of SI generation, which differentiates them from the monoview case. More specifically, different stereo prediction techniques for SI generation are described and compared in terms of prediction quality, complexity and compression efficiency. The proposed DVC approach is evaluated by means of subjective tests and objective quality models.

4.1 Theoretical background

Consider a communication system where two discrete memoryless random sources X and Y must be encoded and transmitted to a common receiver. If X and Y are encoded at rates R_x and R_y respectively, according to Shannon's information theory principles these rates are sufficient to perform noiseless coding if they are equal or greater than the entropies $H(X)$ and $H(Y)$ of the sources (i.e. $R_x \geq H(X)$ and $R_y \geq H(Y)$).

If X and Y are correlated and have a joint probability distribution given by $p_{X,Y}(x, y)$, they can be jointly encoded noiseless at a total rate corresponding to the joint entropy $H(x, y)$ of the sources (obviously $R_x + R_y \geq H(x) + H(Y) \geq H(X, Y)$) (see Fig. 4.1).

Slepian and Wolf [6] showed in 1973 that if X and Y are physically separated and cannot communicate with each other, noiseless coding of X and Y is still achievable if $R_x \geq H(X|Y), R_y \geq H(Y|X)$ and $R_x + R_y \geq H(X, Y)$, where $H(X|Y)$ corresponds to the conditional entropy of X given Y.

This surprising result gives the theoretical foundation of Distributed Source Coding (DSC). It shows that there is in theory no loss in terms of overall rate even though the encoders are separated (see Fig. 4.2).

An extension of the Slepian-Wolf result to the lossy case (with continuous sources) was proposed by Wyner and Ziv, that addressed a particular case of Slepian-Wolf

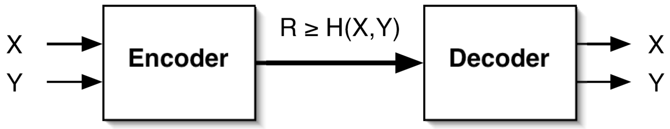


Figure 4.1: Joint source coding: if sources X and Y are correlated, they can be jointly encoded noiseless at a total rate corresponding to the joint entropy of the sources.

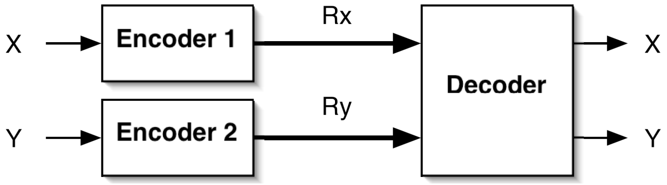


Figure 4.2: Distributed source coding: if X and Y are physically separated, noiseless coding is still possible if $R_x \geq H(X|Y), R_y \geq H(Y|X)$ and $R_x + R_y \geq H(X,Y)$.

coding corresponding to the rate point $(R_x, R_y) = (H(X|Y), H(Y))$, also known as source coding with side information at the receiver [5]. They gave a rate-distortion function $R^*_{WZ}(D)$ for the problem of encoding one source X, guaranteeing an average fidelity of $E\{d(X, \hat{X})\} = D$, assuming that the other source Y (playing the role of side information) is available losslessly at the decoder, but not at the encoder.

DSC theory has resulted to be particularly interesting [32] for the decentralized scenario of wireless sensor network. Due to the spatial proximity of the sensors and the physical properties of the observed phenomenon, the data acquired among the sensors are highly correlated; transmitting this data directly from each sensor to the common receiver could thus imply the communication of a large amount of redundant information. Since the sensors have limited power resources, efficient data transmission is crucial to guarantee the survival of such systems. This particular constraint makes DSC of great interest for wireless sensor network applications.

Distributed Video Coding (DVC) employees DSC in order to allow independent encoding of each frame at the encoder side, while leaving to the decoder the task

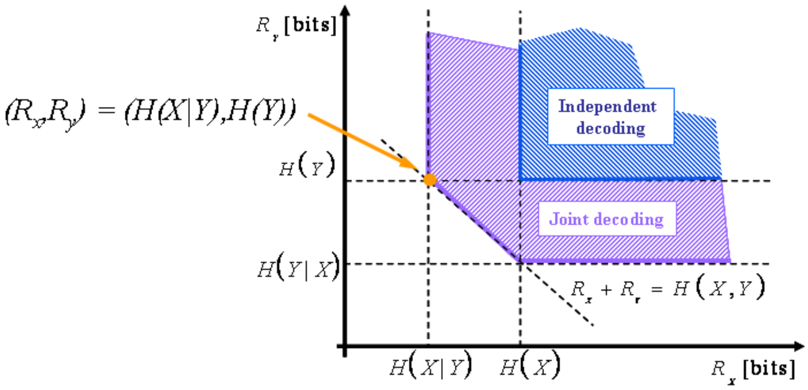


Figure 4.3: Achievable rate region defined by the Slepian-Wolf bounds.

of exploiting the temporal dependencies. DVC [5][6] theory states that, for a given distortion, it is theoretically possible to separately encode and jointly decode two or more statistically dependent sources at the same rate obtained when the same sources are joint encoded and decoded. As already stated, this strategy has been adopted by many authors for the design of high compression rate inter-frame video coding schemes. The common goal is to generate at the decoder a side information that optimally blends temporal and interview data. Figure 4.4 shows a general scheme of Wyner-Ziv video coding.

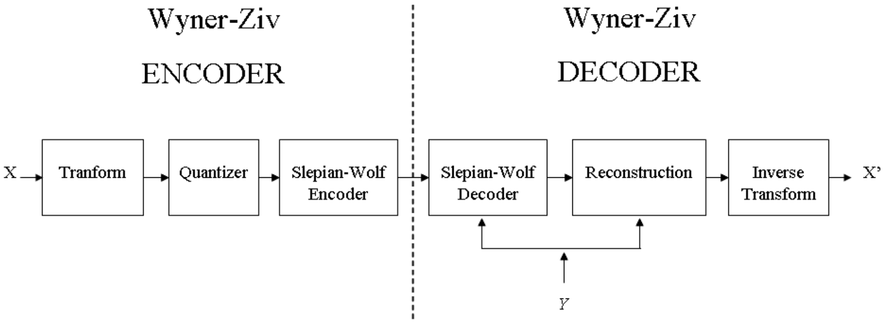


Figure 4.4: General scheme of Wyner-Ziv video coding.

In other terms, while standard video coders exploit the statistical dependencies of the source signal in order to remove spatial and temporal redundancies, in DVC each video frame is encoded independently knowing that some side information will be available at the decoder in order to remove transmission errors and improve the video quality (the side information can typically be a prediction based on previously decoded frames). This approach considerably reduces the complexity of the video encoder by shifting all the complex interframe processing tasks to the decoder. This property can be very interesting for power/processing limited systems such as wireless camera sensors that have to compress and send video to a fixed base station in a power-efficient way. It is normally assumed that the receiver can run a more complex decoder but when the receiver is another complexity-constrained device, a more powerful video transcoder somewhere on the network can be used.

The most important benefits of the DVC paradigm has been recognized in [33]:

- flexible allocation of the overall video codec complexity between encoder and decoder;
- improved error resilience; in fact, DVC behaves as a joint source-channel coding solution where the bits spent work simultaneously to improve quality and recover errors;
- codec independent scalability: the DVC prediction approach between the scalable layers does not require a deterministic knowledge of the previous layers, i.e. the layers may be generated by various, different and unknown codecs;

Scientific community has recently proposed several approaches in this directions. However, although all these approaches are extremely promising, they are still not as efficient as standard video coders in terms of rate-distortion performance due to the fact that distributed source coding techniques rely on a a-priori knowledge of the correlation structure. This knowledge is the only factor that allows to design optimal codes. The estimation of this correlation has proven, however, to be extremely difficult.

Another benefit of the DVC has been identified in the exploitation of multiview correlation. In the case of multi-view coding the amount of raw data acquired by practical systems can be extraordinary large and typically consists of hundreds of different views. A joint encoder could exploit the correlation between the different views but this would require that all the cameras first transmit their data to a common receiver that would have to store it and then perform the joint compression. This would clearly use a big amount of communication resources and storage space, and might not be feasible in some practical settings. For these reasons, it would be preferable to compress the images directly at the encoder stage using distributed compression techniques in order to require a low-complexity encoder at each camera and to considerably reduce the overall amount of transmission necessary from the cameras to the central decoder. Figure 4.5 depicts a possible multiview DVC scenario

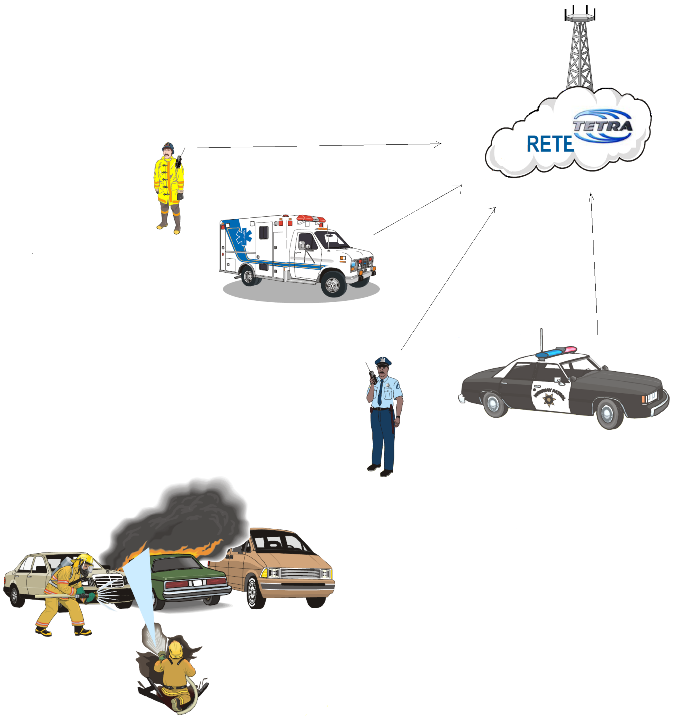


Figure 4.5: DVC applications in TETRA2 network.

in TETRA2: several low power devices separately encode different views of the same scene and transmit the bitstream to the base station; the latter decodes the data by exploiting the spatial and temporal correlation between the views. Moreover, the compressed data could be directly stored at the receiver using optimal memory space: DVC-based encoders do not need to jointly process the various views and thus do not need inter-camera and inter-encoder communication. Nevertheless, in this case the decoder is assumed to be more sophisticated in order to handle the high-complexity joint decoding of the views.
However, multiview DVC approaches are often not simple in practical applications as asymmetric: in fact some cameras need to transmit their full information to provide side information to the decoder while others only transmit partial information. Finally most of the multi-view DVC approaches do not take advantage of the multi-view geometry to improve the performance of their encoders.

4.2 Side Information generation: available approaches

When dealing with distributed stereoscopic video coding, several approaches are feasible:

- Inter-view approach.
  This is probably the simplest implementation of stereo DVC as it is assumed that one view is Wyner-Ziv coded while the other represent the key sequence. However, this systems do not exploit the temporal correlation between the frames of the Wyner-Ziv sequence.

- Intra-view approach.
  Most of these technique are based on a Motion Compensated Temporal Interpolation technique [34] [35], which performs Block-based motion estimation between two key frames and creates the estimated Wyner-Ziv frame by using

halved motion vectors.

DVC monoview schemes can be easily adapted to stereoscopic scenarios, where it is assumed that the different views do not communicate with each other. This approach does not exploit the intraview correlation between the views. On the other side, this approach lead to a balanced system as all the cameras have the same computational complexity, same capacity and share equally the available bandwidth resource. Compared to previous approach, each encoder has to encode both key frames with conventional Intracoding and Wyner-Ziv frames with DVC, thus leading to a higher computational complexity.

Tonomura and Nakachi [36] proposed a dual DVC scheme that utilizes the scalability of the JPEG2000 standard. In particular, the scheme takes the even numbered frames as Key frames and encodes them using the JPEG2000 coder while the odd numbered frames are taken as Wyner-Ziv frames and encoded using a Low Density Parity Code. The system generates the side information before decompressing all bit streams. For each Wyner-Ziv frame the LDPC decoder exploits the side information and the received syndrome bits to recover the Wyner frames.

In [37] the authors efficiently use the wavelet domain to exploit the spatial correlation of a Wyner-Ziv frame. The quantized wavelet coefficients of the Wyner-Ziv frame are reordered using a set partition process similar to the zero-tree generation so as to identify the significant and insignificant coefficients. The significance map is coded with entropy coding and transmitted into decoder with intra coding mode. The significant coefficients are Wyner-Ziv encoded with turbo coder, and only the parity bits are transmitted. At the decoder, a predictive frame generated through motion-compensated prediction from adjacent intra frames is used as side information to recover the significant coefficients to be sent to the turbo decoder. When an acceptable probability of bit error rate is achieved, the frame is supposed successfully reconstructed. This method has also been extended for distributed multiview video coding in [38].

- Fusion approach.

  The most recent works use a side information fusion approach. A good review can be found in [39].

  In [40] the authors adaptively select either the temporal or the interview side information on a pixel by pixel basis. The system uses also a turbo decoder to detect when decoding is successful and no more parity bits need to be requested via the feedback channel. The proposed algorithm has the advantage to be symmetric with respect to the two cameras.

  Yeo and Ramchandran [41] proposed a robust method that exploits inter-view correlation among cameras that have overlapping views in order to deliver error-resilient video in a distributed multiple wireless camera sensors scenario. The system has low encoding complexity and satisfies tight latency constraints, and requires no inter-sensor communication. Each video frame is divided into non-overlapping blocks and the syndrome of each quantized block is transmitted with a cyclic redundancy check (CRC) computed on the quantized block. The encoder at each of the video camera sensors does not need any knowledge about the relative positions of any other cameras. The decoder searches over candidate predictors and attempts to decode using the received syndrome and the candidate predictor as side-information. If the CRC of the decoded sequence checks out, decoding is assumed to be successful. In particular, the decoder first try to decode a block using decoder motion search in the temporal dimension; if that fails, it is then performed a decoder disparity search along the epipolar line in each overlapping camera view.

  One promising stereo DVC scheme has been suggested by Pereira et al in [1]. The authors propose a practical solution for Wyner-Ziv stereo coding that avoids any communication between the low complexity encoders. The method is based on a pixel-level mask fusion of temporal and interview side information. In particular, the first view is coded in conventional way using H.264/AVC, and DVC principles are applied to the coding of the second, dependent view. The system fuses, pixel-by-pixel, the temporal side information created using a motion-based frame interpolation scheme with the interview side information created using a disparity-based frame extrapolation algorithm. Figure 4.6 shows a diagram of this system.

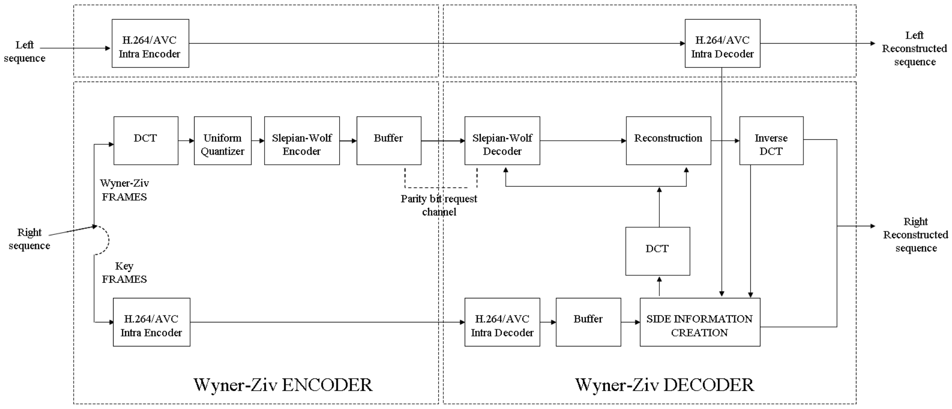


Figure 4.6: Stereo video coder architecture proposed in [1].

- Residual coding approach.

  While in practical Wyner-Ziv video coding, every frame is encoded independently but decoded based on the side information generated from adjacent frames, residual coding, proposed in [42], encodes the residual of a frame with respect to a reference frame, that can be seen as a second side information available both at the encoder and at the decoder, thus allowing the encoder this additional complexity of frame store and frame subtraction.

  In [43] the authors propose an efficient hybrid distributed video coding (HDVC) scheme combining residual coding, SW-SPIHT coding and intra-mode decision technology. The scheme applies a wavelet-domain Wyner-Ziv codec to compress the residual frame between the current frame and its reference frame and conducts an intra-mode decision to exploit the temporal and spatial correlation.

4.3 Syndrome coder approach for lifted TSH: GOP=2

In [44] the authors propose to adopt the distributed coding system to dynamic scenes with motion-compensated temporal wavelets and transform coding of temporal subbands. The system employs nested lattice codes for the transform coefficients and considers disparity-compensated video side information at the decoder.
A modification of [44] is proposed, employing a key-dependent wavelet transform [24] which allows a better adaptation to the frame content. The system encodes independently the left and right frames of the stereoscopic sequence. The decoder exploits the side information to achieve the best reconstruction of the correlated video streams. In particular, a syndrome coder approach based on a lifted scheme has been adopted.

4.3.1 Encoding

Let X and Y be the two considered stereo video sequences and let n_l and n_r denote the left and right frames to be processed where $n = 1, 2, ..., l$ and l is the total number of frames of the sequences.
Let n_l and $(n + 1)_l$ be a couple of frames of the left sequence and n_r and $(n + 1)_r$ be the corresponding frames of the right one. The right frames, assumed as key frames, are H.264/AVC Intra Coded while the left ones are Wyner-Ziv encoded as it is summarized in Fig. 4.7.

The Wyner-Ziv encoding procedure can be described as follows:

1. Given the pair n_l and $(n + 1)_l$, a temporal-spatial decomposition is performed. The temporal decomposition is the 1-D wavelet transform of an input signal along the temporal axis, i.e the input signal consists of pixel values that are collected from the same spatial position in both frame. Assuming each frame

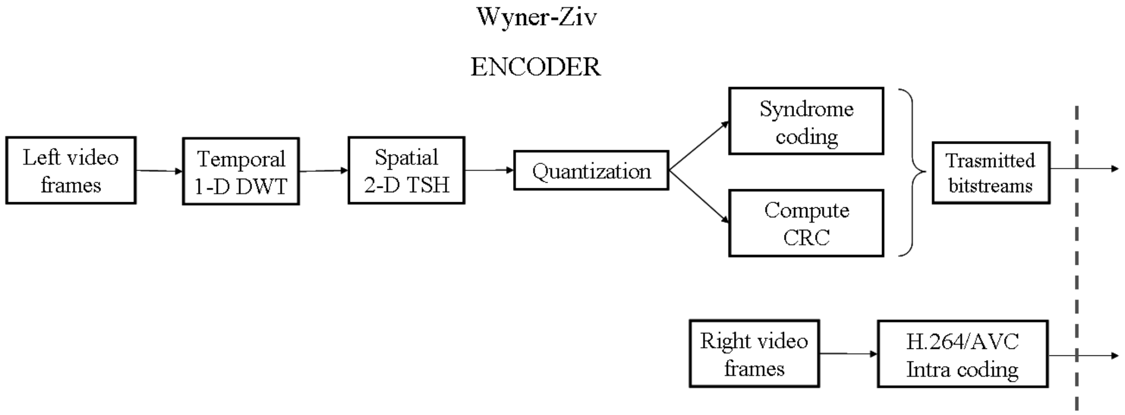


Figure 4.7: Encoding scheme.

has size of N rows and M columns in pixels, then $N \times M$ input signals are 1-D DWT decomposed. A one-level decomposition of the two frames n_l and $(n+1)_l$ results in one temporal high frequency band H_l and one temporal low frequency band L_l.

2. In order to obtain a better frequency localization, the temporal decomposition is followed by the spatial first order decomposition 2-D TSH as shown in Fig. 4.8. Let $\mathcal{L}_l$ and $\mathcal{H}_l$ denote respectively the TSH transforms of L_l and H_l.

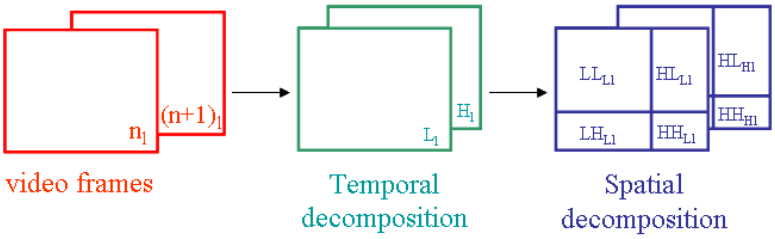


Figure 4.8: Temporal-Spatial decomposition scheme for two frame n_l and $(n+1)_l$ of the left sequence.

3. The coefficients $c_{\mathcal{L}_{l_{i,j}}}$ of $\mathcal{L}_l$ and $c_{\mathcal{H}_{l_{i,j}}}$ $\mathcal{H}_l$ are quantized according to the following

formula:

$$
\begin{aligned}
c'_{\mathcal{L}_{l_{i,j}}} &= sign\left(c_{\mathcal{L}_{l_{i,j}}}\right)\left\lfloor\frac{\left|c_{\mathcal{L}_{l_{i,j}}}\right|}{\Delta}\right\rfloor \\
c'_{\mathcal{H}_{l_{i,j}}} &= sign\left(c_{\mathcal{H}_{l_{i,j}}}\right)\left\lfloor\frac{\left|c_{\mathcal{H}_{l_{i,j}}}\right|}{\Delta}\right\rfloor
\end{aligned}
\tag{4.1}
$$

where a quantization step $\Delta = Q_1$ has been used for the average subbands LL_{L_l} and LL_{H_l} and a quantization step $\Delta = Q_2$ has been used for the details subbands.

4. The transmitted bitstream is composed by the syndrome [45] of the quantized $\mathcal{L}_l$ and $\mathcal{H}_l$ and by a cyclic redundancy check (CRC) word computed on the quantized $\mathcal{L}_l$ and $\mathcal{H}_l$.

4.3.2 Decoding

The side information exploits the inter-view correlation and is created by iteratively using the estimated disparity field for the past decoded frames. The Wyner-Ziv decoding procedure can be summarized as follows (see Fig. 4.9):

1. A temporal-spatial decomposition is performed on n_{SI} and $(n+1)_{SI}$.
2. The decoder exploits the side information and the received syndrome bits to recover the Wyner-Ziv encoded frames. If the CRC of the decoded sequence checks out, decoding is assumed to be successful. Note that in contrast to MPEG, H.26x, etc., it is the decoder's task to do disparity search, as it searches over the space of candidate predictors one-by-one to decode a sequence from the set labeled by the syndrome.
3. The decoded sequences $\hat{\mathcal{L}}_l$ and $\hat{\mathcal{H}}_l$ are dequantized as follows:

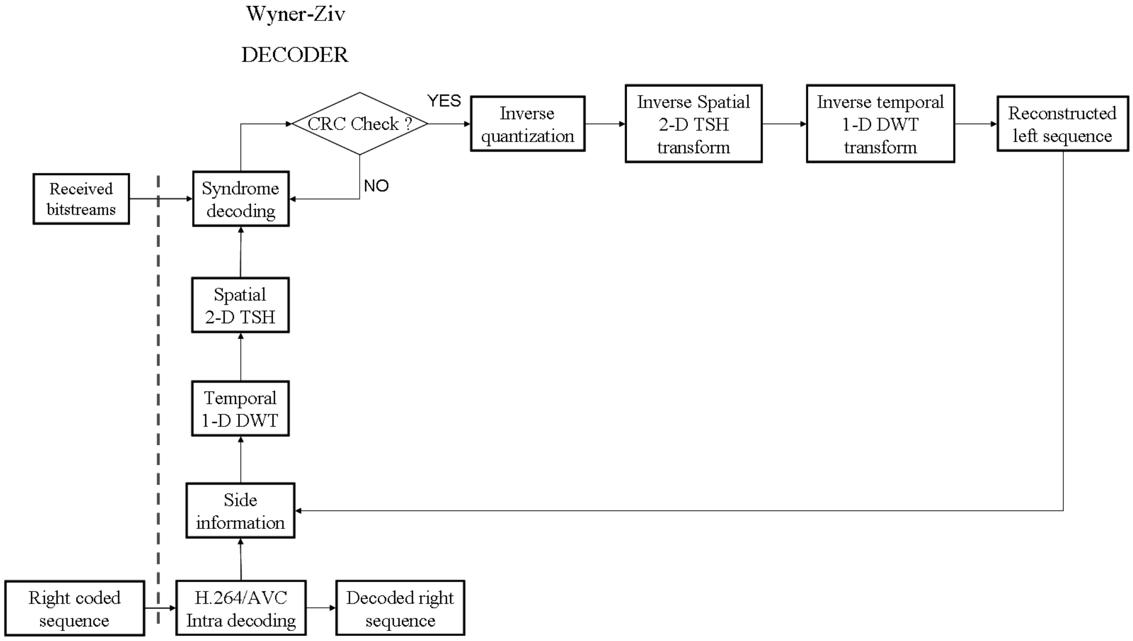


Figure 4.9: Decoding scheme.

$$\begin{aligned}
\hat{c}_{\mathcal{L}_{l_{i,j}}} &= \begin{cases} 0 & c''_{\mathcal{L}_{l_{i,j}}} = 0 \\ sign\left(c''_{\mathcal{L}_{l_{i,j}}}\right)\left(\left|c''_{\mathcal{L}_{l_{i,j}}}\right| + \delta\right)\Delta & c''_{\mathcal{L}_{l_{i,j}}} \neq 0 \end{cases} \\
\hat{c}_{\mathcal{H}_{l_{i,j}}} &= \begin{cases} 0 & c''_{\mathcal{H}_{l_{i,j}}} = 0 \\ sign\left(c''_{\mathcal{H}_{l_{i,j}}}\right)\left(\left|c''_{\mathcal{H}_{l_{i,j}}}\right| + \delta\right)\Delta & c''_{\mathcal{H}_{l_{i,j}}} \neq 0 \end{cases}
\end{aligned} \tag{4.2}$$

4. The inverse spatial-temporal decomposition is performed to reconstruct the estimated left frames $\widehat{n}_l$ and $\widehat{(n+1)}_l$.

4.3.3 Experimental results

In the following the results of the experiments concerning the DIPLODOC 3D ”road stereo sequence” [46] are reported. The TSH transform has been computed using DPV=[100] for the row decomposition and DPV=[150] for the column decomposition. Quantization steps $\Delta = Q_1$ and $\Delta = Q_2$ have been systematically varied.

The more the coefficients are quantized, the fewer quantization levels are available and faster is the syndrome decoder to correctly decode the left stream, thus leading to a lower perceived quality. In order to evaluate the perceptual quality of the estimated frames the Weighted Peak Signal to Noise Ratio (WPSNR) has been computed. This objective quality metric is based on the computation of a Noise Visibility Function (NVF) which depends on a texture masking function. Following [47] a Gaussian model is adopted for estimating the amount of textures in an area of the image. The values of NVF are in the range zero (for extremely textured areas) to one (for smooth areas). The WPSNR can be computed as follows:

$$WPSNR(db) = 10 \log_{10} \left(\frac{L^2}{NVF \times MSE} \right), \tag{4.3}$$

where L represents the maximum luminance value of the images (255 for 8 bit of representation) that are compared and MSE is the Mean Square Error [48].
Figure 4.10 reports the average WPSNR values computed on the whole left sequence versus bit-rate. The numbers in the plot refer to the quantization parameters used. The possibility to adapt the TSH decomposition to the frame content and the possibility to choose a quantization step smaller for the low frequency subband and bigger for the details resulted to be the optimal compromise between computational complexity of the decoder and the quality of the reconstructed frames.

Figure 4.11 shows the original and estimated frames 1_l and 2_l for quantization step $Q_1 = 32$ and $Q_2 = 32$: differences between original and decoded frame can be difficultly noticed, confirming even by visual inspection that the system can efficiently recover the Wyner-Ziv encoded frames.

The reported experimental results have demonstrated that DVC of stereoscopic sequences can be effectively performed with a low computational cost at the transmitter side. The use of the TSH transform with respect to the conventional Haar wavelet allows a finer control of the spatial frequency characteristics of the quantization error. In fact, the TSH decomposition based coder allows to control both quantization step and bandwidth.

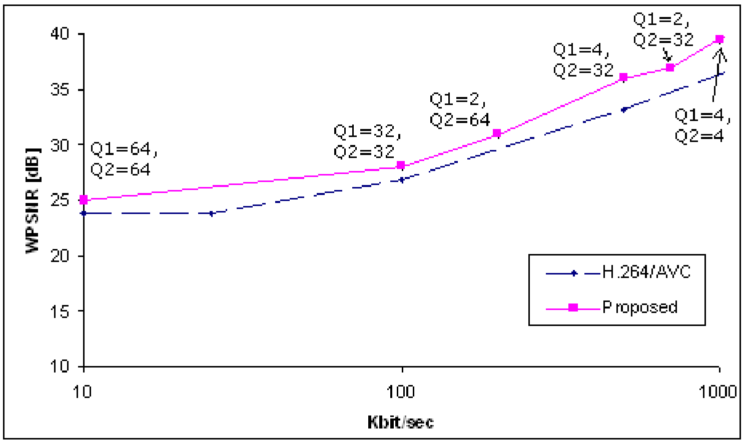


Figure 4.10: Average WPSNR between the original and estimated left frames versus different bit-rates.

