A Family of Stereoscopic Image Compression Algorithms Using Wavelet Transforms

by

Mohd Yunus Nayan

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

Doctor of Philosophy

Department of Computer Science

Loughborough University

December 2004

Supervisors: i. Dr Helmut Bez
ii. Dr Eran Edirisinghe
Director of Research: Professor Roy S. Kalawsky

Abstract

A family of stereoscopic image compression algorithms using wavelet transforms

Mohd Yunus Nayan, December 2004

With the standardization of JPEG-2000, wavelet-based image and video compression technologies are gradually replacing the popular DCT-based methods. In parallel to this, recent developments in autostereoscopic display technology is now threatening to revolutionize the way in which consumers are used to enjoying the traditional 2D display based electronic media such as television, computer and movies. However, due to the two-fold bandwidth/storage space requirement of stereoscopic imaging, an essential requirement of a stereo imaging system is efficient data compression.

In this thesis, seven wavelet-based stereo image compression algorithms are proposed, to take advantage of the higher data compaction capability and better flexibility of wavelets. In the proposed CODEC I, block-based disparity estimation/compensation (DE/DC) is performed in pixel domain. However, this results in an inefficiency when DWT is applied on the whole predictive error image that results from the DE process. This is because of the existence of artificial block boundaries between error blocks in the predictive error image. To overcome this problem, in the remaining proposed CODECs, DE/DC is performed in the wavelet domain. Due to the multiresolution nature of the wavelet domain, two methods of disparity estimation and compensation have been proposed. The first method is performing DE/DC in each subband of the lowest/coarsest resolution level and then propagating the disparity vectors obtained to the corresponding subbands of higher/finer resolution. Note that DE is not performed in every subband due to the high overhead bits that could be required for the coding of disparity vectors of all subbands. This method is being used in CODEC II. In the second method, DE/DC is performed in the wavelet-block domain. This enables disparity estimation to be performed in all subbands simultaneously without increasing the overhead bits required for the coding disparity vectors. This method is used by CODEC III.

However, performing disparity estimation/compensation in all subbands would result in a significant improvement of CODEC III. To further improve the performance of CODEC III, pioneering wavelet-block search technique is implemented in CODEC IV. The pioneering wavelet-block search technique enables the right/predicted image to be reconstructed at the decoder end without the need of transmitting the disparity vectors. In proposed CODEC V, pioneering block search is performed in all subbands of DWT decomposition which results in an improvement of its performance. Further, the CODEC IV and V are able to perform at very low bit rates (< 0.15 bpp). In CODEC VI and CODEC VII, Overlapped Block Disparity Compensation (OBDC) is used with & without the need of coding disparity vector. Our experiment results showed that no significant coding gains could be obtained for these CODECs over CODEC IV & V.

All proposed CODECs in this thesis are wavelet-based stereo image coding algorithms that maximise the flexibility and benefits offered by wavelet transform technology when applied to stereo imaging. In addition the use of a baseline-JPEG coding architecture would enable the easy adaptation of the proposed algorithms within systems originally built for DCT-based coding. This is an important feature that would be useful during an era where DCT-based technology is only slowly being phased out to give way for DWT based compression technology.

In addition, this thesis proposed a stereo image coding algorithm that uses JPEG-2000 technology as the basic compression engine. The proposed CODEC, named RASTER is a rate scalable stereo image CODEC that has a unique ability to preserve the image quality at binocular depth boundaries, which is an important requirement in the design of stereo image CODEC. The experimental results have shown that the proposed CODEC is able to achieve PSNR gains of up to 3.7 dB as compared to directly transmitting the right frame using JPEG-2000.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Dr Eran Edirisinghe and Dr Helmut Bez, for their constant encouragement, guidance, advice, support, and patience shown throughout my research program. I thank my director of research, Prof. Roy Kalawsky, for his time, effort and suggestions. I also thank my fellow members of the Digital Imaging Research Group and Parallelism, Algorithms and Architecture Research Centre (PARC), with whom I had many useful discussions.

My sincere thanks to secretary of the Computer Science Department, Mrs. Judith Poulton, for her cordiality and friendly help regarding department matters.

I am eternally indebted to my family members, my wife, Dayang Rohaya Awang Rambli, my children, Muhammad Faiz, Nur Khairunissa Athirah and Nur Syarafana Amirah, for love and understanding and above all for all the sacrifices made on my behalf during my studies. Also, to both my parents and mother-in-law for their constant prayers, their love, and encouragement which I never be able to repay.

Last but not least, to my employer, University Technology Petronas, for giving me the opportunity to pursue my dream.

Mohd Yunus Nayan   *1st December 2004*

iv

TABLE OF CONTENTS

CHAPTER 1 – An overview

Chapter 2 – Fundamental concepts, issues, and developments related to stereo image/video coding

Chapter 3 – Discrete Wavelet Transform & its application to image processing

Chapter 4 – Wavelet block-based stereo image coding

Chapter 5 -- PBDCPC and OBDC in wavelet-based coding of stereo image pairs

Chapter 6 – Stereoscopic extension of JPEG-2000

List of Figures & Tables

Figures

Tables

Abbreviations & Notations

2D	Two Dimensional
3D	Three Dimensional
3DTV	Three Dimensional Television
ADC	Adaptive Disparity Compensation
AMC	Adaptive Motion Compensation
bpp	Bit per pixel
CODEC	EnCOder/DECoder
dB	Decibel
DC	Disparity Compensation
DCD	Disparity Compensated Difference
DCT	Discrete Cosine Transform
DCTDP	Disparity Compensated Transform Domain Predictive Coding
DE	Disparity Estimation
DFD	Displaced Frame Difference
DPCM	Differential Pulse Coding Method
DV	Disparity Vector
DWT	Discrete Wavelet Transform
EBCOT	Embedded Block Coding Optimized Truncation
EZW	Embedded Zerotree Wavelet
FSMB	Fixed Size Block Matching
FT	Fourier Transform
GS	Gram-Schmidt
HH	High-High
HL	High-Low
HVS	Human Visual System
IDCT	Inverse Discrete Cosine Transform
IDWT	Inverse Discrete Wavelet Transform
JBIG	Joint Binary Image Group
JPEG	Joint Photographic Expert Group
LL	Low-Low
LH	Low-High

LS	Lossless
LSB	Least Significant Bit
MAE	Minimum Absolute Error
MPEG	Motion Picture Expert Group
MSE	Mean Square Error
MSB	Most Significant Bit
MRF	Markov Random Field
NLS	Near Lossless
OBDC	Overlapped Block Disparity Compensation
OBMC	Overlapped Block Motion Compensation
PBDCPC	Pioneering Block-based Disparity Compensated Predictive Coding
PB	Pioneering Block
PBS	Pioneering Block Search
PEI	Predictive Error Image
PSNR	Peak Signal Noise ratio
RASTER	Rate scalable stereo image coding
RD	Rate-Distortion
R/G/B	Red/Green/Blue
SOSU	Sequential Orthogonal Subspace Updating
SNR	Signal Noise Ratio
SPHIT	Set Partitioning in Hierarchical Trees
SPT	Subspace Projection Technique
STFT	Short Time Fourier Transform
WP1	Work Package 1
WP2	Work Package 2
WP3	Work Package 3
WT	Wavelet Transform
WTCQ	Wavelet Trellis Coded Quantization

1 An overview

1.1 Introduction

In a two-dimensional (2D) photograph, we can quite easily perceive the spatial relationship amongst objects portrayed, due to the presence of monocular cues such as occlusion, relative size, shading, and shadows. However, these cues can only provide relative distances between objects, but not the actual distance between or depth to objects. For example, in figure 1.1, we can identify which object is closest to the camera, but cannot determine the actual distance between objects nor the distance between a given object and the camera.

Figure 1.1 Illustration of monocular cues in 2D image

Three-dimensions (3D) enables us to perceive depth in two aspects: actual distance of an object from the viewer/camera (i.e. depth of an object) and actual distance between objects in the scene. In general, 3D depth perception is based on various depth cues: monocular cues such as light, shading, relative size, motion, occlusion, texture gradient and binocular cues such as, disparity, etc. However, one of the most efficient cues is the binocular depth, which is based on the fact that the depth is obtained by

viewing a scene from two slightly different viewing positions. Therefore, objects are seen more sharply in 3D than in 2D images as 3D viewing enables humans to perceive clear contours between objects and background, with the aid of binocular depth information. For example, 3D perception eases our daily tasks such as driving a car or using stairs. In 3D perception, two slightly different 2D perspective views of the same scene are projected onto the retinas of our eyes. These 2D images are processed within the brain and are fused into a single image, allowing us to sense the depth of objects within the 3D scene. This process is called stereoscopic depth perception, stereopsis or binocular vision [87]. We can create a pair of images representing a 3D scene or object, where one image is as seen by our left eye (left image) and the other is as seen by the right eye (right image) by using two identical cameras, horizontally displaced by approximately 65 mm (2.6 in). This distance resembles the typical interocular distance of human eyes. The captured images are called a stereo image pair. In 1832, Sir Charles Wheatstone, the inventor of the first stereoscope, showed that such an image pair when viewed using a Wheatstone Viewer, fused into a single 3D image, visualising binocular depth [83].

In the 1990s, advances in telecommunication, networking and multimedia technologies have enabled real-time multimedia communication in a wide area of applications such as videoconferencing where telepresence is desirable, in applications where precise remote handling is required such as telerobotics, and in telesurgery, virtual reality, space exploration, and entertainment where realism is of paramount important [184]. This has renewed the interest in 3D display systems. Decades of research have produced many 3D display systems including volumetric, holographic and stereoscopic technologies. Stereoscopic display systems are more popular amongst the above as they resemble the human visual system. The recent developments in autostereoscopic displays, which allow direct stereo image viewing without the help of viewing glasses/head mounted devices, have further increased the interest in the use of stereo imaging in 3D applications.

Apart from the capture and display of stereoscopic images, an additional challenge faced by the technology is the two-fold increment required in storage space/

bandwidth as compared to monocular images. Advances in stereoscopic display technology, which allows the use of stereo images in look around systems [68], further worsens this problem. This has created considerable interest among researchers in compression of stereo images and video.

1.2   Brief history: Motivation for stereo image compression

The ability to portray a 3D scene has attracted a lot of interest since ancient times. This can be seen from ancient drawings, which use monocular cues in trying to add a sense of depth to the scene represented [83].

The idea of binocular vision and depth perception commenced with Euclid in 300 B.C., when he realized that depth perception is obtained when each eye received simultaneously one of two similar images [83]. Those who agreed with Euclid, included, Ptolemy in the 2nd century A.D., Alhazen in the 10th century, Roger Bacon, John Peckham and Vitello in the 13th century and Aguilonius and Kepler in the 17th century [218]. However, during this period of time, researchers focused on explaining the geometrical aspect of vision rather than the use of it [83]. Despite these early attempts simple facts about binocular stereoscopic vision were not appreciated until about 160 years ago. Leonardo da Vinci realized that to capture reality in a painting, two paintings, are required, the views seen by each of the artist's eyes. Giovanni Battista della Porta practised this technique in his stereoscopic drawings at the turn of the 17th century [82]. However it remains a mystery how these early stereoscopic drawings were viewed. Later, hand-drawn stereo image pairs were viewed, using a Wheatstone Viewer [84] to create an illusion of 3D. The invention of photography in 1839 eased the creation of stereo image pairs; however, stereo photographs became popular only in the late 19th century after the creation of the refined and reduced version of the Wheatstone Viewer [84]. Ever since, 3D has been an important part of history of photography and film.

The first 3D film showing scenes of New York and New Jersey premiered in New York City in 1915 [218]. Subsequently, 3D imaging was revitalized through the 30's and 40's. In 1933, the Tru-Vue Company introduced a stereoscope using a 35mm film [218]. In 1939, Chrysler Motors paved the way for the projection of full-colour 3D film by showing a 3D film using a polarized material without colour distortion [218]. With the invention of the television in 1939 (London), electronic versions of 3D (based on the anaglyph method – spectators wear a pair of glasses with a red filter covering one eye and a green filter covering the other eye.) were prompted in 1942 [218]. However, in the 1950s, due to disadvantages of anaglyphs methods, 3D survived mostly in the cinema industry rather than TV. During this era, a 3D film was one of the best hopes for the movie industry to reverse the decline in the number of viewers lost to TV audiences. However, these 3D films have also not been widely accepted due to its drawback that viewers have to wear uncomfortable special glasses, which often lead to headaches [68].

However, 3D display technology that doesn't require special glasses, or the more modern polarized version viewing glasses used for IMAX movies, started to appear in application areas ranging from medicine to defence in the early 1990s. "Every month it seems like there are some news about 3D displays," says Chris Chinnock, president of analyst firm Insight Media (Norwalk, CT). "My sense is they're starting to get a little bit of traction in consumer applications."[184]

For several decades, efforts have been made to develop practical 3D systems but 3D technologies including holography, and volumetric displays have not met the demands for realistic displays. Most of the efforts have gone into two-view stereoscopic systems with special glasses, which have limitations in terms of providing high quality or comfortable viewing. Such systems have proved effective in some scientific applications and, to a limited extent, in 3D wide screen cinema, such as IMAX(R) 3D [218]. However, through the 90's, considerable advances in 3D display technologies and innovations in the related fields are opening a new way for the 3D systems without special glasses [184]. This type of stereoscopic display is called an autostereoscopic display. In autostereoscopic displays [84], a parallax barrier added to

a LCD screen delivers separate images to the right and left eyes. So it is the display instrument that is effectively *wearing* the glasses and not the user.

Some other state-of-the-art development around the globe worth mentioning include; Sharp Corp. (Osaka, Japan) announced a mobile phone with 3D LCD stereoscopic display capability in November 2002, released an improved version in June 2003, and is presently working on a 3D computer monitor. In March 2003, the company joined with Sanyo Electric Co. Ltd. (Osaka, Japan), Sony Corp. (Tokyo, Japan), NTT Data Corp. (Tokyo, Japan), and Itochu Corp. (Tokyo, Japan) to form the 3D Consortium to promote the development of 3D displays [84,184,218].

3D imaging could be very appealing for marketing. "If one can somehow create real 3D images of a car or whatever you want to buy on the Internet so that you can look at it in much more of a 3D mode, that's a killer application," [184]. Home entertainment might be the other killer application, says Chinnock. [184] Stereographics (San Rafael, CA) is already marketing a 42-in. autostereoscopic plasma screen, aimed at corporations who want an eye-catching display; although at over USD15,000 it probably won't appear in your average living room in the near future. In United Kingdom, research on 3DTV has been initiated by several projects, such as Television System Technologies (ATTEST) Project [163] at De Montfort University aimed at developing 3D display system targeted specifically at the domestic television market and Cambridge Autostereoscopic 3D Display Project [23] at Cambridge University aimed at developing an autostereoscopic 3D display which generates true 3D pictures in real time. In other part of Europe, projects such as COST230 and DISTIMA, which aimed to develop a system for capturing, coding, transmitting and presenting digital stereoscopic image sequences. The project had been followed up by another project, PANORAMA, which aims to enhance the visual information exchange in telecommunications with 3D telepresence [184]. Therefore, stereoscopic display systems had made vast developments, starting from primitive Wheatstone Viewer to modern autostereoscopic display systems. This has increased the potential applications of stereoscopic images/videos in a wide application area such as telepresence videoconferencing, telemedicine (remote guided surgery) , telerobotics

(remote control, autonomous navigation, surveillance, space exploration), entertainment (interactive 3DTV and cinema) and Virtual Reality.

The obvious price one has to pay for this increased realism in 3D digital imaging systems is, at least the doubling of data size for storage and transmission, as compared to monocular image cases. In general, the problem of increased data can be solved by: (i) increasing data-transmission bandwidth, (ii) mass-storage density (iii) improving channel utilization with efficient protocol or/and (iv) reducing the source itself using efficient compression techniques. Despite rapid progress in (i), (ii), and (iii) the demand continues to outstrip the capabilities of the available technologies. Thus, stereo image/video compression has been attracting considerable attention over last few years.

Analogous to other coding scenarios, compression for stereo images can be achieved by taking advantage of redundancies in the source data, e.g., spatial and temporal redundancies for monocular images and videos. A simple solution for compression is to use independent coding for each image/video with existing compression standards such as JPEG-2000 or MPEG-4. However, in the case of stereo images/videos, further compression can be achieved by exploiting an additional source of redundancy that stems from the similarity, i.e. the strong "binocular redundancy" between two images in a stereo pair, due to stereo camera geometry.

Over the last few years, several successful stereo image/video coding techniques have been proposed by researchers worldwide. The discovery of new compression technologies such wavelets and fractals, enable further improvements to existing algorithms and the introduction of novel stereo image coding techniques. The new developments in the stereo imaging systems and the greater usage of stereo image/video in 3D applications from medicine to entertainment are the main motivation for this research work. As stereo image compression is one of the key players in the increasing success of stereo imaging systems, the research presented in this thesis attempts further contributions in this direction.

1.3   Why use Discrete Wavelet Transform technology and baseline-JPEG standard architecture in this research?

In the past, researchers have successfully proposed a number of stereo image coding techniques (discussed in section 2.5). Due to the popularity/widespread use of Discrete Cosine Transform (DCT)-based baseline-JPEG standard [62] in monocular image compression, most of these algorithms use DCT technology to compress the reference and residual frames. The DCT-based technology is implemented by transforming non-overlapping 8x8 blocks of pixels into a set of spatial frequency coefficients, which are later exploited for redundancy. With block coding, errors tend to line up and accumulate along block boundaries, which are visible as discontinuities at block boundaries particularly at higher compression (low bit-rates). This visible artifacts are called the blocking artifacts, which is a common problem in any DCT-based coding scheme that specifically works at medium-low bit rates. Another drawback of the DCT-based scheme is the distracting lines and patterns known as "mosquito noise". These artifacts are caused when a large section of the higher frequency coefficients are completely lost for a given 8x8 pixel blocks, due to quantization. Although, there exist methods proposed to reduce blocking artifacts [218], they often cause the blurring of normal image edges and are thus unable to completely alleviate blocking artifacts due to discontinuities between blocks [218]. Furthermore, the additional complexity of the de-blocking algorithms does not always correlate well with performance requirements of the compression algorithms and are therefore not often used. However, in spite of these drawbacks, the baseline-JPEG standard has a very simple architecture for hardware implementation, which is one of the factors for its popularity. This motivates us to use a similar architecture in our research.

To cater for the continual expansion of multimedia and Internet applications, image compression technologies have continually improved and evolved in functionality and efficiency. Worldwide efforts in this direction have recently led to the standardization of wavelet-based JPEG-2000 [1], which is expected to complement the DCT-based

image compression technology offered by the baseline-JPEG standard. The DWT has become popular in image/video compression for several reasons:

- The flexibility of the wavelet transforms. This flexibility makes it very efficient in representing images/videos, which are in general non-stationary in nature.

- Wavelets have high decorrelation and energy compaction efficiency.

- The blocking artifacts are absent in wavelet-based CODECs, resulting in subjectively pleasing reconstructed image/video quality.

- Aliasing distortion can be reduced significantly with proper choice of wavelet filters.

- The basis functions match the human visual system (HVS) characteristics.

In addition to the above, the wavelet-based coding schemes, provide many useful features (e.g., progressive coding, region of interest, signal-noise ratio and resolution scalabilities) and have been shown to provide better rate-distortion performance compared to DCT-based schemes [226]. For example, the recent wavelet-based image coding standard, JPEG-2000, outperforms JPEG in terms of compression efficiency while possessing many other attractive features. In video coding, it has been shown that wavelet-based coding can achieve comparable or better coding performance than MPEG-4 coding [96]. Therefore every effort must be taken to explore the possibility of using DWT technology in stereo imaging as it is expected that wavelet-based coding will perform well for stereo image coding.

To this effect, few wavelet-based stereo image compression algorithms have been reported in literature [96,155] during the past few years. These algorithms are a straightforward application of the established monocular DWT-based image compression techniques to stereo imaging, i.e. the reference frame and the pixel domain residual frame are coded directly using the monocular DWT-based compression technique. This thesis proposes a family of novel wavelet-based stereo image coding algorithms, which maximise the flexibility and benefits offered by wavelet transform technology when applied to stereo imaging. In addition the use of a

baseline-JPEG coding architecture would enable the easy adaptation of the proposed algorithms within systems originally built for DCT-based coding. This is an important feature that would be useful during an era where DCT-based technology is only slowly being phased out to give way for DWT and Integer Transform based compression technology.

1.4   Research aim and objectives

1.4.1 Aim

"Design and implement a family of novel algorithms based on wavelet transform technology for efficient compression of stereo image pairs"

1.4.2 Objectives

- Develop a stereo image coding algorithm conforming to baseline-JPEG architecture, but replacing DCT with DWT.
- Develop a stereo image coding algorithm that uses disparity compensated predictive coding in multiresolution DWT domain.
- Develop a stereo image coding algorithm that uses Pioneering Block (PB)-based disparity compensated predictive coding in multiresolution DWT domain.
- Develop a stereo image coding algorithm that uses Overlapped Block Disparity Compensation (OBDC) based disparity compensated predictive coding in multiresolution DWT domain.
- Develop a stereo image coding algorithm based on JPEG-2000 technology that is capable of scalable stereo image coding.

1.5 Research methodology

The research was carried out over a period of 30 months. The areas covered can be categorised into four groups as follows:

1.5.1 Literature survey and selecting the benchmark algorithms

This involved a detailed literature survey on existing techniques for stereo image compression (see section 2.5) and the identification of two algorithms [87,161] to be used as benchmarks for testing the novel algorithms.

The first algorithm selected as a benchmark was, the Disparity Compensated Transform domain Predictive Coding (DCTDP) algorithm [161], proposed by M.G. Perkins in 1992 (see section 4.2), which is widely regarded as one of the best algorithms that gives optimal coding results within a wide range of bit rates and for a wide variety of stereo image pairs, representing, indoor, outdoor, and computer generated scenes. In addition, this algorithm is widely used in the literature as a benchmark and our use of it in the same capacity would enable easy comparison of the performance of the proposed algorithms with that of other state-of-the-art techniques.

The second algorithm chosen as a benchmark is the Pioneering Block-Based Disparity Compensated Predictive Coding (PBDCPC) algorithm [87] proposed by J.Jiang, E.A. Edirisinghe in 1997 (see section 5.2). This algorithm has the unique capability of not requiring the transmission of disparity information as overhead and has thus been proved to be the best existing algorithm [213] for very low bit-rate stereoscopic image coding.

The performances of the proposed algorithms were tested on a collection of widely used test stereo image pairs, downloadable from popular web sites dedicated to stereo image compression research. Further, it was decided that the design and testing of the novel algorithms be limited to stereo image pairs obtained using parallel axis

geometry for reasons explained in detail in section 2.2.1. The experiments were also limited to greyscale image pairs, as colour stereo image compression can be approximately regarded as compression of multiple R/G/B scale stereo image pairs. It was also decided to use the left frame of the stereo image pair as the reference, solely for the consistency and ease of explanation of the theoretical and design aspects.

After the selection of benchmarks and the preliminary set up of test conditions/restrictions, the research work was carried out, under three work packages, as follows.

1.5.2 WP1: Developing a family of stereo image compression algorithms based on DWT technology and baseline-JPEG architecture (Chapter 4)

The benchmark DCTDP algorithm (section 4.2, or [161]) addresses the stereo image compression issue by using the baseline-JPEG still image compression standard as the main compression engine. Therefore it has the advantage of consisting of a simple, standard coding architecture. However it has been proved that DWT technology has a higher data compaction capability and better flexibility in CODEC design as compared to DCT technology. Therefore within this work package the DCT was effectively replaced by DWT, keeping the basic coding structure of baseline-JPEG intact. Further within the same basic design mentioned above, several different disparity compensation based predictive coding algorithms based on both pixel domain and multiresolution wavelet transform were designed, tested and critically analysed.

1.5.3 WP2: Investigating the use of Pioneering Block and Overlapped Block, disparity compensation in the CODECs designed in WP1. (Chapter 5)

Pioneering Block (PB)-based predictive coding (see section 5.2, or [87]) has the unique ability of not needing the transmission of the disparity vectors for the reconstruction of the predicted frame at the decoder end. Therefore it is a highly desirable disparity compensation technique in low bit-rate stereo image coding. The

original PB-based predictive coding algorithm was designed to be used in the pixel domain. Its use in multiresolution DWT domain has not been previously investigated, but promises many advantages over its original implementation. Therefore the first phase of this work package involved the introduction of PB-based predictive coding to the best CODEC designed under WP1.

Overlapped Block Disparity Compensation (OBDC) was originally used in video coding and has received much interest within the stereo image compression research community since its introduction in [220]. OBDC method reduces blocking artifacts by using multiple disparity vectors to estimate a block by linearly combining the prediction, based on the disparity of the block and its neighbouring blocks, as opposed to Fixed Sized Block Matching (FSBM) which depends only on the disparity of the block itself. Within the second phase of WP2, the use of OBDC in multiresolution DWT domain was investigated.

1.5.4 WP3: Extension of JPEG-2000 technology to stereo image coding (Chapter 6)

JPEG-2000 provides an entirely new way of compressing images based on the wavelet transform, with a rate-distortion performance superior to that of baseline-JPEG, especially at low bit-rates. In addition, the DWT which provides multiresolution decomposition of an image, enables the inclusion of new features such as resolution and fidelity scalability, region of interest scalability, random access and etc, which are of critical importance in present day applications. Within this work package a flexible and highly functional stereo image CODEC that uses JPEG-2000 technology was designed, implemented and tested.

1.6   Main contributions of research

The following original contributions have been made. The resulting conference and journal papers are included in Appendix I, and are referenced as A1, A2,..., A8.

1. **Nayan, M.Y.**, Edirisinghe, E.A. and Bez, H.E., "Two Novel Wavelet-block Based Stereo Image Compression Algorithms", *Proceedings of the Second IASTED International Conference: Visualisation, Imaging and Image Processing*, ACTA Press, 2002, pp. 251-256, ISBN 0-88986-345-3.

2. **Nayan, M.Y.**, Edirisinghe, E.A. and Bez, H.E., "Baseline JPEG-Like DWT CODEC for Disparity Compensated Residual Coding of Stereo Images", *Proceedings of Eurographics UK*, IEEE Computer Society Press, De Montfort University, Leicester, UK, 2002, pp. 67-74, ISBN: 0-7695-1518-5.

3. **Nayan, M.Y.**, Edirisinghe, E.A. and Bez, H.E., "Wavelet-based Stereo Image Pair Coding with Pioneering Block Search", *Recent Trends in Multimedia Information Processing*, Word Scientific, Proceedings of the 9th International Workshop on Systems, Signals and Image Processing, 2002, pp. 239-243, ISBN: 981-238-243-7.

4. Edirisinghe, E.A., **Nayan, M.Y.** and Bez, H.E., "A Pioneering Block Based Stereo Image CODEC in Wavelet Domain", *Proceedings of the SPIE International Conference on Stereoscopic Displays and Applications XIV*, 5006, January 2003, pp.399-406, ISBN: 0-8194-4806-0

5. Rajkumar, G., **Nayan, M.Y.** and Edirisinghe, E.A., "RASTER: A JPEG-2000 Stereo Image CODEC", *Proc. of the IASTED International Conference on Visualization, Imaging & Image Processing*, Spain, September 2003, pp. 233-238, ISBN: 0-88989-382-2.

6. **Nayan.M.Y.**, Edirisinghe, E.A. and Bez, H.E., "Transform Domain Overlapped Block Disparity Compensation in Wavelet Coding of Stereo Image Pairs" , *Proceedings of the 2003 joint conference of the 4th International Conference on Information, Communications and Signal Processing and the 4th Pacific-Rim Conference on Multimedia* , IB2.1 , Singapore, December 2003, ISBN: 0-7803-8186-6 .

7. Edirisinghe, E.A., **Nayan, M.Y.** and Bez, H.E., "A Wavelet Implementation of the Pioneering Block-Based Disparity Compensated Predictive Coding Algorithm for Stereo Image Pair Compression", *International Journal on Signal Processing: Image Communication*, 19(1), January 2004, pp. 37-46, ISSN: 0923-5965,
    [WWW] Available from: http://authors.elsevier.com/sd/article/S0923596503001085.

8. **Nayan, M.Y.**, Edirisinghe, E.A., Bez, H.E., "A Family of Wavelet Transform Based Stereo Image Compression Algorithms", Accepted for publication in Int. Journal. of Computers & Applications.

1.7   Thesis overview

This thesis consists of seven chapters.

In chapter 2, we explain the fundamental concepts related to stereo image/video coding. It also consists of a comprehensive literature survey of existing stereo image/video compression techniques. The chapter concludes with a discussion on historical development of stereoscopic display technology.

In chapter 3, we discus the basic theory of wavelets and wavelet transforms, its application in image processing, coefficient extension problem of DWT, three well known DWT-based image compression algorithms and a comparison between the performance of baseline-JPEG and JPEG-2000 in still image coding.

Chapter 4, 5, and 6 details the main contributions of this research. Each chapter describes the proposed CODECs, their design considerations, implementation, testing and a detailed analysis/comparison in relation to the chosen benchmark.

Chapter 7 concludes with a discussion on directions for further developments and extensions of the proposed algorithms.

2 Fundamental concepts, issues and developments related to stereo image/video coding

2.1 Overview

The recent increase in demand for more realistic viewing media, in conjunction with the advances in stereoscopic display technology (see section 2.6), are at present driving the use of stereo imaging in diverse areas of applications such as telepresence videoconferencing, telerobotics, telesurgery, space exploration, virtual reality, entertainment etc. In general, objects are seen more sharply in 3D than 2D images as 3D images enable humans to visualize clear contours between objects and the background using binocular depth information. However, the cost for this increased realism, needs the presence of two/more images as against one image that is sufficient to describe a 2D scene. This brings upon a fundamental problem of data compression necessary for the reduction of transmission bandwidth and/or storage space. In general, efficient transmission can be achieved by exploiting spatial and temporal redundancy in each image/frame as used for monocular image/video. In the case of stereo image pairs, the efficiency can be further improved by exploiting binocular dependency (redundancy) in stereo pairs. Meanwhile, in wavelet domain, spatial dependency between subbands have been exploited for efficient coding of wavelet coefficients, for example, in the EZW algorithm proposed by Shapiro [185]. Similarly, such inter-subband relationships between a stereo image pair transformed into the wavelet domain can be exploited within a disparity estimation and compensation process, for efficient stereo image/video compression.

In this chapter, we discuss the basic concepts of stereo image/video coding and include a detailed survey of existing stereo image/video compression and display technologies.

2.2 Basic concepts and assumptions

This section introduces the basic concepts and assumptions used in stereo image/video coding. These are used in subsequent chapters to explain the operation and performance of proposed algorithms. The assumptions made are typical to the stereoscopic imaging research community.

2.2.1 Camera geometry for stereo image photography

One way to generate a stereo image pair/video is to arrange two cameras such that their axes are in parallel, with the same interocular distance as the human eyes. On the average, the separation of the human eyes is about 65 mm (2.6 in) [196]. By placing the two cameras this distance apart, with coplanar image sensors, will modelled the human visual system (HVS), with relation to the difference in perspective between the two viewpoints [196]. Figure 2.1 shows the basic structure for stereo image formation and *parallel stereo camera geometry* [215]. The centre of the lens is called the camera *focal centre* and the axis extending from the focal centre is referred to as the *focal axis*. The line connecting the focal centres is called the *baseline, b*. The plane passing through one object point and two focal centres is named as *epipolar plane*. The intersection of two image planes with an epipolar plane makes the *epipolar line*. Let (X,Y,Z) denote the real world coordinates of a point. The point is projected onto two corresponding points, (x_l, y_l) and (x_r, y_r), in the left and right images (resembling left and right retinas). The disparity is defined as the difference vector between two points in the stereo images, corresponding to the same point in an object, i.e.,

$$\mathbf{v} = (x_l - x_r, y_l - y_r). \qquad (2.1)$$

A second way to generate a stereo image pair/video is to use the *convergence camera geometry*; where the two cameras with distance 65 mm (2.6 in) apart are slightly rotated towards each other so that the focal axes intersect at the point of convergence. However, convergence camera geometry causes vertical misalignment of

corresponding left and right image points. This misalignment or vertical parallax, is known to be a source of discomfort for the viewer and is called the keystone effect [110]. The keystone effect is due to the image sensors of the cameras being located in different planes, thus, making the left image appear larger at one side than the other side while the right viewpoint appears shorter at one side than the other side. For example, the projected squares are no longer the same size and become trapezoids or other sorts of quadrilaterals. Image processing, algorithms for eliminating the keystone effect have been proposed [116], but it is most easily avoided by using the parallel axes camera setup. The parallel axis camera geometry requires only uni-directional horizontal search and no vertical disparity.
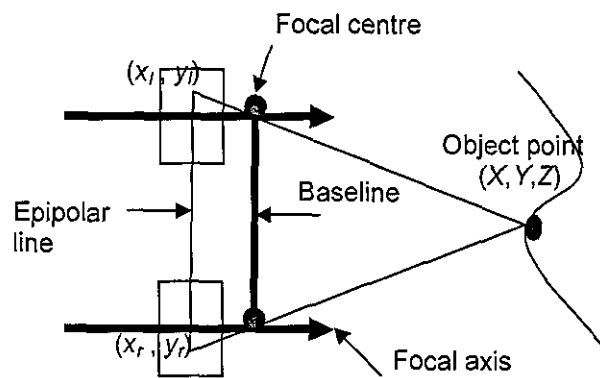
Figure 2.1 Parallel camera geometry

All stereo image pairs used within this thesis for experimentation have been captured by parallax camera geometry described above. Therefore they satisfy the following conditions:

- The search area for correspondence is restricted on the epipolar lines and there is no vertical search.
- Theoretically, the pixel values of corresponding pixels should be the same.
- A pixel in the left image corresponds to only one pixel in the right image as one point in an object is projected onto only one point (pixel) in each image.

- Under the assumption of smooth object surfaces, the disparity varies smoothly in most parts of the image except at object boundaries and occlusion areas.

- The geometry resembles the geometry of the human visual system.

2.2.2 Disparity
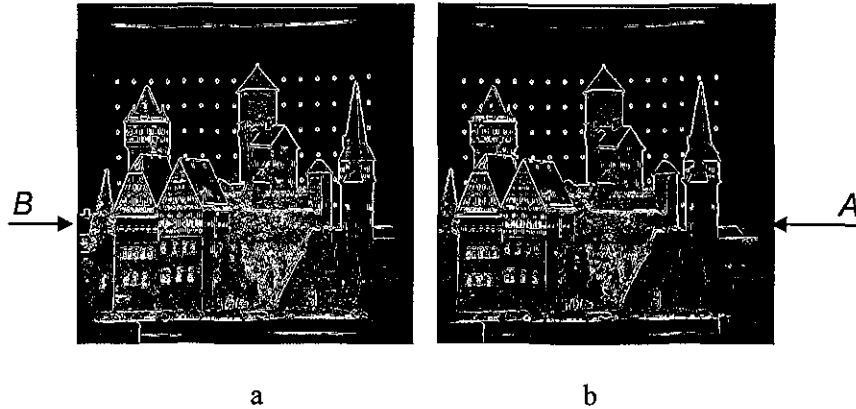
Figure 2.2 (a) Edges of the left image (b) Edges of the right image

The term disparity was first introduced in the human vision literature to describe the difference in location of corresponding features seen by the left and right eyes [84]. In computer vision, disparity is often treated as synonymous with inverse depth [84]. More recently, several researchers have defined disparity as a three-dimensional projective transformation of 3D space (X, Y, Z). The enumeration of all possible matches in such a generalized disparity space can be easily achieved with a plane sweep algorithm [84], which for every disparity d projects all images onto a common plane using a perspective projection. In this study, however, since all images used are taken on a linear path, with the optical axis perpendicular to the camera displacement (parallax camera geometry), the classical inverse depth interpretation will suffice.

Figure 2.2 (a), represents the edges of the left eye view of a stereoscopic scene. Note that more of the left area of the scene is captured in the left image (refer to B and corresponding area in the right image). Figure 2.2 (b), represents the edges of the right eye view of the stereoscopic scene. Note that in this case more of the right area of the

18

scene is seen in the right image (refer to *A* and the corresponding area in the left image). An alternative simple way to understand disparity is to place one image over the other. This would show that the corresponding points in images do not coincide but are displaced from one another. These displacements are called disparity.

The disparity of a pixel is unique in a stereo image pair since each pixel has its own depth (distance from the eyes/cameras) in the scene. The disparity of each object in the image depends on the distance between the object and the camera lenses. For example, most natural images are composed of objects at different distances from the camera lenses. Therefore, no single disparity value could be used to describe the displacement of objects/points between the left and right images. In comparison to this, in a typical video sequence, pixels in the background areas have often got motion vectors equal to zero, since the background in typical video sequence is static. Therefore, more overhead bits will be required for disparity field coding in stereo image as compared to motion vector field coding in video sequence.

In figure 2.3, by triangulation with binocular disparity and given camera geometry, the disparity can be formulated as follows:

$$|d| = \frac{bf}{Z} \qquad\qquad (2.2)$$

where *b* represents the baseline, *f* denotes the camera focal length and *Z* denotes depth. For detail derivation of equation (2.2), we refer the reader to [129]. From equation (2.2), the disparity can be considered as a relative depth as the *disparity is inversely proportional to the depth*. This means that objects closer (smaller depth) to the eyes/cameras will have larger disparity values and vice versa. In [164], research has shown that relative depth between objects in outdoor scenes is larger than that of indoor and close up scenes. Therefore outdoor scenes will typically produce smaller disparity values as compared to indoor and close up scenes.

Therefore in conclusion the following assumptions could be made about the disparity values of a stereo image pair obtained using parallax camera geometry:

- There is no vertical disparity.

- All disparity values have the same sign, i.e. displacements are all in one direction.

- The disparity can be considered as a relative depth as the disparity is inversely proportional to the depth (from equation (2.2))

- Outdoor scenes have smaller disparity values as compared to indoor and close up scenes.
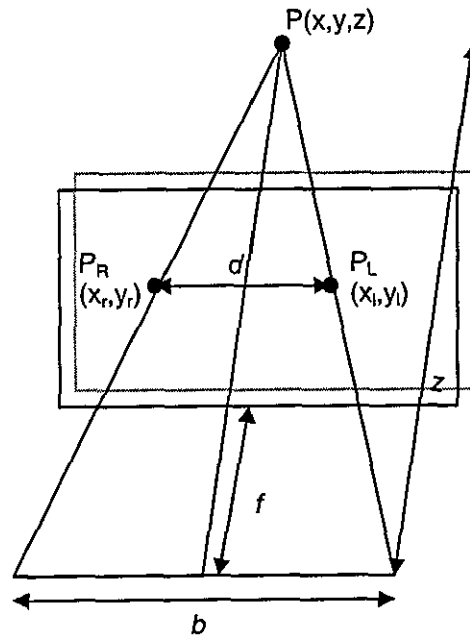
Figure 2.3 Binocular disparity with parallax geometry

2.2.3 Disparity estimation

Disparity estimation is a technique, to estimate the correspondence between pixels or regions of a stereo pair. With disparity estimation, one image is used as a reference image and the other is predicted from the reference image by horizontal shifting of the pixels in the reference image based on the corresponding disparity values. In computer vision applications, many pixel-based [3, 86] and feature-based [95, 198] disparity estimation techniques have been developed, which may be

applicable to stereo image/video coding. However, the direct adoption of these techniques for compression purposes may not be effective for various reasons. Pixel-to-pixel-based methods produced dense disparity fields (requirement in computer vision) but tend to fail because of ambiguities in the correspondence. Various improved estimation techniques have been proposed [3, 86], to overcome these problems, but they require excessive computational power, in order to calculate highly stochastic optimization problems from constraints added, to weaken the noise problem and discontinuities at object boundaries. Thus, pixel-to-pixel-based disparity estimation is inefficient in coding applications, due to its high computational complexity and high cost in transmitting the resulting disparity vector field.

In feature-based methods, local cues such as edges, lines and corners have been used in disparity estimation. Feature-based algorithms attempt to match significant abstract features, such as edges, lines and corners in order to estimate a dense disparity vector field. Matching pairs are found only if the same features are extracted in both images. If a feature-based disparity estimation method is used in applications, then it is necessary to transmit both the coordinates of the features and the corresponding disparities, which make it is inefficient in coding application.

The main emphasis in using stereo views in computer vision is the accurate estimation of disparity (or/and motion analysis) based on pixel-to-pixel-based or feature-based methods. This is required in order to reconstruct the 3D structure of the scene as the disparity corresponds to the distance between cameras and the corresponding object point in the scene. On the other hand, in coding applications, the main focus is the trade off between rate and distortion. Thus, the goal of stereo image/sequence coding is not to estimate the true disparity but rather to achieve a high compression ratio. Thus pixel-to-pixel-based and feature-based disparity estimation is not suitable in coding applications of a stereo image pair.