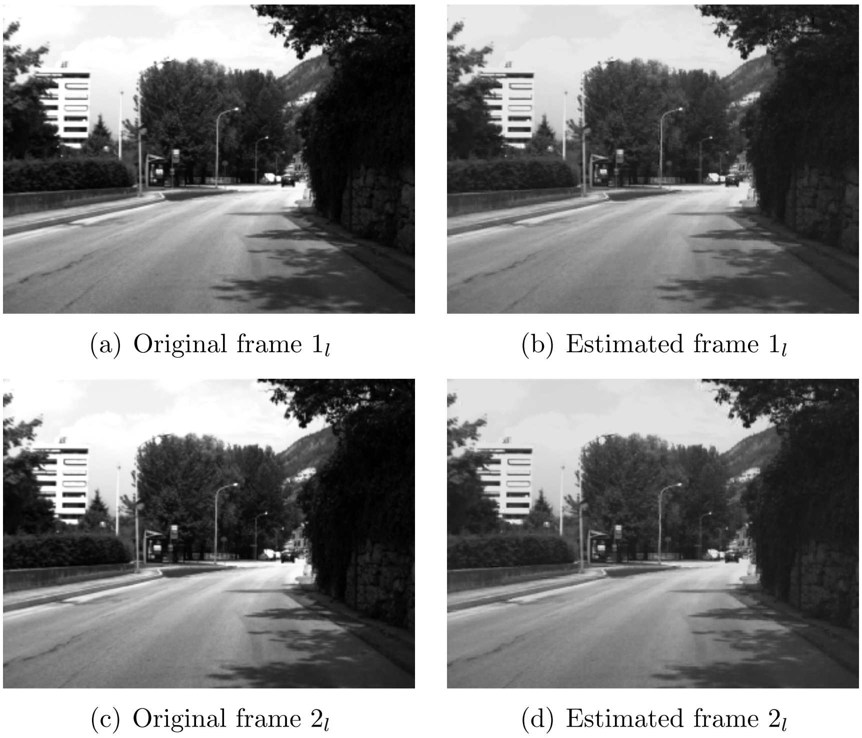
4.4 Turbo code performance with Hybrid TSH and DCT transform

Due to the limited bandwidth available in TETRA2, it is important to find the optimal trade off between the amount of data transmitted and the quality of the decoded stream.

To this aim, the syndrome approach presented in the previous section has been replaced with the use of turbo codes [49] as they allow to send the minimum amount of data while guaranteeing near channel capacity error correcting performance [50].

In the DVC scheme, depicted in Figure 4.12, after a Wyner-Ziv frame is transformed and quantized, it is separated into bit-planes, which are fed one-by-one to a turbo encoder. The turbo encoder consists in a Parallel Concatenation of Recursive Systematic Convolutional Codes (PC-RSC) in addition to a pseudo-random interleaver to spread burst errors. Each RSC encoder produces two output, the systematic bits S_i and the parity bits P_i, where $i = 1, 2$.

The systematic bits of the encoded data are discarded while all generated parity bits are stored at the encoder side in a buffer and transmitted in the decoding phase



(a) Original frame 1_l

(b) Estimated frame 1_l

(c) Original frame 2_l

(d) Estimated frame 2_l

Figure 4.11: Original and estimated frames for quantization step equal to 32 for the subband LL_{L_l} and LL_{H_l} and equal to 32 for the other subbands of the first order TSH decomposition of L_l and H_l.

upon the decoder's request via a feedback channel. As per Figure 4.13, in order to reconstruct the data, the iterative Maximum A Posteriori (MAP) turbo decoder uses the parity bits requested to the encoder and the systematic bits, directly extracted from the side information which can be seen as a corrupted version of the original data. A Laplacian distribution is assumed for the difference between the original data and the side information [51] [37]. The parity bits are requested until they are exhausted or an acceptable probability of symbol error is reached: hence, depending on the accuracy of the systematic bits, additional parity bits are requested, thus leading to an efficient use of the band.

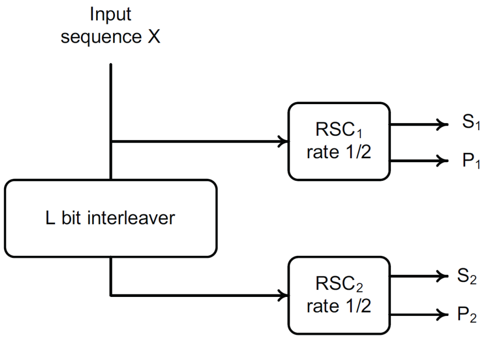


Figure 4.12: Turbo encoder structure in DVC approach [10].

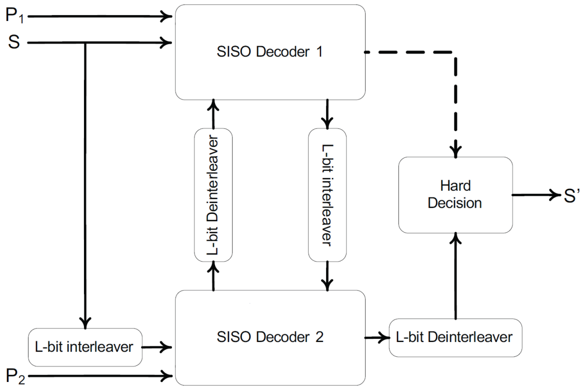


Figure 4.13: Turbo decoder structure in DVC approach [11].

The aim is to design an algorithm where each encoder has high compression performance to minimize transmission costs, low computational complexity to preserve battery life and robustness to avoid effects of channel loss. Traditional video coder as MPEG and H.26x achieve high compression but have high complexity and are sensible to prediction mismatch in case of packet loss. On the other hand MJPEG is robust but has poor compression performance.

Hence, it is proposed a joint turbo code approach, that adopts an hybrid transform domain, which exploits both the TSH and DCT advantages. It has resulted to be the best balance between robustness and quality of decoded data. The combination of DCT and DWT is normally applied in watermarking algorithm to increase robustness of the methods [52] [53] [54] and for compression purposes [55]. In the proposed case:

- the use of TSH transform or alternatively the lifted TSH scheme allows a better adaptation to the frame content. Consequently it is possible to discard the detail subbands while preserving a good quality of the reconstructed data. The amount of data is sensibly reduced.

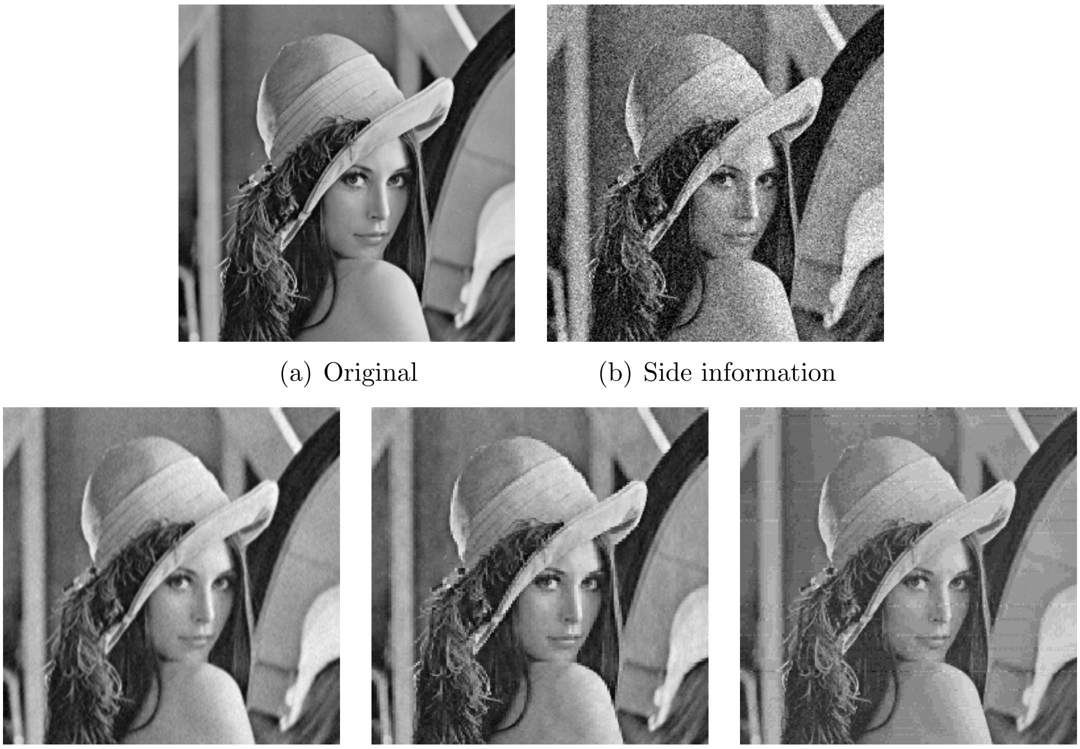
- the use of the DCT guarantees a stronger resiliency to error channels introduced when the Wyner Ziv data stream is transmitted.

As a proof of concept, the image *Lena* of size 256x256 pixels has been decomposed in the (i) DCT domain, (ii) TSH domain by a 1^{st} order decomposition, (iii) TSH domain by a 1^{st} order decomposition followed by DCT decomposition; the bit rates in terms of bpp needed to represent these transformed images have been computed. When the TSH-DCT transforms are both used, the detail subbands of the TSH 1^{st} order decomposition are discarded and only the average subband undergoes the DCT transform; when the inverse TSH is performed, a zero padding on the detail subbands replace the discarded transform coefficients. Table 4.1 reports the bpp computed when WPSNR is fixed to 40dB: a smaller amount of data has to be transmitted when both the transforms are adopted:

Table 4.1: Bpp of DCT, TSH, TSH&DCT versus WPSNR=40dB.

	DCT	TSH DPV=200	TSH DPV=200 DCT of the LL
Bit-rate [bpp]	8	8	4.13

Figure 4.14 shows the turbo decoder performance in the cases depicted above. Side information is a noisy version of the original image when a gaussian noise is applied (Figure 4.14(b)). The parity bits are considered affected by white Gaussian noise, where 8dB is the signal-to-noise ratio per sample. Given the same side information, the use of both the transforms results in a higher quality in terms of WPSNR.



(a) Original

(b) Side information

(c) DCT - WPSNR = 33dB

(d) TSH and DCT of the LL subband - WPSNR =37dB

(e) TSH - WPSNR = 31dB

Figure 4.14: Turbo decoder performance.

Hence, the hybrid transform will be used in the proposed DVC methods since it allows to reduce the amount of data to be transmitted, while guaranteeing a high quality of the reconstructed images.
In the following, several stereo DVC approaches, based on the TSH-DCT transform, are described.

4.5 Stereo DVC for lifted TSH, GOP of variable length and Inter-view Side Information

The system depicted in Figure 4.15 encodes independently the left and right frames of the stereoscopic sequence. The decoder exploits the side information to achieve the best reconstruction of the correlated video streams.

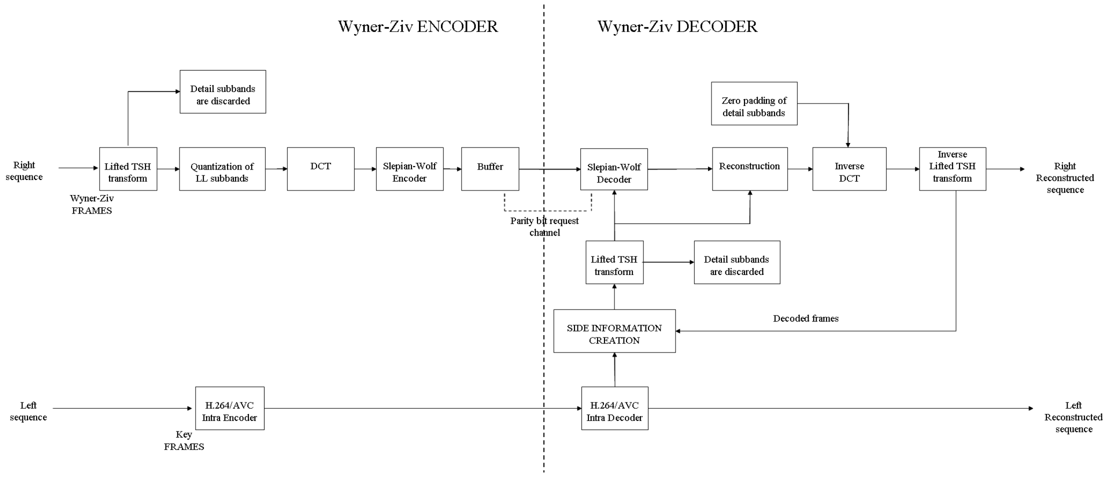


Figure 4.15: Inter-view approach with lifted TSH and GOP of variable length.

4.5.1 Encoding

Let X and Y be the two considered stereo video sequences and let l_i and r_i denote the left and right frames to be processed where $i = 1, 2, ..., F$ and F is the total number of frames of the sequences.
Let a group of frames GOP of k frames be analyzed, i.e. let frames $l_i - l_{i+k-1}$ of the left sequence be considered where l_i is the first frame of the GOP to be processed. Let $r_i - r_{i+k-1}$ denote the corresponding frames of the right sequence.

The left frames constitute the key frames while the right ones are Wyner-Ziv encoded as it is summarized in Fig. 4.15.

The procedure can be described as follows:

1. Given the GOP $r_i - r_{i+k-1}$, a temporal-spatial decomposition is performed. The temporal decomposition is the 1-D TSH of an input signal along the temporal axis, i.e the input signal consists of pixel values that are collected from the same

spatial position in all frames belonging to the same GOP. Assuming each frame has size of N rows and M columns in pixels, then $N \times M$ input signals are 1-D TSH decomposed. The complete decomposition of the GOP $r_i - r_{i+k-1}$ results in one temporal low frequency band and $k-1$ temporal high frequency bands.

2. In order to obtain a better frequency localization, the temporal decomposition is followed by the spatial first order decomposition 2-D TSH as shown in Fig. 4.16. Let $\mathcal{R}_i - \mathcal{R}_{i+k-1}$ denote the $k-1$ 2-D TSH transforms of the temporal subbands. Figure 4.17 shows an example of temporal-spatial decomposition of a GOP $= 7$.

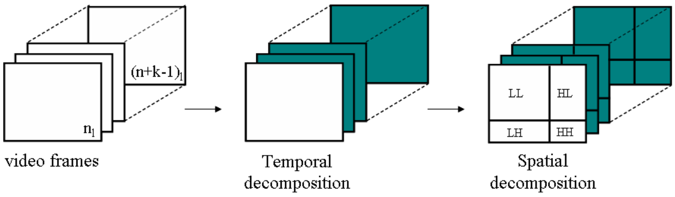


Figure 4.16: Temporal-Spatial decomposition scheme for a GOP beginning with r_i and ending with frame r_{i+k-1} of the right sequence.

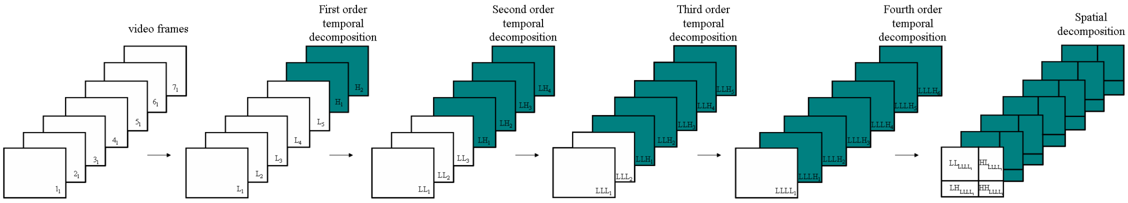


Figure 4.17: Example of Temporal-Spatial TSH decomposition scheme for a GOP of length 7.

3. The coefficients $c_{\mathcal{R}_{s_{i,j}}}$ $(s = i, ..., (i + k - 1))$ belonging to the low frequency subbands of $\mathcal{R}_i - \mathcal{R}_{i+k-1}$ are quantized according to the following formula:

$$c'_{\mathcal{R}_{s_{i,j}}} = sign\left(c_{\mathcal{R}_{s_{i,j}}}\right) \left\lfloor \frac{\left|c_{\mathcal{R}_{s_{i,j}}}\right|}{\Delta} \right\rfloor \tag{4.4}$$

where a quantization step $\Delta = Q_1$ is opportunely chosen.

4. The quantized low frequency subbands of $\mathcal{R}_i - \mathcal{R}_{i+k-1}$ are DCT transformed.

5. The coefficients $c_{\mathcal{R}_{s_{i,j}}}$ $(s = i, ..., (i + k - 1))$ belonging to the detail frequency subbands of $\mathcal{R}_i - \mathcal{R}_{i+k-1}$ are discarded. It is important to underline that this operation allows to sensibly reduce the amount of data that the encoder transmits, further simplifying its complexity.

6. The transmitted bitstream is turbo coded and the parity bits are stored in a buffer.

4.5.2 Decoding

The side information exploits the interview correlation between the view as follow:

1. The disparity field is estimated based on the past decoded GOP - past decoded key frames.

2. The estimated disparity field is applied to the current key frame to estimate the second view's GOP, i.e. the side information SI_{GOP}.

The Wyner-Ziv decoding procedure can be summarized as follows:

1. A temporal-spatial decomposition is performed on the SI_{GOP}.

2. The coefficients belonging to the detail frequency subbands are discarded.

3. The turbo decoder exploits the side information and the received parity bits to recover the Wyner-Ziv encoded low frequency subbands.

4. The reconstructed low frequency subbands are inverse-DCT transformed to form $\widehat{\mathcal{R}_i}$ - $\widehat{\mathcal{R}_{i+k-1}}$.

5. Based on a zero-padding for the detail subbands, the inverse spatial-temporal decomposition is performed to reconstruct the estimated frames $\widehat{r_i}$ - $\widehat{r_{i+k-1}}$.

4.6 Stereo DVC for GOP of variable length and Intra-view Side Information

If it is assumed that the two cameras do not communicate to each other, it is possible to adapt DVC monoview approaches to stereo DVC systems. In this case each camera exploits the temporal correlation of the acquired sequence.

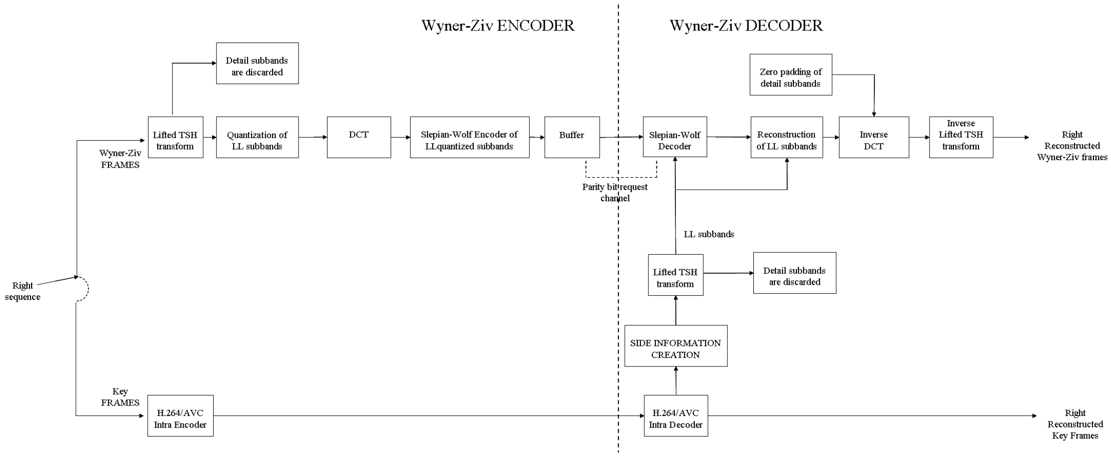

Wyner-Ziv ENCODER
Wyner-Ziv DECODER
Detail subbands are discarded
Zero padding of detail subbands
Lifted TSH transform
Quantization of LL subbands
DCT
Slepian-Wolf Encoder of LLquantized subbands
Buffer
Slepian-Wolf Decoder
Reconstruction of LL subbands
Inverse DCT
Inverse Lifted TSH transform
Right Reconstructed Wyner-Ziv frames
Wyner-Ziv FRAMES
Parity bit request channel
LL subbands
Lifted TSH transform
Detail subbands are discarded
Right sequence
SIDE INFORMATION CREATION
Key FRAMES
H.264/AVC Intra Encoder
H.264/AVC Intra Decoder
Right Reconstructed Key Frames

Figure 4.18: Intra-view approach with lifted TSH and GOP of variable length.

Figure 4.18 shows the procedure that each view undergoes: as in previous section some frames of the sequence are Wyner-Ziv encoded, but key frames belong in this case to the same view. Hence, when exploiting only intraview correlation, the side information is created through a motion compensation interpolation algorithm between the key frames.

The use of a lifted scheme, although leads to a better frequency localization, does not allow to update frame by frame the side information, as the latter is created using only two consecutive key frames. Hence, this scheme has been compared with the one depicted in Figure 4.19 where TSH transform has been adopted, thus allowing to estimate the side information through a motion compensation interpolation algorithm between the past decoded Wyner-Ziv frame and the next key frame.

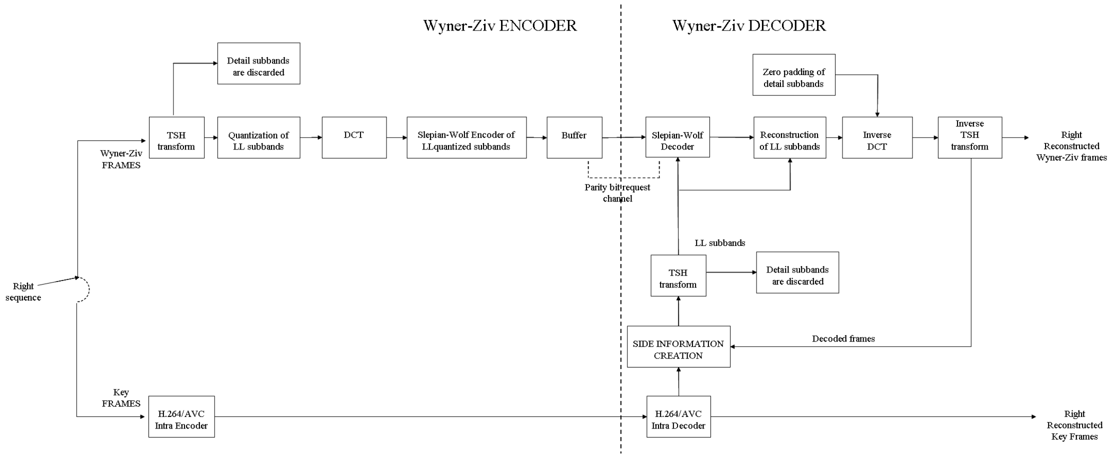

Wyner-Ziv ENCODER
Wyner-Ziv DECODER
Detail subbands are discarded
Zero padding of detail subbands
Wyner-Ziv FRAMES
TSH transform
Quantization of LL subbands
DCT
Slepian-Wolf Encoder of LLquantized subbands
Buffer
Slepian-Wolf Decoder
Reconstruction of LL subbands
Inverse DCT
Inverse TSH transform
Right Reconstructed Wyner-Ziv frames
Parity bit request channel
LL subbands
TSH transform
Detail subbands are discarded
Right sequence
SIDE INFORMATION CREATION
Decoded frames
Key FRAMES
H.264/AVC Intra Encoder
H.264/AVC Intra Decoder
Right Reconstructed Key Frames

Figure 4.19: Intra-view approach with TSH and GOP of variable length.

4.7 Stereo DVC for TSH with residual coding, GOP of variable length and Hybrid Side Information

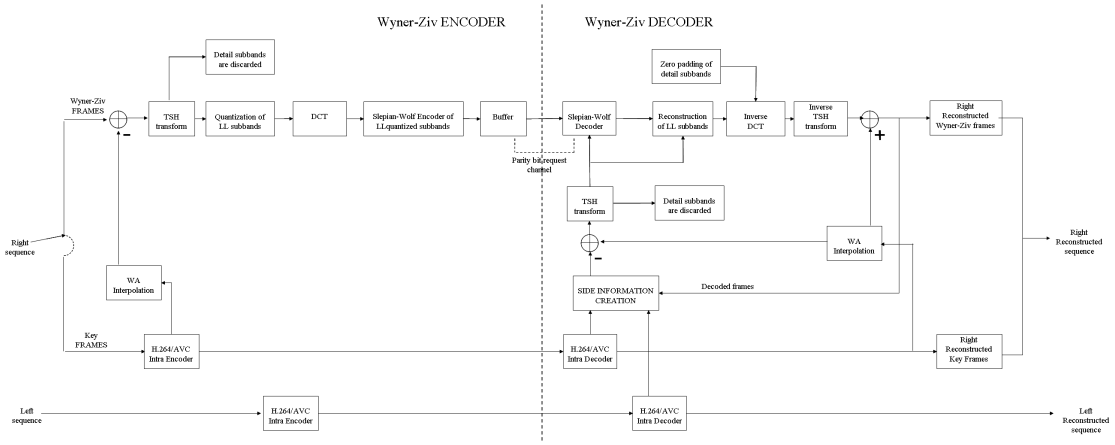

Wyner-Ziv ENCODER
Wyner-Ziv DECODER
Detail subbands are discarded
Zero padding of detail subbands
Wyner-Ziv FRAMES
-
TSH transform
Quantization of LL subbands
DCT
Slepian-Wolf Encoder of LLquantized subbands
Buffer
Slepian-Wolf Decoder
Reconstruction of LL subbands
Inverse DCT
Inverse TSH transform
+
Right Reconstructed Wyner-Ziv frames
Parity bit request channel
TSH transform
Detail subbands are discarded
Right sequence
WA Interpolation
-
WA Interpolation
Right Reconstructed sequence
SIDE INFORMATION CREATION
Decoded frames
Key FRAMES
H.264/AVC Intra Encoder
H.264/AVC Intra Decoder
Right Reconstructed Key Frames
Left sequence
H.264/AVC Intra Encoder
H.264/AVC Intra Decoder
Left Reconstructed sequence

Figure 4.20: Residual coding approach with TSH and GOP of variable length.

The system depicted in Figure 4.20 encodes independently the left and right frames of the stereoscopic sequence. The decoder exploits the side information to achieve the best reconstruction of the correlated video streams.

4.7.1 Encoding

Let X and Y be the two considered stereo video sequences and let l_i and r_i denote the left and right frames to be processed where $i = 1, 2, ..., F$ and F is the total number of frames of the sequences.
Let a group of frames GOP of k frames be analyzed, i.e. let frames $l_i - l_{i+k-1}$ of the left sequence be considered where l_i is the first frame of the GOP to be processed. Let $r_i - r_{i+k-1}$ denote the corresponding frames of the right sequence.

The left frames are H.264 Intra Encoded and are fully available at the decoder while the right sequence is divided in key frames and Wyner-Ziv frames. The overall process is summarized in Fig. 4.20.

The procedure can be described as follows:

1. The reference frame for the current Wyner-Ziv frame r_j, $(j = i, ..., i + k - 1)$, is computed through a weighted average (WA) interpolation as:

$$\begin{array}{l} W_{ref_j} = \alpha_j \cdot r_i + \beta_j \cdot r_{i+k} \\ \alpha_j = 1 - \frac{d_j}{F} \end{array} \tag{4.5}$$

   where r_i and r_{i+k} are the previous and next decoded key frames of the right sequence; d_j is the number of frames between the current frame r_j and the previous key frame r_i within the same GOP.

2. The residual frame $D_j = r_j - W_{ref_j}$ is TSH decomposed. Let $\mathcal{D}_j$ denote the transformed image.

3. The coefficients $c_{\mathcal{D}_{j_{i,j}}}$ belonging to the low frequency subbands of $\mathcal{D}_j$ are quantized according to the following formula:

$$c'_{\mathcal{D}_{j_{i,j}}} = sign\left(c_{\mathcal{D}_{j_{i,j}}}\right) \left\lfloor \frac{\left|c_{\mathcal{D}_{j_{i,j}}}\right|}{\Delta} \right\rfloor \tag{4.6}$$

   where a quantization step $\Delta = Q_1$ is opportunely chosen.

4. The quantized low frequency subbands is DCT transformed.

5. The coefficients $c_{\mathcal{D}_{j_{i,j}}}$ belonging to the detail frequency subbands are discarded. As in previous method, this operation allows to sensibly reduce the amount of data that the encoder transmit, further simplifying its complexity.

6. The transmitted bitstream is turbo coded and the parity bits are stored in a buffer.

4.7.2 Decoding

At the decoder, the side information exploits the interview and the intraview [56] correlation between the view as follow:

1. Temporal estimation r_{jT} of Wyner-Ziv frame r_j is computed by motion compensated temporal interpolation between previously decoded frame r_{j-1} and next key frame r_{i+k};

2. The disparity field, estimated by the past decoded frames r_{j-1} and l_{j-1}, is applied to l_j to spatially estimate the corresponding frame of the right view, named r_{jS}.

3. Frame $l_j^{'}$ is estimated by motion compensated temporal interpolation between l_{j-1} and l_{i+k};

4. Frame $l_j^{'}$ is subtracted by l_j to provide the exact prediction error; this error is thresholded to obtain a binary reliable mask M, that is not sequence-dependent.

5. SI_j is formed as follows: for each pixel, if the mask indicates the pixel as reliable, the intraview side information r_{jS} is used while if the mask indicates the pixel as unreliable, the interview side information r_{jT} is used.

The Wyner-Ziv decoding procedure can be summarized as follows:

1. The residual frame $E_j = SI_j - W_{ref_j}$ is TSH decomposed.

2. The coefficients $c_{\mathcal{E}_{j_{i,j}}}$ belonging to the detail frequency subbands are discarded.
3. The turbo decoder exploits the side information and the received parity bits to recover the Wyner-Ziv encoded low frequency subbands that are inverse-DCT transformed.
4. Based on a zero-padding for the detail subbands, the inverse spatial-temporal decomposition is performed to reconstruct the estimated frames $D_j^{'}$.
5. The recovered Wyner-Ziv frame is computed as $D_j^{'} + W_{ref_j}^{'} = r_j^{'}$

As residual approach leads to an additional computation complexity of frame store and frame subtraction for the encoder, this thesis mainly focuses on the corresponding approach when no residual frame is used as reference frame. This method is shown in Figure 4.21. In the following this method is labeled "hybrid" to differentiate it from the residual coding above described.

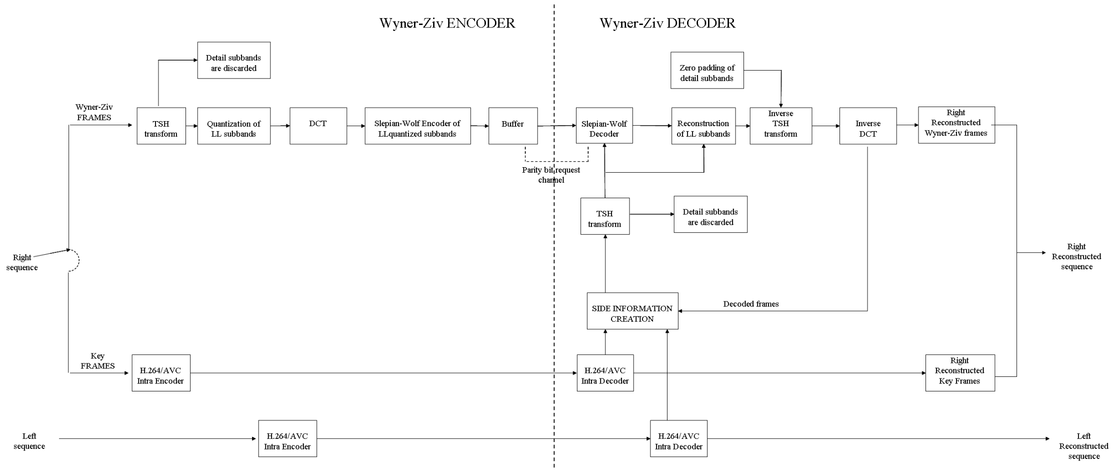

Wyner-Ziv ENCODER
Wyner-Ziv DECODER
Detail subbands are discarded
Zero padding of detail subbands
Wyner-Ziv FRAMES
TSH transform
Quantization of LL subbands
DCT
Slepian-Wolf Encoder of LLquantized subbands
Buffer
Slepian-Wolf Decoder
Reconstruction of LL subbands
Inverse TSH transform
Inverse DCT
Right Reconstructed Wyner-Ziv frames
Parity bit request channel
TSH transform
Detail subbands are discarded
Right sequence
Right Reconstructed sequence
SIDE INFORMATION CREATION
Decoded frames
Key FRAMES
H.264/AVC Intra Encoder
H.264/AVC Intra Decoder
Right Reconstructed Key Frames
Left sequence
H.264/AVC Intra Encoder
H.264/AVC Intra Decoder
Left Reconstructed sequence

Figure 4.21: Hybrid approach with TSH and GOP of variable length.

4.8 Stereo quality evaluation

When dealing with stereoscopic 3D video, the perceived artifacts produce not only visually unpleased results, but also general discomfort on the human visual system.

Due to these reasons, the scientific community is focusing on the definition of a perceptual quality metric that quantifies the typical distortion that could occur. At this aim it is important to identify classes of artifacts which could arise in several scenarios involving stereoscopic content. It is important to underline that visual artifacts could arise at any processing and delivery phase of a stereo video sequence [57] [58]:

- Acquisition and content creation: most un-natural effects in this step derive from an incorrect configuration, calibration or positioning of the camera system. Typical artifacts introduced in this phase are keystone distortion, temporal mismatch and cardboard effect.

- Representation of the acquired data: in particular, 2D to 3D format conversion can cause dense depth video, artifacts as ghosting by disocclusion and temporal and spatial aliasing.

- Encoding phase: stereoscopic color and depth video are normally encoded with multi-view coding schemes or algorithms for 2D video adapted for stereo. This step can alter image details that are important for depth perception.

- Transmission: on one side, packet data loss and channel noise can be sources of a degraded perceived quality of the content; on the other side, the algorithms that attempt to correct these errors, can cause additional problems on their own. Impairments are due to packet loss, jitter and color bleeding.

- Visualization: stereo video quality is strongly dependent on the adopted approach to 3D visualization, i.e. on the artifacts that characterize the 3D displays. Flickering, cross-talk, puppet theater effect and shear distortion can occur at this step.

Considering that, from a quality of experience point of view, the attention has been focused on conventional video coders, stereo video artifacts introduced by stereo distributed video coders available in the literature have been analyzed by means of subjective experiments.

4.8.1 Artifacts introduced in stereo video coding

The presence of artifacts related to the encoding of a stereoscopic video sequence depends hardly on the used algorithms and how the decoders copes with the channel errors [59][60].
In conventional 2-D video coding, the introduced monoscopic artifacts comprise all the typical artifacts of 2D images as blurring, noise, blocking and other structural changes. In the viewing of a stereo video, the final user could recognize "2D artifacts" but still having a perfect perception of the depth; obviously, larger distortion could damage the binocular view.
On the other side, the stereoscopic artifacts change the relation between the two views and thus forbid the brain to have the proper binocular depth view. Such artifacts can change the disparity information of a scene or cause any other structural changes. Other annoying effects could convey unnatural information to the brain inducing so eye-strain and visual discomfort [61].
When dealing with stereo video content artifacts, 4 groups of impairments based on how they are perceived by human brain have to be considered: *structure, color, motion* and *binocular* [58]. Structure includes those distortions that can impact on structural changes (i.e. contours and texture); motion and color, those that can affect motion and color vision. Finally, binocular impairments can degrade the binocular depth perception when perceived as a stereo-pair (cannot be noted with a single eye). Based on this classification, the attention has been focalized on artifacts introduced in the phase of coding.
In traditional and stereoscopic video coding, quantifying the artifacts in terms of the visual impact is a difficult task. In fact, the perceived distortion is not only related to the absolute quantization error but it is depending on local, global spatial inter-view and temporal characteristics of the video sequences [62]. Consequently, it is not possible to provide a specific bit-rate at which anyone artifacts is showed. Hence, the

discussion will take into account the descriptions of some possible artifacts as *blocking effect, blurring, ringing, staircase effects* and *mosaic patterns* for the category **structure**; *color bleeding* for the category **color**; *motion compensation artifacts* and *mosquito effect* for the category **motion** and *cross-distortion, cardboard effect* for the category **binocular**.

The *blocking* effect is a discontinuity between consecutive blocks in a picture. The severity of blocking effect is subject to the coarseness of the quantization of DCT coefficients of either one or both adjacent.

The *blurring* manifests as a lack of spatial details in moderate to high spatial activity regions of pictures, such as in roughly textured areas or around scene object edges. For intra-frame coded macro blocks, blurring is directly related to the suppression of the higher order AC DCT coefficients through coarse quantization, representing the content of a block only through lower order coefficients.

The *ringing* effect is most evident along high contrast edges in areas of generally smooth texture in the reconstruction, and appears as a shimmering or rippling outwards from the edge up to the encompassing block's boundary. The higher the contrast of the edge, the greater the level of the peaks and troughs of the rippling.

The *staircase* effect is linked to both the blocking and mosaic pattern effects in terms of the manifestation of discontinuities between adjacent blocks. When a diagonal edge is represented within a string of consecutive blocks, the consequence of coarse quantization is the reconstruction of the diagonal edge as a number of horizontal or vertical steps. It produces staircase edges.

The *mosaic pattern* effect manifests as the apparent mismatch between all, or part, of the contents of adjacent blocks; this has a similar effect to using visually ill-fitting square tiles in a mosaic. This may mean a block with a certain contour or texture dissimilar to the neighboring blocks, or a block used in the representation of an object which does not blend satisfactorily with the other constituent blocks.

The *color bleeding* is due to a coarse quantization of high frequency chrominance coefficients. It results in the representation of the chrominance components with only the lower frequency coefficients.

The *mosquito* effect is a temporal artifact seen mainly in smoothly textured regions as a fluctuations of luminance/chrominance levels around high contrast edges, or moving objects, in a video sequence. This effect is related to the high frequency distortions introduced by both the ringing effect, and the prediction error produced by the motion compensated mismatch artifacts.

Cross-distortion is caused by asymmetrical video coding in both temporal or spatial domains. In the case of temporal asymmetric video coding, one channel has lower frame-rate than the other; otherwise, in case of spatial asymmetric video coding one channel has lower resolution than the other.

The *cardboard* effect is typically caused by image acquisition or compression parameters resulting in a coarse quantization of the disparity or depth maps. Due to it, the objects appear flat as if the scene is divided into discrete depth planes. The flattening of the objects in a scene evokes an unnatural depth percept.

More unusual coding artifacts that have impact on depth perception are *depth bleeding* and *depth smoothing*. Depth bleeding affects the depth channel, it is similar to color bleeding. Depth smoothing is due to asymmetric compression or resolution of the depth channel.

4.8.2 Quality metrics

As illustrated in the previous sections, stereo video stream can be subject to several distortions during the the capturing, representation, coding, transmission or visualization steps. Any of these phases may result in a degradation of visual quality. At the aim of evaluating the quality of a stereo video content, objective and subjective quality assessments can be performed. The goal of objective video quality assessments is to develop a quantitative measure that can automatically predict perceived video quality.

In the following, some objective and subjective quality metrics used for evaluation of video quality will be discussed. These metrics are generally used for modeling 2D video quality but since there are not specific quality metrics for stereoscopic video, the conventional quality metrics, PSNR and VQM, will be adopted for objective evaluation; at the contrary, the mean opinion score, MOS, will be analyzed for subjective assessments [63].

A useful and often used metric is Peak Signal-to-Noise Ratio, PSNR. Given two images x and y, it is calculated as:

$$PSNR = 10 \cdot \log_{10} \frac{MaxErr^2 \cdot w \cdot h}{\sum\limits_{i=0,j=0}^{w,h} (x_{i,j} - y_{i,j})^2} \tag{4.7}$$

where *MaxErr* is the maximum absolute value of color components difference, w is the video width and h represents the video height. Generally, this metric is equivalent to Mean Square Error, but it is more convenient because of the use of a logarithmic scale. This metric is appealing because it is simple to calculate and has clear physical meaning. However, it does not match very well the perceived visual quality. Along with this metric, it is important to consider other objective metrics as VQM.

The Video Quality Metric adopted here, is derived by Watson's DCT-based metric (DVQ) [64] [65] video quality evaluation. Only the luminance of the video sequence is considered; the image is then decomposed in the DCT domain so that input images are split into different spatial frequency components. Each DCT coefficients is converted to local contrast; this operation converts each DCT coefficients to a number between $[-1, 1]$ that represents the magnitude of the corresponding basis function as a fraction of the average luminance of that block. Then a human Spatial Contrast Sensitivity Function (SCSF) matrix is applied. Finally, a measure of visual error is performed via Minkowski metric to define the mean and the maximum distortion, then VQM is calculated as:

$$VQM = (Mean\_dist + 0.005 \cdot Max\_dist) \tag{4.8}$$

whereas $Mean\_dist$ represents the mean distortion and Max_dist represents the maximum distortion and 0.005 is a weight parameter chosen on several primitive psychophysics experiments.

It is clear that it may not be possible to fully characterize system performance by objective means; consequently, it is necessary to supplement objective measurements with subjective measurements. In this case, the mean opinion score (MOS) provides a numerical indication of the perceived quality of received stereo video content after compression and/or transmission. The MOS is expressed as a single number that can range from 1 to 5 or from 1 to 100. The lowest number represents the worst case while the biggest number corresponds to the highest perceived video quality. MOS tests for video are specified by ITU-R BT.500-11 [2].
The MOS is generated by averaging the results of a set of standard, subjective tests where a number of viewers rate the viewed video quality of test sequences.

4.9 Experimental results

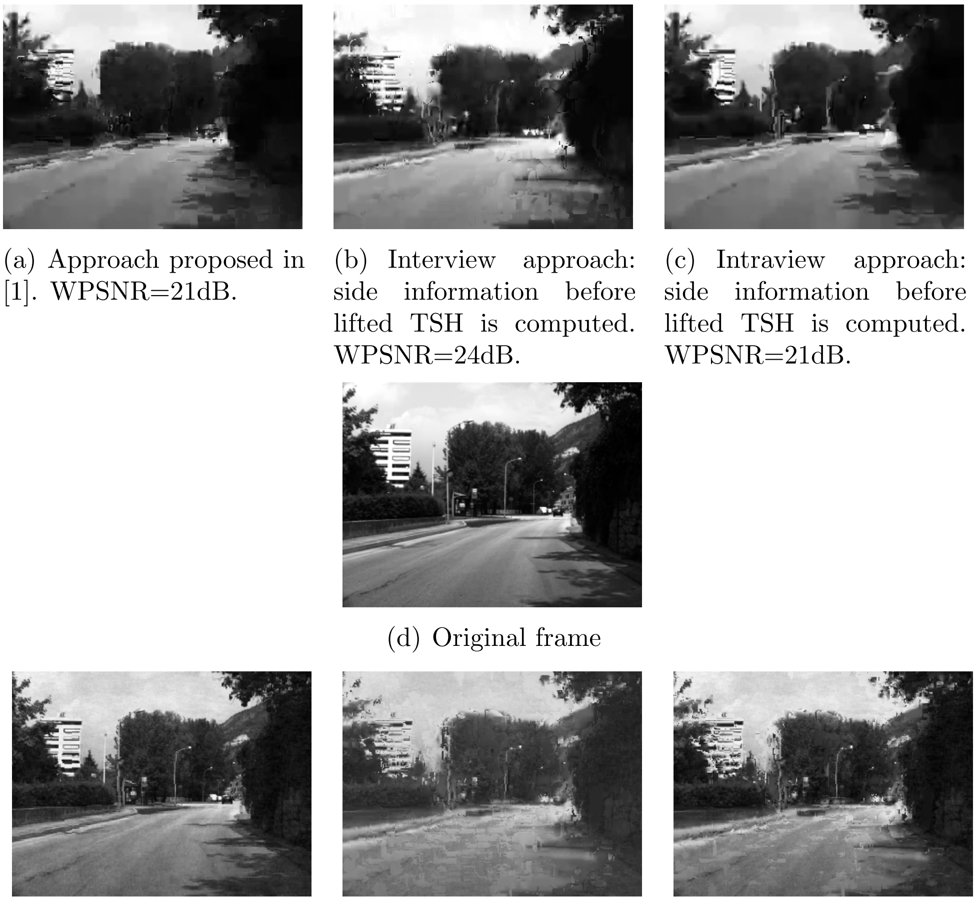
In the following, the shown results refer to the DIPLODOC 3D "road stereo sequence" [46], that is 240x320 pixels and 15 frames per second; 90 frames were used for the sequence. TSH transform has been carried out using DPV=[32] for the row decomposition and DPV=[80] for the column decomposition.
In the DVC architecture, a GOP equal to 9 has been analyzed: one key frame is followed by 8 Wyner-Ziv ones; in the interview and hybrid approaches the right view has been Wyner-Ziv coded while the left view is coded with a conventional H.264/AVC. On the contrary, both the views are Wyner-Ziv encoded when the intraview approach is exploited. As usual for Wyner-Ziv coding, only luminance data has been coded; the total bit-rate includes the luminance rate for the Wyner-Ziv frames and key frames for the right view to be coded since the left view is always the same.

The algorithm adopted for computing the disparity map is inspired to [66]. Initially an estimation of the disparity at each pixel in the image is computed by choosing a reference image and sliding the other image across it. In this step, the intensity values and the gradient information (spatial derivatives) of the images are subtracted:

the combinations of both the factors gives better accuracy, especially on surfaces with texture. Hence, far-away objects go dark (meaning they line up in the two images) at different times than close-up objects. The offset when the difference is the smallest as well as the value of the difference are recorded.
This slide-and-subtract operation is carried out from right-to-left (R-L) and left-to-right (L-R). Then bad pixels are removed in two ways: (i) disparity from the R-L step or the L-R step is used depending on which has the lowest matching difference; (ii) all points where the R-L disparity is significantly different from the L-R disparity are considered bad estimation.
The image information is then combined with the pixel disparities to get a cleaner disparity map: after the reference image is segmented by a "Mean Shift Segmentation" technique (image is broken into "tiles" of constant color), thus resulting in a very "blocky" version of the original image, then the associated pixel disparities is analyzed and each segment is assigned to have the median disparity of all the pixels within that segment.
Given the reference frame and the disparity map, the image corresponding to the second view can be estimated.
On the contrary motion estimation has been carried out according to [67] as it requires low computational complexity.

The stereo DVC methods, described in previous sections, have been evaluated in terms of side information (SI) quality: Figure 4.22 compares SI accuracy for the right frame number 5, encoded at 100 Kbit/sec. Frame 5 has been chosen as it represents the most unpredictable frame within a GOP of length 9 due to its position; in fact it is as far from the previous key frame as from the next one.
Given that SI intraview is more reliable than SI interview due to occlusions, SI computed for Intraview approach, where lifted TSH is adopted, results to have a lower quality than the corresponding scheme where TSH only is used, because the method

does not allow an update of the SI frame by frame; this means that the SI corresponding to a GOP is computed all at once from the key frames and undergoes a lifted TSH transform. Hence, although lifted schemes are usually preferred because they allow a better frequency localization, lower quality is due to the propagation of errors. On the contrary, the use of THS only leads to a more accurate SI as this is computed by motion compensation interpolation between the last decoded key frame and the next key frame.



(a) Approach proposed in [1]. WPSNR=21dB.

(b) Interview approach: side information before lifted TSH is computed. WPSNR=24dB.

(c) Intraview approach: side information before lifted TSH is computed. WPSNR=21dB.

(d) Original frame

(e) Intraview approach: side information before TSH is computed. WPSNR=25dB

(f) Residual approach: side information before reference frame is subtracted. WPSNR=24dB

(g) Hybrid approach: side information before TSH is computed. WPSNR=26dB

Figure 4.22: Side information comparison.

WPSNR corresponding to SI of residual coding has a lower value when compared to the proposed hybrid scheme due to the operation of discarding the details of the TSH decomposition of the SI. In fact, discarding the high frequency subband of the residual frame has a higher visual impact in reconstruction when compared to the discarding of the same frequency band of the frame itself; as the SI is formed via an iterative procedure, the SI results to be affected by the propagation errors of the reconstructed Wyner-Ziv frames.

The proposed hybrid approach results also in a higher quality of the SI when compared with the delayed mask of [1].

Figure 4.23 shows the mask used to generate the SI in the hybrid scheme: white areas indicate that interview/spatial correlation is used, while black pixels indicate that temporal/intraview information is exploited. It is important to underline that interview information is exploited in high motion areas while intraview information is exploited in object area where occlusions are due to the movement.

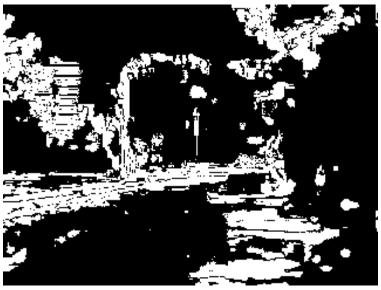


Figure 4.23: SI generation mask related to hybrid approach.

As illustrated in previous section, subjective evaluation testing is used to measure the effect of distributed video coding artifacts on the perceived quality of the reconstructed stereoscopic sequence. The obtained results have also been compared with the quality evaluated by using three 2-D video objective quality models namely PSNR and VQM [63].

16 non-expert observers (8 males and 8 females) participated in the experiments

and were asked to rate the video sequence perceived quality according to the method proposed in [2]: the subjective ratings for the coded stereoscopic sequences have been scaled into a linear opinion score scale, which ranges from 0 (bad quality) to 100 (excellent quality). The observers were also asked to evaluate the kind of annoying visual artifacts for each shown sequence.

The bit-rate of the 3-D coded sequence have been systematically varied and 4 reference bit-rates were considered: 10, 25, 100, 500 Kbit/sec, i.e. ranging from the low bit-rate transmission case nowadays available in TETRA1 to the higher bit-rate that TEDS will offer. The goal is to analyze whether DVC approach is more suitable than conventional H.264/AVC coder for data transmission over TETRA2 channel. For each reference bit rate, the H.264/AVC 3-D coded sequence, the DVC 3-D coded according to [1] and the DVC 3-D stereo sequence encoded with the proposed hybrid scheme have been considered.

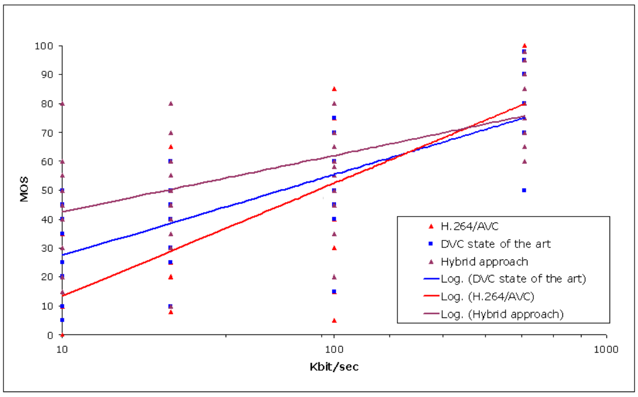


Figure 4.24: MOS scores for perceived stereo video quality.

The stimulus set contains 12 coded sequences and the original, uncompressed sequence is used as the reference in the evaluation test. The set is randomized and presented sequentially.

Figure 4.24 shows MOS scores for the overall perceived quality, while Figures 4.25 and Figure 4.26 show the rate distortion (RD) performance for the analyzed Wyner-Ziv stereo coding architectures, by respectively considering the PSNR and VQM quality models. These metrics have been adopted to evaluate the quality of the decoded 2-D right sequence.
It is important to underline that, in order to simulate TETRA2 channel error, the parity bits are affected by white Gaussian noise before turbo decoding is carried out, where 8dB is the signal-to-noise ratio per sample.

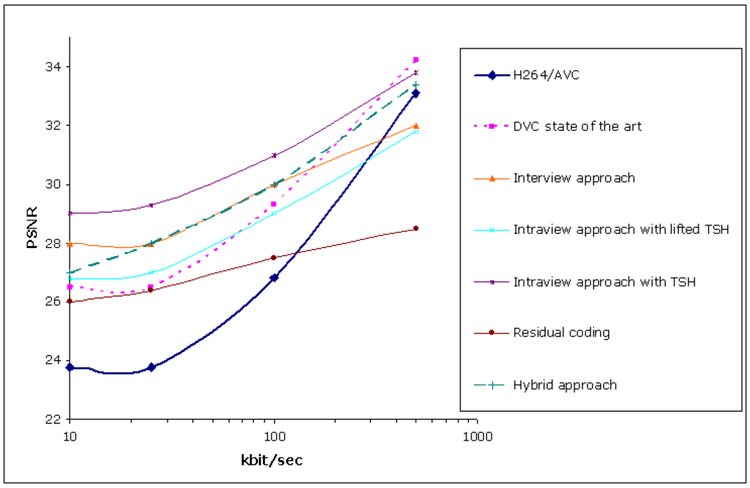


Figure 4.25: RD performance by PSNR evaluation. PSNR is averaged on the whole right sequence.

A number of interesting conclusion can be drawn. Given that DVC schemes are more suitable for low bit-rate channels because less amount of data need to be transmitted, Figures 4.24, 4.25 and 4.26 supports the above statement from a visual quality point of view. On the contrary, the reversal of the trend at about 400 Kbit/sec for the proposed hybrid scheme shows that a conventional H.264/AVC coder results to be more appropriate at high bit-rate even if DVC approach would be still preferred in some cases due to the advantage of low-complexity encoders. The reversal of the trend for [1] is at a higher bit rate compared with the proposed hybrid scheme due

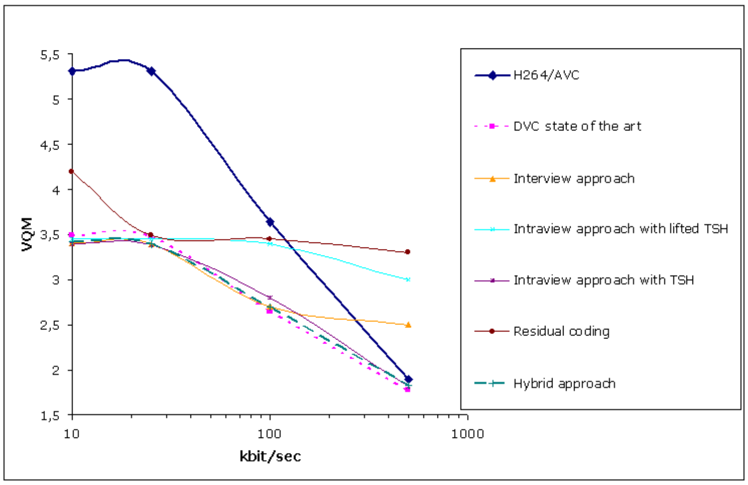


Figure 4.26: RD performance by VQM evaluation. VQM is averaged on the whole right sequence.

to the lack of details in the reconstruction stage of the proposed method that is noticeable at about 400 Kbit/sec and not at lower bit rates.

As already stated, the residual coding has low performance when compared with hybrid scheme because of the discarding of detail subbands: the zero-padding of the TSH detail subbands of the frames in the hybrid scheme guarantees a high quality of the reconstructed frames; on the contrary, if some packets of the residual frames are lost or discarded, the effect is in a lower quality. It is important to note that the residual has the additional disadvantage of higher encoder complexity of frame store and frame subtraction when compared with hybrid scheme.

Hence, for bit rates allowed by TETRA2, i.e. lower than 400 Kbit/sec, the proposed hybrid scheme gives better results when compared with [1] and with conventional coder H.264/AVC from an objective point of view and also with regard to the perceived quality. In order to highlight the above statement, Figure 4.27 shows the histogram of the MOS scores, while Table 4.2 shows that MOS values corresponding to the hybrid approach are higher than the corresponding values computed for state of the art approach and for H.264/AVC coding.

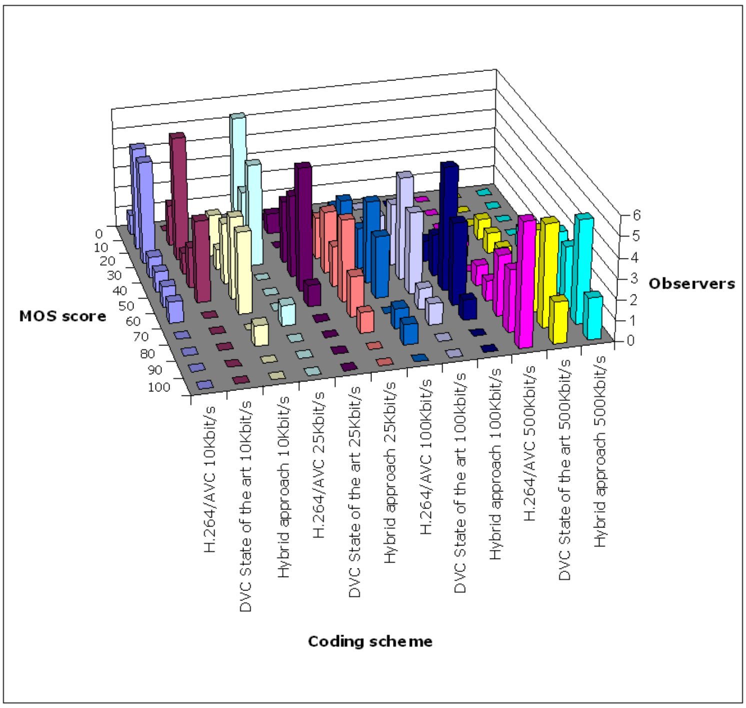


Figure 4.27: MOS scores histogram.