In stereo image/video coding, a disparity field represents the redundant information between two images. Therefore, the aim of disparity estimation and compensation is to use the disparity field to remove the stereo redundancy. The idea is to use a

disparity corrected version of one image as a predictor of its partner image [128]. Since the disparity information should be transmitted as overhead, computing a sparse disparity field is the key to compression [189]. Thus in coding of stereo images/video, block-based (fixed or variable size block) techniques have been widely used, even though the true disparity/motion fields are not blockwise constant. Among block-based methods, fixed size block matching (FSBM) is the most popular. This is due to its simplicity in implementation and effectiveness in terms of rate-distortion (RD). The gain in RD, obtained from exploiting the redundancy on the disparity field with a regular structure, does not require additional information for the structure of the disparity field to be transmitted. In addition, our visual system has the ability to fill the missing information from the sparse disparity field for depth perception.

Meanwhile, due to the similarity between stereo images and videos, many of the intuitions and techniques used in video coding are applicable in stereo image coding [215]. Motion estimation in video coding is used to exploit the temporal redundancy whereas disparity estimation is used in stereo coding. This temporal redundancy removal is possible, as consecutive images in a video sequence tend to be similar. Thus in general, disparity estimation is similar to motion estimation in the sense that they are both used to exploit the similarity between two (or more) images in order to reduce the bit rates. It can be carried out in the spatial [128, 129, 161, 218] or in the transform domain [96, 189, 208]. However, the motion estimation schemes developed in video coding may not be efficient unless geometrical constraints for stereo imaging are taken into account. For example, if the camera met the epipolar constraint, the direction of the disparity is predominantly horizontal. In comparison, motion vectors can take any direction in the 2D plane. Note that unlike in a video sequence, stereo images are projected onto two different cameras and due to variations in lighting, camera gains and reflections, the intensity levels between two images of a stereo pair may tend to be slightly different. Note also that regardless of the fact that an object in a scene is moving or not, due to binocular parallax, an object could be occluded in one image of a stereo image pair. As a result, the Disparity Compensated Difference (DCD) may have a higher residual energy. Thus a typical DCD frame may require relatively more bits as compared to a typical Displaced Frame Difference (DFD)

frame in video coding. Consequently, the efficiency of disparity compensated coding can be greatly reduced due to the above reasons. Thus typically, the coding gain obtainable by disparity compensation is relatively smaller as compared to that obtainable from motion compensation.

2.2.4 The influence of block size on disparity estimation

The Human Visual System (HVS) is quite good in filling in missing information. Thus sparse disparity vector fields can effectively convey the desired stereo cues. Hence in applications where stereo image pairs are meant for human viewing rather than computer vision, computing sparse disparity vector fields is the key to compression. As a result, Fixed-Sized Block Matching (FSBM) algorithms have been a popular approach to disparity estimation and compensation. A main advantage of FSBM method is simple implementation and effectiveness in terms of rate-distortion (RD) performance. The latter is due to the exploitation of binocular redundancy using a regular structure, which eliminates the need for any overhead to specify the location of each block. Therefore, it is only necessary to store or transmit the disparity vector (DV) field and the corresponding DCD image.

The basic idea of FSBM is to segment the target image into fixed size non-overlapping blocks and find for each block the corresponding block that provides the best match from the reference image. The difference between the disparity compensated and the original target image is called predicted error image (PEI). In [164], it had been shown that a smaller block size produces low energy (i.e., better disparity estimation) in the PEI as compared to larger block sizes. However for small block sizes, although the entropy of the residual image is reduced, the overhead required to handle the disparity vector field becomes too large. On the other hand, using a larger block size will reduce the overhead bits needed for the disparity vector field coding, but results in an increased energy level of the PEI. The commonly used block size in most literature is 8x8 or 16x16. In our research, for consistency, block size of 8x8 was used in all experiments.

2.2.5 Types of redundancy in stereo images/sequences

The basic idea behind all image/video compression methods is to take advantage of the various data redundancies present within images and video. As for stereo images/sequences, the use of two views implies at least doubling the amount of visual information for transmission/storage. Thus it is essential to make use of compact representations of this information. The representation should, on one hand, take advantage of all of the redundancy sources present in stereoscopic images, and on the other hand, only exploit the information, which is strictly necessary to the human observer. Three types of redundancy could be identified in relation to stereo images/video: intra-image redundancy, inter-image redundancy and binocular view redundancy.

a. Intra-image redundancy

This type of redundancy is within the image itself, therefore is the basis used in all monocular image compression methods.

If a pixel is selected in the image at random, there is a high probability that its neighbours will have the same or a close value to the pixel selected. Monocular image compression is therefore based on the fact that neighbouring pixels are highly correlated, which can be exploited to achieve compression. This type of redundancy is also called spatial redundancy.

In monocular image compression methods, the way the spatial redundancy is used to compress the image results in two types of compression: lossless and lossy. In lossless image compression, a pixel is predicted from a group of neighbouring pixels of the image, and the pixel domain prediction error is directly coded using a lossless entropy coding process. Typically a compression ratio 2:1-3:1 is obtainable through lossless image compression methods. JPEG-LS and Lossless JPEG-2000 are typical examples of lossless image compression standards. However the first is a pixel-based compression technique whereas the second is based on wavelet transforms and a

subsequent lifting scheme that is cable of providing lossless compression prior to entropy coding.

Lossy image compression methods fall into two categories, namely, predictive-based and transform-based. In predictive-based algorithms, a pixel (e.g. JPEG-NLS) or a group of pixels (Fractal Image Compression) is predicted by a group of pixels similar pixels. The predicted image thus obtained is quantized prior to entropy coding and transmission. In transform-based algorithms, the pixels of the image are transformed to an intermediate representation (e.g. DCT, DWT, Integer Transform) where they are de-correlated to eliminate redundancy. If the transformed pixels are decorrelated, certain pixel values become common, thereby having large probabilities, while others are rare. This results in smaller entropy. Thus quantizing the transformed values can produce efficient lossy image compression. Note that in lossy compression the basic cause for the loss of data is associated with the rounding-off process associated with quantizing. Typical image compression standard that represent lossy image compression are baseline-JPEG [62] and lossy JPEG-2000 [1].

b.   Inter-image redundancy

Inter-image redundancy is between images and is frequently used in monocular video compression methods. A video is a collection of digitised snapshots of a scene, taken at fixed time intervals in the temporal domain (time scale). Therefore, in an image sequence/video, background objects do not generally move from one frame to the next, and only a small percentage of the scene undergoes motion. The displacement of these moving objects usually does not exceed a few pixels. Thus, a lot of redundant information exists between adjacent images (frames) of a video sequence. Video compression standards such as MPEG-1, MPEG-2, MPEG-4, H.261, H.263 and H.264 exploit inter-image redundancy, along with intra-image redundancy present in the spatial domain of individual frames, to achieve compression. With the knowledge that corresponding pixels typically change only slightly from frame to frame, the complete I frames (Intra coded frames) need to be transmitted only occasionally (1-2 per second in MPEG-1/2, much less in H.261/.263)

by some monocular image compression method by exploiting intra-frame redundancy. The predicted *P* frame and bi-directional *B* frames are synthesised at the receiver from the decoded *I* frame, plus a few dynamic coefficients (motion vectors) that tell the receiver how to estimate and interpolate the motion dynamics, plus perhaps also a residual (reconstruction error) image stream. This residual error is coded compactly because its amplitude distribution is typically much narrower than that of the original image (the entropy of the residual image is much smaller than the entropy of the original image).

c. *Binocular redundancy*

In a stereo image pair, there exists a third type of redundancy named binocular redundancy, due to the similarity between the left and right images, since both images represent the same scene viewed from slightly different viewpoints. Thus most parts of a scene are visible from the two viewpoints, and projects on to similar regions of the left and right image planes. In theory, this geometrical redundancy can be exploited for compression purposes. A possible strategy consists of putting into correspondence regions of pre-established size and shape (blocks, objects, or regions) of a stereoscopic pair, and transmitting disparity vectors allowing the reconstruction of one viewpoint from the other. This prediction strategy is referred to as *disparity estimation and compensation.*

It is obvious, by exploiting both the intra-image, inter-image and binocular view redundancies present in stereo image pairs/video, a much higher efficiency can be achieved in compressing the residual image. In figure 2.4, we illustrated that most of the compression of a stereo image pair comes from the residual image; as it mostly consists of black areas (pixel value equal 0).

In using wavelet technology, after the wavelet transform is applied, an image will be decomposed into a collection of subbands. These subbands represent the same image but at different frequency or with a different geometric property. The relationship between the subbands can be considered as spatial redundancy. This has been

26

effectively exploited, for example, in EZW [185], SPIHT [16], and EBCOT [43] coding algorithms to efficiently code the resulting wavelet coefficients.

a   b
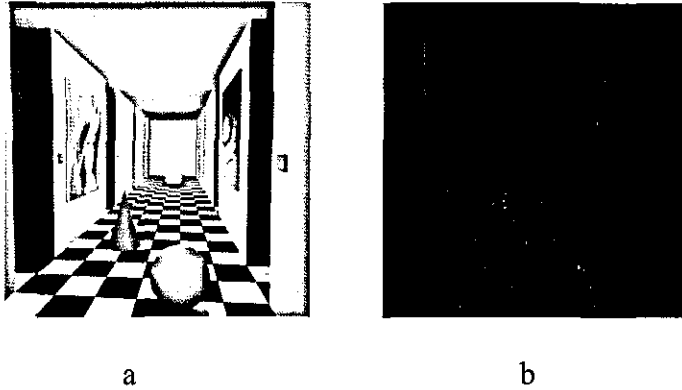
Figure 2.4 (a) The left frame (reference frame) (b) The residual frame

2.2.6 Occlusion

Occlusion is a difficult phenomenon to be dealt with by any stereo image coding technique. The problem with occlusions plays a role in stereo matching, as some basic assumptions, for example uniqueness and smoothness, and the fundamental triangulation geometry are invalid in occluded regions. With any reasonably complex scene there exist occluded pixels that have no correct match. Occlusion can happen for two main reasons: finite viewing areas and depth discontinuities. Finite viewing areas occur on the left side of the left image and the right side of the right image where the respective eye can see objects that the other eye cannot. Depth discontinuities are due to overlapping objects in the image; certain portions can be covered from one eye on which the other eye has direct sight.

The residual of those blocks that are occluded because of depth discontinuities display different characteristics from the other parts of the image. As noted in [196], the occluded blocks are more correlated. By detecting such blocks, and coding them differently from the rest of the residual image blocks, can improve the efficiency [196]. Woo [215], in order to avoid occlusion effects, adopted an occlusion indicator and then segmented the block selected as an occlusion candidate into smaller blocks.

If the small block (or pixel) has still larger disparity estimation error than a specified threshold, it is classified as an occlusion, Drawback; a considerable amount of bits are used for encoding occluded areas to increase the PSNR. Woo claims that *"the occlusions do not affect on perceiving 3D structure, i.e., even though we do not send any side information for occlusion areas, we can perceive 3D from stereo images.* Therefore, another way to reduce this kind of overhead is to interpolate occluded areas from their neighbouring blocks. Stereo image coding techniques that deal with occlusion are discussed further in section 2.5.

2.3 Characteristics of a disparity compensated residual image

Yamaguchi [70] discussed the characteristics of disparity compensated residual images such as local correlation and residual image intensity distribution characteristics to be used with an entropy encoder. However, no experimental compression performance results are presented.

In [139], it is claimed that disparity compensated stereo residuals contain more structure than motion compensated pairs in a video sequence. The residual image has two special properties [196]. First, much of the residual image has intensity values near zero [0=black, as in figure 2.4 (b)]. Second, the remaining contents are narrow vertical edges. These are the consequence of strictly horizontal apparent motion between the two views. Incompleteness in disparity estimation appears as horizontal alignment errors that cause the vertical features in the residual. These two features explain why the residual image tends to have small pixel values and can be coded efficiently. There have been various research efforts to code residual images [139, 196], based on these different characteristics of residual images as compared to natural images. For example, [196] shows that the correlation across block boundaries in the residual image, diminishes, suggesting that the coding of these blocks individually, might be preferable. Thus in [196], a concept of mixed image transform was used by applying DCT on well matched blocks and Haar wavelet on occluded

blocks. However, generally, the residual image is coded using a method similar to the one used to code the reference image.

2.4 Theory of human stereoscopic vision

Throughout this thesis, the reference image is maintained at certain high quality level while the PEI is compressed from high to low quality level. This procedure is based on a theory of stereoscopic vision [30], the suppression theory [53], and mixed resolution theory [30].

2.4.1 Theory of stereovision

Viewing a true, three-dimensional scene normally simulates the sense of stereovision. It is possible to stimulate the sense of stereovision artificially by acquiring two pictures of the same scene from separated positions, and by presenting the left picture to the left eye and the right picture to right eye. Two pictures acquired in this manner form a stereopair. One of the most important ideas in the study of stereopairs is that of disparity, discussed in section 2.2.2. Depending on the camera geometry being used, each component of the disparity can either be positive or negative. When negative disparity occurs, the scene being viewed will appear to be floating in the space between viewer's eyes and monitor. This type of imagery cannot be produced without the aid of stereoscopic devices such as shutter glasses. For another case of parallel axis camera geometry, the vertical component of disparity is always zero and the horizontal component of the disparity is always positive. This implies that the parallel axis geometry processes a simple mathematical relationship between the disparity of a point pair and the distance to the object it represents. In general, the disparity vector d can be used to predict one image of a stereopair from the other. For example, given a luminance level of left image at position p, $I_L(p)$, the luminance level of its corresponding right picture can be calculated as,

$$I_R(p) = I_L(p + d), \qquad (2.3)$$

where d denotes the disparity vector whose direction is from left to right.

2.4.2 The suppression theory

When the eyes are presented with two similar images, the result is a single percept of the scene. This phenomenon is known as binocular fusion [83]. On the other hand, when the eyes are presented with two distinctly different images, instead of being combined the two images will compete against each other. This is known as binocular rivalry [83]. The final percept is unstable shifting between the patterns of each eye. Because of this unstable combination of the two images, certain sections of one image will dominate certain sections of the other image. This alternation of patterns results from a shift in dominance between each eye. Information in the less dominant image will be suppressed, while the information in the dominant image will be visible [208]. These lead to the following compression scheme: one image of the stereo image (the reference image) retains the details of the scene, while the second image (the predicted image) retains the disparity information. Hence, the second image can be highly compressed without affecting the depth information in the compressed stereo images [81]. This compression scheme has been used in the majority of stereo image coding methods [51, 81, 129, 161, 189, 215] including the ones proposed in this thesis.

2.4.3 Theory of mixed-resolution coding

Mixed-resolution coding is a perceptually justified technique for compressing stereopairs. The compression is achieved by presenting one eye with a low-resolution image and the other eye with a high-resolution image. Psychophysical experiments [30] have shown that a stereo image with one high-resolution and one lower resolution image is sufficient to provide the almost identical stereoscopic depth as that of a stereo image with two high-resolution images. Thus, the eye/brain can easily fuse

such stereopairs and perceive depth in them. In summary, the concept of mixed-resolution coding can be symbolically represented by the following equations:

(Stereo image) ≈ (High resolution left image) + (Low resolution right image)

where (Low resolution right image) = (High resolution left image) + (LR Predictive Error image)

The mixed-resolution coding is able to significantly reduce the bit-rate required to transmit a stereo image with two high-resolution images.

2.5 Survey of stereo image coding techniques

Conventional monoscopic compression techniques JPEG [62], JPEG-2000 [1] or 3D-DCT [81] can be applied to stereo image coding by treating the left and the right images independently. However exploiting the redundancy between the two images can yield better performance than independent coding of the stereo image pair. Literature pertaining to stereo image compression is still scares as compared to that of monoscopic still image and video compression. Most stereo image coding techniques [51,81,87,129,161,189,215] use a predictive coding framework with disparity estimation and compensation to un-correlate the two images. In the early form of these algorithms, it was suggested to encode an image (reference image) and the difference between two images [161]. To further uncorrelate the two images, M.G.Perkins [161] proposed to code the sum and difference of the two images provided the left and right images have first and second order statistics. However, these techniques were inefficient in the presence of large/variable disparity values.

A modification to the sum-difference coding is to shift one of the images horizontally to the point where the cross correlation between the images of stereo pair reaches it maximum value [70] (the shift approximately equals the mean disparity of the scene). The shifted image is then subtracted from its partner to remove the redundancy, and

the difference is encoded [70]. This method, referred to as global translation method, is inefficient since it is based on a wrong assumption that objects in the scene have similar disparity values. Another approach called enhanced correlation is to translate the so-called *row blocks* instead of the whole image. The amount of translation is determined either by cross correlation statistics (correlation enhanced technique) [81] or using the most interesting feature in the row block [206].

Lukacs [128] was the first to introduce proper predictive disparity estimation and compensation in stereo image coding, a technique, which has been adopted in most proposed stereo image coding techniques. The idea is to use disparity information to obtain the predicted disparity compensated image blocks from the reference image to construct the predicted image block by block.

In coding applications, where the compressed stereo image pairs are meant for human viewing, block-based disparity compensated is most popularly used. With reference image (left image) independently coded, the right image is divided into non-overlapping blocks. Either fixed (no information about the block locations and size need to be transmitted) or variable size blocks (information on block locations and size need to be transmitted) are used. Based on the camera parallel axis geometry assumption, each of these blocks is shifted horizontally and compared to the corresponding blocks in the coded reference image using some measure such as Mean Squared Error (MSE) or Mean Absolute Error (MAE) to determine the similarity between the two blocks. The most similar block is the disparity compensated prediction and the corresponding translation is the disparity for the block.

Given the right image that is to be encoded and a disparity vector field, there are a variety of coding strategies that may be used. For example, the residual of the disparity compensated prediction may be encoded and transmitted. This method is referred to as Disparity Compensated Residual Coding and was first proposed by H.Yamaguchi et al. [70].

M.G.Perkins [161] proposed Disparity Compensated Transform Domain Predictive Coding (DCTDP) to further improve the work of Lukacs [128]. This method

followed the disparity compensation method proposed by Lukacs and obtained the best match for a given block of the right image from within the reference left image. However the actual prediction error is calculated in the DCT transform domain using a linear predictor. The transformations remove the redundancies within samples in both images prior to prediction. The transform domain error block is subsequently quantized, entropy coded and transmitted.

This coder by M.G.Perkins [161], which works with all types of stereo images, is widely used as a benchmark by the stereo image research community. Furthermore, the DCTDP framework which, consist of the following components: disparity estimation/compensation, transformation, quantization and entropy coding has been adopted as the basic framework of most proposed stereo image coding techniques. In addition each of the above components of the framework has been the subject of further research within stereo image research community.

The first component, disparity estimation, is the most important part in stereo image analysis/coding based on its popular use in computer vision and as a method of data redundancy exploitation in image coding. The disparity estimation techniques generally employ a block-based (fixed size block or variable size block) approach. Fixed size block matching (FSBM) scheme has the unique advantage of eliminating the overhead that is required to specify the block locations. Thus the target image can be reconstructed based on the reference image, residual image and disparity field only. In [51, 87], a technique named Pioneering Block-based Disparity Compensated Predictive Coding (PBDCPC) was proposed that eliminates the need to transmit the disparity vector field. With PBDCPC, the target image can be reconstructed only based on the reference and residual images. We use this technique (explained in detail in chapter 5) as our second benchmark to compare the performance of CODEC IV and CODEC V.

As explained earlier, FSBM has a drawback in that the overhead required to transmit the disparity map becomes larger as the block size becomes smaller and fails to provide a good matching result as the correspondences are locally ambiguous due to

noise, occlusion and repetition or lack of texture. In addition, using large block sizes at low bit-rates, where fewer bits are used to encode the residual (error) images, would result in blocking artifacts in the decoded target (predicted) image. Thus in order to increase the efficiency of the FSBM scheme, techniques to reduce the blocking artifacts have been proposed in literature.

A method proposed in [216] uses a the so-called Overlapped Block Disparity Compensation (OBDC) to reduce the blocking artifacts. In this method, a Markov Random Field (MRF) model-based disparity estimation process helps in providing a smooth disparity field, while maintaining the entropy of the disparity compensated difference low. The smooth disparity field is used in applying selective OBDC to reduce the blocking artifacts in the reconstructed image [216]. An improvement of up to 1.5 dB is reported over the FSBM method without OBDC.