The advantages of the proposed hybrid approach can be summarized as:

- A GOP of variable length can be used while standard algorithms normally limit its length to 2. Hence, while in classical case 1 key frame is followed by 1 Wyner-Ziv frame, in the proposed system 1 key frame, encoded with high computational complexity H.264/AVC coder, is followed by variable N Wyner-Ziv frames, encoded with low computational complexity, where $N + 1$ is the total length of the GOP. This further reduces the complexity of the encoder.

- The use of a joint TSH-DCT transform domain is proposed. The TSH decomposition is exploited due to the possibility to vary the decomposition and the subbands size according to the content of the frame. In this way it is possible to discard the detail subbands, thus considerably reducing the amount of data to be transmitted while guaranteeing an acceptable quality of the reconstructed

Table 4.2: MOS values.

Method	MOS
H.264/AVC 10Kbit/s	21
DVC State of the art 10Kbit/s	28
Hybrid approach 10Kbit/s	46
H.264/AVC 25Kbit/s	22
DVC State of the art 25Kbit/s	41
Hybrid approach 25Kbit/s	48
H.264/AVC 100Kbit/s	45
DVC State of the art 100Kbit/s	50
Hybrid approach 100Kbit/s	57
H.264/AVC 500Kbit/s	85
DVC State of the art 500Kbit/s	77
Hybrid approach 500Kbit/s	78

frame. On the other hand, DCT is applied to increase the robustness of the algorithm over the error-prone TETRA2 channel. This combination of the two transforms allows to further reduce the complexity of the encoder as only the low frequency subband is transmitted.

- No additional computational complexity is given to the encoder, when the proposed method is compared with residual coding approach.
- The SI exploits efficiently either spatial and temporal correlation leading to an accurate SI.

A *flickering* effect has been noticed on the DVC coded sequence according to [1] only at low bit-rate due to the alternate of low-quality decoded H.264/AVC key frames and higher quality decoded Wyner-Ziv frames. Obviously this effect is not anymore noticeable at high bit-rate when the quality of the decoded key frames and the quality of the decoded Wyner-Ziv frames becomes similar.

In order to avoid the flickering effect, a motion compensation temporal interpolation has been performed to refine the key frames once the Wyner-Ziv frames are decoded. Hence, the observer did not notice the flickering when asked to evaluate the quality of the hybrid scheme but perceived a block artifact in the sequence due to the motion compensation interpolation algorithm that has been used.

For the reasons described above, the *blocking* effect typical of a conventional H.264/AVC at low bit-rate was less noticeable in the DVC decoded sequence. In fact this effect only affects the key frames of the right sequence and the left sequence. A symmetric approach where both the views have a limited amount of key frames would probably not let the observer perceive a blocking effect.

A *blurring* effect has been also noticed on the DVC coded sequence due to the particular Wyner-Ziv coder that is used. In the analyzed scheme, a turbo decoder with puncturing rate equal to 1/3 has been used. Note that the turbo decoder performance is strictly dependent on the amount of parity bit planes used to reconstruct the Wyner-Ziv frames. The adopted turbo decoder was always able to reconstruct the Wyner-Ziv frames even if in very uniform areas to be decoded a blurring effect was noticed by few careful observers. This effect is to be related not only with the used Wyner-Ziv decoder but also with the accuracy of the SI: this is the actual challenge in the most recent stereo DVC approaches that have to exploit the temporal and inter-view correlation.

The combination of the above effects let few participants note a *jerkiness*-like effect for low bit-rate cases: this corresponds to the perception of originally continuous motion as a sequence of distinct "snapshots". In fact, when a sequence of still frames is perceived by the human brain at a continuous rate, intermediate images are interpolated and the observer subjectively appears to see continuous motion that in reality does not exist.

Finally, in the lowest bit-rate case, some observers perceived a loss of stereo vision either in H.264/AVC coded sequences and in Wyner-Ziv ones.

4.10 Chapter summary

When dealing with stereo content, DVC approach can exploit either interview or intraview redundancies or both. According to this classification, several approaches have been proposed in literature but SI creation remains the most challenging factor as it mostly affect the quality of the reconstructed sequence.

In this chapter several approaches have been compared in terms of objective quality evaluation of the decoded stream and the hybrid scheme has resulted the most suitable method that can be used for data transmission over TETRA2 channel.

A stereoscopic video quality assessment has been conducted for the evaluation of the proposed distributed video coding scheme. In particular, the hybrid distributed stereo video coding approach has been compared with conventional H.264/AVC and with current method available in literature.

The objective evaluations showed that DVC has a better quality than H.264/AVC for lower bit-rate; at the contrary, for higher bit-rate conventional stereo video coders result more powerful. These results have been validated by subjective tests.

Subjective video quality experiments have been carried out to evaluate video artifacts introduced in a stereo distributed video coding system. The most relevant artifacts that have been noticed are flickering, blocking and blurring.

Chapter 5

Multiple Description Video Coding

This chapter focuses on Multiple Description Coding approach in the error-prone TETRA2 channel. After a brief introduction of the MDC principle, a novel method is proposed that aims to achieve the highest possible quality, while preserving the minimum overhead.

5.1 Joint source-channel coding

One of the objectives of TEDS communication systems designers is the delivering of multimedia data in a fast, reliable, and low cost way. However, the cited goals are very difficult to be achieved: network congestions, channel variability, attacks, software errors, hardware faults, and other factors, are increasing the probability that the transmitted stream will be received partially impaired. On the other end, new services require high fidelity and the demand for quality of the communication by the users is increasing.

Several error resilient techniques have been devised to combat transmission errors. For example TETRA copes with this problem by applying specific codes as Shortened Reed Muller codes, cyclic codes and Rate Compatible Punctured Convolutional codes depending on the scenario/application [68]. In general error correction codes can be grouped into [3] (i) those introduced at the source and channel coder to make

the bit stream more resilient to potential errors; (ii) those invoked at the decoder upon detection of errors to conceal the effects of errors, and (iii) those which require interactions between the source encoder and decoder so that the encoder can adapt its operations based on the loss conditions detected at the decoder.
The error resiliency problem is very challenging: in fact, a single bit error of the compressed video streams can propagate in space and time and can cause the decoder to loose synchronization because of the use of predictive coding and variable-length coding (VLC) by the source encoder; moreover, both the video source and the channel conditions are time varying, and therefore, it is not possible to derive an optimal solution for a specific transmission of a given video signal.
Another technique that deals with these problems is the use of retransmission based transport protocols; unfortunately, especially in case of multimedia communications, the retransmission of data can be useless. In the case of real time communications, such as video conferences or live audio and video streaming, the data information becomes useless if it is received with delay. This is especially true when many independent users, equipped with low power consumption terminals (as TETRA2 devices), send real time MJPEG2000 video to an access point (AP) in case, for example, of emergency situations as flooding or traffic congestion. The peculiarities of this scenario, low power terminals, error prone transmission channel, and shared communication system, are perfectly tuned to the application of an advanced video coding technique, as joint source-channel coding (JSCC).

It is well known that the source coder should compress a source to a rate below the channel capacity while achieving the smallest possible distortion, and the channel coder should add redundancy through forward error correction (FEC) to the compressed bit stream to enable the correction of transmission errors. However, the assumptions on which separation theory is based (infinite length codes, delay, and complexity) may not hold in a practical system. This leads to the joint source-channel coding (JSCC) to deal with the time-varying channel and time-varying source. In fact, JSCC does not aim to remove the source redundancy completely but should use it and regard it as an implicit form of channel coding. The redundancy is added according to the application requirements such as computational capacity, delay requirements, and channel characteristics and should prevent error propagation, limit the distortion caused by packet losses, and facilitate error detection, recovery, and concealment at the receiver.

Techniques such as reference picture selection (RPS), intra-MB insertion, independent segment decoding, and video redundancy coding (VRC) and Multiple Description Coding (MDC) are designed for this purpose. A different approach towards error resilience is to add redundancy at the entropy coding level. Examples include reversible VLCs (RVLCs), resynchronization and data partitioning techniques, which can help limit the error propagation effect to a smaller region of the bit stream once the error is detected. Another type of error resilient source coding system is named flexible macroblock ordering (FMO). Finally, scalable coding or layered video coding, although designed primarily for the purpose of transmission, produces a hierarchy of bit streams, where the different parts of an encoded stream have unequal contributions to the overall quality and unequal error protection (UEP) is provided for different layers with different importance. This approach is commonly referred to as layered coding with transport prioritization.

5.2 Multiple Description Coding

MDC is one promising technique for improving streaming media quality. Firstly developed for speech communication over the telephone network, it is now applied for delivering information over noisy and unreliable channels. The basic idea is to split and to encode the source message into two (or more) complementary descriptions, which are independently transmitted to the receiver by using separate channels or paths. The decoder can reconstruct the original message also from a single descriptor but with lower quality, such as the baseline quality video. The more descriptors are received, the higher the reconstruction quality. To this aim, a certain amount of redundancy is added to each descriptor, increasing the overall bit rate. Figure 5.1 shows the block diagram of MDC when two descriptions of the source are adopted.

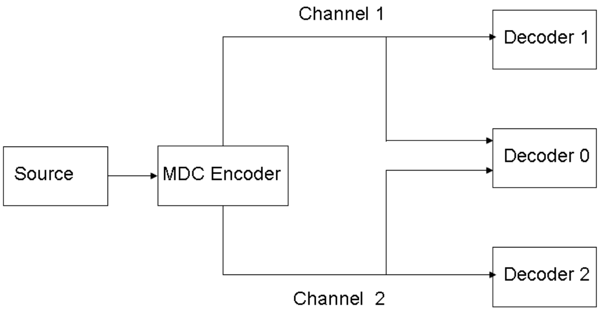


Figure 5.1: Multiple description source coding.

MDC works differently than conventional (SDC) Single Description Coders (MPEG, H.261, H.264, ...) which produce a single data stream. The main benefits in using MDC are obtained when combined with path diversity: in this case the different descriptions are explicitly sent over different routes to a client. The choice of using path diversity links is based on the fact that while network link can be affected by packet loss, the probability that two or more network paths simultaneously suffer from losses is lower. The analysis of network statistics show that losses on the different links are likely to be uncorrelated. As a consequence, MDC combined with path diversity is useful for delay sensitive, real-time applications such as streaming media, where data losses, especially consecutive ones, are strongly impairing the application.
Note that if there are no packet loss, a progressive coding is preferable; however, when packets are lost over noisy channels, using MD coding can get a useful image to the user more quickly because it is not needed to wait the retransmission of lost packets to get a consistent improvement in image quality.
As already stated, the cost of this channel resiliency is in the increased amount of data to be delivered and in increased computational complexity both at the source

and at the receiver side. The more redundancy in each descriptor, the higher the quality when a subset of delivered descriptors is received. According to the specific application (off-line or real time), the trade-off between complexity and efficiency of the method must be considered.

Starting from the pioneering works on speech communications, many researcher have approached the MDC framework. The theoretical aspects were firstly formalized by Wyner et Al. back in 1979 [69] with the definition of the well known 'multiple description problem'. This analysis has been improved and extended with a two channels analysis [70], [71], or multiple-channel one [72], [73], [74]. From an application point of view, several studies have been performed to use the MDC scheme with the state of the art standard video coder as MPEG, or H.264 [75], [76], [77], exploiting spatial or temporal correlation to obtain the descriptions. A good review of the state of the art methods for video MDC can be found in [78]. Video communication schemes, based on MDC, have been studied exploiting spatial or temporal correlation to obtain the descriptions [79].

Even if the main use of MDC scheme is in signal communication frameworks, other applications of MDC have been proposed in literature. In particular, MDC has been used with data hiding methods by exploiting MDC resiliency to transmission errors. Data hiding is a technique proposed for embedding a message in a host signal. According to the particular application, the processed data may or not reveal the presence of hidden data resulting in a visible or invisible modification of the original data [80]. In most cases data hiding is proposed for protecting the ownership of the digital data (watermarking); the performances of the scheme are evaluated by considering, among others, the reliability of hidden data detection even if the host signal has been corrupted by wanted or un-wanted transformations, such as compression, cropping, rotation, interpolation, noise, quantization, etc.

To improve the robustness performance, many authors have exploited the main features of MDC to data hiding. An example is in [81] where the authors adopt a MDC

of the hidden data. At the transmitter, the watermark is encoded by balanced two-description scalar quantizers in the wavelet transform domain. These descriptors are then embedded in the host image in the spatial domain. At the receiver, the multiple description decoder combines the information of each description and reconstructs the original secrete data.
Besides content protection many other uses have been proposed as data hiding, image quality monitoring, fingerprinting, and others [82] [83].
The approach proposed in this thesis exploits the data hiding peculiarities for improving the MDC scheme. The main idea is to reduce the redundancy introduced by a MDC scheme, or alternatively to improve the quality of the received video when a single descriptor is received.

5.3 Proposed MDC Method

The proposes method, inspired to M-JPEG2000 coder scheme, creates two balanced descriptions of the video by intra-coding each frame separately with no motion-prediction/compensation. The goal is to achieve the highest possible quality, while preserving the minimum overhead.

To this aim:

- the redundancy reduction is obtained by exploiting a wavelet based data hiding technique; in particular, the low frequency subband of the second level decomposition of each frame is embedded into the remaining detail subbands. As the low frequency subband concentrates most of the total energy of the image, corresponding to the biggest absolute values of the transform, the amount of data to be sent is sensibly reduced.
- The rate-distortion rate can be adapted to the frame content thanks to the the use of the reversible integer TSH transform which allows to vary the size of the

subbands decomposition. According to a uniformity (or activity) indicator the granularity of the details (an consequently of the LL) subbands can vary.

Let X be a frame of the video sequence and $\mathcal{X}$ its second order TSH decomposition. This decomposition originates seven subbands labeled as illustrated in Figure 5.2.

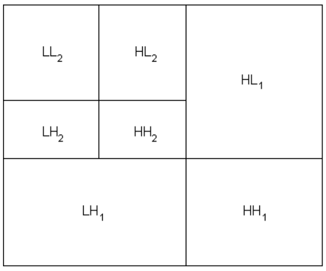


Figure 5.2: Second order TSH decomposition.

Let $\mathcal{X}_{Q_s}$, where $s = 1..3$, denote the quantized version of $\mathcal{X}$, computed as follows:

$$d_{i,j} = sign\left(c_{i,j}\right)\left\lfloor \frac{|c_{i,j}|}{\Delta} \right\rfloor \tag{5.1}$$

where $c_{i,j}$ is the coefficient to be processed and Δ is the quantization step performed according to the JPEG2000 standard. $\mathcal{X}_{Q_s}$ are computed by using three different quantization step Δ_s, $s = 1..3$ with $\Delta_1 < \Delta_2 < \Delta_3$.

To generate two balanced description, the subbands HL_2, LH_2 and HH_2 of $\mathcal{X}_{Q_1}$ and the corresponding subbands of $\mathcal{X}_{Q_2}$ are split each in two halves. The first description A is composed by half HL_2 of $\mathcal{X}_{Q_1}$ and half of the same subband from $\mathcal{X}_{Q_2}$. At the same time, the corresponding subband of description B will be composed by $\mathcal{X}_{Q_2}$ for the first half and by $\mathcal{X}_{Q_1}$ for the other half.

The same procedure is performed on HL_1, LH_1 and HH_1 of $\mathcal{X}_{Q_1}$ and $\mathcal{X}_{Q_3}$ to originate the corresponding subbands of the descriptions A and B. The non quantized subband LL_2 is duplicated in both descriptions. The HL, LH and HH subbands are encoded by using the parent-children partition as depicted in Figure 5.3.

For each description l with $l = A, B$:

- The LL_2 of dimension nxm is represented by 8 bit-planes, each bit-plane is

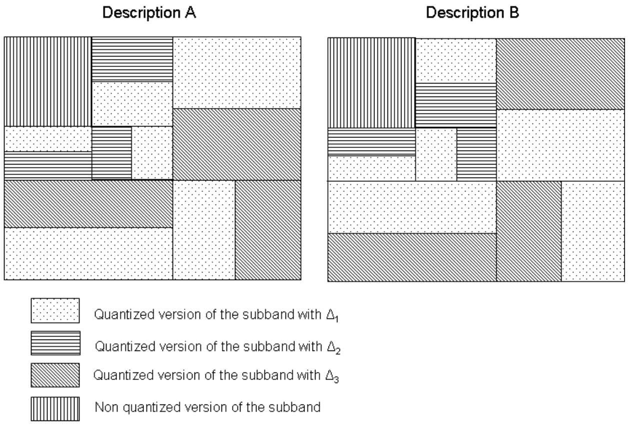

Description A
Description B
Quantized version of the subband with Δ_1
Quantized version of the subband with Δ_2
Quantized version of the subband with Δ_3
Non quantized version of the subband

Figure 5.3: Balanced description generation procedure.

represented by the column vector, and they are stacked together to a long vector W.

- The binary sequence W is embedded in the detail subbands of the description. The procedure used to insert the watermark is a quantization index modulation [84]:

$$\begin{aligned} d'_{i,j_l} &= d_{i,j_l} + \alpha q \\ q &= Q_\Lambda \left\{ d_{i,j_l} - \Lambda \left(\frac{W_h}{D} + k_h \right) \right\} - \left(d_{i,j_l} - \Lambda \left(\frac{W_h}{D} + k_h \right) \right) \end{aligned} \tag{5.2}$$

  where d_{i,j_l} is the coefficient to be modified, α is the strength of the watermark, Q_Λ denotes scalar uniform quantization with step size Λ, W_h is h-th element of the watermark binary sequence with $h = 1, ..., mxn$, D is the dimension of the alphabet that composes the watermark (binary in this case, i.e. $D = 2$) and $k_h \in [0, 1)$ is the watermark key.

  The possibility to vary the size of the subbands of the decomposition joint with the possibility to embed the average subband into the details subbands allows to considerably reduce the size bitstream to be coded; in fact LL_2 is not inserted in it as shown in Figure 5.4.

- The bitstream is coded by variable-length Huffman coding.

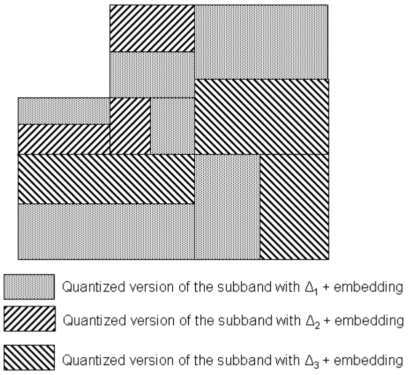

Quantized version of the subband with Δ1 + embedding
Quantized version of the subband with Δ2 + embedding
Quantized version of the subband with Δ3 + embedding

Figure 5.4: The non quantized LL_2 is inserted in all the detail subbands. The resulted bit stream that does not contain the subband LL_2 is Huffman coded. Description A is depicted.

5.3.1 Decoding process: side decoding

When one descriptor l is received, the reconstruction is as follows:

- The received bit stream is decoded by Huffman decoding.

- The binary watermark is extracted from the detail subbands of description without referring to the original description or to inserted LL_2:

$$\begin{aligned} u_h &= Q_\Lambda \left\{ d''_{i,j_l} - k_h\Lambda \right\} - \left(d''_{i,j_l} - k_h\Lambda \right) \\ \widehat{W}_h &= \begin{cases} 0, & if \quad |u_h| \leq {}^{\Lambda}\!/_{2} \\ 1, & if \quad |u_h| > {}^{\Lambda}\!/_{2} \end{cases} \end{aligned} \tag{5.3}$$

  where d''_{i,j_l} is the coefficient supposed to contain the watermark. It is clearly needed in this stage the knowledge of the watermark key k_h and the step size of the quantization Λ. Then the sequence is rearranged to reconstruct the $\widehat{LL}_{2_l}$.

- The coefficients belonging to the detail subbands are de-quantized as follows:

$$\hat{d}''_{i,j_l} = \begin{cases} 0 & d''_{i,j_l} = 0 \\ \left\lfloor \left(d''_{i,j_l} + \delta \right) \Delta \right\rfloor & d''_{i,j_l} > 0 \\ & \\ \left\lceil \left(d''_{i,j_l} - \delta \right) \Delta \right\rceil & d''_{i,j_l} < 0 \end{cases} \tag{5.4}$$

where δ is a parameter conveniently selected. For each $\mathcal{X}_{Q_s}$ the corresponding Δ_s is used.

- The inverse TSH decomposition of the description containing the reconstructed $\widehat{LL}_{2_l}$ is performed.

5.3.2 Decoding process: central decoding

If both description are received, the procedure is as follows:

- Both the streams are Huffman decoded.
- $\widehat{LL}_{2_A}$ and $\widehat{LL}_{2_B}$ are separately extracted by the received descriptions and the average subband between $\widehat{LL}_{2_A}$ and $\widehat{LL}_{2_B}$ is generated.
- For each subband, apart from $\widehat{LL}_2$, the best representation is selected.
- The inverse quantization of the coefficients is carried out by using quantization step Δ_1 (equation (4)).
- The inverse TSH decomposition of the reconstructed frame is performed.

5.4 Experimental results

The proposed method allows to generate two balanced descriptions, each approximately encoded at rate $(R_A+R_B)/2$, where R_A and R_B are respectively the encoding rate of description A and B and the overall bit rate being $R = (R_A + R_B)$.
The algorithm has been tested on 90 frames of the two video sequences $Foreman$ and $Stefan$ in CIF-YUV 4:2:0 format. Each frame, of size 288 x 352 pixels, has been TSH decomposed by using $DPV_1 = [26, 21, 166, 42]$ and $DPV_2 = [7, 135, 232, 10]$ to perform the rows and the columns decomposition. Consequently the LL_2 is of size 106 x 106 pixels. The classical Haar wavelet decomposition results in a 72 x 88 size. Δ_1 has been chosen equal to 2, while Δ_2 and Δ_3 have been varied; δ has been chosen

equal to 0.5. The used QIM parameters have been $\alpha = 1$, $\Lambda = 3$ and $k = 1$ as the optimum trade-off between invisibility and robustness.
In the following the results for the sequence $Foreman$ are shown. Note that the standard JPEG2000 coding computed with wavelet $(9, 7)$ represents the reference SDC to be compared with the proposed method. To evaluate the improvement obtained with the data hiding technique, it has been shown the performance corresponding to the case in which TSH decomposition is adopted to create the descriptions and the LL_2 is transmitted and is not embedded in the detail subbands. In the Figures, labels "no DH" represents the case where data hiding is not adopted.

The central distortion D_c is reported in Table 1 in terms of the central WPSNR [48] averaged on all the frames. It is a function of quantization step Δ_1 only; in fact when both descriptions are received, all the subbands are extracted from the bitstream encoded at maximum rate. The corresponding quality is that of a standard JPEG2000 coding, encoded at rate $R^* = 1.9$.

Table 5.1: Central distortion D_c.

WPSNR *Descr. A* + *Descr. B*	WPSNR *Descr. A* + *Descr. B* no DH	WPSNR JPEG2000
49 dB	51 dB	50 dB

On the other hand, if a description is lost, the received one still gives an inferior but acceptable quality. Assuming that the distortion is additive [85], side distortion D_s is the arithmetic average of the distortions yielded by the two descriptions. Figure 5.5 shows side-WPSNR averaged on all the frames versus the overall bit rate: the trend highlights the gain obtained by the data hiding. When for example side-WPSNR is equal to 45, the description created with the use of the data hiding needs to transmit 2.8 bpp while, if data hiding is not used, 4.8 bpp are required to obtain the same quality.

Based on the above considerations, the extra rate needed to obtain satisfactory

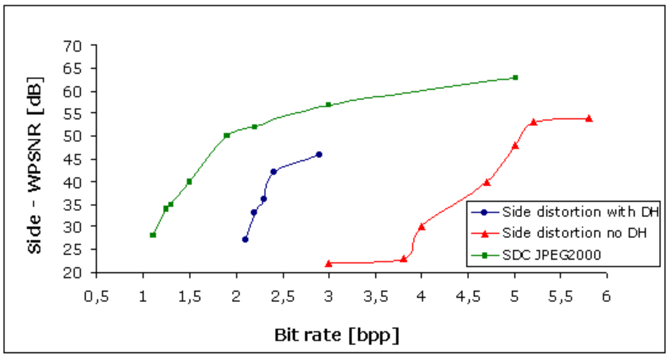


Figure 5.5: Side distortion versus the overall bit rate.

side quality is the redundancy of the proposed MDC scheme, evaluated as $\rho = R - R^*$ where $R = R_A + R_B$ is the resulting rate of the MD coder with the central distortion D_c and R^* is the best single description coder for the given central distortion D_c. The concept of introducing redundancy in a description is to reduce the distortion when only one description is received. Table 2 compares the redundancy values computed when data hiding is adopted with the values obtained when LL_2 is included in the bitstream to be coded. The adopted scheme allows to considerably reduce the redundancy: when bpp= 2, ρ becomes equal to 5%.

Table 5.2: Redundancy of the system.

	Bit-rate [bpp]	SIDE WPSNR [dB]	CENTRAL WPSNR [dB]	Redundancy ρ [%]
Proposed method	2.4 2	42 27	48 48	20 5
Proposed method no DH	4.7 4	40 30	51 51	59 51

5.5 Chapter summary

In this chapter a novel scheme for multiple description video coding has been presented. The evaluated perceptual impact versus the overall bit rate has demonstrated that the use of a data hiding technique to insert a variable-size average subband in the TSH transform domain allows to sensibly reduce the bit rate needed to transfer the multimedia signal with a low perceptual distortion. If a high quality of the received video is requested, the system can be tuned to improve it keeping fixed the amount of overhead; in low bit rate communication channel case, such as TETRA2, the proposed scheme allows to reduce the amount of data to be transmitted while preserving an acceptable perceived quality.

Chapter 6

Security aspects for a reliable data transmission

This chapter addresses the security requirement in data transmission over TETRA channel. Two commutative watermarking and ciphering schemes for digital images are presented to avoid unauthorized use to intercept and manipulate confidential information.

6.1 TETRA security

A fundamental aspect when dealing with public safety systems is the security of the communications, that include security of the user and security of the network operator; the two are complementary, not mutually exclusive [86]. The user's concerns primarily focus on the availability of service and the confidentiality of their communications, while the network operator's lie more with the ability to control access to the system and its resources in a way that ensures the revenue stream is safeguarded.

In the public safety arena three classes of treads should be taken into account [87]:

- Message related threats, as eavesdropping the air interface, eavesdropping in equipment rooms, eavesdropping with stolen radios , masquerading, false transmitters, replay of recorded old messages, etc.

- System related threats, as jamming, IP network attacks, denial of service, accidental damage to cables, natural disasters, equipment/system faults, etc.
- User related threats, as traffic monitoring and analysis, etc.

TETRA deals with the security principles as follows:

- **Authenticity**: this states that it is fundamental to identify univocally if the person is who he claims to be; on one hand, TETRA supports the single *authentication* algorithm of a Mobile Station to control access to the network and the mutual authentication of the SwMI to let the MS verify that the network is trusted; on the other hand, the authenticated users undergo a *authorization* process for different types of access or activities depending on the administration rights/communication rights, etc.
- **Confidentiality**: ensures that the information is protected through *Air Interface Encryption* between the terminal and base station, *End-to-End Encryption* between the terminals end-to-end without any intervention by the infrastructure rather than its role as a bit carrier (see Figure 6.1), *Encryption within the infrastructure*, *Authentication Key Distribution*, that is a secure transfer from terminal factory to system, and *Physical security* to handle sensitive data information.
- **Integrity**: ensures that correct data is delivered without any impairment; this principle is guaranteed through a *Transmission across air interface*, that is the use of channel coding, error correction codes, appropriate *Network elements* as protocols with error detection and re-transmission capability, data recovery mechanisms, storage of consistent data in different databases, *Dispatching data*,

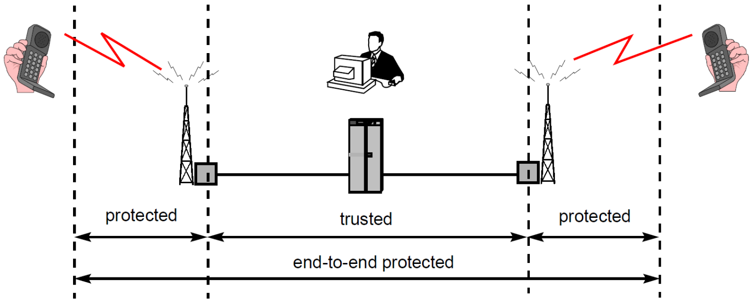


Figure 6.1: Air Interface Encryption and End-to-End Encryption.

i.e. tracking the status of the radio terminals to give correct picture, information on talk group members, periodic registrations, etc.

- **Availability**: information is always available thanks to *Fault tolerant computing*, i.e. recovery system, automatic replacement of faulty unit and no service breaks, *Transmission link monitoring*, i.e. automatic re-sending, rerouting and recovery, *Priority access* so that important users always have access to the system, priorities, *Base station fall-back capability* and *Direct Mode Operation*, that guarantee communication between the users when some or all the network features are not available.

- **Non-repudiation**: Original source of information needs to be always traceable; this is realized through *Recording and playback* capabilities, archive recording of all control room traffic and monitoring of calls, *Logging of management events*, i.e. all actions are written to system logs and *Accounting capability* that log all the used resources.