At the price of increased complexity, projection based methods have been proposed which try to improve the disparity estimation/compensation ability for scenes with occluded objects and photometric variations. The Subspace Projection Technique (SPT) [72] and Sequential Orthogonal Subspace Updating (SOSU) [183] are two examples. SPT generates a set of orthogonal basis vectors, from four blocks for every image block, using the Gram-Schmidt (GS) orthogonalization procedure. Three out of the four blocks are fixed: one block contains all ones (1's) to compensate for the intensity differences while two blocks provide tilt in each of the horizontal and vertical directions, respectively. The fourth block is the one for which the disparity vector is estimated using a block matching method. Although, the orthogonalization process in this method is not computationally demanding (since three blocks are fixed and only one is changing depending on the content of the image block), the fixed blocks may not be adequate to compensate for different types of mismatching problems that occur [183]. In SOSU, the basis vectors are formed from an extended set of neighbouring blocks as well as some typical edge blocks for a more flexible and accurate representation. These image-dependent basis vectors are well matched to the shape and edge structures within an image block. The SOSU uses an iterative approach in order to find a set of best basis vectors that minimizes the MSE for each

block. At every stage of the algorithm, the remaining blocks are orthogonalized using the GS procedure and the residual image is projected onto the orthogonalized subspace and the coefficient corresponding to the largest projection is extracted. The process continues until certain quality measure is satisfied.

An alternative to FSBM approach is to use a segmentation approach [50, 187, 189]. However, not all segmentation techniques are suitable due to the excessive number of bits required to describe the shape and the location of each region. Quadtree decomposition provides a relatively economical and effective solution to this problem. The idea is to segment the blocks depending on the mean-squared error (MSE) criterion. Blocks with mismatching problems or different depth objects, which usually lead to large MSE, are partitioned into smaller blocks to increase the homogeneity inside the block. A regular quadtree approach, where a block can be divided only at the midpoint of its sides has been implemented in [72]. Also, a generalized quadtree can be considered where a block can be divided at $2^k - 1$ locations (k is the number of bits that can be allocated per side per node) [189]. The division takes place at the permitted location that lies closest to a sharp intensity discontinuity. In general, variable block size schemes outperform the FBSM schemes [218].

As compared to research in disparity estimation/compensation, significantly less effort has been put into the coding of residual images. These images display characteristics that are different from that of natural images. The correlation across block boundaries in the residual image is low, suggesting that individually coding these blocks might be preferable [196]. In [196], it is proposed to use mathematical transforms that take into account the correlation properties of the residual image as well as the block-based nature of the disparity estimation used. Occluded blocks are often difficult to estimate. Residuals of occluded blocks therefore are different from blocks that are well matched by the DE process. These two blocks are treated differently. DCT is used on blocks that are well matched by disparity estimation and Haar filtering is used on the occluded blocks, resulting in a mixed image transform. In [138], specialized coding methods for the residual images have been proposed. The algorithms make use of theoretically expected and experimentally observed

characteristics of the disparity compensated stereo residual to select transforms and quantization methods. Exploiting the directional characteristics within a DCT framework provides the above algorithm's best performance, below 0.75 b/pixel for 8-bit greyscale imagery. In a wavelet algorithm, roughly a 50% reduction in bit rate is possible by encoding only the vertical channel, where much of the stereo information is contained. Other techniques aimed at reducing the computational cost of full search disparity estimation include the fast-three step search algorithm [104], hierarchical disparity estimation [188] and adaptive reduction of search area [115].

The majority of the above methods are based on DCT technology, followed by the lossy coding of the DCT coefficients. However, one of the major drawbacks of the DCT-based methods, are annoying and even non-tolerable blocking artifacts at low bit rates. In stereo imaging applications such artifacts result in viewer discomfort and are a serious problem. The main cause of the blocking artifacts comes from the fixed length basis vector function used in DCT-based coding schemes.

On the other hand, wavelet-based coding schemes use variable length basis functions, have no blocking artifacts while providing many useful features such as embedded coding, region of interest capability, signal-to-noise ratio and resolution scalabilities. On going research have shown in many applications wavelet-based still image coding schemes provide better performance than other coding scheme specifically in rate-distortion performance and functionality terms. Thus it is expected that wavelet-based coding techniques would perform well in stereo image coding. However, very little work has been reported in applying wavelet transform technology to stereo image compression [30, 96, 155, 188, 208]. This is possibly due to the popularity of DCT-based technology, while wavelet-based image compression technology, is still at its early age.

In [208], a stereo image coding algorithm that does not involve disparity estimation has been proposed based on the suppression theory and multiresolution analysis of a stereo image. In this method, after the left and right images of the stereo pair have been decomposed by a wavelet transform, one image is considered the reference

image and the other is represented by a low resolution version of itself. Based on suppression theory, 3D perception still occurs as a result of presenting different resolution images to each eye. To enhance the depth perception and to reduce the suppression zone, selected subbands of the high frequency information are added to the low resolution image. This is made possible, since in wavelet transforms, the image is decomposed into subbands of low and high frequencies. This is impossible to achieve in DCT-based methods, since in obtaining the low resolution image, all the high frequency information would be destroyed . However, non-use of disparity compensation and the use of inefficient Lloyd-Max quantizers, limit the efficiency of this scheme.

In using disparity estimation and compensation, wavelet transforms enable multiresolution-based hierarchical disparity estimation to be used within a stereo image coding scheme [30, 188, 189]. In these schemes, after applying a 3-level DWT on the left and right images, two pyramids for hierarchical block matching were constructed using the coarse images at each level. Disparity estimation is performed starting from the level-3 coarse image up to level-1 coarse image (coarse-to-fine disparity estimation), by propagating the disparity estimation of previous level as initial disparity estimate of the current level. At level-1, the same disparities of the coarse image, are used as disparities for all subimages (detail images) in that level. Thus at the decoder, the right image is reconstructed based on disparities of level-1, its four subimages at level-1 and the reference image (left image). Meanwhile, in [30], wavelet-based multiresolution compression technique for 3D stereoscopic images is based on a mixed-resolution psychophysical experiment. According to the mixed resolution psychophysical experiment, the block based disparity estimation is performed on low resolution image only. The computational complexity of the low resolution disparity estimation becomes smaller due to the smaller search area at the lowest resolution. At the receiver, the low resolution right subimage is estimated using the disparity from the low resolution left subimage. A full-size reconstruction is obtained by upsampling a factor of 4 and reconstructing with the synthesis low pass filter. However, in both methods [30], and [188] performing disparity estimation and compensation only on selected subimages, managed to reduce the complexity of

disparity estimation in wavelet multiresolution domain but at the price of reduction in the quality of the predicted image. Meanwhile in [167], disparity compensation is used in the wavelet domain followed by a subspace projection technique for the coding of wavelet coefficients. However, this method is computationally complex, since a different basis has to be constructed for the representation of each block of wavelet coefficients.

Apart from [208], all other wavelet-based stereo image coding schemes discussed above implements disparity estimation and compensation in the wavelet transform domain. However, the performance of these coders are inefficient mainly due to the fact that they do not use all the subbands in the disparity estimation process. Wavelet-based stereo image coding [96,155] can also be performed by disparity estimation in the pixel domain as commonly done in DCT or other transform-based techniques found in literatures. However the drawback of this approach is that the artificial block boundaries in the PEI could cause undesirable, high frequency coefficients that would reduce the compression efficiency. This problem is discussed further in proposed CODEC I in chapter 4.

To overcome this deficiency, in [96], and [155] the use of high performance embedded coding schemes such as EZW [185], SPHIT [16] or EBCOT [43] were proposed. These coding techniques are able to compensate the ineffectiveness of performing DWT on the whole PEI by efficiency coding the wavelet coefficients.

2.6 Stereoscopic image displays

From the first practical stereoscopic display device, invented by Sir Charles Wheatstone in 1832, to modern era of stereoscopic (need viewing eyewear) and autostereoscopic (no need of viewing eyewear) display devices, the basic principle used is the same:

The principle of any stereoscopic display devices are to make sure the left image is viewed by the left eye and right image by the right eye.

This separation can be accomplished through optical, color, polarization or temporal separation techniques [83]. To assist in the separation process, many of the display devices, require the observer to wear some kind of eyewear/headgear. These systems are generally referred to as being stereoscopic [82]. Displays that do not require such aids, and can be freely viewed while still being stereoscopic are known as autostereoscopic [82].

2.6.1 Stereoscopic displays

a.   *The Stereoscope*

This display device is an adaptation of the early stereoscope such as the Wheatstone Viewer, in which electronic ones replaced the photographic stereo images. Basically, the optical framework of stereoscope is as illustrated in figure 2.5. A medium such as mirrors, prisms and convex lenses are used in a stereoscope to assist the viewer's left eye to see the left image and the right eye to see right image. Modern stereoscopes range from relatively cheap pocket size models to more complex models such as the head mounted displays, which are used in a wide area of applications.
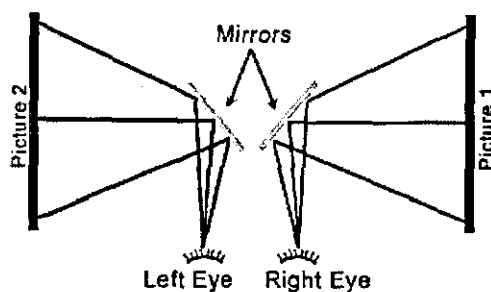
Figure 2.5 Framework of Wheastone's Mirror Stereoscope

b.   *Anaglyph stereoscopy*

In anaglyph stereoscopy, color-separation technique is used in order to achieve stereopsis. The picture for one eye is presented in red and for the other in green (or blue). The viewer wears glasses containing red and green (or blue) filters or lenses, so that one eye sees only the green (or blue) picture and the other eye only the red as shown picture figure 2.6. This approach has been used in entertainment such in cinema, television and in print [80]. Limitation of this method, when using color imagery, is that it causes a "ghosting effect", which limit its effective use to black and white imagery. Also, there is some loss in information due to light loss when using the filters. Finally, the glasses could cause eye fatigue to the user [83].

Figure 2.6 Eyeglass for anaglyph viewing

c.   *Polarized display*

In this stereoscopic system, the left eye and right eye images are projected onto the same polarisation preserving screen through two projectors. A pair of orthogonal polarised filters which are oppositely oriented are placed at the output lens of the two projectors. The display is viewed through cross-polarized spectacles so that each eye sees only one image. They work in much the same way as the coloured anaglyphic glasses. Polarised display allows for the display of full-color stereoscopic images to a large audience and the per unit cost of the polarised glasses for viewing are relatively low [219]. Drawback is that each eye sees a faint version of the other eye's image in addition to its own. This effect increases if the viewer tilts the head so that the axis of polarization of the glasses is no longer aligned with that of the projected image [83]. This could cause eye strain and fatigue.

d.   *Frame sequential display*

This display system incorporates shuttering glasses, a normal display, stereo image generator, and a system for synchronizing the components. Typically the images are presented sequentially on the display using odd and even fields of the screen interlace. The monitor is viewed using special liquid-crystal lenses that obscure the eyes alternately so that each eye only sees one of the images or through binocular shutters that open alternately at the same rate [83]. Drawback of this system is bad flickering caused by the alternate opening of the shutter.

Studies have shown that observers tend to dislike wearing any invasive equipment over their eyes, or wearing anything that impairs their general ambient visual acuity [35]. Therefore, stereoscopic display systems resembling the human visual system are more popular among the 3D display systems with the development of autostereoscopic displays.

2.6.2 Autostereoscopic display

In autostereoscopic displays, all of the work of stereo separation is done by the display, so that the observers need not wear special eyewear. A number of researchers have developed autostereoscopic displays. Most of them are variations of the parallax barriers or lenticular arrays methods [109]. In these methods, a fine grating or a lenticular lens is placed in front of a display screen, that allow the separation of viewing by the left and right eye. To accommodate arbitrary viewer positions and orientations, in [109], the fixed parallax barrier (the fine grating or lenticular lens) is replaced by a dynamic parallax barrier consisting of very large stripes. These large stripes, which are made of many slender stripes, are moved rapidly across the image so that the viewer cannot perceive the individual stripes. The dynamic parallax barrier and retroreflective camera based tracking device enables the viewer to look around an object, without wearing any special eyewear [109]. The introduction of autostereoscopic display technology and its subsequent developments has led to a renewal of interest in stereo imaging/video technology.

2.7 Compression performance measures

Throughout this thesis, the compression efficiency is measured in bits per pixels (bpp) and has been defined as follows:

$$Bit\,Rate(BR) = \left[\frac{Tot\_Bits_{comp}}{NxM}\right]bpp.$$

where, N and M represent the image dimensions and Tot_Bits_{comp} is the total number of bits required to represent a given image in its compressed format. For benchmark 1, CODEC I, CODEC II, and CODEC III, Tot_Bits_{comp} includes the bits that are required to code the disparity vector field, in addition to the bits requirement to encode and transmit the residual image. However, for benchmark 2, CODEC IV, CODEC V and CODEC VII, Tot_Bits_{comp} represents only the bits requirement to encode the error image.

The objective image quality is measured in terms of the Peak Signal to Noise Ratio (PSNR). The PSNR (in dB) of the reconstructed image is defined as,

$$PSNR = 10*\log_{10}\left[\frac{255^2}{MSE}\right].$$

The Mean Square Error (MSE) between the original image blocks $Org(i,j)$ and the reconstructed image blocks $Rec(i,j)$, pixel values, with both of size NxM is given by,

$$MSE = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}(Org(i,j)-REc(i,j))^2}{NxM}.$$

3 Discrete Wavelet Transform & its application to image processing

3.1 Overview

Over the past decade, the Discrete Wavelet Transform (DWT) has been widely used in signal processing research, particularly, in image compression. Images are non-stationary signals. Therefore, capturing the transients and their particularities, performing a localised time-frequency analysis, focusing on the singularities, and providing a multiresolution signal representation are basic needs in image processing that are only met by the DWT [9]. This is due to the ability of DWT to obtain a localised time-frequency analysis of the signals, which Fourier Transform (FT) or other similar transforms, fail to provide. In addition DWT resembles the human visual system, research in neurophysiology provides evidence that the receptive fields in the primary visual cortex and peripheral regions such as the retina approximate wavelet functions and perform similar filter operations [132]. Thus in many applications, wavelet-based schemes achieve better performance than other coding schemes. For example, the recently launched new digital still image standard, JPEG-2000, which is wholly wavelet-based, proved these facts by providing rate-distortion performance and subjective image quality far superior to the existing standard, baseline-JPEG [see section 3.7]. JPEG-2000 also enables features and functionalities that current standards can either not address efficiently or in many cases cannot address at all [24]. For example, the multiresolution domain of the wavelet transform, enables scalability in fidelity (PSNR) or resolution in JPEG-2000 based on a single embedded codestream, which are absent in baseline-JPEG. With this in mind, our research aims at proposing wavelet-based stereo image CODECs that outperform existing DCT-based stereo image CODECs, similar to the outperformance of baseline-JPEG by JPEG-2000, especially at low bit-rates.

In this chapter, a brief discussion of wavelets and their applications to image compression is provided. For a more detailed discussion on wavelet theory, we refer

the reader to publications presenting the wavelet theory and many of the perspectives it can be looked at [39, 63, 121,128].

3.2   Brief history of wavelet

The word "wavelet" was first mentioned in Alfred Haar's thesis of 1909 and emerged as a new tool in image processing during the past ten years. Although wavelets are a new tool in image processing, its mathematical foundations date back to the work of Joseph Fourier in the nineteenth century. In the early 1800's, Joseph Fourier discovered he could approximate periodic functions using linear combinations of periodic complex exponential functions. Many years after this remarkable discovery, the ideas were generalized to non-periodic functions, and then to periodic and non-periodic discrete time signals. Fourier laid the foundations with his theories of frequency analysis, which proved to be enormously important and influential.

The equation

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt \qquad (3.1a)$$

is generally called the Fourier Transform (FT). The equation

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{j\omega t} d\omega \qquad (3.1b)$$

is called the inverse Fourier transform.

In the FT equation (3.1a), the integration is from minus infinity to plus infinity over time. So, no matter when the component with frequency z appears in time, it will affect the result of the integration equally as well. The lack of time information is one serious weakness of the Fourier Transform. Thus, for many decades, scientists have wanted more appropriate functions than the exponential, which comprise the basis of Fourier analysis, to approximate physical situations where the signal contains

discontinuities and sharp spikes (such as edges in an image). This is because; the exponential functions are non-local (i.e., stretch out to infinity), and are not the most efficient way of approximating local abrupt changes such as those representing edges in an image.

To solve the above problem, Denis Gabor proposed the Short-Time Fourier transform (STFT) [142]. The basic idea is to divide the signal into small enough segments/windows, where these segments can be assumed to be stationary. The width of this window must be equal to the segment of the signal where this assumption is valid. But, the STFT has several setbacks. If we used a window of infinite length, we again get the Fourier Transform, which gives perfect frequency resolution with no time information. On the other hand, in order to obtain a stationary sample, we must have enough windows in which the signal is stationary. The narrower we made the window, the better the time resolution, and better the assumption of stationarity, but poorer the frequency resolution, which is a main drawback of STFT [142].

The attention of researchers gradually turned from frequency-based analysis to scale-based analysis using wavelets, when it started to become clear that an approach measuring average fluctuations at different scales which prove less sensitive to noise. Wavelets are functions defined over a finite interval. The basic idea of the wavelet transform is to represent an arbitrary function $f(x)$ as a linear combination of a set of such wavelets or basis functions. These basis functions are obtained from a single prototype wavelet called the *mother wavelet* by dilations (scaling) and translations (shifts). The purpose of wavelet transform is to change the signal from time domain to time-scale domain.

The concept of wavelets in its present theoretical form was first proposed by Jean Morlet and the team at the Marseille Theoretical Physics Center working under Alex Grossman in France [142]. The main algorithm dates back to the work of Stephane Mallat in 1988 [128]. Since then, research on wavelets has become international. Such research is particularly active in the United States, where it is spearheaded by the work of scientists such as Ingrid Daubechies, Ronald Coifman, and Victor

Wickerhauser [142]. The wavelets are now recognized as a new powerful tool in signal processing.

3.3   Wavelet theory

3.3.1 What is a wavelet?

A signal is a function of one or more real variables.

An image is a positive function on a plane. The value of this function at each point specifies the luminance or brightness of the picture at that point. A digital image is a sampled version of such a function, where the value of the function is specified only at discrete locations on the image plane, known as pixels [18].

A wavelet is a signal of a small wave as shown in figure 3.1b. "Small" refers to a common but not necessary feature of wavelets- they exist only in a finite interval and is zero or almost zero, outside this interval.

a   b
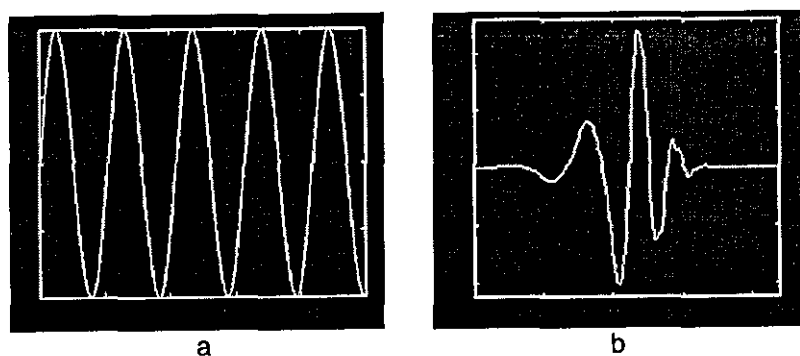
Figure 3.1 Comparison of the basic function used in (a) Fourier Transform (b) Wavelet Transform

Technically, wavelets are functions generated from one single function $\psi(x)$ called the *mother wavelet* by dilations (stretching or shrinking) and translations (shifting)

46

$$\psi_{s,t}(x) = \frac{1}{\sqrt{s}}\psi(\tfrac{x-t}{s}), \qquad s \in R^+, t \in R, \hspace{2cm} (3.2)$$

where the shift parameter "t" gives the position of the wavelet, whereas the dilation parameter "s" controls its frequency.

The translation of a wavelet $\psi(x)$ by t is the wavelet $\psi(x-t)$.

The dilation of a wavelet $\psi(x)$ by s is the wavelet $\psi(\tfrac{x}{s})$, where $s > 0$.

The mother wavelet $\psi(x)$ can be any real or complex continuous function having the following properties:

- The total area under the curve of the function is zero,

$$\int_{-\infty}^{\infty} \psi(t)\,dt = 0. \hspace{3cm} (3.3a)$$

This implies that $\psi(x)$, oscillates above and below the x-axis (a wave-like appearance).

- The total area of $|\psi(x)|^2$ is finite,

$$\int_{-\infty}^{\infty} |\psi(t)|^2\, dt < \infty.$$

This implies that $\psi(x) \in L^2(R)$. \hspace{1.5cm} (3.3b)

- Admissibility condition

$$\int_{-\infty}^{\infty} |\Psi(\omega)|^2\, |\omega|^{-1}\, d\omega < \infty \hspace{2cm} (3.3c)$$

47

where Ψ is the Fourier transform of ψ.

For $s > 1$, the wavelet $\psi_{s,t}$ is a very much stretch out version of the mother wavelet, with frequency content mostly in the low-frequency range. Conversely, for $0 < s < 1$, $\psi_{s,t}$ is very much shrink version of the mother wavelet and has mostly high frequencies.

Grossman and Morlet [128] showed that any square integrable function $f(x) \in L^2(R)$ can be represented as a combination of the dilations and translations of a single wavelet.

The continuous wavelet transform (CWT) performs this transformation to convert a signal $f(x)$, which is a function of x, into a function $W_f(s,t)$ of dilation/scale s and translation t. Specifically,

$$W_f(s,t) = \int_{-\infty}^{\infty} f(x) \psi_{s,t}(x) dx \qquad (3.4a)$$

where

$$\psi_{s,t}(x) = \frac{1}{\sqrt{s}} \psi(\frac{x-t}{s}). \qquad (3.4b)$$

The factor of $\frac{1}{\sqrt{s}}$ normalises the energy of $\psi_{s,t}$ for difference scales s. A wavelet is defined in terms of this transform- any function $\psi(x)$ for which $f(x)$ can be reconstructed from W_f qualifies as a "valid" wavelet.

In practice the CWT is highly redundant- it contains repetition of the information in the original signal. The dyadic discrete wavelet transforms (DWT) is only calculated for scales (m) which are integers powers of two and at translations (n) which are integers multiples of the current scale factor, thus eliminating this redundancy [128].

It is defined by,

$$DW_f(m,n) = \int\limits_{-\infty}^{\infty} f(x)\psi_{m,n}(x)dx \qquad (3.5a)$$

where

$$\psi_{m,n}(x) = 2^{\frac{m}{2}}\psi(2^{-m}(x-n)), \qquad m,n \in Z. \qquad (3.5b)$$

It can be seen that this represents the calculation of $W_f(s,t)$ at values $s = 2^m$ and $t = 2mn$. The conversion of a function of one variable into a function of two discrete variables may not seem to be particularly productive. However, this representation of f can illuminate important features, which are hidden in its original format. Small scale versions of $\psi(x)$ are compressed, and consequently contain high frequencies. These respond to localise high frequency characteristics of the signal. Those with a larger scale are stretched wavelets, and consequently contain lower frequencies, which respond to the low frequency of the signal. Thus by analysing with respect to both scale and position, W_f can provide information about the frequencies present at a certain position in the signal, or about the positions in a signal at which a certain frequency is present.

3.3.2 Theory of multiresolution analysis

The basic idea of multiresolution analysis is to represent any arbitrary function $f \in L^2(R)$ as a set of successive approximations of the original signal at different resolutions. Without affecting the generality of the proposed framework, Mallat considered particular resolutions of the form 2^j (dyadic) [128]. Let $V_{2^j} \subset L^2(R)$ be the vector space of all possible approximations at resolution 2^j of the functions in $L^2(R)$, and A_{2^j} the linear/projection operator in V_{2^j}. Among all the possible approximations of $f(x)$ at resolution 2^j, $A_{2^j}f(x)$ is the function which is the most

similar to $f(x)$ in the least square sense. Furthermore, the approximate of $f(x)$ at the resolution 2^{j+1} contains all the necessary information to construct the approximation of the same signal at the smaller resolution 2^j. Since A_{2^j} is the projection in V_{2^j}, this principle is equivalent to [128]:

$$V_{2^j} \subset V_{2^{j+1}}, \quad \forall j \in Z. \tag{3.6}$$

As shown in [Mallat 1989], there exist a function $\phi_{2^j} \in L^2(R)$, called a scaling function, such that if $\phi_{2^j} = 2^j \phi(2^j x)$ is the dilation at the scale 2^j of $\phi(x)$, then $\sqrt{2^{-j}} \phi_{2^j}(x - 2^{-j}n), n \in Z$ is an orthonormal basis in V_{2^j}. With the inner product of two functions $f(t)$ and $g(t)$ defined as follow:

$$\langle f(t), g(t) \rangle = \int_{-\infty}^{\infty} f(t) g^*(t) dt. \tag{3.7}$$

The * denotes the complex conjugate. The continuous approximation of $f(x)$ at the resolution 2^j can be computed as the orthogonal projection of the signal on this orthonormal basis:

$$\forall f(x) \in L^2(R), \quad A_{2^j} f(x) = 2^{-j} \sum_{-\infty}^{\infty} \langle f(u), \phi_{2^j}(u - 2^{-j}n) \rangle . \phi_{2^j}(x - 2^{-j}n). \tag{3.8}$$

The inner products of the form:

$$A_{2^j}^d f(n) = \langle f(u), \ \phi_{2^j}(u - 2^{-j}n) \rangle_{n \in Z} \tag{3.9}$$

constitute the *discrete approximation* of $f(x)$ at the resolution 2^j. These approximations can be computed using a fast algorithm, called the pyramidal algorithm, which was proposed by Burt and Adelson [128] in the context of image

coding, and has been extrapolated to the iterative calculation of the wavelet transform by Mallat [128]. Define the discrete filter h whose impulse response is given by:

$$\forall n \in Z, \quad h(n) = \langle \phi_{2^{-1}}(u), \; \phi(u-n) \rangle, \qquad\qquad (3.10)$$

and $\bar{h}$ the mirror filter given $\bar{h}(n) = h(-n)$ The discrete approximation $A_{2^j}^d f$ can be computed by convolving $A_{2^{j+1}}^d f$ with $\bar{h}$ followed by downsampling by 2 (i.e. keeping every other sample of the output) [128]:

$$A_{2^j}^d f(n) = \sum_{-\infty}^{\infty} \bar{h}(2n-k) A_{2^{j+1}}^d f(k). \qquad\qquad (3.11)$$

Let O_{2^j} be the orthogonal complement of V_{2^j} in $V_{2^{j+1}}$, satisfying $O_{2^j} \oplus V_{2^j} = V_{2^{j+1}}$ The difference of information between the discrete approximation at the resolution 2^{j+1} and the approximation at the coarser resolution 2^j is determined by identifying an orthonormal basis in O_{2^j} and projecting the signal $f(x)$ onto this basis. Mallat demonstrated that such an orthonormal basis could be constructed by scaling and translating a function $\psi(x)$, which is called an orthogonal wavelet [128]. The Fourier transform of the wavelet $\psi(x)$ is given by:

$$\Psi(\omega) = G(\tfrac{\omega}{2}) \Phi(\tfrac{\omega}{2}) \qquad\qquad (3.12)$$

where $\Phi(\omega)$ is the Fourier transform of $\phi(x)$, and:

$$G^*(\omega) = e^{-j\omega} H(\omega + \pi), \text{ with } H(\omega) = \sum_{-\infty}^{\infty} h(n) e^{-jn\omega}. \qquad\qquad (3.13)$$

With these definitions, it can be shown that $(\sqrt{2^{-j}} \psi_{2^j}(x - 2^{-j} n))_{(n,j) \in Z^2}$ is an orthonormal basis in O_{2^j} [128]. Projecting $f(x)$ onto this basis yields:

$$P_{O_{2^j}} f(x) = 2^{-j} \sum_{-\infty}^{\infty} \langle f(u), \ \psi_{2^j}(u - 2^{-j}n) \rangle . \psi_{2^j}(x - 2^{-j}n). \qquad (3.14)$$

The *discrete detail* of $f(x)$ at resolution 2^j is defined by the set of the inner products in (3.14)

$$D_{2^j} f(n) = \langle f(u), \ \psi_{2^j}(u - 2^{-j}n) \rangle_{(j,n) \in Z^2}. \qquad (3.15)$$

It can be shown that the discrete filter given by $g(n) = \langle \psi_{2^{-j}}(u), \ \psi(u-n) \rangle$ has the transfer function $G(\omega)$ defined by (3.13). The discrete detail $D_{2^j} f$ can be computed via the pyramidal algorithm by convolving $A_{2^{j+1}}^d f$ with $\overline{g}(n) = g(-n)$ followed by downsampling by 2 [128]:

$$D_{2^j} f(n) = \sum_{-\infty}^{\infty} \overline{g}(2n - k) A_{2^{j+1}}^d f(k). \qquad (3.16)$$

The h and g filters are low-pass and high-pass (band-pass) filters respectively, with their impulse responses satisfying [128]: $g(n) = (-1)^{1-n} h(1-n)$.

The pyramidal algorithm formalised by (3.11) and (3.16) decomposes the discrete approximation $A_{2^{j+1}}^d f$ into an approximation at a coarser resolution $A_{2^j}^d f$ and detail signal $D_{2^j} f$. Repeating in cascade, this algorithm for $-J \leq j \leq -1$ yields the Discrete (dyadic) Wavelet Representation (DWT) on J levels of the original signal $A_0^d f$:

$$A_{2^{-J}}^d f, (D_{2^j} f)_{-J \leq j \leq -1}. \qquad (3.17)$$

The block diagram of the wavelet decomposition process for 1-D signals is illustrated in figure 3.2a. In the inverse process, the discrete approximation $A^d_{2^{j+1}} f$ at the higher resolution 2^{j+1} is reconstructed from $A^d_{2^j} f$ and the detail signal $D_{2^j} f$ as:

$$A^d_{2^{j+1}} f(n) = 2 \sum_{k=-\infty}^{\infty} h(n-2k) A^d_{2^j} + 2 \sum_{k=-\infty}^{\infty} g(n-2k) D_{2^j} f(k). \qquad (3.18)$$
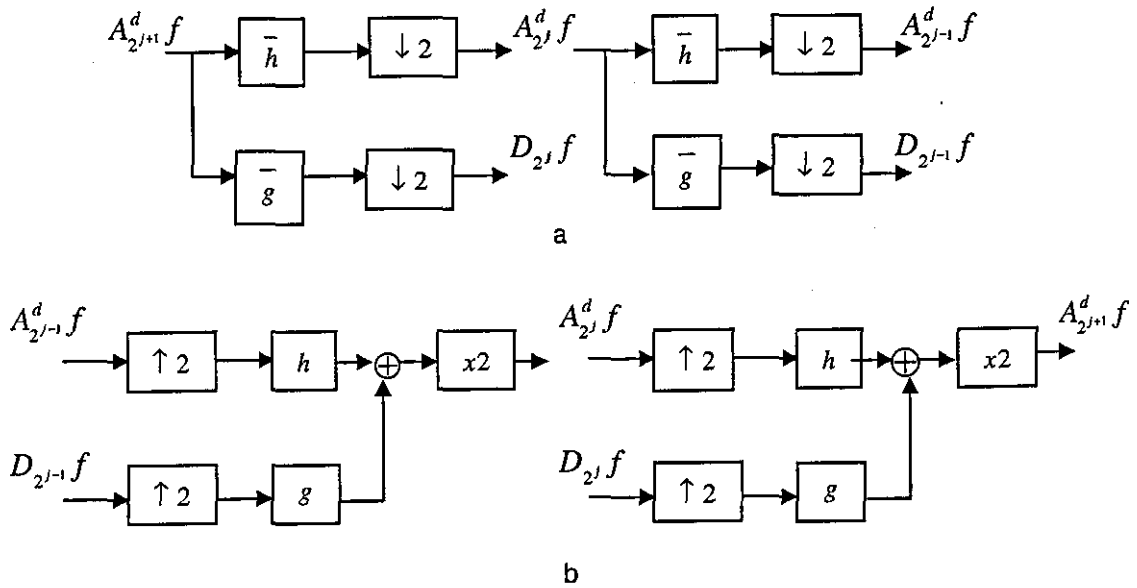
Figure 3.2. **(a) 1-D Forward Discrete Wavelet Transform**
(b) 1-D Inverse Discrete Wavelet Transform

Thus, the Inverse Discrete Wavelet Transform (IDWT) determines $A^d_{2^{j+1}} f$ by upsampling the signals $A^d_{2^j} f, D_{2^j} f$ through insertion of zeros, followed by the application of the synthesis filter h, g. The original signal is iteratively reconstructed from its wavelet representation as shown in figure 3.2b

In the 2-D case, a multiresolution decomposition on J levels of an image f results into a set of $3J +1$ subbands, defining the DWT representation on J levels of f :

$$(A^d_{2^{-J}} f, HL_{2^j} f, LH_{2^j} f, HH_{2^j} f)_{-J \le j \le 1}. \qquad (3.19)$$

The 2-D wavelet decomposition consists of successively applying the one-dimensional DWT, using a row-column approach. We first convolve the rows of $A_{j+1}^d f$ with an one-dimensional filter, retain every other sample, convolve the columns of the resulting signals with another one-dimensional filter and retain every other sample; for details, the reader is referred to [128]. A detailed discussion of applying DWT to an image is discussed in the section 3.4.

3.4 Applying DWT on an image and inter-subband relationship

An image may be decomposed into different decomposition levels using a DWT. In applying DWT, the coefficients at each level of decomposition are generated by applying a cascade of two-channel filter banks to the image. Each decomposition level contains a number of subbands, which consist of coefficients that contain the horizontal and vertical spatial frequency characteristics of the original image. *Power of two* decompositions is allowed in the form of a dyadic decomposition, as shown in figure 3.3.
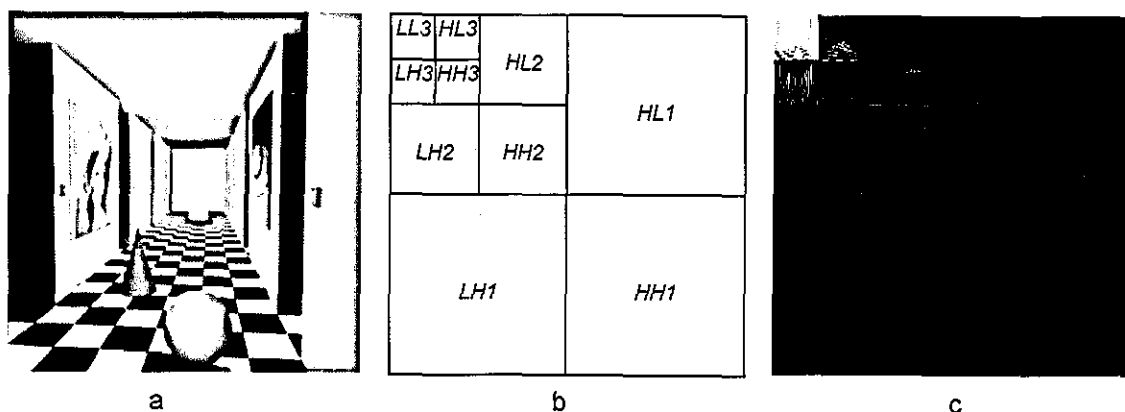
Figure 3.3 (a) Original 'Corridor' image (b) Subband naming convention (c) 3-level DWT of 'Corridor' image

Each image is decomposed into a collection of 3S+1 subband images, where S is the number of levels. Specifically, a one-dimensional low pass filter (L) and a one-dimensional high pass filter (H), each based on a mother wavelet, are applied horizontally and vertically to the original image, and the outputs are subsampled by a factor of two, to create four subband images. For example, for a three level decomposition, the first DWT stage (level 1) decomposes the image into four subbands, denoted by *LL1* (low-resolution version of original image), *HL1* (vertical features of the image), *LH1* (horizontal features of the image) and *HH1* (diagonal features of the image). The next DWT stage decomposes the *LL1* subband into four more subbands, denoted *LL2, LH2, HL2* and *HH2*. Lastly at level 3, the *LL2* subband is decomposed further into four more subbands, denoted by *LL3, LH3, HL3* and *HH3* as shown in figure 3.3b. In general, for S greater than 3, this process continues for some number of levels, S, producing a total of 3S +1 subbands whose samples represent the original image. The total number of samples in all subbands is identical to that in the original image. An example of a three level dyadic decomposition into subbands of the left image 'Corridor' is illustrated in figure 3.3c.

Regardless of the particular wavelet filter used, the image is decomposed into subbands, such that lower subbands correspond to higher image frequencies (they are the highpass levels) and higher subbands correspond to lower image frequencies (lowpass level), where most of the image energy is concentrated. This is why we can expect the detail coefficients to get smaller as we move from high to low levels. Also, there are spatial similarities among the subbands. For example an image part, such as an edge, occupies the same spatial position in each subband. In our proposed coders, these relationships among subbands of the wavelet decomposition are exploited in order to perform DE simultaneously in all subbands. In addition, based on the inter-subband relationship introduced above, we are able to implement all our proposed CODECs using a baseline JPEG-like framework.

3.5 The coefficient / border extension problem

Different wavelet filters produce different results when applied to the same image [174]. Although, no claims have been made for an optimal wavelet filter design for image compression, the desired properties of wavelet families have been stated in [152]. These properties are, compact support, orthonormality and symmetry.

Compact support of the wavelet filters is needed to capture effectively the salient information of an image through local variations such as local changes (non edges) over a large spatial area and abrupt changes (edges) over a small spatial span of the image. The orthogonal wavelet filter can be viewed as projecting the input signal onto a set of orthogonal basis functions. If the wavelet filters are also normalized, then this results in energy preserving (no loss of information) in the representation. However, the standard orthogonal wavelet transform has one shortcoming, called *coefficient (border) extension* problem, where the total number of wavelet coefficients after wavelet transform and downsampling, is greater than the number of input samples. In general, if the input signal is of length N (say even), and wavelet filter is of length M (even), the outputs of the filters h (low pass) and g (high pass) will be of length $N+M-1$, and after downsampling, the output will be of length $(N+M)/2$. Thus, after one level of wavelet transform, the length of the output, $L= N+M$, is greater than N. This increase will continue in a similar way at the next level of decomposition. Thus, coefficient expansion is a problem for a coding system, where the fundamental aim is to reduce, but not increase, the amount of information to be coded.

The coefficient extension (border distortion) problem can be solved by *symmetric periodic extension* [193]. More precisely, if N is the length of an input signal X, then the extended signal is $ext(X) = [X(N-M + 1:N) \ X \ X(1:M)]$ where M is the length of the filter. Consequently, convolution and downsampling operations are done, followed by keeping the central part of length $N/2$. Thus, after one level of decomposition, the total number of wavelet coefficients is N (i.e., no coefficient expansion). However, in symmetry periodic extension, the symmetry needs to be preserved across scales in order to enable the downsampling half of the wavelet coefficients by symmetry. Failure to maintain the symmetry across scales results in

large wavelet coefficient magnitudes, caused by border discontinuities. Symmetry can be preserved across scales by using symmetric (linear phase) wavelet filters, h (low-pass) and g (high-pass). However, for compactly supported orthogonal wavelets, only the Haar wavelet is symmetric. The lack of linear phase filters in orthogonal wavelets had led to research and discovery of a more general form of wavelets known as biorthogonal wavelets. But, the biorthogonal wavelets differ from orthogonal ones in that the forward wavelet transform is equivalent to projecting the input signal onto non orthogonal basis functions. This means, it is not energy preserving. However, there exist biorthogonal wavelets, which are close to being orthogonal, such as the bior9/7 filter. In [193], it has been shown that the energy loss of by bior9/7 filter is about 2%. In chapter 4, experimental results show that bior9/7 (compactly supported, 'closely' orthogonal, and symmetric) gave better performances than orthogonal wavelets, such as db7 implemented in MATLAB. This is the reason for the bior9/7 wavelet to be used in lossy JPEG-2000 standard. We have used this DWT in the design of all stereo image CODECs proposed in this thesis.

3.6 Coding of wavelet coefficients

There are three well known techniques which are used for the efficient coding of wavelet coefficients; Embedded Zerotree Wavelet Coding (EZW) proposed by Shapiro in 1993 [179], Set Partitioning in Hierarchical Trees (SPIHT) proposed by Said and Pearlman in 1996 [16] and Embedded Block Coding with Optimized Truncation (EBCOT) proposed by Taubman in 2000 [46]. In this section, we discuss briefly discuss the operation of these coding techniques.

A common feature of all above coding techniques is their ability to exploit the multiresolution nature of the wavelet decomposition to produce embedded bit streams. The embedded bit stream is based on the principle that large wavelet coefficients contain more information, thus are more important and should be placed at higher priority order in the resultant coded bit stream. Note that, embedded/scalable compression refers to the generation of a bit-stream, which contains embedded

subsets, each of which represents an efficient compression of the original image at a reduced resolution or increased distortion. The term "resolution scalability" and "SNR scalability" are commonly used in connection with this idea. The EZW and SPIHT are SNR scalable whereas EBCOT is SNR and resolution scalable. This difference in scalability is mainly due to the way the wavelet coefficients are coded to generate the bit streams.