In general, when describing the TETRA security functions, the following categories can be identified [88]:

- Security mechanisms: functions that aim to achieve a specific security objective such as confidentiality of information or authentication of mobile terminals. Figure 6.2 shows a simple diagram of these functions.

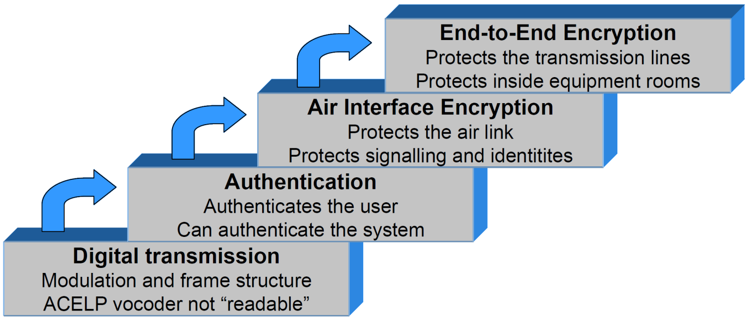

End-to-End Encryption
Protects the transmission lines
Protects inside equipment rooms
Air Interface Encryption
Protects the air link
Protects signalling and identitites
Authentication
Authenticates the user
Can authenticate the system
Digital transmission
Modulation and frame structure
ACELP vocoder not “readable”

Figure 6.2: Security mechanisms.

- Security management features: functions used to control/manage security mechanisms and to realize interoperability of the security mechanisms over different networks. The keys (authentication key/air interface encryption keys) often form the interface between the security management and the security features; the security management is responsible for dealing with the keys in a secure and flexible way.

- Standard cryptographic algorithms: standardized system specific mathematical functions used to provide an adequate security level for the security mechanisms and the security management features.

- Lawful interception mechanisms: functions used to provide the lawfully required access to information and communication, with the aim to fulfil national regulatory requirements.

Given that TETRA is an increasing need for transmission of multimedia data with high level of security, research has addressed encryption synchronization mechanism for transmitting encrypted video over TETRA [89]. This chapter presents two commutative watermarking and ciphering schemes for digital images to protect the data to be transmitted.

The commutative property of the proposed methods allows to cipher a watermarked image without interfering with the embedded signal or to watermark an encrypted image still allowing a perfect deciphering. Both operations are performed on parametric transform domains: the first method addresses gray scale images and is based on the TSH transform; the second approach is based on the Fibonacci Haar transform domain and addresses color images.

The key dependence of the adopted transform domains increases the security of the overall system. In fact, without the knowledge of the generating key it is not possible to extract any useful information from the ciphered-watermarked image.

6.2 Watermarking and cryptography

Multimedia communications and information security are two active areas in both academia and industry. These two separate worlds are expected to continue playing important roles in the information era. The trend shows a fusion between them to allow a secure delivery of multimedia data. According to ITU-T, Rec. X.800 [7] and IETF RFC 2828 [8], the security of data is pursued by assuring, among others, the following services: authentication, to verify the identity claimed by or for any system entity; data confidentiality, to protect data against unauthorized disclosure; data integrity, to verify that data have not been changed, destroyed, or lost in an authorized or accidental manner. To satisfy these constraints several methods have been proposed in literature, such as watermarking and cryptography.

As introduced in section 5.2, watermarking techniques are suitable for copyright protection: before distributing the data, the owner embeds an invisible signature, the watermark, into the host source (audio, text, image, or video) using a secret key (see Figure 6.3). In most applications, the existence of the signature is kept secret and the secret key, previously shared on a secure channel, is used to verify the presence of the embedded sequence in the detection phase. The design of a watermarking scheme is based on some important requirements: imperceptibility of the hidden data, robustness against data processing, capacity of hiding as many bits as needed, and granularity. As widely demonstrated in literature, such constraints are often in

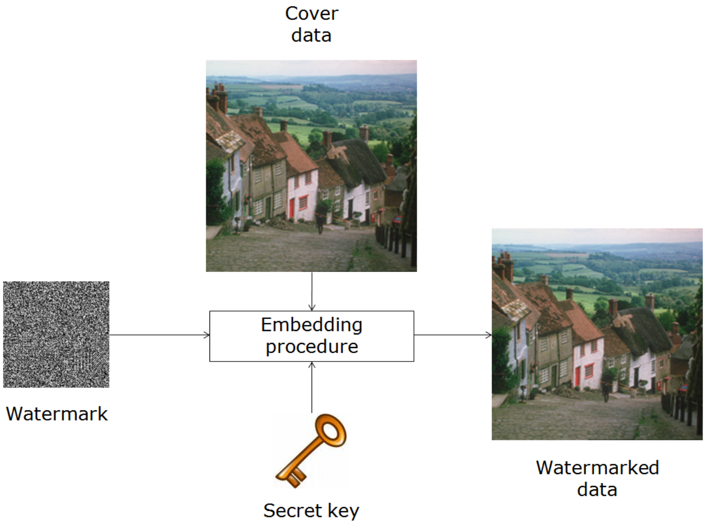


Figure 6.3: General watermarking scheme.

contrast to each other, forcing the designer to find a trade off among them. As far as robustness is concerned, the watermark must be detectable even after modifications, editing, or transmission of the cover data. Therefore, several techniques insert the watermark into the most significant portions of the digital data, so that it cannot be removed without impairing the original content.

A different approach for protecting data is given by cryptography, whose aim is to make the to-be-protected data not intelligible to any unauthorized user who might intercept the message. In this case, the data content is kept secret and the security of the methods lays in the secret key involved in the process. The digital data has to be decrypted in order to *extract* its information, being vulnerable to attacks, and manipulations. Obviously, protection vanishes after decryption. Actually, encrypted data need an additional level of protection in order to keep control on them after the decryption phase. In fact, when the ciphered data is deciphered by the authorized user, it is unprotected and it can be easily modified, tampered, or stolen.

It is important to underline that the principle defined by Kerckhoffs [90, 91], for cryptography also stands for watermarking: the effectiveness of a cryptographic system should only depend on the secrecy of the key. The knowledge of the ciphering, or of the watermarking algorithm, should not allow an unauthorized user to decrypt the message or to have information about the existence of hidden data. Discovering an hidden message should only be possible with the knowledge of the secret key. Besides this analogy, the two techniques are complementary rather than overlapping and can be combined to increase protection of the message.

The scientific community started focusing on the possibility of providing both security services simultaneously. The two protection levels can be combined in many different ways. Bas *et al.* [92] give an overview on the possible scenarios where the combination of both level of protection can be exploited, while Merhav [93] presents a theoretic analysis of this problem.

Although the most challenging goal would be to process the data directly in the encrypted domain, this is a very ambitious task and therefore it is important to distinguish between secure watermark detection and secure watermark embedding [94]. In the first case the aim is to avoid removal of the watermark by a cheating verifier, since the crucial weakness of the detection procedure is to require the knowledge of the secret key used for embedding [92]. The two proposed solutions are asymmetric watermarking [95], [96] and zero-knowledge detector [97], [98]. In the second scenario, the goal is to protect both the original content and the watermark even if the insertion is operated by an un-trusted embedder. Consequently, the challenge is to avoid the access to the non watermarked data by jointly deciphering and fingerprinting sensitive data [99], [100].

In the last decade a slightly different approach for content protection has been introduced. To overcome the computation complexity and cost of encrypting digital images or videos, and to deal with the high transmission rate vs. low available bandwidth, many authors proposed the idea of applying cryptographic methods only to relevant portions of the data. Basically the protection level of the data is adapted according to the particular application. As a general guide, some constraints should be satisfied: the original content (full resolution or quality) could not be obtained without correctly deciphering the data, sufficient security for the considered application should be provided, the bitstream size should not be increased, and the computational complexity with respect to full encryption should be reduced.

Selective encryption of digital data may refer to the partial encryption of the data (i.e., only some bands of the image wavelet decomposition, or some bit plane of its binary representation [101] are ciphered), to localized encryption of portion of the data (e.g., edges, regions, faces) [102], or to partial modification of the bit stream or of its structure [103]. A vulnerability study of a selective encryption method can be found in [104], where the author demonstrates that, with respect to total encryption, several attacks requiring computational complexity much lower than brute force attacks can produce good quality content. In [101] an analysis of the security of selective encryption of the bit planes image decomposition is performed. The authors demonstrate that, when more than the most significant bit (MSB) plane is selected for the ciphering process, the reconstructed image, obtained by the replacement attack, is severely affected. A good overview of existing selective encryption methodologies can be found in [105].

Despite the difficulties to realize effective algorithms which combine simultaneously watermarking and selective cryptography, some solutions have been proposed. Puech and Rodrigues [106] encrypt the secret key with an encryption method based on public-private keys. Then, this secret key is embedded in the encrypted image by using a DCT based watermarking method. The same authors in [107] propose a lossless joint crypto-data hiding method for medical image in which the image is decomposed in bit planes: the first semi-pixel image (the four Most Significant Bit planes) is compressed with a similar Run Length Encoding algorithm and stenographed with the patient information; then, this image is ciphered with a secret-key and scrambled with the remaining semi-pixel image (the four Least Significant Bit (LSB) planes).

In [108] an hybrid image protection algorithm is proposed. A pre-positioned secret sharing scheme is used to reconstruct encryption secret keys by communicating different activating shares. The activating share is used to carry copyright or usage rights data that are embedded in the content as a visual watermark. A Singular Value Decomposition (SVD) based watermarking scheme is used to insert the watermark. When the encryption key needs to be changed, the data source generates a new activating share and embeds the corresponding watermark into the multimedia stream. Before transmission, the composite stream is encrypted with the key constructed from the new activating share. Once both the activating share and the encrypted content are obtained, each receiver is able to reconstruct the decryption key, decrypt the content and extract the watermark.

In [109] the image is divided into blocks of size 16x16 pixels and the DCT of each block is computed. The watermark is embedded into the encrypted LSBs of High-DCT-data (the highest and the second highest frequency coefficients) of each block in order to replace one bit every 8 with one bit of the watermark; the encryption is performed with RSA [110] algorithm by using a private key. Then the watermarked encrypted LSBs is decrypted using the corresponding private key and then the watermarked DCT coefficients are obtained combining High-DCT-data with original Low-DCT-data.

In [111] and [112] by Lian *et al*, the two operations have been presented in a commutative way. This strategy allows to cipher a watermarked image without interfering with the embedded signal or to watermark an encrypted image still allowing a perfect deciphering. In [112] a detailed scenario concerning video is presented, while in [111] the original image is wavelet transformed and some subbands are ciphered while some others are both ciphered (with sign encryption) and watermarked. In [113] the author proposes a compressed data encryption and/or watermarking by the use of meta source coding and bit insertion for watermarking and protecting data from visualization while decoding.

6.3 TSH based approach for gray scale images

A commutative watermarking and encryption method is proposed, that is based on a secret decomposition of the image. The two procedures provide different levels of security as highlighted by Cox *et al.* [9] and they are transparent to each other thus allowing to watermark encrypted data, or to encrypt watermarked data. Correspondingly, it is possible to detect the watermark into encrypted or decrypted data. Moreover, the security is increased by using the TSH transform [24], which depends on the secret DPV. In the literature, secret domains are exploited to encrypt visual data [114], to both compress and encrypt multimedia data [115], or to insert watermarks [116], [117]. Both the encryption with the Advanced Encryption Standard (AES) [118], and the watermark embedding, via Quantization Index Modulation (QIM) scheme [119], [84], are performed in the secret wavelet subband decomposition. Notice that the secret DPV is necessary (but not sufficient) for both the decryption and the extraction procedure.

In the following the design of a layer architecture for cryptography and watermarking is described. Given a digital image X, let f_W be the function used to hide the digital watermark W into X, i.e.,

$$X_W = f_W\left(X, W\right), \tag{6.1}$$

and let f_C be the function employed to cipher the original data X

$$X_E = f_C\left(X, \xi\right), \tag{6.2}$$

given an encryption key ξ. Watermarking and encryption commute whenever the two functions f_W and f_C satisfy the following rule:

$$\begin{aligned} X_{W,E} &= f_W\left(f_C\left(X, \xi\right), W\right) \\ &= f_C\left(f_W\left(X, W\right), \xi\right). \end{aligned} \tag{6.3}$$

A simple commutative watermarking and encryption scheme can be obtained by separately watermarking a subset of the bit-planes of the image representation in a suitable transformed domain and encrypting the remaining bit-planes. However, correlation between watermarked and encrypted bit-planes can introduce some security breach in the form of plain text attacks. To enhance security it is proposed to pseudo-randomly select the representation domain among a wide set of parametric transformations. Knowledge about the actual transform is conveyed by an additional secret key shared by means of a secure channel. In particular, here, the discrete TSH transform is employed [24].

Since all the operations - watermark embedding and extraction, as well as encryption and decryption - are performed in the TSH domain, the knowledge of the chosen DPV becomes crucial. In fact, this becomes a further element to increase the security of the overall method, though the length of such a secret key is not arbitrary (the maximum size of the vector correspond to the size of the signal that has to be decomposed).

6.3.1 Watermark embedding and encryption

In the following, the general scheme for the commutative watermark insertion and encryption is presented. Let X be the original image, $\mathcal{W}$ be the watermark (binary pseudorandom matrix of the same size of X), TSH$_{DPV}$ be the TSH decomposition depending by the secret key DPV, N be a fixed integer number, and B the number of bits used to represent each TSH sample. Figure 6.4 shows the case in which $N = 3$. The watermarking function f_W of Eq.(6.1) is defined by the following operations:

1. Computation of $\mathcal{X}$, the n-th order TSH$_{DPV}$ decomposition of X.

2. Quantization of the real coefficients of $\mathcal{X}$, resulting into the bit-planes BP_l with $l = 1, 2, ..., B$ and the binary matrix $S_{\mathcal{X}}$ that contains the sign of $\mathcal{X}$.

3. Substitution of the least significant bit-plane BP_B with $S_{\mathcal{X}}$.

4. Watermark insertion in the $(B - N - 1)$ bit-planes from BP_{N+1} to BP_{B-1}. The coefficients $c_{i,j}$ of the matrix resulting from the conversion binary-to-decimal of

the such planes are quantized by using a QIM scheme [84]:

$$c'_{i,j} = c_{i,j} + \alpha \cdot q$$
$$q = Q_{\Delta}\left\{c_{i,j} - \Delta \cdot \left(\frac{\mathcal{W}_{i,j}}{D} + k_{i,j}\right)\right\} - \left(c_{i,j} - \Delta \cdot \left(\frac{\mathcal{W}_{i,j}}{D} + k_{i,j}\right)\right)$$

where $\mathcal{W}$ is the watermark, α is the strength of the watermark, Q_{Δ} denotes scalar uniform quantization with step size Δ, D is the cardinality of the alphabet that composes the watermark (2 in this case), and $k_{i,j} \in [0, 1)$ is the watermark key.

5. Reconstruction of the watermarked data X_W, by performing the digital-to-analog conversion and the inverse TSH$_{DPV}$.

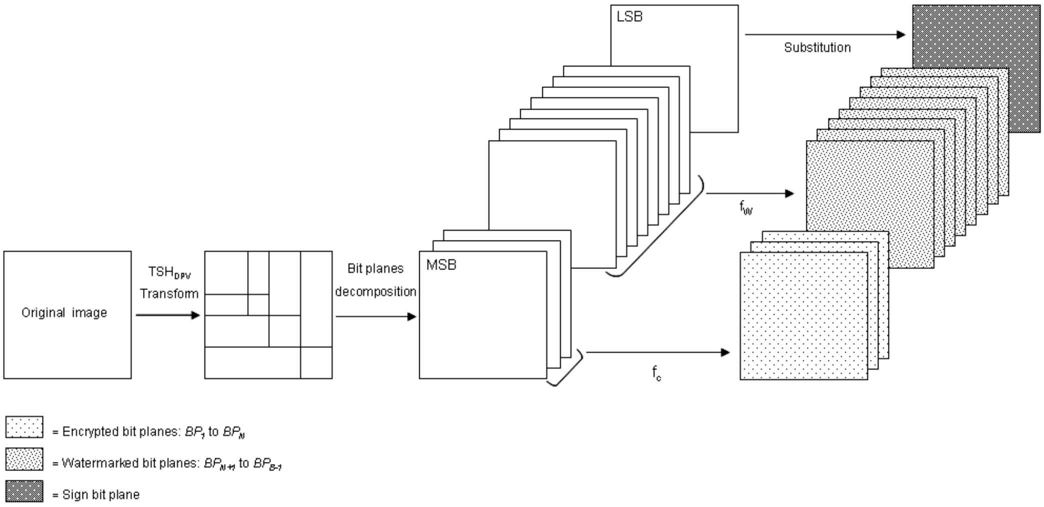


Figure 6.4: Commutative watermarking and encryption procedure.

As far as the encryption function f_C of Eq.(6.2) is concerned, it is defined by the following operations:

1. Computation of $\mathcal{X}$ (as for f_W).

2. Decomposition of $\mathcal{X}$ in BP_l bit planes with $l = 1, 2, ..., B$ and the binary matrix $S_{\mathcal{X}}$ that contains the sign of $\mathcal{X}$ (as for f_W).

3. Substitution of the least significant bit-plane BP_B with $S_{\mathcal{X}}$ (as for f_W).

4. Encryption of the N most significant bit-planes BP_1 to BP_N. These planes are individually encrypted by using the AES. This block cipher is used since it is widely accepted. However, any other secure block cipher can be used for encryption. In the performed tests 128-bit AES keys (k_{AES}) is used.

5. Reconstruction of the ciphered data X_E, by performing the digital-to-analog conversion and the inverse TSH$_{DPV}$.

As it can be noticed, the proposed system is based on the encryption of those bit-planes that are not used for the embedding, and viceversa. Thus, the order of embedding or ciphering is not relevant. Consequently, the proposed method is compliant with the commutative property of Eq.(6.3).

6.3.2 Watermark extraction and decryption

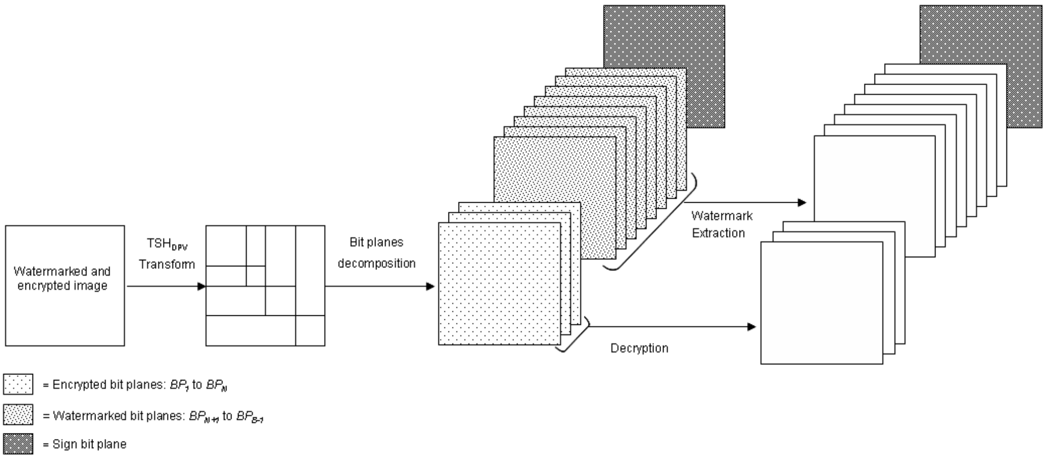

Watermarked and encrypted image
TSH$_{DPV}$ Transform
Bit planes decomposition
Watermark Extraction
Decryption
= Encrypted bit planes: BP_1 to BP_N
= Watermarked bit planes: BP_{N+1} to BP_{B-1}
= Sign bit plane

Figure 6.5: Commutative extraction and decryption procedure.

Given an image $\mathcal{X}_{W,E}$ that has been both watermarked and encrypted according to Section 2.1, the extraction of the watermark and the decryption procedures obtained by inverting the embedding and encryption schemes are as follows (see Figure 6.5 for the case $N = 3$):

1. Computation of $\mathcal{X}_{W,E}$, the n-th order TSH$_{DPV}$ decomposition on the watermarked and encrypted image $X_{W,E}$.

2. Quantization of each coefficient of $\mathcal{X}_{W,E}$ resulting in the bit-planes BP_l with $l = 1, 2, ..., B$.

3. Decryption with k_{AES} of BP_1 to BP_N.

4. The $(B - N - 1)$ bit-planes, BP_{N+1} to BP_{B-1}, used for watermarking are converted in decimal representation. The inverse QIM is then performed to recover the inserted watermark $\mathcal{W}'$ according to the following formula:

$$u_{i,j} = Q_\Delta\{c^{''}_{i,j} - k_{i,j}\Delta\} - (c^{''}_{i,j} - k_{i,j}\Delta)$$
$$\mathcal{W}'_{i,j} = \begin{cases} 0 & \text{if } |u_{i,j}| \leq \Delta/2 \\ 1 & \text{if } |u_{i,j}| > \Delta/2 \end{cases}$$

   where $c''_{i,j}$ is the coefficient supposed to contain the watermark. The knowledge of the watermark key $k_{i,j}$ and of the quantization step Δ are required.

5. The normalized correlation coefficient ρ between the original watermark and the extracted one is computed as follows:

$$\rho = \frac{\sum_i \sum_j \left(\mathcal{W} - \overline{\mathcal{W}}\right)\left(\mathcal{W}' - \overline{\mathcal{W}'}\right)}{\sqrt{\sum_i \sum_j \left(\mathcal{W} - \overline{\mathcal{W}}\right)^2 \left(\mathcal{W}' - \overline{\mathcal{W}'}\right)^2}}. \tag{6.4}$$

   where $\overline{\mathcal{W}}$ and $\overline{\mathcal{W}'}$ are the average values of $\mathcal{W}$ and $\mathcal{W}'$.

It is worth noticing that, besides the knowledge of the secret keys k_{AES} and $k_{i,j}$, also the DPV chosen for the definition of TSH$_{DPV}$ is needed to perform both the extraction and/or the decryption procedures, thus increasing the overall security of the proposed approach. Furthermore, it is important to underline that the two procedures do not need to be simultaneous, allowing the extraction of the watermark without deciphering the data (and viceversa). The overall proposed scheme for the reconstruction of the watermarked image is shown in Figure 6.6.

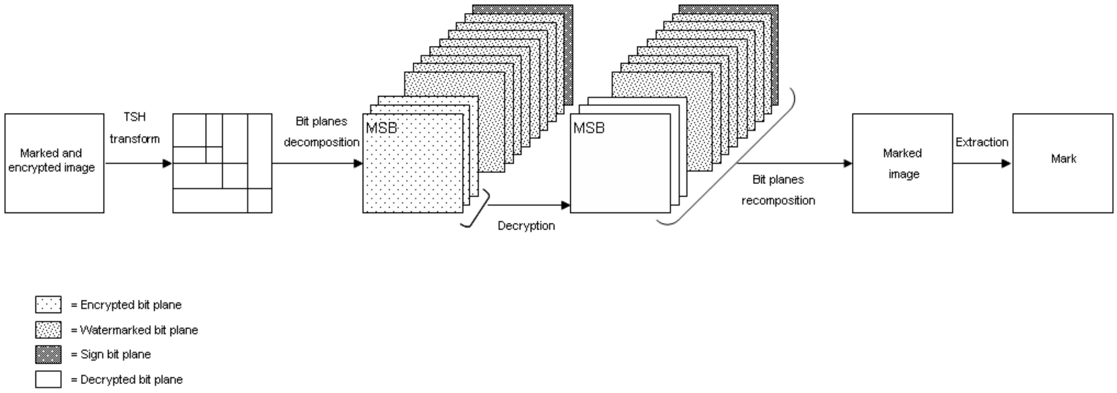


Figure 6.6: Reconstruction of the watermarked image.

6.3.3 Experimental results

Simulations have been performed on a database of 25 gray-scale images of size 512×512 pixels; in the following the results obtained for the image *Lena* are reported. In the simulation, it has been used the 3^{rd}-order decomposition of TSH$_{[120,43]}$, $B = 12$ bits of precision, k_{AES} a pseudorandom scalar value drawn from a uniform distribution on the unit interval, $\alpha = 1$, $\Delta = 30$, and $k_{i,j} = 0.5$.

Several tests have been carried out for evaluating the effectiveness of the proposed approach. First, it is considered the effect of the encryption procedure f_C on the perceived quality of the image. The number of bit-planes N, involved in the watermarking and ciphering procedures, has been systematically varied to determine the optimum value N for protecting the image content from non-authorized users, while reducing the computational complexity. As stated before, the robustness of the system and the non intelligibility of the processed image are inversely proportional: encrypting more bit planes results in a smaller number of bit planes available for watermarking, thus reducing its robustness.

As already described, the security of the selective encryption strictly depends on the application. In most of cases, as for video on demand, database retrieval, etc..., a low quality and non usable version of the data, still protected with a robust watermark, is desirable. This can be obtained by encrypting a smaller number of significant bit planes. On the other hand severe visual degradation is needed for sensitive data (satellite images and videos, video surveillance, etc...): in this case a bigger number of significant bit planes is encrypted to guarantee the non intelligibility of the content.

As analyzed by Podesser et al. in [101] it is possible to verify that, for a 512×512 pixels image with 8 bpp precision, the AES encryption of the two most significant bit-planes grants severe damages in the image even if the encryption of the four MSB provides a high confidentiality level. To assess the security of the bit-plane selective encryption, two different ciphertext-only attacks have been performed: the replacement attack and the reconstruction one. The results of the experiments show that, as expected, the security increases with the number of encrypted bit-planes and that, as stated before, the encryption of only the MSB is not sufficient to grant the cryptographic security of the system. In the case of the replacement attack, it can be prevented by the encryption of the four MSB bit-planes while the reconstruction attack fails when the two MSBs are encrypted.

In addition to this, the parametric nature of the TSH transform allows to increase the security of the method. The mutual information I between the original matrix resulted from the conversion binary-to-decimal of the $(B - N - 1)$ bit-planes used for watermarking and the corresponding matrix computed in the extraction-decryption step is computed. For the mutual information between two binary i.i.d sequences X

Table 6.1: Mutual information between the original matrix resulted from the conversion binary-to-decimal of the $(B - N - 1)$ bit-planes used for watermarking and the corresponding matrix computed in the extraction-decryption step. Different DPV are chosen.

	DPV [120,43]	DPV [119,43]	DPV [50,150]
8 Bit Planes	inf	0,18	0,042
10 Bit Planes	inf	0,37	0,08

and Y of the same length M, the following formula has been applied:

$$I(X,Y) = E_{X,Y}\left[\log\frac{p(x,y)}{p(x)p(y)}\right] = 1 - \hat{P}\log\frac{1}{\hat{P}} - (1-\hat{P})\log\frac{1}{1-\hat{P}}, \qquad (6.5)$$

where

$$\hat{P} = \frac{d_H(X,Y)}{M}. \qquad (6.6)$$

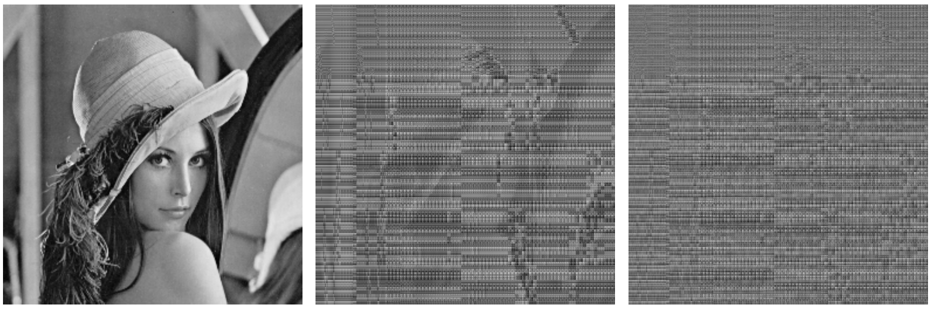
Table 6.1 shows that if a non correct DPV is used in the extraction-decryption step, the mutual information decreases: in particular, the more different is the DPV and consequently the TSH transform, the lower is the mutual information.
Figure 6.7 shows the results obtained for the image *Lena*: by encrypting only the most significant bit-plane ($N = 1$) (see Figure 6.7 (b)) it is possible to disclose the original content while the encryption of the three most significant bit-planes ($N = 3$) allows to obtain a complete visual degradation of the image (see Figure 6.7 (c)).

In the following the experiments obtained selecting $N = 1$ and $N = 3$ are compared.

The watermarked and the watermarked-encrypted versions of the image *Lena* in the two cases $N = 1$ and $N = 3$ are shown in Figure 6.8. It can be noticed that in both cases the watermark insertion does not affect the perceived quality of the original image (see Figure 6.8 (b)-(d)).