In EZW, to generate the embedded bit stream, the subbands are scanned in the following fixed order; for a N-scale transform, the scan begins at the lowest frequency subband, denoted as LL_N, and scans subbands HL_N, LH_N, and HH_N, at which point it moves on to scale N-1, etc. Each coefficient within a given subband is scanned before any coefficient in the next subband. During the scan, the significance map is represented based on a data structure called the *zerotree*. A zerotree is based on the hypothesis that if a wavelet coefficients at a coarse scale is insignificant with respect to a given threshold T, then all wavelet coefficients of the same orientation in the same spatial location at finer scales are likely to be insignificant with respect to T. If this is the case, the root of the zerotree is coded with symbol ZTR and the remaining coefficients, which are considered insignificant, are not transmitted. Meanwhile, the significant coefficients are placed in the bit stream according to the scanning order in descending order, which allow for SNR scalability. However, the fixed scanning order through all the subbands makes EZW non scalable, in terms of resolution.

In SPIHT, the wavelet coefficients are partitioned into a number of sets T_k, using a special data structure called a spatial orientation tree. This structure also exploits the spatial relationships between the wavelet coefficients in the different levels of the subband pyramid. In general the number of sets is equal to three quarters of the number of coefficients of the highest level subband. Each set goes through a significant test, which results in a series of 1's or 0's, based on the following:

$$S_n(T) = \begin{cases} 1, & \max_{(i,j)\in T} |c_{i,j}| \geq 2^n \\ 0, & otherwise \end{cases}. \qquad (3.20)$$

58

If the result is "1", then T_k is partitioned by both encoder and decoder, using the same rule, into subsets and the same significance test is performed on all the subsets. This partitioning is repeated until all the significant sets are reduced to size 1 (i.e., contain one coefficient each, and that coefficient is significant). As in EZW, based on the fact that the largest wavelet coefficients contain most important information and should be sent first. However, SPIHT differs from EZW in that bit-plane coding is used, where the most significant bits of the binary representation of the wavelet coefficients are transmitted first as compared to bits with lesser significance.

The EBCOT algorithm [42], differ from EZW and SPHIT. The principle of EBCOT is to divide each subband into blocks (termed code-blocks) that are independently coded using a bit-plane coder. No inter subband relationship (figure 3.4) is exploited in EBCOT as in EZW and/or SPHIT. A detailed description of EBCOT is provided in chapter 6.
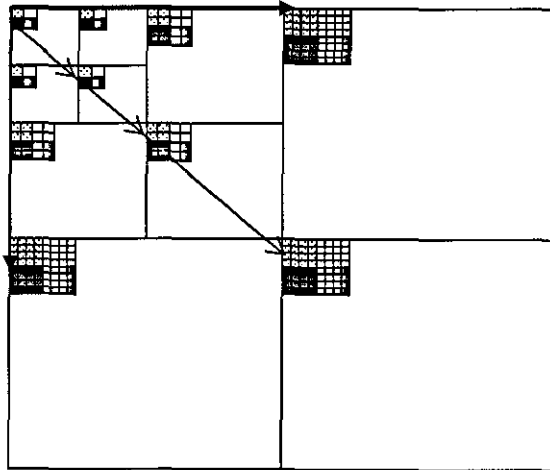
Figure 3.4 Inter-subband relationships in multiresolution DWT domain

3.7 Comparative evaluation of JPEG-2000 versus baseline-JPEG

Baseline-JPEG, a DCT-based technology, became a still image compression standard in 1992 [61]. At that time, DCT-based transform techniques were well established as compared to wavelet based image analysis and coding techniques [193]. However, since the introduction of the baseline-JPEG standard, the wavelet-based technologies advanced significantly. In addition to this, with the continual expansion of multimedia and Internet applications, the needs and the requirements of the technologies used, grew and evolved. Thus a new still image compression standard was needed to cater for these new demands. As a result of this, in March 1997, a new call for contributions was launched for the development of a new standard for the compression of still images, known as JPEG-2000 [42]. The standardization process, which was coordinated by the JTC1/SC29/WG1 of ISO/IEC, made JPEG-2000 a new International Standard (IS) in December 2000 [174].

JPEG-2000 provides an entirely new way of compressing images based on the wavelet transform, with rate-distortion performance characteristics superior to that of baseline-JPEG, especially at low bit rates. In addition, the DWT, which provide multiresolution decomposition of an image, enables new features such as resolution and fidelity scalability, region of interest coding, random access, etc., which are important in present day applications [174].

Figure 3.5, illustrates the subjective image quality of 'Lena' image, compressed at compression ratios; 30:1 and 200:1 using JPEG-2000 and baseline-JPEG. At the higher bit rate (30:1), there seem to be no significant difference between JPEG-2000 and baseline-JPEG coded images, as illustrated by figures, 3.5 (a) and (b). At the compression ratio 200:1, JPEG-2000 coded image (fig. 3.5(d)) has better edge quality as compared to the baseline-JPEG coded image (fig.3.5(d)). This is due to the excessive blocking artifacts of DCT based coding, taking predominance over the ringing artifacts of DWT coding. Similar results can be observed in figure 3.6 for the image 'Barb' and in figure 3.7 for the image 'Boat'. Thus it could be concluded that even though there is not that much of a difference in subjective quality of JPEG-2000

and baseline-JPEG coded images at high and medium bit rates, at low bit rates, JPEG-2000 outperforms baseline-JPEG.

a   b

c   d

Figure 3.5 Comparison of image quality of 'Lena' at different compression ratios (a) JPEG, 30:1   (b) JPEG-2000, 30:1   (c) JPEG, 200:1 (d) JPEG-2000, 200:1

Figure 3.6 Comparison of image quality of 'Barb' at different compression ratios
(a) JPEG, 30:1   (b) JPEG-2000, 30:1 (c) JPEG, 200:1 (d) JPEG-2000, 200:1
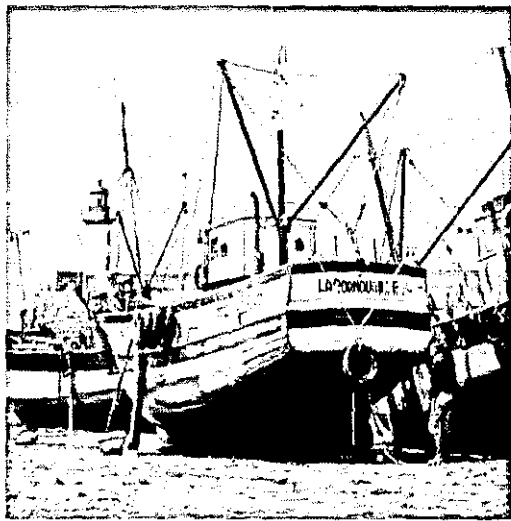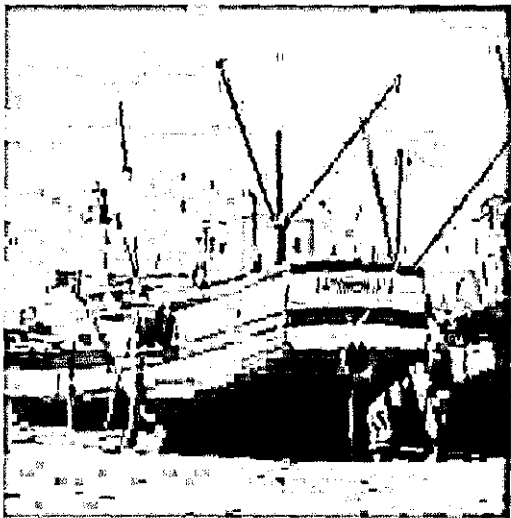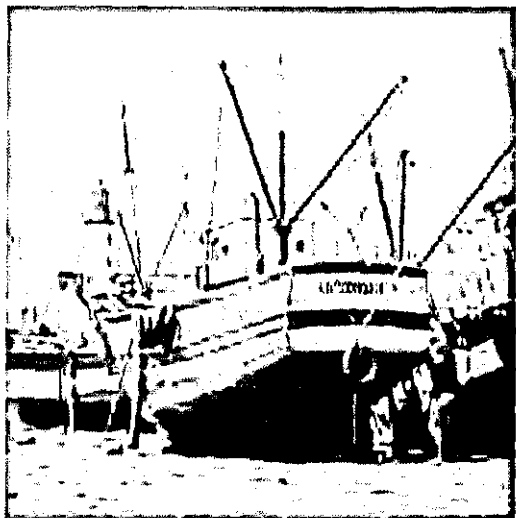
Figure 3.7 Comparison of image quality of 'Boat' at different compression ratios
(a) JPEG, 30:1   (b) JPEG-2000, 30:1 (c) JPEG, 250:1 (d) JPEG-2000, 250:1

4 Wavelet-block based stereo image coding

4.1 Overview

This chapter proposes three novel algorithms for wavelet-block based stereo image coding and discusses the operational features of the benchmark (section 4.2) chosen to evaluate their performance. Section 4.3 introduces the basic technology used within the designs of the three novel CODECs proposed in this chapter. Section 4.4 includes the design, implementation, experimental results and analysis of the DWT- based coder (CODEC I) proposed first. It applies DWT to the reference frame and the predictive error frame obtained by pixel domain disparity estimation & compensation. In addition it uses an architecture similar to that of baseline-JPEG for subsequent coding of the reference and predictive error images. Section 4.5 discusses the design, implementation and testing of CODEC II and CODEC III, where, in contrast to the pixel-based approach used in CODEC I, the disparity estimation and compensation is performed between the corresponding subbands of the DWT decomposed left and right images. The operational difference between CODEC II and CODEC III is that in CODEC II, disparity estimation and compensation is performed only between the lowest resolution subbands whereas in CODEC III, coefficients of all subbands belonging to a given area is considered in the disparity estimation and compensation process. The chapter includes an introduction to the *wavelet-block* concept used throughout the designs. The effect of using different wavelets in the designs is also investigated.

4.2   The benchmark – Disparity Compensated Transform Domain Predictive Coding (DCTDP)

The benchmark algorithm, Disparity Compensated Transform Domain Predictive Coding, is one of the most frequently used benchmark algorithms [49, 50,

85] in literature. This is mainly due to its efficiency in coding stereo image pairs with varying statistics at a wide range of bit rates. The main structure of this scheme is illustrated in figure 4.1.

The Encoder: Figure 4.1, represents the block diagram of the encoder of the benchmark algorithm. The reference image (i.e. original left), L_{Ori}, is independently transmitted using the baseline-JPEG compression algorithm [61]. The right image, R_{Ori}, is sub-divided into non-overlapping 8x8 blocks. These blocks are subsequently used to search the reference image, for their best match within a specified search window area. As we assume that the stereo pair has been obtained with parallel axis camera geometry, theoretically, the vertical disparity would be negligible. Therefore the above search can be limited to the horizontal direction. The best match is selected based on minimum Mean Square Error (MSE), which corresponds to the highest correlation between the matched blocks.

Let the right image block to be encoded, be denoted by R_{blk} and it's best match obtained following the above procedure be denoted by, L_{BM_blk}. The DCTDP algorithm calculates the predicted value, $\hat{R}_{blk}$, of the block to be encoded, using a linear predictor, as follows:

$$Tr(\hat{R}_{blk}) = A \otimes Tr(L_{BM\_blk}) + B \qquad (4.1)$$

where, Tr represents DCT transformation of a given block, A and B are constant matrices and $\otimes$ represents *element-by-element* multiplication of two matrices.

Subsequently, if PE_{blk}, represents the prediction error block, it is calculated in the transform domain as follows:

$$Tr(PE_{blk}) = Tr(R_{blk}) - Tr(\hat{R}_{blk}) = Tr(R_{blk}) - (A \otimes Tr(L_{BM\_blk}) + B). \qquad (4.2)$$

The above is subsequently compressed, entropy coded and transmitted following a scheme similar to that used in baseline-JPEG.

The above process is repeated until the prediction error blocks for all non-overlapping blocks of the right image have been obtained in the transform domain, quantized, encoded and transmitted to the decoder. The resulting disparity vector field is coded using a fixed length code and losslessly transmitted. In addition the matrices A and B should be coded and transmitted to the decoder.

The *Decoder*: The losslessly transmitted/received disparity vectors are used to locate the best match for each block to be decoded, in the decoded reference image. The block thus located is added to the corresponding reconstructed prediction error block to form the decoded block. The process is repeated until all blocks of the right image have been decoded.

Figure 4.1 The DCTDP Encoder

The inclusion of the linear predictor, $A \otimes Tr(L_{BM\_blk}) + B$, can be justified as follows [155]. In the DCTDP scheme, the disparity estimation is performed in the pixel domain, yet the prediction difference is calculated in the DCT domain. In [155], M.G.Perkins points out that the correlation between low frequency components of the matching blocks is greater than the correlation between the high frequency components [155] and that an improvement in the compression performance of DCTDP encoder can be obtained by the use of matrix A in the linear predictor above. In a typical case, the A matrix consists of values closer to 1, for elements corresponding to DC and low frequency components. The values of matrix A elements decrease with increasing frequency (i.e. down the zig-zag order). The constant matrix B has been included in the linear predictor to compensate for the differences in gain of the two cameras and their associated digitization circuitry [155]. As the differences are mainly in the gains of the digitization circuits, the matrix B affects only the DC component of the transformed matrix.

The calculation of matrices, A and B is a challenging task, which is typically carried out by initially training the encoder on a collection of stereo image pairs. Our experiments showed that such a training process does not guarantee the calculation of a set of matrices A and B that would apply well for all stereo image pairs. In addition to the extra computational burden of this process, one has to transmit the matrices A and B to enable correct decoding. Given the above facts, the marginal compression gain obtainable when using the linear predictor does not justify the additional computational burden of the process. Therefore in selecting our benchmark, we have excluded the use of the linear predictor. Thus the transform of the prediction error is calculated as follows:

$$Tr(PE_{blk}) = Tr(R_{blk}) - Tr(\hat{R}_{blk}) = Tr(R_{blk}) - Tr(L_{BM\_blk}) = Tr(R_{blk} - {}_{BM\_blk}) \quad (4.3)$$

4.3 The basic technology

All three novel CODECs proposed in this chapter have differing design features but are based on DWT technology and a baseline-JPEG like CODEC architecture. This section introduces this technology in detail.

4.3.1 Why replace DCT by DWT?

Preserving the quality of the edges is an important criterion in stereo image coding as edges are one of the main binocular cues in 3D perception. Image compression techniques based on the DCT, such as the benchmark; suffer from homogeneous blocking artifacts at low bit rates that make it unsuitable for low bit rate stereo image coding. These blocking artifacts cause severe distortion of block edges, which could affect the quality of object edges within the scene. However, the benchmark has the advantage of having a simple, well-established JPEG-like coding architecture, which may be beneficial in certain applications and easy hardware implementation. Alternatively, DWT, which is the core underlying technology of JPEG-2000 still image coding standard, offers much improved rate-distortion performance characteristics as compared to the DCT-based baseline-JPEG standard [61]. Even though ringing artifacts in the form of distortions at edges are present in DWT-based coding, the said distortions only affect viewing quality at very low bit rates. In other words, at any given bit rate (especially at lower rates), DWT-based coding offers better image edge quality than DCT-based coding. Unfortunately the subsequent quantization, entropy coding techniques used with typical DWT coding techniques, complicates the design of a straightforward stereoscopic version of a typical monoscopic image CODEC. Thus, all the CODECs use a simple architecture similar to that of a DCT-based baseline JPEG-CODEC [61] for easy hardware implementation, but effectively replaces the DCT technology by the more efficient DWT technology for improved rate-distortion performance.

It should be noted that the use of quantization and entropy coding techniques typically associated with DWT coding would be straightforward in the design of the proposed CODECs and should lead to further improvement of the rate-distortion performances

of all CODECs proposed in this thesis. To this effect, the main concentration of this thesis and the associated research is developing effective Disparity Estimation and Compensation techniques that could be associated with DWT coding.

4.3.2 Formation of wavelet-blocks

In applying DWT to an image, the coefficients are generated by applying a cascade of two-channel filter banks. Figure 4.2(a) represents a three-level wavelet decomposition of an image. Usually in DWT subband coding procedures, the coefficients are grouped into subband-oriented groups (common subband, different spatial location) whereas in DCT, they are grouped into blocks (common spatial locations, different subbands). In order to use an architecture similar to the benchmark, a so called *wavelet-block* structure could be formed. The idea behind wavelet-blocks is to group the DWT coefficients into blocks as illustrated in figure 4.2, so that the grouping is similar to that used by a DCT-based subband coding procedure.

As illustrated in figure 4.2, the final result is the transfer of the input image into a re-organized data structure of original image size, but consisting of a two-dimensional array of non-overlapping wavelet-blocks, each representing different subband elements from the location corresponding to the block. In other words, for a S-level DWT, blocks of $2^S \times 2^S$ samples are constructed. Each wavelet-block is subsequently scanned into a vector (see figure 4.2(b)) in order to be processed by the remaining parts of the baseline JPEG-like coding procedure.

A single wavelet-block

a   b

Figure 4.2 Formation of a single wavelet-block from a 3-level DWT decomposed image (a) Decomposed image (b) Scan order for a single wavelet-block

4.4 CODEC I – DWT-based stereo image coding with pixel domain disparity estimation & compensation

The proposed CODEC I is based on performing disparity estimation and compensation in the pixel domain and the subsequent DWT based coding of the reference and predictive error frames.

4.4.1 Design details

Figure 4.4 represents the block diagram of the proposed CODEC I. Note that it is assumed here that the disparity values are suitably coded and transmitted losslessly to the decoder end.

At the encoder end, the right image is divided into non-overlapping 8×8 pixel blocks. For each block, a search is performed within a windowed area of maximum likelihood on the left image (considered as reference image) to find the best predictor.

Assuming that parallel axis camera geometry is used to obtain the stereo image pair, one could limit the above mentioned search window in only a horizontal direction, spanning to the left of the corresponding block (of the block to be encoded) in the left image. The best predictor is found using the minimum MSE as the selection criteria,

71

i.e., the matching block that provides the least MSE is selected as the predictor. The prediction errors of all the non-overlapping right frame blocks are then combined to form the PEI.

The entire left (L) and PEI images then undergo 3-level dyadic DWT separately and are subsequently converted into their wavelet-block representations, by following the procedure described in detail in section 4.3.2. Each wavelet-block of the two representations then undergo scanning as illustrated in figure 4.2(b) as against the zig-zag scanning procedure normally adapted by baseline-JPEG [61]. This modification of scanning procedure is essential as in the wavelet-block representation, for a given block; there are coefficients that not only belong to different subbands, but also multiple coefficients which belong to the same subband. Disregard of this fact and the use of the typical zig-zag scanning order of DCT coefficients could severely affect coding efficiency.

After the scanning, the ordered coefficients undergo scalar quantization. We use the strategy adapted in [167] to determine a suitable quantization table, that provides results equivalent to the 64-entry, uniform, fixed quantization table used by the baseline-JPEG standard shown in figure 4.3. For easy reference we provide a brief description of this Quantizer design as follows:

Let the block size be $N \times N$ ($N=2^S$) and the step size be Δ_{ij} for $0 \le (i,j) \le N-1$. The step size used in the proposed scheme is then calculated as,

$$\Delta_{ij} = \frac{A}{H\left(\frac{\sqrt{i^2+j^2}}{N}\right)}, \qquad\qquad (4.4)$$

where A is a scaling parameter that is used to control the bit rate. The model for $H(x)$ is defined as,

$$H(x) = \left(a_0 + a_1 x + a_2 x^2 + a_3 x^3\right) e^{\left(a_4 x + a_5 x^2\right)} \qquad\qquad (4.5)$$

The parameters are then optimized for $S=3$ in such a way as to minimize the error between the achieved table and the luminance quantization table provided by baseline- JPEG. The parameters thus found are,

$$a_0 = 0.794, \ a_1 = -1.639, \ a_3 = 0.614, \ a_3 = 0.470, \ a_4 = 4.277 \text{ and } a_5 = -4.892.$$

Due to the use of DWT in its block-based representation, in a wavelet-block there will be several coefficients in the same band and the model developed for the DCT will have different step sizes for different coefficients in a block. In order to solve this problem the quantizer steps obtained above are averaged, on a per-subband basis and all coefficients in a particular subband are replaced by this average value. Figure 4.4 compares the Quantizer table obtained with $A=6.7$ using the strategy described above, with the standard luminance quantizer table of baseline-JPEG.