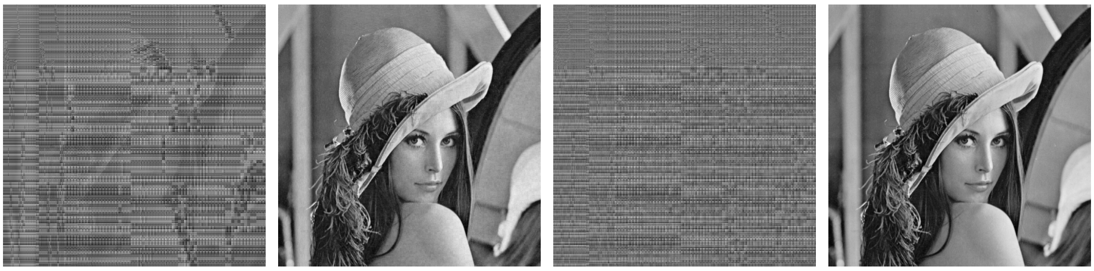
To evaluate the invisibility of the watermarking technique f_W, also the WPSNR has been computed. The average values obtained for the watermarked images when no attacks are performed are:

- $N = 1$: PSNR=38 dB and WPSNR=40 dB.



(a) (b) BP_1 encryption (c) BP_1-BP_2-BP_3 encryption

Figure 6.7: Original image (a), encryption of the MSB (b), and encryption of the three MSBs (c) of the image *Lena*.



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

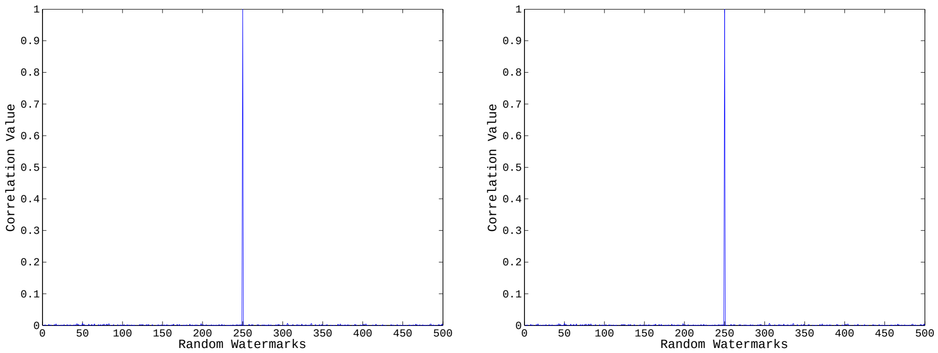
Figure 6.8: Watermarked and encrypted image *Lena* for $N = 1$ (a) and $N = 3$ (c); watermarked *Lena* image for $N = 1$ (b) and $N = 3$ (d).

- $N = 3$: PSNR=40 dB and WPSNR=45 dB.

Finally, the distortion introduced by the quantization step in the f_W and f_C procedures is negligible. The average PSNR values after this step is $\simeq 75$ dB.

In order to demonstrate the robustness of the watermarking procedure, Figure 6.9 reports the detector response when 500 random watermarks are presented to the detector and no attacks are performed. The highest value corresponds to the original embedded watermark. The correlation between the inserted watermark and the extracted one is shown. Figure 6.9 shows the case $N = 1$ and $N = 3$.

Moreover the resistance to some of the most common watermarking attacks has been tested by the StirMark Benchmark 4.0 [120],[121]. The following attacks have



(a)

(b)

Figure 6.9: Detector response when 500 random watermarks are presented for *Lena*. Both cases $N = 1$ (a) and $N = 3$ (b) are shown. No ECC used.

been considered:

- Gaussian: Gaussian noise as by Stirmark to X_W.
- Motion: approximation of the linear motion of a camera by 5 pixels, with an angle of 10 degrees in a counterclockwise direction through a two dimensional filter.
- Blurring: using a circular averaging filter within the square matrix of size 5.

In order to improve the robustness of the digital watermark some Error Correction Codes (ECC) can be used. Recent studies [122] [123] have shown that, at the moment, Reed-Solomon (RS) codes have the highest error correction capability compared to Hamming and BCH codes: in fact they are able to reconstruct the original message for error rates of up to 5%. Therefore, in the following, the results obtained for $N = 1$ and $N = 3$ without ECC are compared with the case where RS codes are adopted. It is worth noticing that the more the length of the codeword is, the more the robustness of the method is guaranteed and less capacity for the watermark is available.

The average results obtained on the whole database, are summarized in Table 6.2. It shows the average highest correlation peaks obtained by presenting 500 random watermarks to the detector.

Table 6.2: Simulation results.

Attack	Gaussian	Motion	Blurring
8 Bit Planes no RS	0.78	0.08	0.11
8 Bit Planes RS	0.41	0.13	0.13
10 Bit Planes no RS	0.91	0.18	0.12
10 Bit Planes RS	0.97	0.19	0.14

The experimental results have been computed considering four possible scenarios:

- watermarking of 8 bit planes and encryption of 3 bit planes;
- watermarking of 8 bit planes and encryption of 3 bit planes with the application of the RS error correcting codes;
- watermarking of 10 bit planes and encryption of 1 bit plane;
- watermarking of 10 bit planes and encryption of 1 bit plane with the application of the RS error correcting codes.

The watermarked images were compressed using the JPEG standard with increasing quality factors from 50 to 100 with step 10. For each quality factor, the detector has been tested with 500 random watermarks: the highest correlation peak corresponding to the original watermark and the second highest correlation value are plotted. Figure 6.10 shows a comparison between the four analyzed cases. The experimental results, as expected, show a more robust behavior in the case when 10 bit planes are used for the watermarking and the error correcting codes are applied.

The proposed scheme has also been tested with the JPEG 2000 standard compression. In particular the compression ratio has been increased from 0.1 to 1 with step 0.1. As for JPEG standard attack, Figure 6.11 shows a comparison between the four cases. As expected, also in this case, the encryption of one most significant bit plane and the watermarking of 10 bit planes gives the best results in terms of robustness.

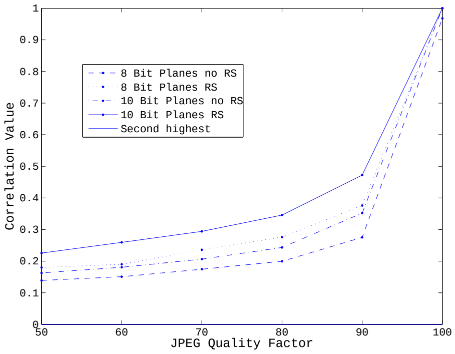


Figure 6.10: JPEG robustness test for the test image *Lena*: first and second highest correlation peaks when 500 random watermarks are presented at the detector are shown for the four analyzed cases.

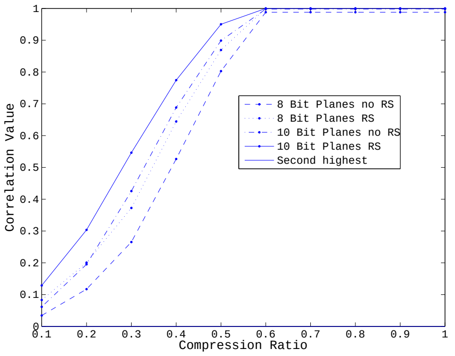
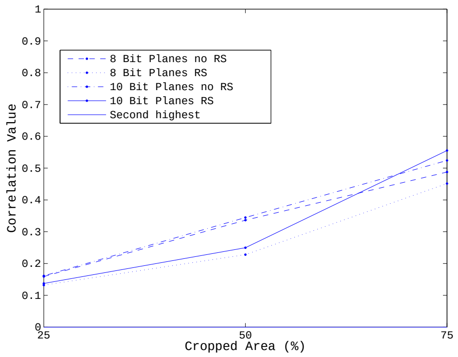


Figure 6.11: JPEG 2000 robustness test for the test image *Lena*: first and second highest correlation peaks when 500 random watermarks are presented at the detector are shown for the four analyzed cases.

The robustness of the proposed techniques against geometric manipulations, e.g. cropping has also been analyzed. In the cropping attack, the watermarked image, of size 512 × 512 pixels, has been cropped to different sizes; in particular the cropped area (external frame of the image) has been replaced with the corresponding original frame [124]. Three different cropping dimensions have been considered, 25%, 50%, and 75% portion of the image. The detector response has been tested with 500 random watermarks. The experimental results are shown in Figure 6.12.

In order to highlight the importance of the secret key DPV in the whole process,



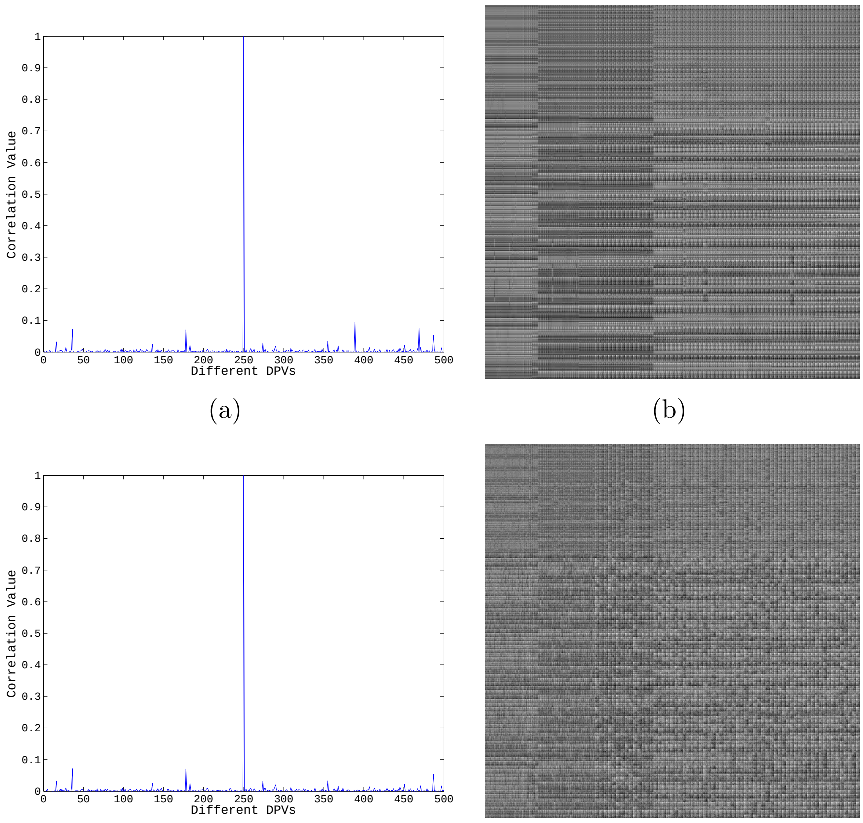
(a)

Figure 6.12: Simulation results for the cropping attack: first and second highest correlation values when 500 random watermarks are presented at the detector are shown for the four analyzed cases.

some results obtained by performing the extraction of the watermark and the decryption of the image with different DPVs are finally presented. Figure 6.13 (a) and (c) show the detector response when 500 random DPVs are used to decompose the image in order to extract the watermark respectively for $N = 1$ and $N = 3$; the peak corresponds to the DPV used by the sender in the first stage of the transmission. In Figure 6.13 (b) and (d) an example of decryption trial of the images Lena with a random DPV is shown. Such experiments demonstrate that in an attacker tries to use random DPVs he will get no information about the hidden watermark as well as on the original image. A brute force attack would be necessary in order to find the correct DPV.

6.4 FHT based method for color images

Based on some preliminary work [125], a joint encryption and watermarking scheme for color images is considered to allow watermark insertion and extraction without interfering with the encryption scheme and viceversa, thus providing both levels of security. As described in [126], it is possible to increase the overall system security by utilizing a layered approach where watermarking and cryptography are



(a)

(b)

(c)

(d)

Figure 6.13: Detector response computed with 500 different DPVs for the test images *Lena* for $N = 1$ (a) and $N = 3$ (c). No ECC used. Examples of decrypted images computed by performing a different decomposition of the received image*Lena* with a random DPV are shown for $N = 1$ (b) and $N = 3$ (d). No ECC used.

simultaneously used as shown in Figure 6.14.

To increase the security of the method, a key-dependent transform domain, the Fibonacci-Haar transform (FHT), is used for both procedures. The crucial point is that no operation like watermark embedding or detection as well as encryption and decryption can be performed without the knowledge of the secret key used to perform the subbands decomposition. The embedding is based on SVD of the Fibonacci-Haar subband decomposition because of the well known SVD properties: stability, scale invariance, rotational invariance, translation, and transposition invariance, which are suitable to counteract attacks like rotation, scaling, noise addition, and others.

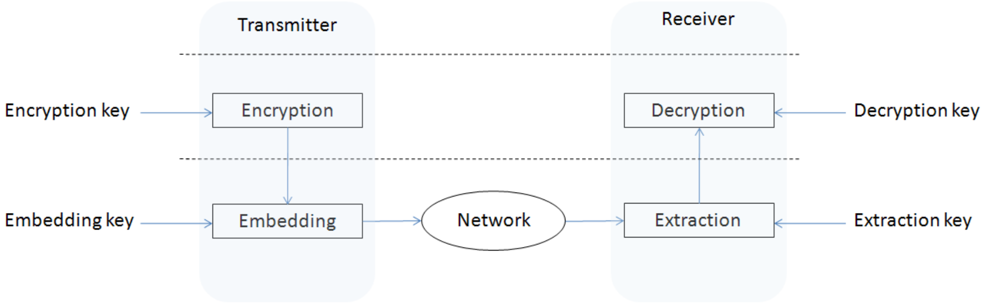


Figure 6.14: Layered security approach.

6.4.1 Color image watermarking background

In the literature several methods have been proposed to protect digital color images by using watermarking. However only few of them are considering the relation existing among the color components. Most of the proposed systems have been designed for gray scale images and then independently applied on the color components.

SVD has been proposed as a tool for watermarking color images in [127] and [128]. Xing and Tan in [127] propose to partition each color component of the cover image (the image to be protected) in non overlapping blocks and to embed one bit of the watermark, using an additive scheme, to the greatest singular value of each block. The block size is adapted to the amount of information to be hidden. The robustness of the whole scheme is increased by scrambling the watermark before insertion through the Arnold transformation. The basic idea is adapted to the wavelet transform of the green component by Yin et al. in [128]. In this case, the SVD of the chaotic scrambled watermark is embedded with additive scheme, into the SVD of the high frequencies sub band components (LH, HL, HH). This method shows an improved resistance to some of the most common attacks as JPEG compression, cropping, median filtering, resizing, and additive gaussian noise. Different approach is presented in [129] where the Discrete Fourier Transform (DFT) based watermarking scheme is applied to different color representations of the cover image (RGB, YUV, and YC_bC_r) to highlight the advantages and disadvantages of each color space. The color space YIQ, adopted in the NTSC color TV system, is considered in [130]: the watermark is embedded in the DWT of both Y and Q components. The system is robust against JPEG compression, filtering, cropping, and additive noise.

DCT has also been used by several authors as a suitable watermarking domain. Among them, the scheme proposed by Ahmidi et al. in [131], is based on the permutation and adaptation of the watermark before its embedding in the the middle DCT frequencies of a block of the image. Li et al. in [132] redundantly embed the watermark into the DCT of the three color components (RGB) of the image by applying a Direct Sequence Spread Spectrum (DSSS) technique. This method grants robustness in case of transmission in noisy paths.

More recent and advanced methods consider the correlation among the color channels. Tsui et al. in [133] present two watermarking schemes. The first one inserts the watermark by performing the *spatiochromatic* Discrete Fourier Transform (SCDFT) in the CIE-L*a*b* color space. To satisfy the invisibility constraint the characteristics of the Human Visual System (HVS) are exploited. The second scheme they propose operates in the L*a*b* color space by using the Quaternion Fourier transform. Both schemes are resistant to different attacks. As can be noticed, all the methods are using some sort of scrambling to increase the security of the system. In the following, some works that are going further in this direction are reviewed, by adopting also cryptographic techniques.

6.4.2 Fibonacci-Haar transform domain

The Fibonacci-Haar transform is a generalization of the Haar transform [24] in which the subband decomposition depends on the particular Fibonacci *p-sequence* $F_p(n)$ defined by the following recursive formula:

$$F_p(n) = \begin{cases} 0, & n < 0; \\ 1, & n = 0; \\ F_p(n-1) + F_p(n-p-1), & otherwise. \end{cases} \tag{6.7}$$

Different values of p define different *p-sequences.* For example, if $p = 0$ the sequence of Fibonacci is obtained *0-numbers*:

$$1, 2, 4, 8, 16, 32, 64, 128, 256, ...$$

i.e the sequence of power of two. If $p = 1$ the Fibonacci *1-numbers* are obtained:

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ...$$

that is the classical Fibonacci sequence. It can be demonstrated[134] that, by using a Fibonacci *p-sequence*, with $p \geq 0$, any positive natural number N can be always represented as follows:

$$N = \sum_{i=p}^{n-1} c_i F_p(i). \tag{6.8}$$

where $c_i = \{0, 1\}$, n is the number of bits needed to represent N with the chosen p-sequence, and $F_p(i)$ are the generalized Fibonacci *numbers.*

These sequences have been used to define the Fibonacci-Haar transformation matrices $H^{(p,n)}$:

$$H^{(p,n)} = \begin{bmatrix} \bar{H}^{(p,n-p-1)} & 0_{F(p,n-p-2)\times F(p,n-p-2)} \\ 0_{[F(p,n-p-1)-F(p,n-p-1)]\times F(p,n-p-1)} & \widehat{H}^{(p,n-p-2)} \\ \hat{H}^{(p,n-p-1)} & 0_{[F(p,n-p-1)-F(p,n-p-2)]\times F(p,n-p-2)} \\ 0_{[F(p,n-p-2)-F(p,n-p-3)]\times F(p,n-p-1)} & \tilde{H}^{(p,n-p-2)} \end{bmatrix} \tag{6.9}$$

where, for $n \leq p$, $H^{(p,n)} = [1]$, $H^{(p,p+1)} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, $\bar{H}^{(p,n-p-1)}$ is a rectangular matrix obtained from $H^{(p,n-p-1)}$ by taking its first $F(p, n-p-2)$ rows, i.e.:

$$\bar{H}^{(p,n-p-1)}_{i,j} = H^{(p,n-p-1)}_{i,j}, \quad i = 1, ..., F(p, n-p-2), \quad j = 1, ..., N;$$

$\overset{\frown}{H}^{(p,n-p-2)}$ is a rectangular matrix obtained from $H^{(p,n-p-2)}$ by taking its first $F(p,n-p-1)-F(p,n-p-2)$ rows,

$\overset{\frown}{H}^{(p,n-p-2)}_{i,j} = H^{(p,n-p-2)}_{i,j}, \quad i=1,...,F(p,n-p-1)-F(p,n-p-2), \quad j=1,...,N;$

$\hat{H}^{(p,n-p-1)}$ is a rectangular matrix obtained from $H^{(p,n-p-1)}$ by taking the last $F(p,n-p-1)-F(p,n-p-2)$ rows,

$\hat{H}^{(p,n-p-1)}_{i,j} = H^{(p,n-p-1)}_{i,j}, \quad i=N-F(p,n-p-1)-F(p,n-p-2)+1,...,N, \quad j=1,...,N;$

$\tilde{H}^{(p,n-p-2)}$ is a rectangular matrix obtained from $H^{(p,n-p-2)}$ by taking the last $F(p,n-p-2)-F(p,n-p-3)$ rows,

$\tilde{H}^{(p,n-p-2)}_{i,j} = H^{(p,n-p-2)}_{i,j}, \quad i=N-F(p,n-p-2)-F(p,n-p-3)+1,...,N, \quad j=1,...,N$

and $0_{q\times r}$ is a $(q\times r)$ zero matrix.
Given a column vector $\mathbf{x}$ of size N, its Fibonacci-Haar transform $\mathbf{t}$ is

$$\mathbf{t} = H^{(p,n)}\mathbf{x}. \tag{6.10}$$

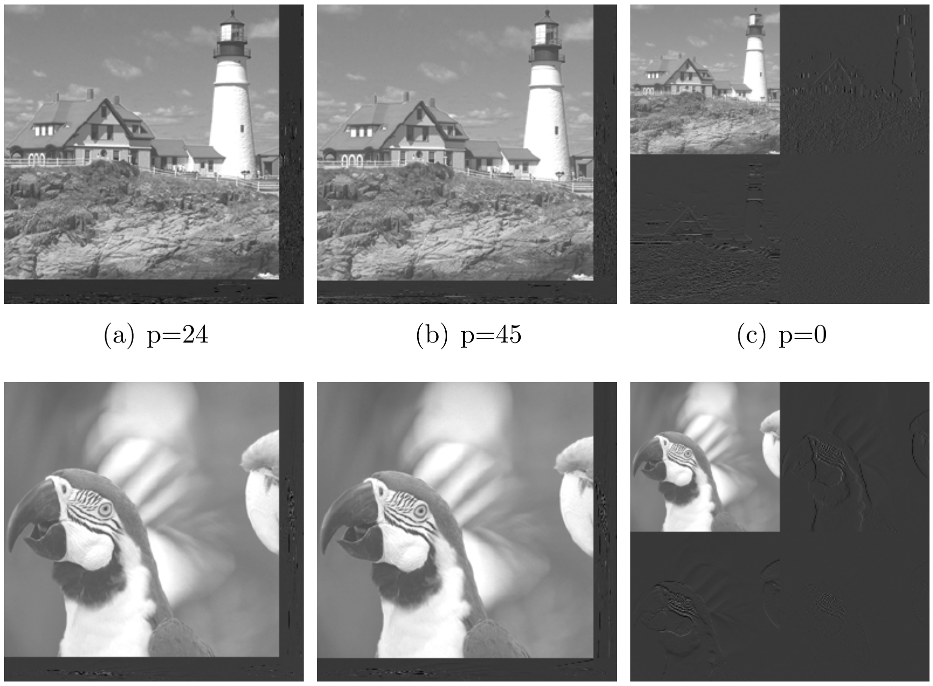
The *p-sequence* used in the embedding process is the secret key which is crucial for the security of the method. For example, the number 256 is the 46^{th} element of the 24-*sequence*, the 66^{th} element of the $p=45$ sequence, the 9^{th} element of the 0-*sequence* (corresponding to the classical Haar decomposition).

As illustrated in Figure 6.15, for the two images Lighthouse and Parrot, the decompositions vary with p.

In general, if the cover image size is $N\times N$, where N is a Fibonacci number, the subband sizes are:

$$\begin{array}{llcll} LL: & N_{n-1} & \times & N_{n-1} & \text{pixels;}\\ LH: & N_{n-1} & \times & N_{n-p-1} & \text{pixels;}\\ HL: & N_{n-p-1} & \times & N_{n-1} & \text{pixels;}\\ HH: & N_{n-p-1} & \times & N_{n-p-1} & \text{pixels.} \end{array} \tag{6.11}$$

where N_{n-1} is the number preceding N in the *p-sequence* and N_{n-p-1} is the number in $p-1$ positions before N in that sequence.



(a) p=24 (b) p=45 (c) p=0

(d) p=24 (e) p=45 (f) p=0

Figure 6.15: First-level decomposition of the Lighthouse and Parrot images: (a)-(d) $p = 24$, (b)-(e) $p = 45$, and (c)-(f) $p = 0$.

For instance, for an image of size 256 × 256 pixels decomposed by using the 24-*sequence*, the size of the four subbands are: LL: 235×235, LH: 235×21, HL: 21×235, and HH: 21×21 respectively since the Fibonacci-Haar 24-sequence is:

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19,$$

$$20, 21, 22, 23, 24, 25, 26, 28, 31, 35, 40, 46, 53, 61, 70, 80, 91,$$

$$103, 116, 130, 145, 161, 178, 196, 215, 235, 256, ...$$

6.4.3 Watermarking insertion and encryption

Let a color image X be considered and let X_c, where $c = R, G, B$, denote its color components. In the following the subscript c is used to denote the generic color component. It is proposed to perform both watermarking and encryption of each color component in the Fibonacci-Haar transform domain. Let the watermark W be a binary sequence of length N_w. To increase the security of the watermarking system, W is spread in the three color components, namely W_R, W_G, and W_B. In this way even if an attacker succeeds in deciphering one color component, he will not get enough information to extract or to modify the whole watermark.

The embedding-encryption procedure, performed on each X_c, can be summarized as follows (see Figure 6.16):

1. The first order decomposition of Fibonacci-Haar transform of X_c is computed according to the chosen p_c*-sequence* (different p_c-values can be used for each component). Let $\mathcal{X}_c$ indicate the correspondent transform.

2. The LL_c subband is encrypted by using the symmetric block cipher AES with a 128-bit key.

3. Each subband LH_c, HL_c and HH_c of $\mathcal{X}_c$ is partitioned into B_c blocks of size N_{n-p_c-1}xN_{n-p_c-1} pixels, where

$$B_c = \left\lfloor \frac{N_{n-1}}{N_{n-p_c-1}} \right\rfloor \cdot 2 + 1, \tag{6.12}$$

   $\lfloor z \rfloor$ denotes the largest integer smaller than z, and N_{n-1} and N_{n-p_c-1} are the larger and the smaller dimensions of LH_c and HL_c, respectively. For each color component B_c depends on the chosen p_c value.

4. Each block is decomposed through the SVD. According to this representation every real matrix A can be expressed as product of three matrices:

$$A = USV^T \tag{6.13}$$

   where U and V are orthogonal matrices and S is a diagonal matrix whose singular values $\{s_1, ..., s_{N_{n-p_c-1}}\}$, are disposed in decreasing order. Since the largest singular values have a stronger impact on the perceived image quality, and the smallest ones are extremely sensitive to noise, the middle singular values

$\{s_{l_c}, ..., s_{m_c}\}$ $(l_c > 1, m_c < N_{n-p_c-1})$ are selected for watermark insertion. Notice that the maximum capacity N_{w_c} for each color component is given by

$$N_{w_c} = B_c(m_c - l_c + 1). \tag{6.14}$$

5. For each block A_i $(i = 1, ..., B_c)$ the embedding is performed in the corresponding S_i diagonal matrix, according to the SVD watermarking scheme proposed in [135]:

$$\begin{cases} \tilde{s}_{i_j} = s_{i_{j-1}} - 1.25\,\Delta, & if \;\; W_{c_{j+(i-1)(m_c-l_c+1)}} = 1, \quad s_{i_{j-1}} - s_{i_j} < 1.25\,\Delta \\ \tilde{s}_{i_j} = s_{i_j}, & if \;\; W_{c_{j+(i-1)(m_c-l_c+1)}} = 1, \quad s_{i_{j-1}} - s_{i_j} > 1.25\,\Delta \\ \tilde{s}_{i_j} = s_{i_{j-1}} - 0.25\,\Delta, & if \;\; W_{c_{j+(i-1)(m_c-l_c+1)}} = 0, \quad s_{i_{j-1}} - s_{i_j} > 0.75\,\Delta \\ \tilde{s}_{i_j} = s_{i_j} & if \;\; W_{c_{j+(i-1)(m_c-l_c+1)}} = 0, \quad s_{i_{j-1}} - s_{i_j} \leq 0.75\,\Delta \end{cases} \tag{6.15}$$

   where s_{i_j} and $\tilde{s}_{i_j}$ are the singular values of the original and watermarked block, respectively, $j = l_c, ..., m_c$, and Δ is the selected detection threshold.

6. Find the smallest singular value $a = \tilde{s}_{i_z}$, with $l_c < z < m_c$; in order to maintain the decreasing order of the singular values, find the first $b = s_{i_h} < a$, where $h > m_c$. Replace the singular values $[s_{i_{z+1}}, s_{i_{h-1}}]$ by linear interpolation between a and b.

7. The inverse SVD of each block is computed.

8. The inverse Fibonacci-Haar is computed according to the p_c-sequence in order to obtain the c component of the watermarked-encrypted image.

Figures 6.17 and 6.18 show the RGB components before and after the encryption-watermarking process for the two images Lighthouse and Parrot.

6.4.4 Watermarking extraction and decryption

The extraction of the watermark and the decryption procedures are performed individually by analyzing the RGB components of the watermarked-encrypted image

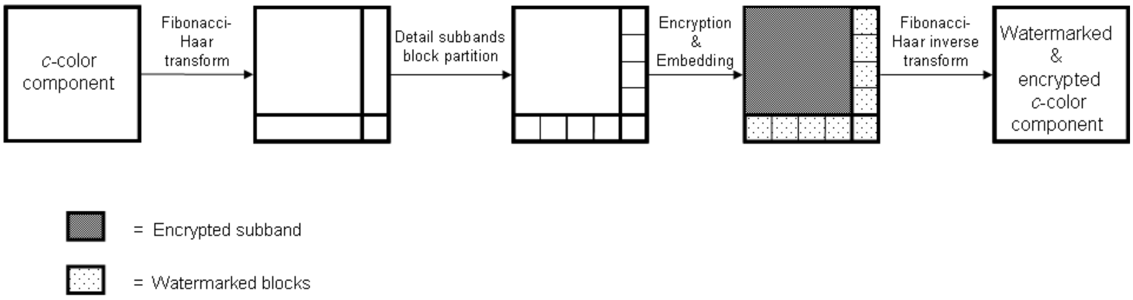

c-color component
Fibonacci-Haar transform
Detail subbands block partition
Encryption & Embedding
Fibonacci-Haar inverse transform
Watermarked & encrypted c-color component
= Encrypted subband
= Watermarked blocks

Figure 6.16: Watermarking and encryption method for each color component.