After the quantization, the quantized coefficients are subjected to entropy coding (Runlength/Huffman coding) as in baseline-JPEG. The decoder operations are essentially the inverse processes of the encoder processes and are clearly illustrated in figure 4.4.

$$Q_{JPG} = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix} \quad Q_{DWT} = \begin{bmatrix} 8 & 7 & 8 & 8 & 34 & 34 & 34 & 34 \\ 7 & 7 & 8 & 8 & 34 & 34 & 34 & 34 \\ 8 & 8 & 12 & 12 & 34 & 34 & 34 & 34 \\ 8 & 8 & 12 & 12 & 34 & 34 & 34 & 34 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \end{bmatrix}$$

Figure 4.3 Comparison of Quantizer tables
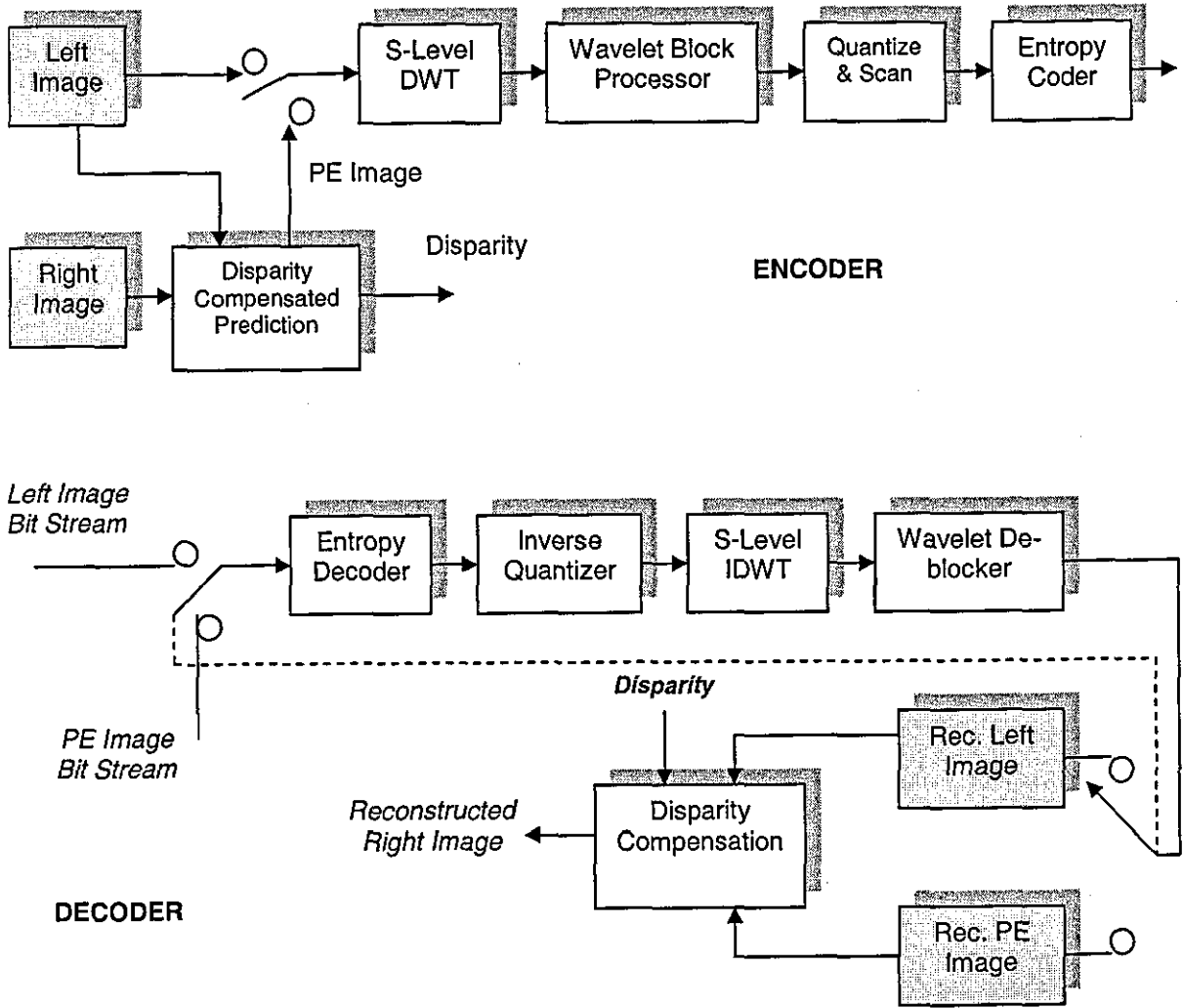
Figure 4.4 Proposed wavelet-based stereo Image CODEC I

4.4.2 Experimental results & analysis

The proposed CODEC I was implemented using purpose built MATLAB routines and was tested on a set of six commonly used test stereo image pairs representing indoor, outdoor, natural and synthetic scenes. All test stereo image pairs were of size 512×512 or 320×320 or 256x256.

Parallel axis camera geometry has been used to acquire these images. Thus a windowed search area representing a disparity range of 0-7 was used. The disparity vector field was coded using a fixed length code, requiring $\log_2(a+1)$, bits per block

where $a = 7$. The use of different wavelets in the coding process was investigated. A Daubachies compactly supported orthornormal filter (db7), a Daubachies biorthogonal (nearly orthogonal and symmetric) filter (bior 9/7) and different length Daubachies filters for different levels of decomposition (db7 for level 1, db5 for level 2 and db3 for level 3) were used as the wavelet transform for the reference image and the PEI.

In order to evaluate the performance of the proposed coder, in figure 4.5, we compare the rate-distortion performance of the proposed coder when using different wavelets, against the rate-distortion performance of the benchmark algorithm that uses DCT (i.e. baseline-JPEG) as the basic compression engine.

The objective image quality is measured in terms of Peak Signal to Noise Ratio (PSNR), whereas the compression efficiency is measured in bits per pixels (bpp). The compression efficiency (or Bit Rate) is measured as follows:

$$BR = \left[\frac{Tot\_Bits_{comp}}{N \times M} \right] \quad bpp$$

where, N and M represent the image dimensions and Tot_Bits_{comp} is the total number of bits required to represent a given image in its compressed format.

For a fair comparison, the objective quality of the left image (reference) was set at equivalent levels for both CODEC I, and the benchmark. For example for SScastle, Castle and Bottle the reference image was set at 41.32 dB, 39.38 dB and 35.76 dB, respectively. The rate distortion performance graphs clearly indicate the performance of CODEC I, for all the reconstructed right images bit rates based on different wavelets used. When the biorthogonal wavelet bior9/7 is used, for the SScastle image (see figure 4.5 (a)), CODEC I provides enhanced RD performance at all bit rates as compared to the benchmark. Contrary to our expectations, the orthogonal wavelet db7 performs worse compared to the benchmark at bit rates lower than 0.38 bpp and better at bit rates above this value. The combination of db7, db5 and db3 wavelets provides an RD performance identical to the benchmark at bit rates lower than 0.38 bpp, but performs marginally better above this value. For the Castle image (see figure 4.5 (b)),

when using the wavelet bior9/7, CODEC I provides a marginally better performance as compared to the benchmark, below the bit rate 0.27 bpp. As expected, CODEC I performs worse above this bit rate. Further, the benchmark outperforms CODEC I at all bit rates when the orthogonal wavelet db7 or combination of db7, db5 and db3 wavelets are used. As for the Bottle stereo pair, the performance of CODEC I when using bior9/7 is marginally better than the benchmark for bit rates below 0.5 bpp, but worse above this rate. The use of the remaining two filters in CODEC I results in a performance that is worse as compared to the benchmark at all bit rates.

Results of the above experiments show that among the wavelets used in CODEC I, biorthogonal wavelet filter 9/7 gave the best performance. This is due to the periodic/symmetric extension used to solve the coefficient expansion problem at the boundary of subbands (see Chapter 3) is better solved by the symmetric property of the bior9/7 wavelet as compared to non symmetric property of orthonormal wavelet filter db7 or the combination of filters, db7, db5 and db3. In addition, figure 4.5 (a) and (b), show that using shorter length wavelet filters at higher decomposition levels (db7 for level 1, db5 for level 2 and db3 for level 3), produces better result than using same length wavelet (db7) throughout all decomposition levels. This is due to the fact that shorter length wavelet filters are more appropriate to smaller sized subbands as filtering distortions are minimized.

Figure 4.6, 4.7 and 4.8 compare the subjective image quality of the reconstructed right images when using CODEC I (based on bior9/7 only) and benchmark techniques. It illustrates that the subjective image quality achieved by the proposed CODEC I, is equivalent to that of the benchmark. This proves that the marginal PSNR improvement of CODEC 1 at lower bit rates is not significant enough for a subjectively noticeable difference.

The overall analysis of the above results show that even though one would expect the DWT to perform better than the DCT at low bit rates for all images, results in figures 4.5, a, b and c does not directly imply this. Apart from the negative affect of the coefficient expansion problem discussed above a second reason that results in the

degraded performance of CODEC I is as follows: Due to the block based disparity estimation in the pixel domain, the PEI contains of artificial block boundaries. When DWT is applied on such an image, the artificial block boundaries result in an overall increase of information in the higher resolution subbands. This results, in the reduction of efficiency in the coding of DWT coefficients and hence the R-D performance of CODEC I.
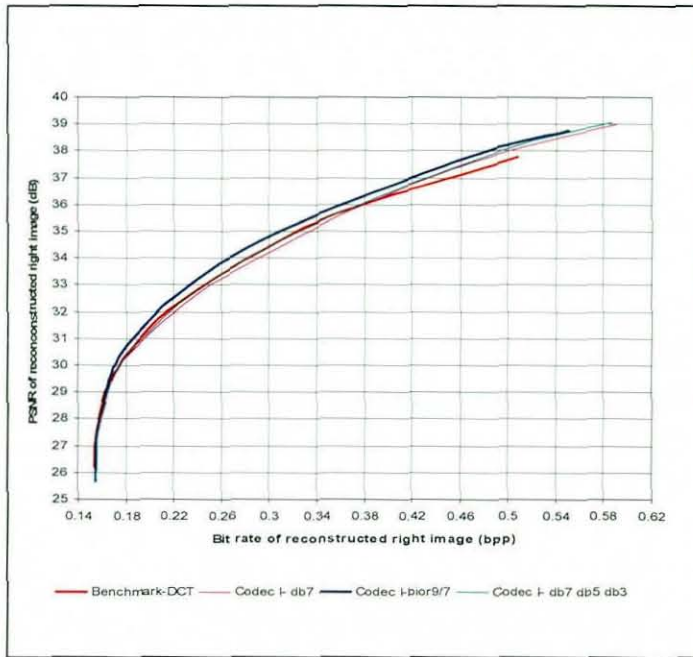
a
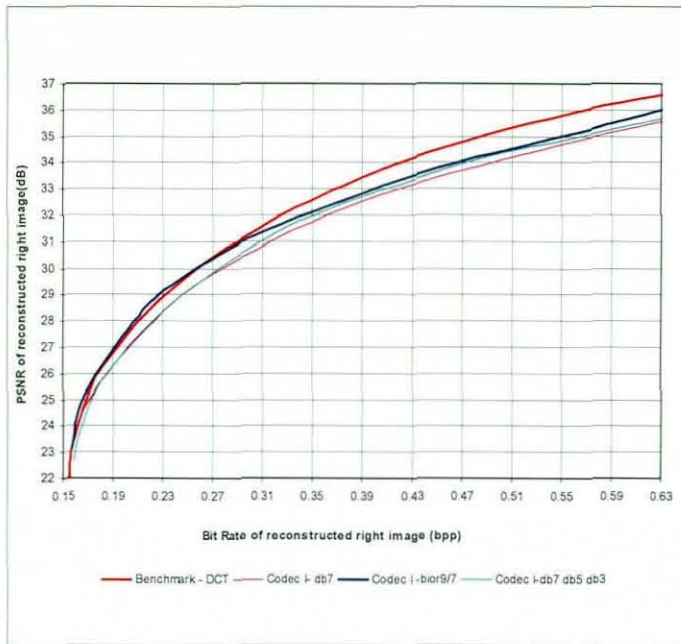
b
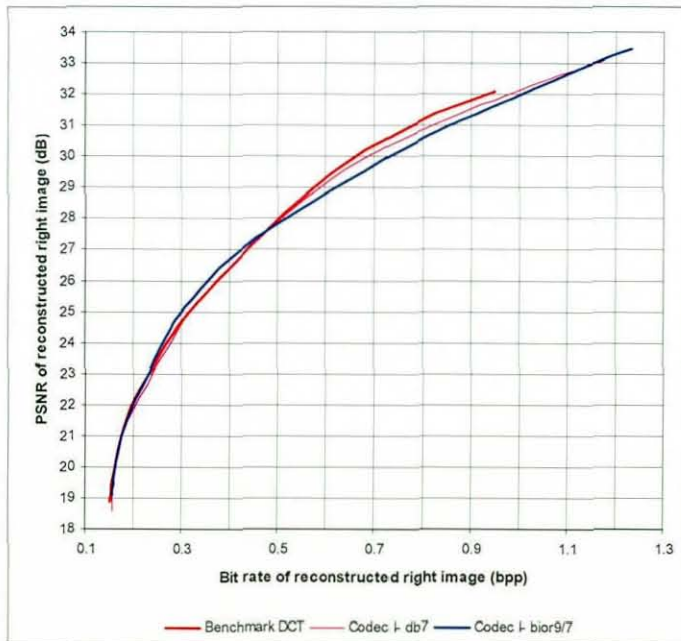
c

Figure 4.5 Comparison of rate-distortion performance
graphs for reconstructed right image of
(a) SScastle (b) Castle (c) Bottle

Figure 4.6 Subjective quality evaluations for SScastle
(a) Original left image (b) Original right image (c) Benchmark reconstructed right image , 0.1836 bpp, 30.54 dB (d) CODEC I reconstructed right image 0.1836 bpp, 30.87 dB (bior9/7)
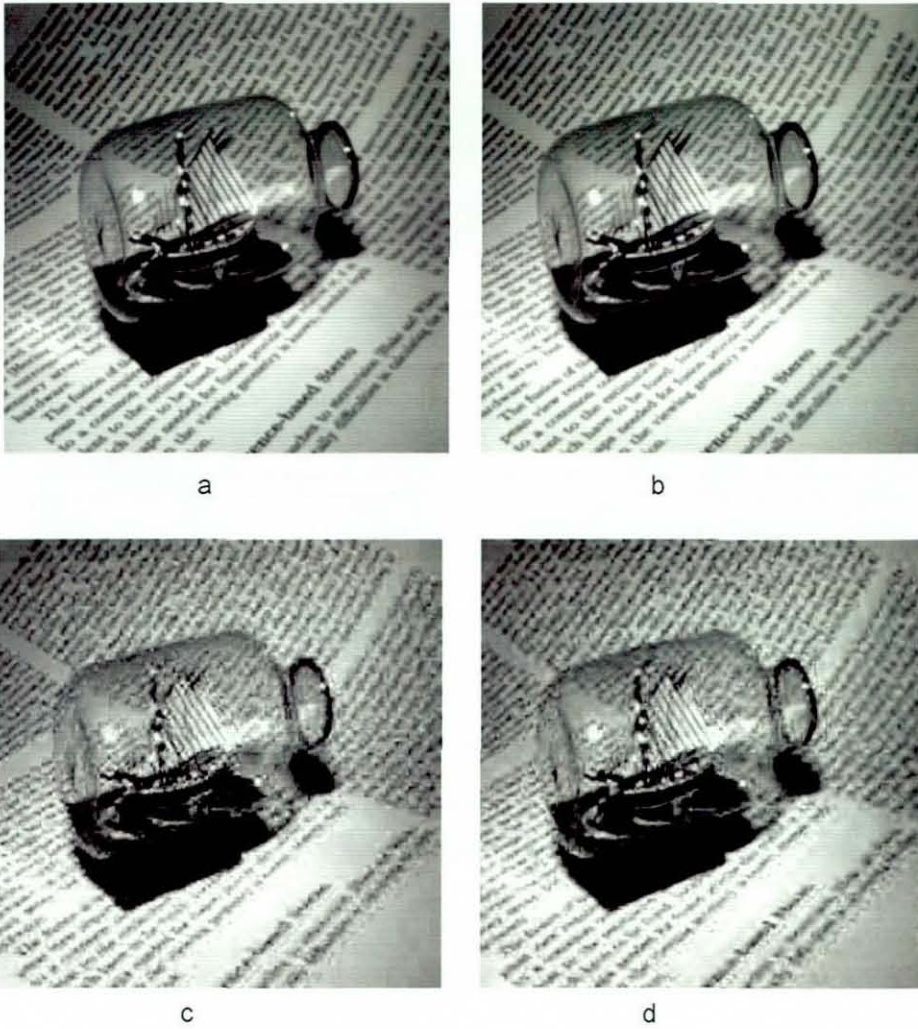
Figure 4.7 Subjective quality evaluations for Bottle
(a) Original left image (b) Original right image (c) Benchmark reconstructed right image, 0.2785 bpp, 24.14 dB (d) CODEC I reconstructed right image 0.2785 bpp, 24.45 dB
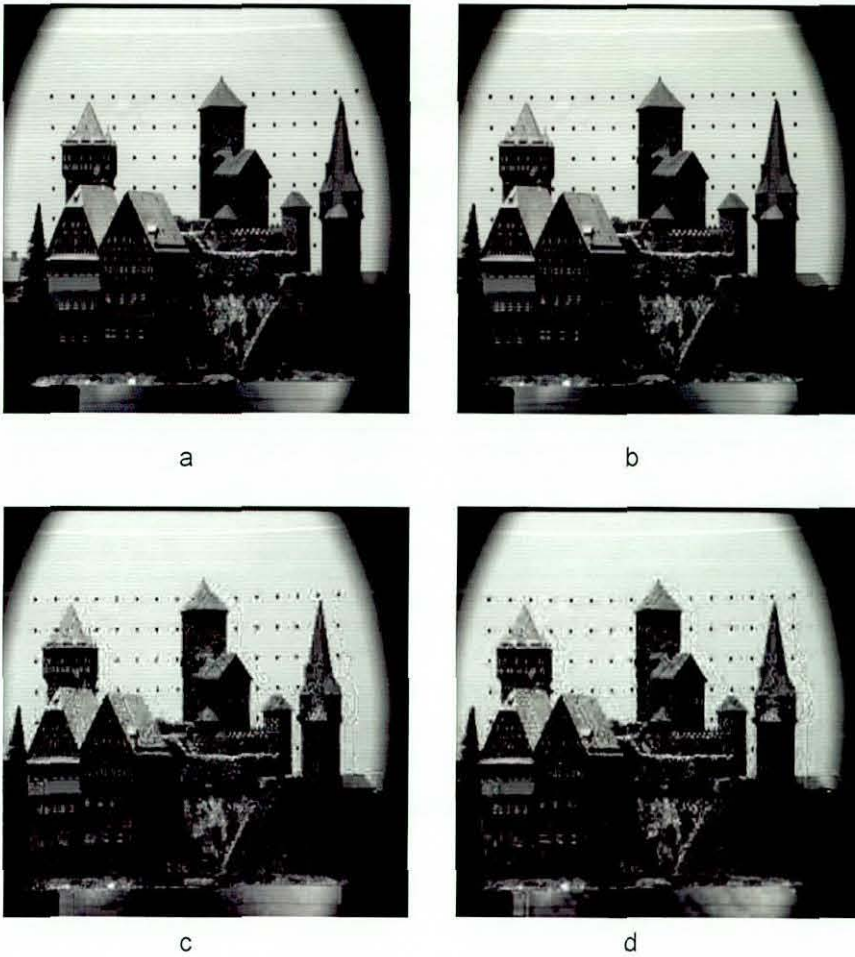
Figure 4.8 Subjective quality evaluations for Castle
(a) Original left image (b) Original right image (c) Benchmark reconstructed right image, 0.183 bpp, 26.31 dB (d) CODEC I reconstructed right image, 0.183 bpp, 26.52 dB

To further analyse the effect of filter length vs. image sample size relationship, we modified the CODEC I so that the PEI is first blocked to 8x8 pixel samples and each block is then transformed to the wavelet domain using a db7 filter and the simplest (shortest) wavelet filter 'Haar', db1. Figure 4.9 compares the two reconstructed images obtained. It shows the adverse effect of the db7 filter length (i.e. greater than the sample length) when applied to an 8x8 block. However the simple 'Haar' transform does not cause the coding artifacts of figure 4.9 (a). However due to the short length of the filter, Haar Transform has a lower signal compaction capability and hence results in inefficient data compression.

a   b
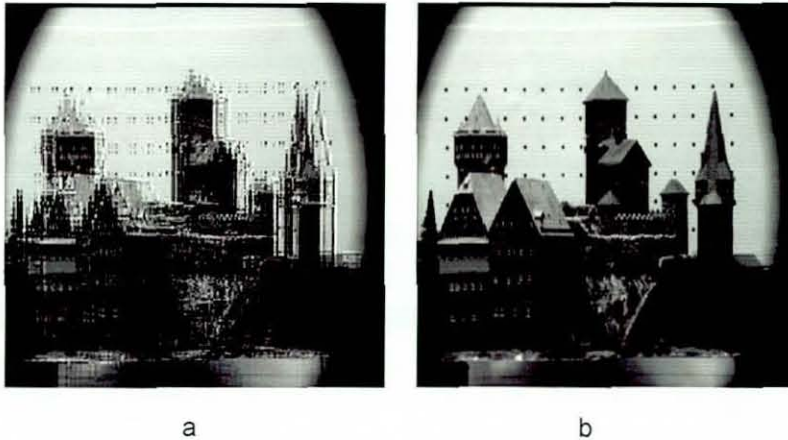
Figure 4.9 Comparing the effect of applying DWT on 8x8 pixel blocks of the PE image
(a) Using db7, 0.9269 bpp, 17.88 dB (b) Using db1 ('Haar'), 0.9018 bpp, 35.38 dB

4.4.3 Conclusion

In addition to the marginally better performance achieved by the proposed CODEC I, it also provides a simple architecture for DWT-based coding of stereo images. As the design is based on the building blocks of a baseline-JPEG CODEC, much use could be made of the software, hardware modules that are already available for baseline-JPEG coding in a software/hardware implementation of the proposed CODEC I. This could be particularly important at a time, image and video coding technology is undergoing a steady change, from DCT-based coding schemes to DWT based coding schemes.