$\hat{X}$. The following steps are performed on each color component $\hat{X}_c$:

1. The first order Fibonacci-Haar decomposition is performed according to the secret key p_c yielding to $\hat{\mathcal{X}}_c$, which allows the receiver to recover the Fibonacci sequence used in the embedding-encryption procedure.

2. The $\hat{LL}_c$ subband undergoes an inverse AES performed with the shared 128-bits secret key.

3. Subbands $\hat{LH}_c$, $\hat{HL}_c$ and $\hat{HH}_c$ are partitioned into the B_c blocks as described in section 6.4.3 and each block is decomposed through the SVD. The watermark $\hat{W}_c$ is extracted from the singular values of each block as follows:

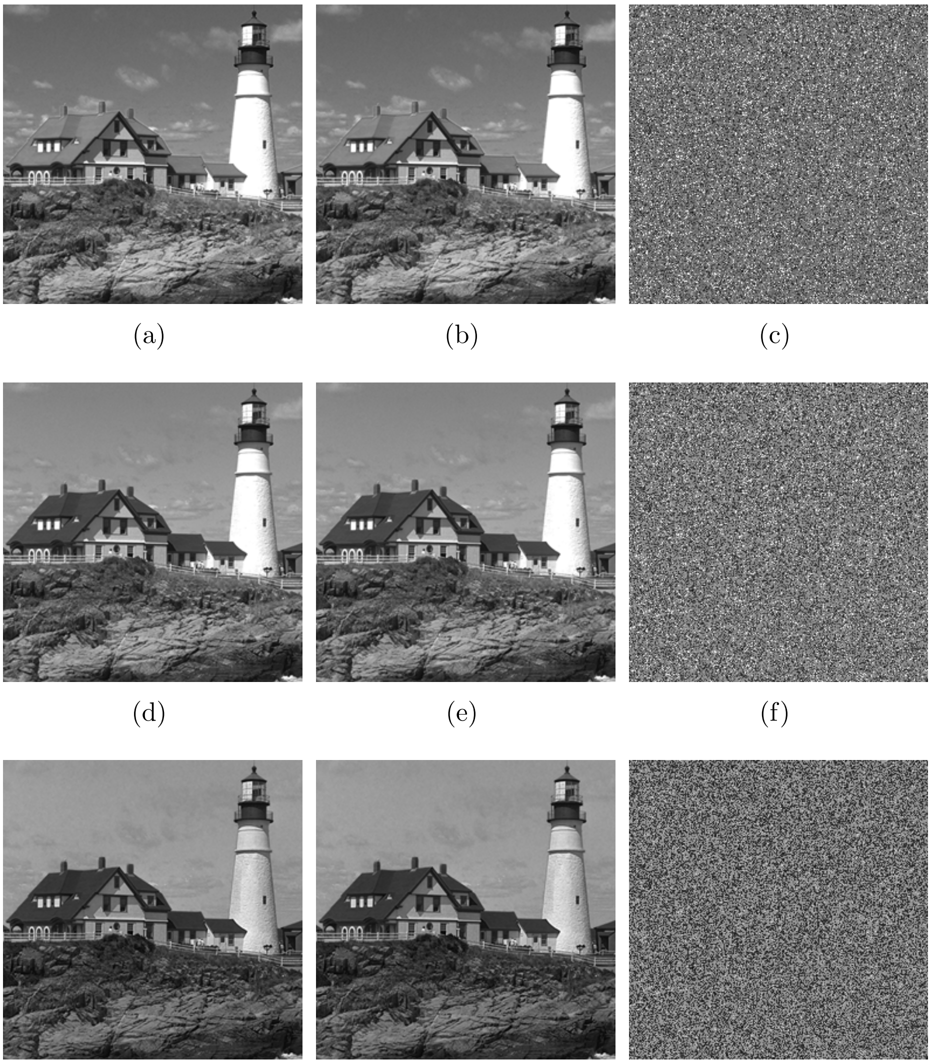
$$\begin{array}{lll} if & \hat{s}_{i_{j-1}} - \hat{s}_{i_j} > \Delta, & \hat{W}_{c_{j+(i-1)(m_c-l_c+1)}} = 1, \\ if & \hat{s}_{i_{j-1}} - \hat{s}_{i_j} \leq \Delta, & \hat{W}_{c_{j+(i-1)(m_c-l_c+1)}} = 0. \end{array} \tag{6.16}$$

   where $i = 1, ..., B_c$, $j = l_c, ..., m_c$, and Δ is the detection threshold.

4. The three extracted watermark components $\hat{W}_c$ are combined recovering $\hat{W}$.

6.4.5 Experimental results

In this Section the results obtained in the experimental tests are shown. In particular, the goal is to show that the proposed method is compliant with both the robustness and the invisibility constraints.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 6.17: Color components of the Lighthouse image: original components (a)-(d)-(g), decrypted-watermarked components (b)-(e)-(h), encrypted-watermarked components (c)-(f)-(i).

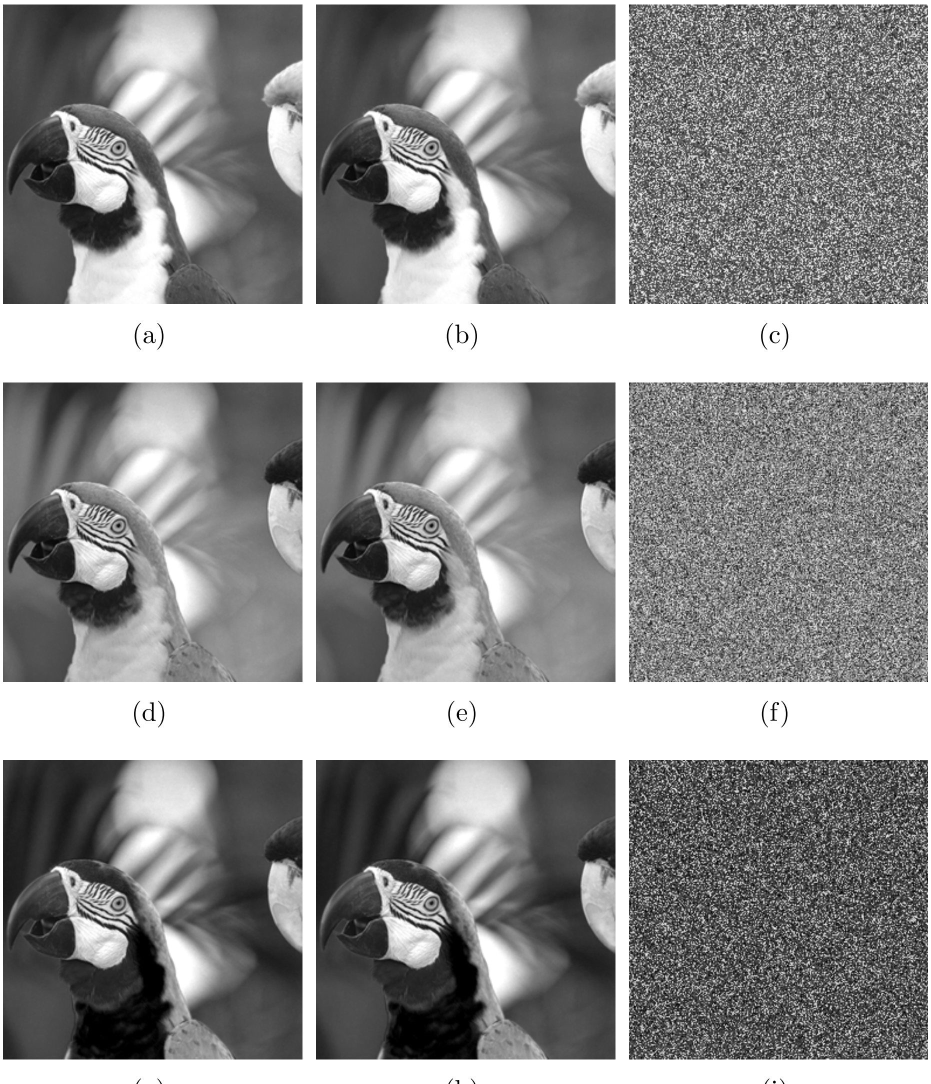


Figure 6.18: Color components of the Parrot image: original components (a)-(d)-(g), decrypted-watermarked components (b)-(e)-(h), encrypted-watermarked components (c)-(f)-(i).

The experimental tests have been performed on 25 color images (256x256 pixels) that have been extracted from the database freely available on the Internet (25 reference images, 17 types of distortions for each reference image, 4 levels for each type of distortion)[136][137]. The watermark is a pseudo-random binary matrix generated by a secret seed.

Watermark invisibility

In the design of a watermarking scheme there is always the need to find the best tradeoff between capacity and perceived quality of the watermarked data. In particular, considering color images, subjective tests show that the human perception is less annoyed by artifacts introduced in the blue component of the RGB color space [138]. A possible motivation lies in the higher spatial and temporal sensitivity of the red−green (Long−Middle wavelength regions) opponent mechanism with respect to the blue−yellow mechanism. Based on this consideration, since the proposed method allows to select the amount of data that can be embedded inside the color components of the host image, a larger part of the watermark is inserted in the blue component.

The *p_c-sequences* used in the reported results are $p_R = 45$, $p_G = 24$ and $p_B = 0$. This choice allows to decompose each color component in blocks:

- the decomposition corresponding to $p_R = 45$ allows to obtain 23 blocks; in each block $l_R = 2$, $m_R = 17$, and $\Delta = 10$ are set;
- the decomposition corresponding to $p_G = 24$ allows to obtain 23 blocks; in each block $l_G = 2$, $m_G = 17$, and $\Delta = 10$ are set;
- the decomposition corresponding to $p_B = 0$ allows to obtain 3 blocks; in each block $l_B = 10$, $m_B = 105$, and $\Delta = 10$ are set.

To allow the reader to visually verify the perceptual quality of the decrypted-watermarked images, Figure 6.19 and Figure 6.20 report the original, the encrypted-watermarked, and the decrypted-watermarked images for two cases extracted from the cited database (images Lighthouse and Parrot). The encryption of the most perceptually significant subbands LL_c results in an image that is non intelligible. The Fibonacci-Haar decompositions obtained with the selected p_c are shown in Figure 6.15 for both images.

In the performed experiments the first order FHT decomposition has been employed, since it is the best trade off between intelligibility of the ciphered data and perceptual quality of the watermarked data. This is due to the fact that, when the FHT decomposition order increases, the size of the LL subband (the amount of information encrypted) decreases while the embedding capacity increases. In the performed simulations it has been exploited the maximum embedding capacity, according to Eq.6.14, for an image of size 256 × 256 pixels, by embedding a watermark of size 32 × 32 pixels.

To evaluate the watermark invisibility in the decrypted-watermarked images, the PSNR and the WPSNR [47] have also been computed.

Table 6.3: Quality evaluation of the watermarked components.

	R component		G component		B component	
	PSNR(dB)	WPSNR(dB)	PSNR(dB)	WPSNR(dB)	PSNR(dB)	WPSNR(dB)
Lighthouse	42	27	42	28	55	57
Parrot	45	30	44	31	55	49

The perceptual impact of the watermark insertion in the test images Lighthouse and Parrot, in absence of attacks, is shown in Table 6.3. As can be noticed the PSNR and WPSNR show good performances concerning the imperceptibility requirement. Figure 6.19 (c) and Figure 6.20 (c) show the decrypted-watermarked image. In this work it has also been tested the possibility to recursively apply the FHT to the original image to increase the embedding capacity. For example the second level decomposition results in a smaller LL subband and in six middle-high frequency subbands (as shown in Figure 6.21).

Experimental results show that, the encryption effectiveness is preserved since it

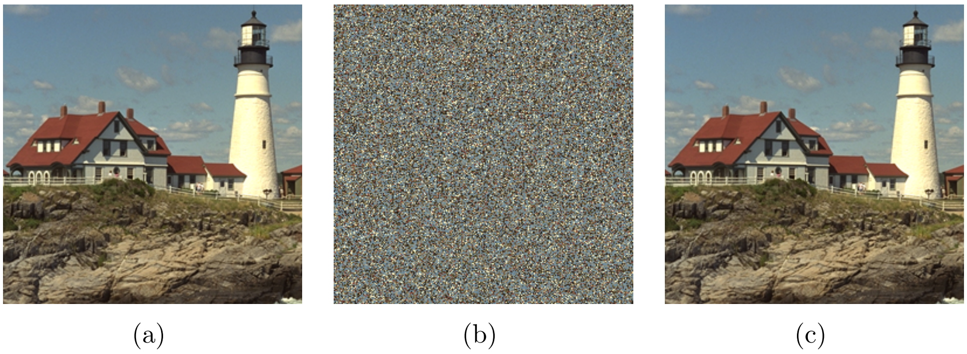


Figure 6.19: Visual impact of the proposed encryption-watermarking scheme for the image Lighthouse: (a) original image; (b) encrypted-watermarked image; (c) decrypted-watermarked image.

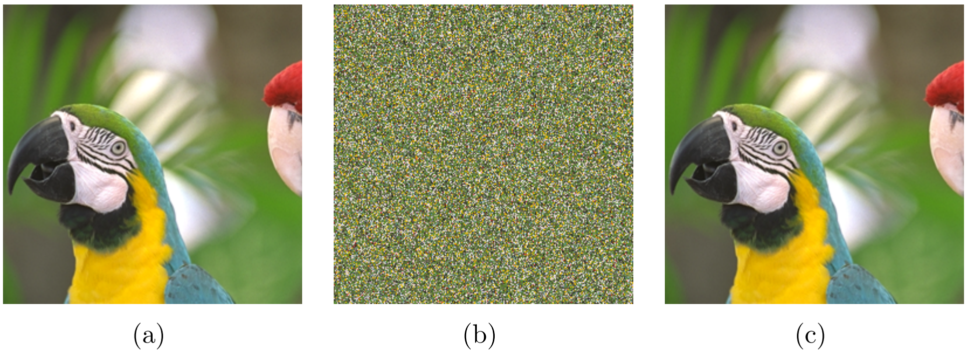


Figure 6.20: Visual impact of the proposed encryption-watermarking scheme for the image Parrot: (a) original image; (b) encrypted-watermarked image; (c) decrypted-watermarked image.

is not possible to visually understand the image content, moreover a smaller LL subband corresponds to an increased embedding capacity. The quality of the deciphered-watermarked image is still good. The average PSNR value for the three color components on the whole database, when the maximum capacity is used (56×34 bits for a 256×256 pixel image), is around 45 dB.

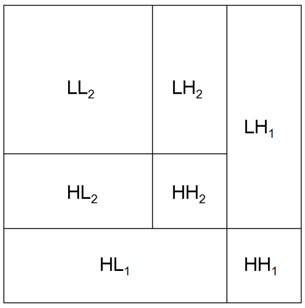
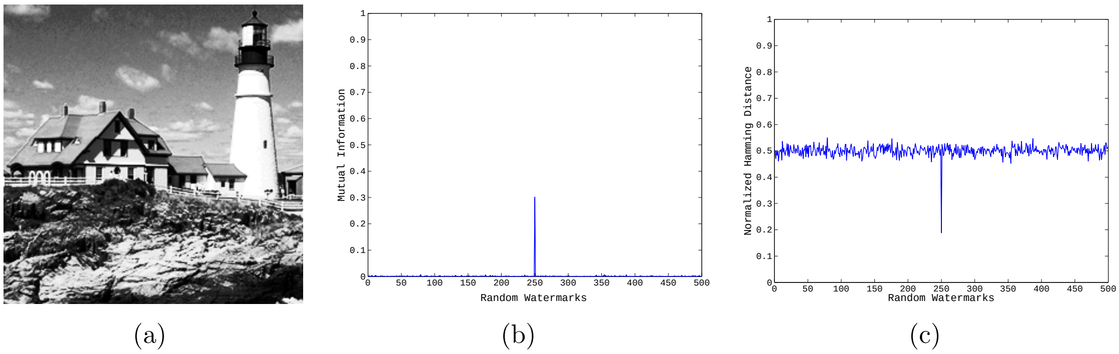


Figure 6.21: Second Level FHT Decomposition.

Robustness

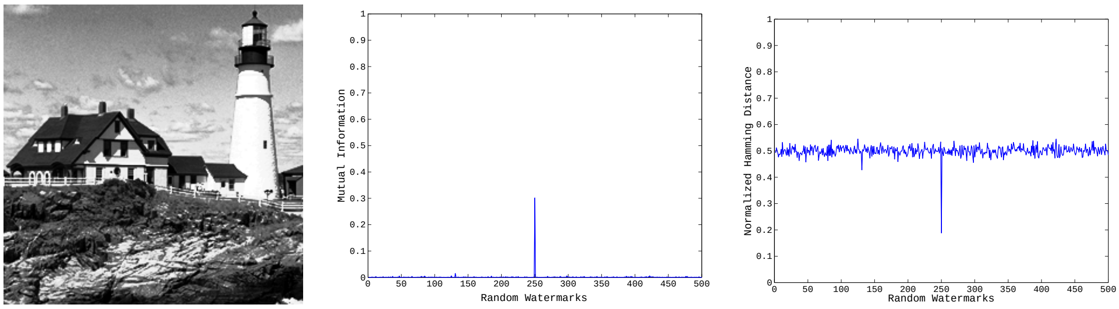
To evaluate the robustness of the embedding method the Stirmark [139] system has been used to attack the watermarked images. Three kind of attacks have been performed: each addressing a single color component. This choice is motivated by the fact that it is infeasible for an attacker to recover all the keys needed for joint decryption and watermark extraction from each component X_c. Simulations show similar performances for the three color components when the same attack is performed.



(a) (b) (c)

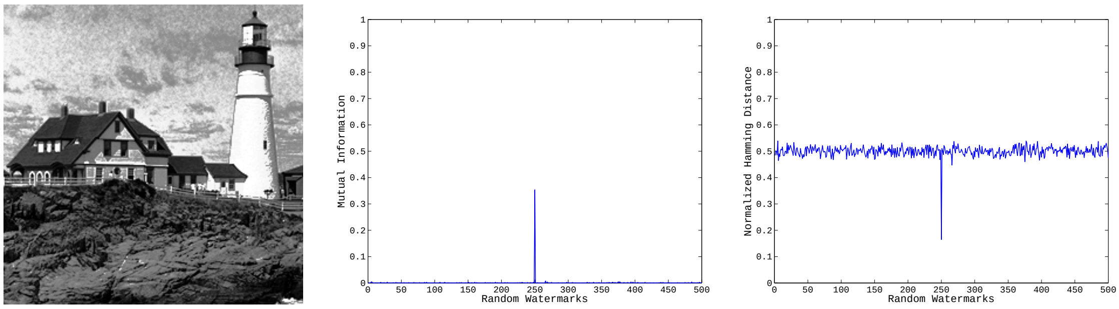
Figure 6.22: Equalization attack tested on a dictionary of 500 randomly generated watermarks: (a) red component after equalization attack; (b) mutual information; (c) normalized Hamming distance.

As an example of this common behavior in Figures 6.22, 6.23, and 6.24, the results obtained for the equalization attack on the image Lighthouse are reported.



(a) (b) (c)

Figure 6.23: Equalization attack tested on a dictionary of 500 randomly generated watermarks: (a) green component after equalization attack; (b) mutual information; (c) normalized Hamming distance.



(a) (b) (c)

Figure 6.24: Equalization attack tested on a dictionary of 500 randomly generated watermarks: (a) blue component after equalization attack; (b) mutual information; (c) normalized Hamming distance.

In the plots the Hamming distance $d_H(\hat{W}, W)$ between the watermark $\hat{W}$ extracted from the attacked image and the original watermark W, normalized with respect to the watermark size, is compared to the the Hamming distance $d_H(\hat{W}, V^{(i)})$ between $\hat{W}$ and the elements $V^{(i)}$, where $i = 1, ..., 500$, of a watermark dictionary randomly generated. It can been noticed that, for all the three attacks, $d_H(\hat{W}, W)$ is rather smaller than $d_H(\hat{W}, V^{(i)})$, where $i = 1, ..., 500$, thus assuring very good watermark detection performance. For completeness the mutual information $I(\hat{W}, W)$ between $\hat{W}$ and W, as well as the mutual information I$(\hat{W}, V^{(i)})$ between $\hat{W}$ and $V^{(i)}$,

Table 6.4: I values between the original watermark and the extracted one after attacks.

Attack	Parameters	Embedded
Gaussian	mean=0, and standard deviation $\frac{3}{4}$=1	0.29
Sharpening	3-by-3 contrast enhancement filter	0.28
Motion	linear motion of a camera by 10 pixels	0.27
Blurring	using a circular averaging filter within the square matrix of size=5	0.26
Median	using a median filter within the square matrix of size=3	0.28

are plotted. It can be noticed that the highest peak corresponds to W while for the other random watermarks the mutual information is practically null.

As mentioned before, the behavior for the three color components is similar; therefore, in the following, the results obtained for the green component averaged over the whole database are discussed. Several attacks have been considered; for each manipulation, Table 6.4 reports the mutual information MI between the embedded watermark and the watermark extracted after the attack. For JPEG and JPEG2000 compression attacks the sensitivity of the mutual information to the compression ratio and quality factor has been assessed.

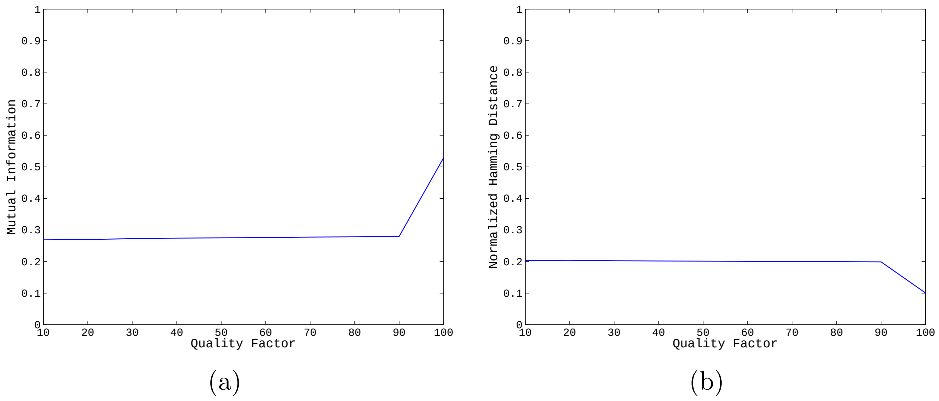


Figure 6.25: Detector response to JPEG compression attack: (a) mutual information; (b) normalized Hamming distance.

Results are respectively shown in Figures 6.25 and 6.26. In particular, quality

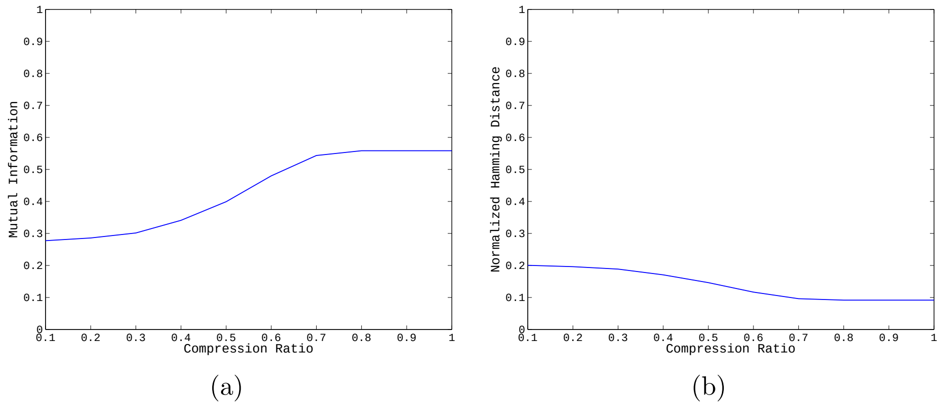
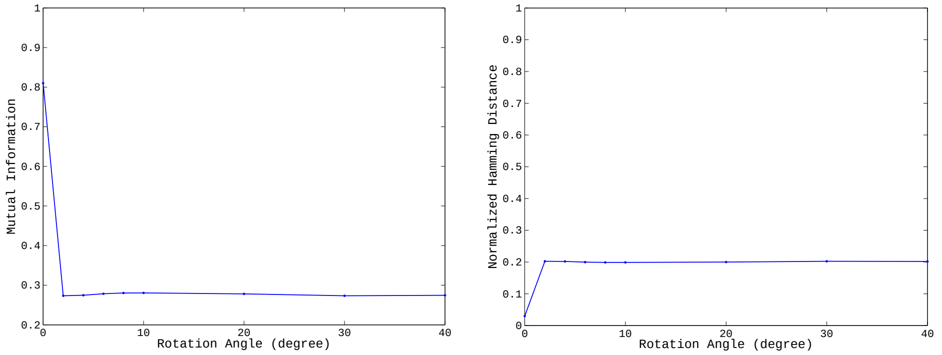


Figure 6.26: Detector response to JPEG2000 compression attack: (a) mutual information; (b) normalized Hamming distance.

factors from 10 to 100 with step 10 for JPEG, and compression ratios varying from 0.1 to 1 with step 0.1 for JPEG2000, have been employed. Mutual information between the detected watermark after decoding and 500 random watermarks is practically null and therefore has not been displayed. Even in this case, the presented values are the average values computed for the whole set of images contained in the database. Rotation attack has also been performed with rotation angle increasing from 0 to 40 degrees. In Figure 6.27 the average values of the mutual information are depicted. Once again mutual information between the restored watermark and 500 randomly selected watermarks is practically null.

Results show that it is always possible to extract the inserted watermark, thus verifying the robustness of the proposed method. As already stated, the proposed method is used to increase the security of the whole system, by further protecting the information with the data hiding technique, after the image decryption. The main security constraint is in the knowledge of the encryption keys used for the AES procedure. From a crypto-analysis point of view, the strength of the whole procedure strictly depends on the security of the AES algorithm[140]. As mentioned before, once the encrypted and watermarked image has been decrypted, the content is still protected thanks to the watermark presence. To this aim the importance of the



(a)

(b)

Figure 6.27: Detector response to rotation attack: (a) mutual information; (b) normalized Hamming distance.

secret key p_c is crucial. To demonstrate this, the watermark has been extracted by choosing a different $\tilde{p_c}$ from the one used in the embedding-encryption procedure and the performances has been evaluated in terms of mutual information between the original and the extracted watermark. For example, by using $\tilde{p_R} = 24$, $\tilde{p_G} = 0$, and $\tilde{p_B} = 45$ the mutual information value decreases from 0.92 (computed by using the correct p_c, that is $p_R = 45$, $p_G = 24$ and $p_B = 0$) to 0.03.

6.5 Chapter summary

In this chapter a commutative watermarking and encryption system has been presented for gray scale digital images, based on a layered scheme and on the key dependent TSH transform domain. Although the proposed method presents some similarities with [111] and [112], there are important differences with the cited works. First of all, here it is proposed to modify the same wavelet coefficients with both the AES encryption and the watermarking, achieving the commutative property in a different way. Moreover, the key dependent TSH transform domain improves the overall security of the system.

The proposed method grants the authenticity of the transmitted data, thanks to the watermarking technique, and the privacy, obtained through the encryption procedure. RS codes have been also used to improve the robustness of the digital watermark. The security system is extremely flexible since the decryption and the watermark extraction can be performed simultaneously or in different stages. Several experimental tests have shown the effectiveness of the proposed method.

A new joint watermarking and encryption technique for color images is also proposed, which exploits the Fibonacci-Haar wavelet transform domain to increase its security. The three RGB color components are ciphered with the standard block cipher AES, and watermarked via a SVD-based blind watermarking method. The intrinsic security of the method is in the AES scheme. Several experimental tests have been performed to verify the impact on the perceived quality of the watermark insertion, and to verify the robustness of the adopted watermarking procedure. The performances have been evaluated in terms of mutual information and normalized Hamming distance.

Chapter 7

Conclusions

7.1 Summary and contributions

This thesis has addressed the problem of data transmission over error-prone, low bit rate TETRA2 channel. Although the advent of TEDS will improve the data rate up to about 500 Kbit/sec, it is very challenging to find the optimal trade off among the redundancy added to protect the data, the efficiency of the compression algorithm and the computational complexity of the encoding procedure carried out by low power TETRA2 devices, while guaranteeing a high level of data security.

In this dissertation, new video coding approaches have been designed, based on Distributed Video Coding and Multiple Description Coding. Both the techniques exploit the content-adaptive feature of the discrete Tree Structured Haar (TSH) transform, recently developed.

The proposed DVC method for stereo sequences has been evaluated in presence of channel errors and has resulted to give better rate distortion performance when compared to conventional video coding standard H.264/AVC and to state of the art method [1] at bit rates lower than 400 Kbit/sec. In particular the main benefits of the proposed scheme are as follows:

- A GOP of variable length can be used while standard algorithms normally limit its length to 2. Hence, while in classical case 1 key frame is followed by 1 Wyner-Ziv frame, in the proposed system 1 key frame, encoded with high computational complexity H.264/AVC coder, is followed by variable N Wyner-Ziv frames, encoded with low computational complexity, where $N + 1$ is the total length of the GOP. This allows to reduce the complexity of the encoder.

- A joint TSH-DCT transform domain is used. The TSH decomposition is exploited due to the possibility to vary the decomposition and the subbands size according to the content of the frame. In this way it is possible to avoid the transmission of the detail subbands, thus considerably reducing the amount of data to be transmitted while guaranteeing an acceptable quality of the reconstructed frame. On the other hand, DCT is applied to increase the robustness of the algorithm over the error-prone TETRA2 channel. This combination of transforms allows to further reduce the complexity of the encoder as only the low frequncy subband is transmitted.
- No additional computational complexity is given to the encoder, when the proposed method is compared with residual coding approach.
- The side information exploits efficiently either spatial and temporal correlation leading to an accurate side information.

Informal subjective evaluation testing has been used to measure the effect of distributed video coding artifacts on the perceived quality of the reconstructed stereoscopic sequence: MOS values corresponding to the hybrid approach are higher than the corresponding values computed for state of the art approach and for H.264/AVC coding.

MDC based schemes has been studied to limit the effect of packet losses: TETRA copes with this problem by applying specific codes as Shortened Reed Muller codes, cyclic codes and Rate Compatible Punctured Convolutional codes depending on the scenario/application.

The proposes MDC method, inspired to M-JPEG2000 coder scheme, achieves the highest possible quality, while preserving the minimum overhead as follows:

- the redundancy reduction is obtained by exploiting a wavelet based data hiding technique; in particular, the low frequency subband of the second level decomposition of each frame is embedded into the remaining detail subbands. As the low frequency subband concentrates most of the total energy of the image, corresponding to the biggest absolute values of the transform, the amount of data to be sent is sensibly reduced.

- The rate-distortion rate can be adapted to the frame content thanks to the the use of the reversible integer TSH transform which allows to vary the size of the subbands decomposition. According to a uniformity (or activity) indicator the granularity of the details (an consequently of the LL) subbands can vary.

The evaluation of the perceptual impact versus the overall bit rate has demonstrated that the use of a data hiding technique to insert a variable-size average subband in the TSH transform domain allows to sensibly reduce the bit rate needed to transfer the multimedia signal with a low perceptual distortion. If a high quality version of the received video is requested, the system can be tuned to improve it keeping fixed the amount of overhead; in low bit rate wireless communication channel case, as for example TETRA2, the proposed scheme allows to reduce the amount of data to be transmitted while preserving an acceptable perceived quality.

Image secure transmission has been finally addressed: in order to guarantee a high level of security, two novel commutative watermarking and ciphering schemes have been presented, based on a layered scheme and on key dependent transform domains. The first approach is suitable for gray scale images and modifies the TSH coefficients with both the AES encryption and the watermarking, thus achieving the commutative property. Hence, the security system is extremely flexible since the decryption and the watermark extraction can be performed simultaneously or in different stages.

Moreover, the key dependent TSH transform domain improves the overall security of the system. The proposed method grants the authenticity of the transmitted data, thanks to the watermarking technique, and the privacy, obtained through the encryption procedure. RS codes have been also used to improve the robustness of the digital watermark. Several experimental tests have shown the effectiveness of the proposed method.

A new joint watermarking and encryption technique has been proposed for color images, which exploits the Fibonacci-Haar wavelet transform domain to increase its security. The three RGB color components are ciphered with the standard block cipher AES, and watermarked via a SVD-based blind watermarking method. The intrinsic security of the method is in the AES scheme. Several experimental tests have been performed to verify the impact on the perceived quality of the watermark insertion, and to verify the robustness of the adopted watermarking procedure. The performances have been evaluated in terms of mutual information and normalized Hamming distance.

Future research will focus on further improvements of the proposed techniques to:

- optimize TSH parameters depending on the GOP content;
- design MDC systems for multiview applications;
- apply the hybrid DVC approach to multiview systems.

Bibliography

[1] J.D. Areia, J. Ascenso, C. Brites, and F. Pereira. Wyner-Ziv stereo video coding using a side information fusion approach. In *IEEE 9th Workshop on Multimedia Signal Processing, MMSP 2007*, pages 453–456, October 2007.

[2] ITU-R BT.500-11. Methodology for the subjective assessment of the quality of television pictures. In *Recommendations of the International Telecommunications Union - Radiocommunication sector*, 2002.

[3] A. Bovik. *The Essential Guide to Video Processing.* Elsevier Academic Press, 2009.

[4] M. Nouri, V. Lottici, R. Reggiannini, D. Ball, and M. Rayne. Teds: A high speed digital mobile communication air interface for professional users. *IEEE Vehicular Technology Magazine*, 1:32 – 42, December 2006.

[5] A.D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 22(1):1–10, January 1976.

[6] D. Slepian and J.K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19:471–480, July 1973.

[7] ITU-T. Rec. x.800 security architecture for open systems interconnection. 1991.

[8] R. Shirey. Internet security glossary. *RFC 2828, GTE/BBN Technologies*, May 2000.

[9] I.J. Cox, G.Döerr, and T. Furon. Watermarking is not cryptography. *5th Int. Workshop on Digital Watermarking (IWDW)*, November 2006.

[10] S. Klomp, Y. Vatis, C. Brites, X. Artigas, L. Torres, F. Dufaux, D. Kubasov, and M. Dalai. Discover deliverable 5: Reference video codec specification from the literature, tech. rep.

[11] Mourad Ouaret. *Selected Topics on Distributed Video Coding.* PhD thesis, École Polytechnique Fèdèrale de Lausanne, 2009.

[12] ETSI EN 300 392-2. Terrestrial trunked radio (TETRA); voice plus data (v+d); part 2: Air interface - v2.3.2. June 2003.

[13] ETSI EN 300 392-7. Terrestrial trunked radio (TETRA); voice plus data (v+d); part 7: Security - v2.2.1. February 2001.

[14] Peter Stavroulakis. *Terrestrial Trunked Radio - TETRA: A Global Security Tool.* Springer, 2007.

[15] Liaison statement (ls) providing TETRA description to mesa. *European Telecommunications Standards Institute ETSI EPT25 (05) 34*, March 2005.

[16] S. Bakaric, M. Borzic, D. Bratkovic, and V. Grga. TETRA (terrestrial trunked radio) - technical features and application of professional communication technologies in mobile digital radio networks for special purpose services. *TETRA seminar at ExpoCommMexico*, February 2007.

[17] ISO/IEC JTC1/SC29/WG11 recommendation MPEG-7. *Coding of Moving Pictures and Audio, Document N6828*, 2004.

[18] ISO/IEC JTC1/SC29/WG11 recommendation MPEG-21. *Coding of Moving Pictures and Audio*, 2002.

[19] A. Grilo, M. Nunes, A. Casaca, Redol, F. Presutto, and I. Rebelo. Communication network architecture for mobiles surveillance in an airport environment.

Jissa journees internationales sur les senseurs et systemes de surveillance aeroportuaire, June 2005.

[20] Martin Steppler. Tetris: A simulation tool for TETRA systems. *Mobile Kommunikation*, 135:403–410, September 1995.

[21] L. Thornton, M. Chakraborty, and J. Soraghan. Video over TETRA. *IEE Seminar on TETRA Market and Technology Developments (Ref. No. 2000/007)*, pages 2/1–2/9, February 2000.

[22] Yoong Choon Chang, M. Salim, and Ting Fook Tang. Performance evaluation of MPEG-4 visual error resilient tools over a mobile channel. *IEEE Transactions on Consumer Electronics*, 49(1):6–13, February 2003.

[23] Peter Meerwald and Andreas Uhl. A survey of wavelet-domain watermarking algorithms. *Proceedings of SPIE, Electronic Imaging, Security and Watermarking of Multimedia Contents III*, 4314:505–516, 2001.

[24] K. Egiazarian and J. Astola. Tree-Structured Haar Transform. *Journal of Mathematical Imaging and Vision, Kluwer Academic Publishers*, 16(3):267–277, May 2002.

[25] M. Cancellaro, M. Carli, K. Egiazarian, and A. Neri. Access control to hidden data by biometric features. *Proceedings SPIE Defense and Security, Mobile Multimedia/Image Processing For Military And Security Applications*, April 2007.

[26] E. Pogossova, Karen Egiazarian, and Jaakko Astola. Fast algorithms and applications of TSH transform. *Proceedings Second Int. Workshop on Spectral Methods and Multirate Signal Processing (SMMSP)*, September 2002.

[27] M. Cancellaro, M. Carli, K. Egiazarian, and J. Astola. Perceptual data hiding in Tree Structured Haar transform domain. In E.J. Delp and P.W. Wong, editors,

Proceedings SPIE Security, Steganography, and Watermarking of Multimedia Contents IX, volume 6505, San Jose, California, USA, February 2007.

[28] M.D. Adams, F. Kossentini, and R.K. Ward. Generalized s transform. *IEEE Transactions on Signal Processing*, 50:2831– 2842, November 2002.

[29] C.T.E.R. Hewage, H.A. Karim, S. Worrall, S. Dogan, and A.M. Kondoz. Comparison of stereo video coding support in MPEG-4 MAC, H.264/AVC and H.264/SVC. *Proceedings of the 4th IET Int. Conf. on Visual Information Engineering (VIE'2007)*, July 2007.

[30] MIUR far project 2004-2006, advanced three dimensional multi-band system (3D-MBS) for airport surface movement surveillance and control integrating stereoscopic techniques.

[31] M. Cancellaro, M. Carli, and A. Neri. A distributed coding approach for stereo sequences in the Tree Structured Haar Transform domain. *Proceedings SPIE International Conference on Electronic Imaging 2009, Image Processing: Algorithms and Systems VII*, January 2009.

[32] N. Gehrig. *Distributed source coding of multi-view images.* PhD thesis, University of London, 2007.

[33] Fernando Pereira, Luis Torres, Christine Guillemot, Touradj Ebrahimi, Riccardo Leonardi, and Sven Klomp. Distributed Video Coding: Selecting the most promising application scenarios. *Signal Processing: Image Communication*, 23:339–352, 2008.

[34] Sung-Hee Lee, Bong-Soo Hur, Shin-Haeng Kim, and Rae-Hong Park. Weighted-adaptive motion-compensated frame rate up-conversion. *IEEE International Conference on Consumer Electronics, ICCE*, 17:342–343, June 2003.

[35] R. Martins, C. Brites, J. Ascenso, and F. Pereira. Refining side information for improved transform domain Wyner-Ziv video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(9):1327–1341, September 2009.

[36] Y. Tonomura and T. Nakachi. A new framework for distributed video coding based on JPEG 2000. *IEEE International Conference Acoustics, Speech and Signal Processing ICASSP Proceedings*, 3, May 2006.

[37] Xun Guo, Yan Lu, Feng Wu, and Wen Gao. Distributed video coding using wavelet. *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 5427–5430, 2006.

[38] Xun Guo, Yan Lu, Feng Wu, Debin Zhao, and Wen Gao. Wyner-Ziv-based multiview video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(6), June 2008.

[39] Thomas Maugey, Wided Miled, Marco Cagnazzo, and Beatrice Pesquet-Popescu. Fusion schemes for multiview distributed video coding. *17th European Signal Processing Conference EUSIPCO*, August 2009.

[40] M. Tagliasacchi, G. Prandi, and S. Tubaro. Symmetric distributed coding of stereo video sequences. In *IEEE Int. Conf. on Image Processing, ICIP 2007*, volume 2, October 2007.

[41] C. Yeo and K. Ramchandran. Robust distributed multi-view video compression for wireless camera networks. In *Proceedings SPIE Visual Communications and Image Processing*, January 2007.

[42] Anne Aaron, David Varodayan, and Bernd Girod. Wyner-Ziv residual coding of video. *Proceedings Picture Coding Symposium PCS-2006*, April 2006.

[43] Anhong Wang, Yao Zhao, and Jeng-Shyang Pan. An efficient hybrid distributed video coding. *IEICE Electronics Express*, 5(17):650–656, 2008.

[44] M. Flierl and P. Vandergheynst. Distributed coding of dynamic scenes with motion-compensated wavelets. volume 1, September 2004.

[45] S. Sandeep Pradhan and Kannan Ramchandran. Distributed source coding using syndromes (DISCUS): Design and construction. In *DCC '99: Proceedings of the Conference on Data Compression*, page 158, Washington, DC, USA, 1999. IEEE Computer Society.

[46] available at http://tev.fbk.eu/databases/.

[47] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T.y Pun. A stochastic approach to content adaptive digital image watermarking. In *Proceedings of the Third International Workshop on Information Hiding*, September 1999.

[48] A. Watson, G. Yang, J. Solomon, and J. Villasenor. Visibility of wavelet quantization noise. *IEEE Transactions on Image Processing*, 6(8), August 1997.

[49] C. Berrou, A. Glavieux, and P. Thitimajshima. Near Shannon limit error-correcting coding and decoding: Turbo codes. *Proceedings IEEE International Conference on Communications*, pages 1064–1070, May 1993.

[50] A. Neri, D. Blasi, L. Gizzi, and P. Campisi. Joint security and channel coding for OFDM communications. *16th European Signal Processing Conference EUSIPCO*, August 2008.

[51] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero. Distributed video coding. *Proceedings IEEE, Special Issue Advances Video Coding*, 93(1):71–83, January 2005.

[52] Saeed K. Amirgholipour and Ahmad R. Naghsh-Nilchi. Robust digital image watermarking based on joint DWT-DCT. *International Journal of Digital Content Technology and its Application JDCTA*, 3(2):42–54, June 2009.

[53] Serkan Emek and Melih Pazarci. A cascade DWT-DCT based digital watermarking scheme. *13th European Signal Processing Conference EUSIPCO*, September 2005.

[54] Ali Al-Haj. Combined DWT-DCT digital image watermarking. *Journal of Computer Science*, 3:740–746, September 2007.

[55] H.S. Dee and V. Jeoti. On image compression: a DWT-DCT algorithm. *Sixth International Symposium on Signal Processing and its Applications*, 2:553–556, 2001.

[56] X. Artigas, E. Angeli, and L Torres. Side information generation for multiview distributed video coding using a fusion approach. In *Proceedings of the 7th Nordic Signal Processing Symposium NORSIG 2006*, pages 250–253, June 2006.

[57] A. Gotchev, S. Jumisko-Pyyko, A. Boev, and D. Strohmeier. Mobile 3DTV system: quality and user perspective. In *4th Int. Mobile Multimedia Communications Conf. MobiMedia*, 2008.

[58] A. Boev, D. Hollosi, A. Gotchev, and K. Egiazarian. Classification and simulation of stereoscopic artifacts in mobile 3DTVù content. In *Proceedings of SPIE, Stereoscopic Displays and Applications XX*, Feb. 2009.

[59] A. Boev, A. Gotchev, and K. Egiazarian. Stereoscopic artifacts on portable auto-stereoscopic displays: what matters? In *4th Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics-VPQM*, 2008.

[60] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, and G. B. Akar. Towards compound stereo-video quality metric: a specific encoder-based framework. In *Proceedings of the 7th Southwest symposium on image analysis and interpretaion*, pages 218–222, March 2006.

[61] L. Meesters, W. Ijsselsteijn, and P. Seuntiens. Survey of perceptual quality issues in three-dimensional television system. In *Proceedings of SPIE, Stereoscopic Displays and Virtual Reality Systems X*, volume 5006, 2003.

[62] M. Yuen. *Coding artifacts and visual distortion.* H.R. Wu and K.R. Rao, Digital Video Image Quality and Perceptual Coding, CRC Press, 2005.

[63] C.T.E.R. Hewage, S.T. Worrall, S. Dogan, and A.M. Kondoz. Prediction of stereoscopic video quality using objective quality models of 2-d video. *Electronics Letters*, 44(16):963–965, 31 2008.

[64] Feng Xiao. DCT-based video quality evaluation. In *MSU Graphics and Media Lab (Video Group)*, Winter 2000.

[65] A.B. Watson. Towards a perceptual video quality metric. In *Human Vision, Visual Processing, and Digital Display VIII*, volume 3299, pages 139–147, 1998.

[66] A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. *18th International Conference on Pattern Recognition, ICPR*, 3:15–18, 2006.

[67] S. Erturk and Tae Gyu Chang. Wavelet domain one-bit transform for low-complexity motion estimation. *Proceedings IEEE International Symposium on Circuits and Systems, ISCAS*, 2006.

[68] F. Gutierrez and A. Valdovinos. Performance of channel coding techniques for the digital mobileradio system TETRA (trans european trunked radio). *IEEE 47th Vehicular Technology Conference*, 2:480 – 484, May 1997.

[69] J.K. Wolf, A.D. Wyner, and J. Ziv. Source-coding for multiple descriptions. Technical Report 59, Oct. 1980.

[70] A.E. Gamal and T. Cover. Achievable rates for multiple descriptions. *IEEE Transactions on Information Theory*, 28(6):851–857, November 1982.

[71] Zhen Zhang and T. Berger. New results in binary multiple descriptions. *IEEE Transactions on Information Theory*, 33(4):502–521, Jul 1987.

[72] Fang-Wei Fu and R.W. Yeung. On the rate-distortion region for multiple descriptions. *Proceedings IEEE International Symposium on Information Theory*, 2000.

[73] Xuguang Yang and K. Ramchandran. Optimal subband filter banks for multiple description coding. *IEEE Transactions on Information Theory*, 46(7):2477–2490, Nov 2000.

[74] R. Puri, S.S. Pradhan, and K. Ramchandran. n-channel symmetric multiple descriptions: new rate regions. *Proceedings IEEE International Symposium on Information Theory*, 2002.

[75] O. Campana and S. Milani. A multiple description coding scheme for the H.264/AVC coder. *Proceedings Int. Conf. Telecommun. Comput. Netw. IADAT-TCN 2004*, December 2004.

[76] A. El Essaili, S. Khan, W. Kellerer, and E. Steinbach. Multiple description video transcoding. *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, 6:VI –77–VI –80, 16 2007-Oct. 19 2007.

[77] Nicola Conci and Francesco G. B. De Natale. Multiple description video coding using coefficients ordering and interpolation. *Image Commun.*, 22(3):252–265, 2007.

[78] Y. Wang, A.R. Reibman, and S. Lin. Multiple description coding for video delivery. *Proceedings of the IEEE*, 93(1):57–70, January 2005.

[79] T. Tillo, M. Grangetto, and G. Olmo. Multiple description coding with error correction capabilities: An application to motion JPEG 2000. pages V: 3129–3132, 2004.

[80] Ingemar Cox, Matthew Miller, Jeffrey Bloom, Jessica Fridrich, and Ton Kalker. *Digital Watermarking and Steganography.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007.

[81] Mohsen Ashourian and Yo-Sung Ho. Multiple description coding for image data hiding in the spatial domain. *Computer and Information Sciences - ISCIS 2003*, 3349, 2003.

[82] F. Battisti, M. Cancellaro, G. Boato, M. Carli, and A. Neri. Joint watermarking and encryption of color images in the Fibonacci-Haar domain. *Hindawi - Academic Publisher for Open Access journals*, 2009.

[83] Chowdary .b. Adsumilli, Mylene Christine Queiroz de Farias, Marco Carli, and Sanjit K. Mitra. A robust error concealment technique using data hiding for image/video transmission over lossy wireled/wireless channels. *IEEE Transactions on Circuits An Systems for Video Technology*, 15:1394 – 1406, 2005.

[84] J. J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod. Scalar Costa scheme for information embedding. *IEEE Transaction on signal processing*, 51(4), April 2003.

[85] Majid Rabbani and Diego Santa Cruz. The JPEG2000 still-image compression standard. October 2001.

[86] D. W. Parkinson. TETRA security. *BT Technology Journal*, 19(3):81–88, 2001.

[87] Risto Toikkanen. TETRA and security. *ELMAR, 47th International Symposium*, pages 307–310, June 2005.

[88] G. Roelofsen. TETRA security. *Information Security Technical Report*, 5(3):44 – 54, 2000.

[89] M.I. Samarakoon, B. Honary, and M. Rayne. Encrypted video over TETRA. *IEE Seminar on TETRA Market and Technology Developments (Ref. No. 2000/007)*, pages 3/1 – 3/5, 2000.

[90] A. Kerckhoffs. La cryptographie militaire. *Journal des sciences militaires*, 9:5–83, January 1883.

[91] A. Kerckhoffs. La cryptographie militaire. *Journal des sciences militaires*, 9:161–191, February 1883.

[92] S. Katzenbeisser et al. First summary report on hybrid systems. *European Project IST-2002-507932, ECRYPT - Network of Excellence in Cryptology, Deliverable D.WVL.5*, January 2005.

[93] N. Merhav. On joint coding for watermarking and encryption. *IEEE Transactions on Information Theory*, 52(1), January 2006.

[94] A. R. Sadeghi. The marriage of cryptography and watermarking - beneficial and challenging for secure watermarking and detection. *Proceedings of the 6th International Workshop on Digital Watermarking (IWDW2007)*, December 2007.

[95] G. Boato, F. G. B. De Natale, and C. Fontanari. An improved asymmetric scheme suitable for copy protection. *IEEE Transaction on Signal Processing*, 54(6), July 2006.

[96] M. Malkin and T. Kalker. A cryptographic method for secure watermark detection. *Proceedings of Information Hiding Conference*, 2007.

[97] A. Adelsbach and A. R. Sadeghi. Zero knowledge watermark detection and proof of ownership. *Proceedings of the 4th International workshop on Information Hiding*, 2001.

[98] A. Adelsbach, S. Katzenbeisser, and A. R. Sadeghi. Watermark detection with zero knowledge disclosure. *ACM Multimedia System Journal*, 9(3), September 2003.

[99] M. Celik, A. Lemma, S. Katzenbeisser, and M. van der Veen. Secure embedding of spread-spectrum watermarks using look-up tables. *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP07)*, 2007.

[100] S. Katzenbeisser, B. Skoric, M. Celik, and A. R. Sadeghi. Combining tardos fingerprinting codes and fingercasting. *Proceedings of Information Hiding Conference*, 2007.

[101] M. Podesser, H.P. Schmidt, and A. Uhl. Selective bitplane encryption for secure transmission of image data in mobile environments. *5th Nordic Signal Processing Symposium*, 2002.

[102] R.M. Scopigno and S. Belfiore. Image decomposition for selective encryption and flexible network services. *IEEE Global Telecommunications Conference*, 2004.

[103] S. Li, G. Chen, A. Cheung, B. Bhargava, and K. T. Lo. On the design of perceptual MPEG-video encryption algorithms. *IEEE Transactions on Circuits and Systems for Video Technology*, 17.

[104] A. Said. Measuring the strength of partial encryption schemes. *IEEE International Conference on Image Processing*, September 2005.

[105] A. Massoudi, F. Lefebvre, C. De Vleeschouwer, B. Macq, and J.-J. Quisquater. Overview on selective encryption of image and video: challenges and perspectives. *EURASIP Journal on Information Security*, vol. 2008, 2008.

[106] W. Puech and J.M. Rodrigues. A new crypto-watermarking method for medical images safe transfer. *Proceedings 12th European Signal Processing Conference*, pages 1481–1484, September 2004.

[107] J.M. Rodrigues, W. Puech, and C. Fiorio. Lossless crypto-data hiding in medical images without increasing the original image size. *Proceedings 2nd International*

Conference on Advances in Medical Signal and Information Processing, pages 358–365, September 2004.

[108] X. Xu, S. Dexter, and A.M. Eskicioglu. A hybrid scheme for encryption and watermarking. *Proceedings of IST/SPIE Electronic Imaging*, pages 358–365, January 2004.

[109] K. Kuroda, M. Nishigaki, M. Soga, A. Takubo, and I. Nakamura. A digital watermarking using public-key cryptography for open algorithm. *Proceedings International Conference on Information Technology and Applications*, November 2002.

[110] R. L. Rivest, A. Shamir, and L. Adelman. A method for obtaining digital signatures and public-key cryptosystem. *Commununication of the ACM*, 21(2):120–126, 1978.

[111] S. Lian, Z. Liu, and H. Wang. Commutative watermarking and encryption for media data. *Optical Engineering Letters*, 45(8), August 2006.

[112] S. Lian, Z. Liu, Z. Ren, , and H. Wang. Commutative encryption and watermarking in video compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(6), June 2007.

[113] B.A. Assanovich. Embedding bits in compressed data for selective encryption and watermarking. *IEEE Region 8 International Conference on Computational Technologies in Electrical and Electronics Engineering,SIBIRCON*, 2008.

[114] G. Unnikrishnan and K. Singh. Double random fractional fourier-domain encoding for optical security. *Optical Engineering*, November 2000.

[115] A. Pommer and A. Uhl. Selective encryption of wavelet-packet encoded image data - efficiency and security. *Special Issue on Multimedia Security of ACM Multimedia System*, 2003.

[116] I. Djurovic, S. Stakovic, and I. Pitas. Digital watermarking in the fractional fourier transformation domain. *Journal of Network and Computer Applications*, 2001.

[117] W. M. Dietl and A. Uhl. Robustness against unauthorized watermark removal attacks via key-dependent wavelet packet subband structures. *Proceedings IEEE International Conference on Multimedia and Expo (ICME04)*, June 2004.

[118] J. Deamon and V. Rijmen. The design of rijndael. AES - the Advanced Encryption Standard. *Springer - Verlag*, 2002.

[119] B. Chen and G. Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory*, 47(4), May 2001.

[120] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn. Attacks on copyright marking systems. *Information Hiding, Second International Workshop, IH'98*, April 1998.

[121] F.A.P. Petitcolas. Watermarking schemes evaluation. *IEEE Signal Processing*, 17(5):58–64, September 2000.

[122] N. Terzija, M. Repges, K. Luck, and W. Geisselhardt. Digital image watermarking using discrete wavelet transform: performance comparison of error correction codes. *Proceedings of IASTED 2002*, September 2002.

[123] N. Terzija, M. Repges, K. Luck, and W. Geisselhardt. Impact of different reed-solomon codes on digital watermarks based on DWT. *Multimedia and Security Workshop at ACM Multimedia 2002*, 2002.

[124] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Trans. on Image Processing*, 6(12):1673–1687, 1997.

[125] F. Battisti, M.Cancellaro, M. Carli, G. Boato, and A. Neri. Watermarking and encryption of color images in the fibonacci domain. *Proceedings of IST/SPIE Electronic Imaging*, January 2008.

[126] I.J.Cox, G. Döerr, and T.Furon. Watermarking is not cryptography. *Proceedings of 5th International Workshop on Digital Watermarking*, pages 1–15, November 2006.

[127] Y. Xing and J. Tan. A color watermarking scheme based on block-SVD and Arnold transformation. *Proceedings of the Second Workshop on Digital Media and its Application in Museum & Heritage*, pages 3–8, October 2007.

[128] C.Q. Yin, L. Li, A.Q. Lv, and L. Qu. Color image watermarking algorithm based on DWT-SVD. *Proceedings of IEEE International Conference on Automation and Logistics*, pages 2607–2611, August 2007.

[129] R. Ridzon and D. Levicky. Robust digital watermarking in color images. *Proceedings of 15th International Conference on Systems, Signals and Image Processing*, pages 425–428, June 2008.

[130] G. Sun and Y. Yu. DWT based watermarking algorithm of color images. *Proceedings of 2nd IEEE Conference on Industrial Electronics and Applications*, pages 1823–1826, May 2007.

[131] N. Ahmidi and R. Safabakhsh. A novel DCT-based approach for secure color image watermarking. *Proceedings of International Conference on Information Technology: Coding and Computing*, 2:709–713, April 2004.

[132] X. Li and X. Xue. Improved robust watermarking in DCT domain for color images. *Proceedings of the 18th International Conference on Advanced Information Networking and Application*, pages 53–58, March 2004.

[133] Tsz Kin Tsui, Xiao-Ping Zhang, and D. Androutsos. Color image watermarking using multidimensional Fourier transforms. *IEEE Transactions on Information Forensics and Security*, 3(1):16–28, March 2008.

[134] V.E. Hoggatt. Fibonacci and Lucas numbers. *The Fibonacci Association*, 1969.

[135] J. Liu, X.Niu, and W.Kong. Image watermarking based on Singular Value Decomposition. *Proceedings of International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 457–460, December 2006.

[136] N. Ponomarenko, V. Lukin, K. Egiazarian, J. Astola, M. Carli, and F. Battisti. Color image database for evaluation of image quality metrics. In *Proceedings of International Workshop on Multimedia Signal Processing*, October 2008.

[137] N. Ponomarenko, V. Lukin, K. Egiazarian, J. Astola, M. Carli, and F. Battisti. Tampere Image Database TID2008. *http://ponomarenko.info/tid2008.htm.*

[138] G. R. Cole, T. Hine, and W. Mcllhagga. Detection mechanisms in L-, M-, and S-cone contrast space. *Journal of the Optical Society of America*, 10(1):38–51, 1993.

[139] Fabien A. P. Petitcolas, Ross J. Anderson, and Markus G. Kuhn. Attacks on copyright marking systems. pages 218–238. Springer-Verlag, 1998.

[140] James Nechvatal, Elaine Barker Lawrence Bassham, Morris Dworkin, James Foti, and Edward Roback. Report on the development of the Advanced Encryption Standard AES. *Journal of Research of the National Institute of Standards and Technology*, 106(3):511–577, 2001.