As discussed above, the main drawback of CODEC I is that the PEI is obtained in the pixel domain resulting in a blocky image, which compresses ineffectively due to the presence of artificial edges along block boundaries. This situation would not arise if the left and right images were first decomposed to a DWT multi-resolution representation and then the disparity estimation was done in the DWT domain.

4.5 CODEC II & III – DWT-based stereo image coding with wavelet domain disparity compensation and prediction

To overcome the problems associated with pixel domain disparity estimation and compensation, in DWT-based stereo image coding (discussed in section 4.4), the same could be carried out in DWT domain. This section proposes 2 such algorithms.

DWT domain disparity estimation and compensation techniques have been adopted by previous techniques proposed in literature. In [182], a scheme of hierarchical block matching is used to reduce the complexity of disparity compensation and to obtain smoother disparity vector fields. In [30], by combining both the mixed resolution coding and the disparity compensated technique, first, the left image (reference) independently coded at high resolution. Based on mixed resolution theory [30], disparity compensation is performed at low-resolution images to predict the low resolution right from the low resolution left images. Both low-resolution images are obtained by wavelet decomposition.

To keep the complexity of disparity compensation process low and to reduce the overhead needed for the coding of disparity information, in the above techniques, not all the subbands are used/transmitted and DE/DC is performed only on the low-resolution images. However each subband in the wavelet decomposition contributes to the quality of the reconstructed image [201]. In addition our experiments showed that the assumptions made by the above algorithms [30,182,201] on the relationship between disparity vectors of subbands with the same orientation (i.e. vertical,

horizontal or diagonal) are not always true. In the design of CODEC II and CODEC III, the above observations have been taken into account.

4.5.1 CODEC II – Multiresolution disparity estimation & compensation

Figure 4.10 represents the block diagram of the proposed CODEC II. At the encoder end, the left image and the right image undergo independent, 3-level dyadic DWT. For a given block of each subband in level-3 (i.e., at the third level of decomposition) of the right image, a search is performed in the corresponding subband of the left image, within a maximum likelihood area for the best predictor. The prediction errors blocks thus obtained are used to form the four subbands of the third level decomposition of the predictive error image.

For level-2 and level-1 detailed images (HL, LH, HH), the predicted best matches are found by using the disparity vectors of the corresponding detailed images of level-3 obtained above. For level-2, the best match is found by multiplying the disparity vector of level-3 by two while for level-1 the disparity value is obtained by multiplying the corresponding level-3 disparity value by four. Thus in CODEC II, we only need to transmit the disparity vectors of blocks for all subbands in level-3. The prediction errors of all the ten subbands are then used to form the Prediction Error Image (PEI), which would initially be in the form of a 3-level decomposed image. The PEI is subsequently converted to the wavelet-block domain as explained in section 4.3.2 and is subsequently coded using a procedure similar to that used in CODEC I.

The decoder design follows the usual inverse process design of the encoder and is not therefore explained in detail.

It should be noted that the novelty of the above design is not in the multiresolution disparity estimation and compensation. As discussed previously, such schemes have been proposed in the past for stereoscopic image coding. However the application of wavelet-block based coding to the subsequent PEI is a novel approach.
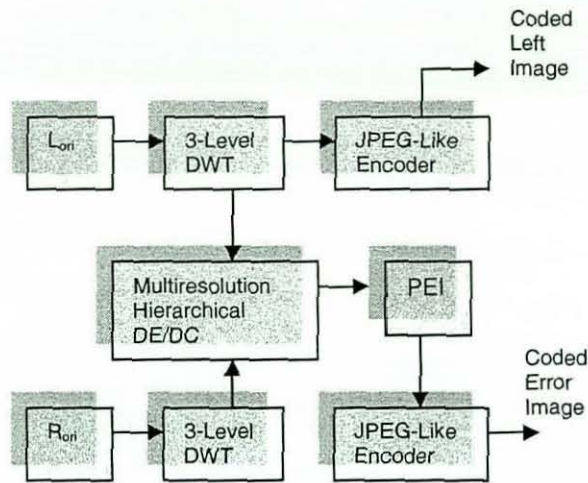
Figure 4.10 The CODEC II- Encoder

4.5.2 CODEC III – Disparity estimation & compensation in the wavelet block-domain

Figure 4.11 represents the block diagram of the proposed CODEC III-encoder. At the encoder end, the left image (reference image) and the right image (predicted image) undergo a 3-level dyadic DWT separately and are subsequently converted into their wavelet-block images (WL and WR respectively). For each block in WR a search is performed in WL for best matching block, within a horizontally oriented windowed area of maximum likelihood on WL. Note that due to the nature of the wavelet-block domain, the above search is performed with 8 pixel jumps, considered as one unit block shift as in figure 4.12. The best predictor is found using the minimum MSE as the selection criteria. The prediction errors of all the non-overlapping right image wavelet-blocks are then used to form the PEI, which would be in wavelet-block domain. Finally, each wavelet-block in WL and PEI undergo scanning, scalar quantization and entropy coding as in CODEC I.

The decoder operations are essentially the inverse of the processes of the encoder and are clearly illustrated in figure 4.13.
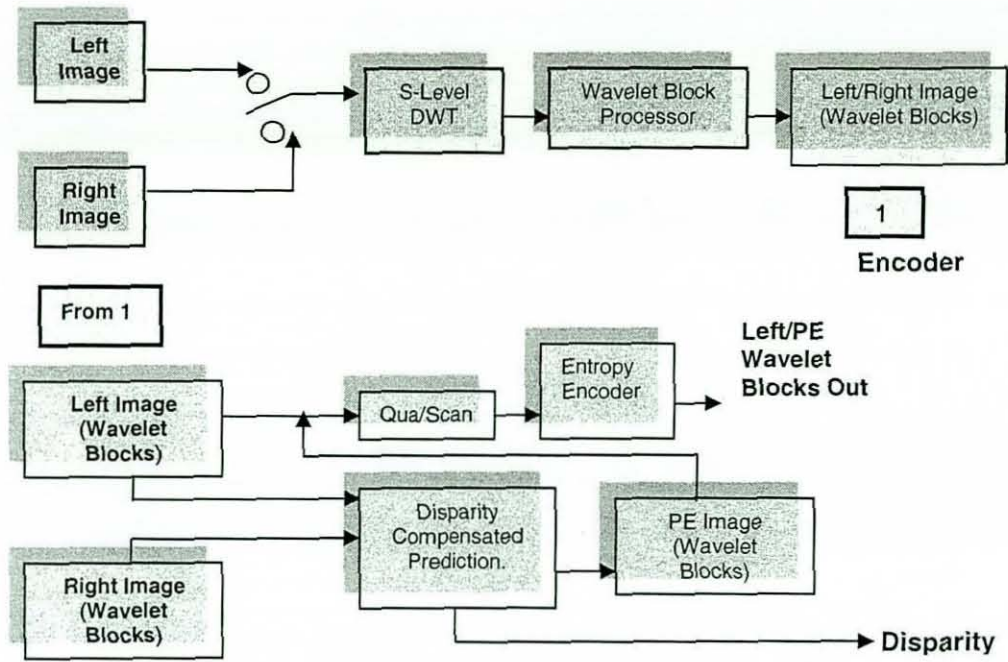
Figure 4.11 The encoder of CODEC III
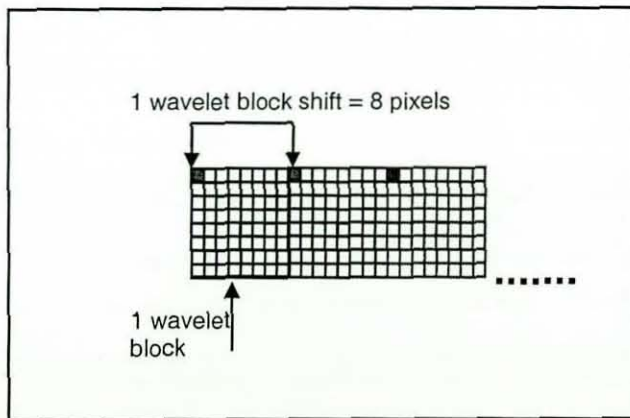
Figure 4.12 Illustrates one wavelet-block shift is equivalent
to eight pixels jump
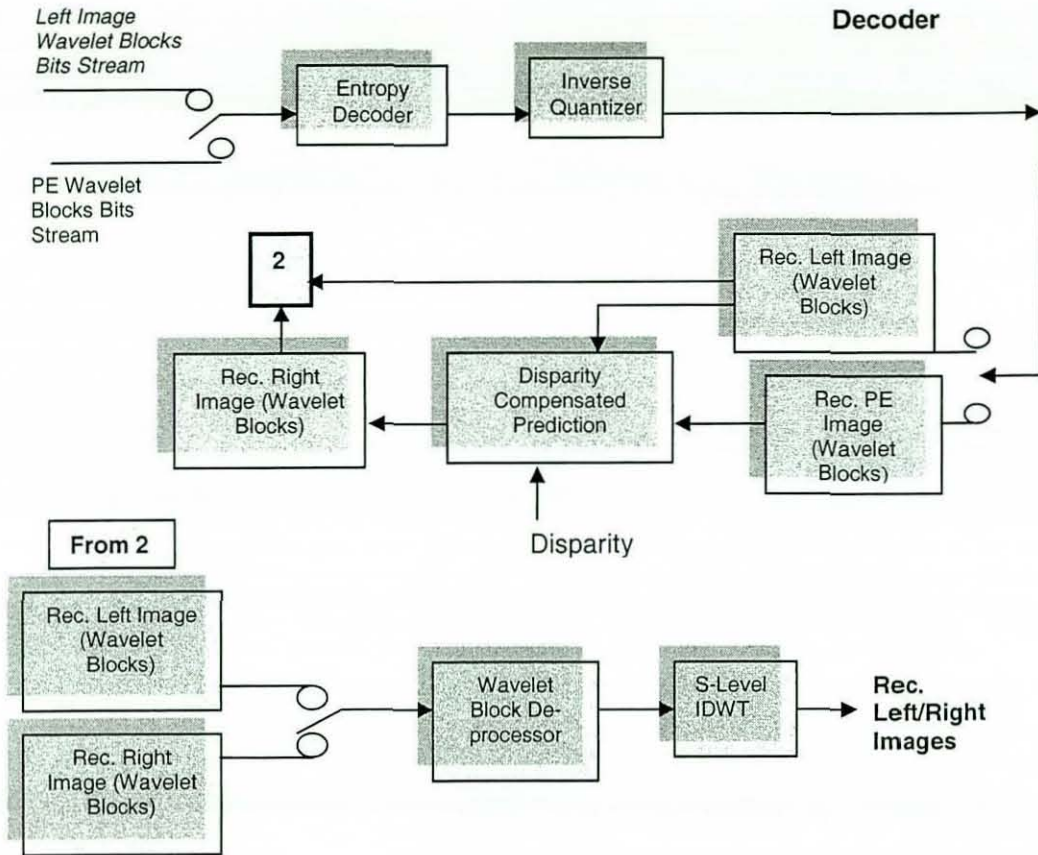
Figure 4.13 The decoder of CODEC III

4.5.3 Experiment results and analysis

As in CODEC I, the proposed CODEC II and CODEC III were implemented using purpose built MATLAB routines and were tested on commonly used test stereo image pairs representing indoor, outdoor, natural and synthetic scenes: Bottle (320x320), Room (256x256), Castle (512x512), SScastle (512x512) and Tunnel (512x512). All these images were acquired using parallel axis camera geometry. For CODEC II, windowed search area representing a disparity range 0-7 was used in all four, level-3 subbands. For CODEC III, since the search in being performed in the wavelet-block domain, the disparity value of k is equivalent to k times 8 pixels jump, where $k = 0,1,2...7$.

A Daubechies compactly support biorthogonal wavelet filter bior9/7 is used as the wavelet transform for the left and the right images in all decomposition levels. In order to evaluate the performance of the proposed coder, in figure 4.14 (a), 4.15 (a) and 4.16 (a), we compare the rate-distortion performance of CODEC II and CODEC III against the rate-distortion performances of the benchmark and CODEC I.

The objective image quality is measured in terms of Peak Signal to Noise Ratio (PSNR), whereas the compression efficiency is measured in bits per pixels (bpp). For a fair comparison, the objective quality of the left image (reference) was set at equivalent levels for CODEC II and CODEC III, the benchmark, and CODEC I throughout the experiments (as previously stated in CODEC I).

Results of the experiments, show mixed performance of CODEC II, over the benchmark and CODEC I. For SScastle, CODEC II performance is poor compared to both benchmark and CODEC I. Further experimental analysis showed that this was due to the presence of a high proportion of noise in this image pair, which resulted in sub-optimal matches at higher resolution subbands, as in CODEC II, the disparity estimation is only done on the third level subbands, i.e. at a low resolution level. In other words this proves that in the presence of high frequency noise, the assumption that the motion vectors of blocks of a particular subband is twice that of the subband from the lower resolution level with same orientation, is false. Unfortunately the CODEC II is designed based on this said assumption. Hence it performs inferior for the noisy, SScastle image. For Castle image, the rate-distortion performance graphs show that the CODEC II performs better than the benchmark and CODEC I in the range of 0.2 bpp and 0.4 bpp. For the Bottle image, CODEC II performance is worse compared to the benchmark and CODEC I, beyond 0.3 bpp. Thus in general, CODEC II is unable to perform better than the benchmark and CODEC I beyond bit rates of 0.3 bpp. The slight performance improvement that is possible at lower bit rates is due to the significance of disparity vector bit budget at these low bit rates as compared to the bit budget of the PEI. Note that CODEC II requires significantly lesser amount of bits for coding the disparity vectors as compared to the benchmark and CODEC I and would thus perform better at very low bit rates.

Meanwhile, results of the experiments clearly show the superiority of performance of CODEC III over the benchmark, CODEC I and CODEC II. An improved rate-distortion performance is indicated for all bit rates of the reconstructed right images. For example, from the rate-distortion performance graphs in figure 4.14 (a) for SScastle, the PSNR improvement obtainable is up to 3 dB. In Figure 4.15 (a) for Castle, the PSNR improvement obtainable is up to 5 dB. Similar results have been obtained for the Bottle image (see figure 4.16 (a)).

The performance of CODEC III can also be observed subjectively in figure 4.14 (b), (c), figure 4.15 (b), (c), and figure 4.16 (b), (c) for SScastle, Castle, and Bottle reconstructed images respectively. At the same bit rate, the reconstructed right images of CODEC III have higher PSNR than CODEC I. Highlighted areas in these figures show some of the better subjective quality areas of the images of CODEC III when compared to similar highlighted areas of reconstructed image by CODEC I.
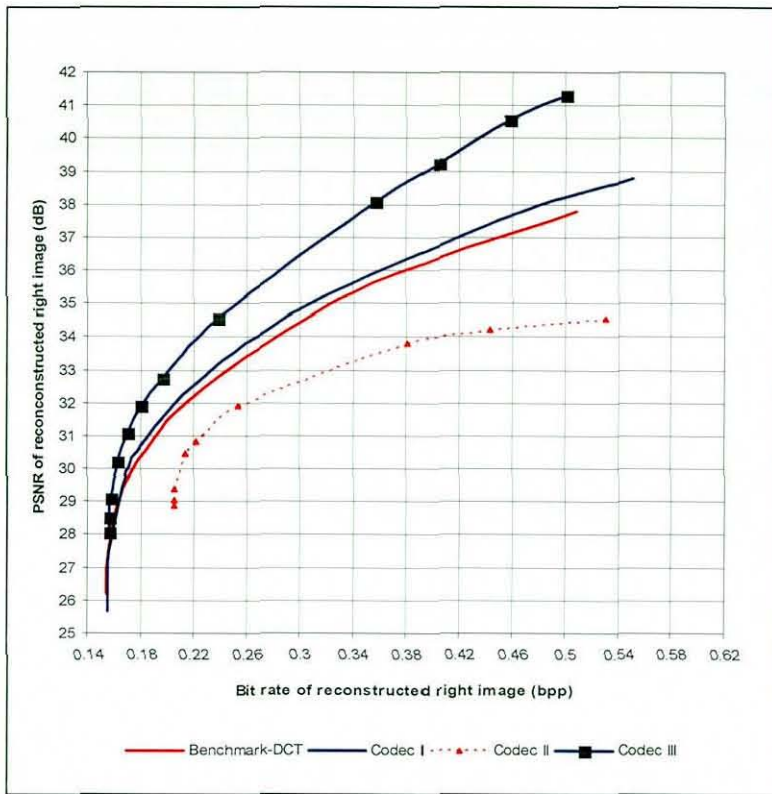
a

b   c

Figure 4.14 Reconstructed right 'Sscastle' image using bior9/7
(a) Comparison of rate-distortion graphs between the benchmark, CODEC I,
CODEC II and CODEC III (b) Using CODEC I, 0.1836 bpp, 30.87 dB (c) Using
CODEC III, 0.1836 bpp, 31.44 dB
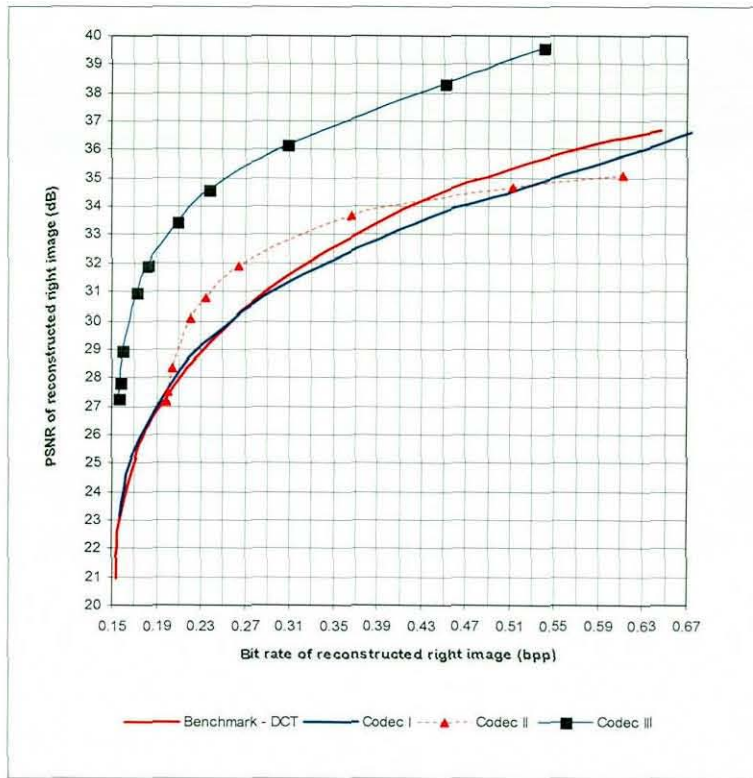
a
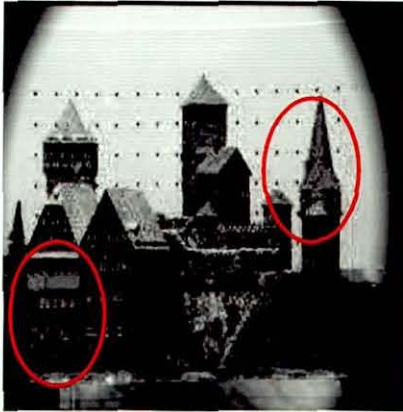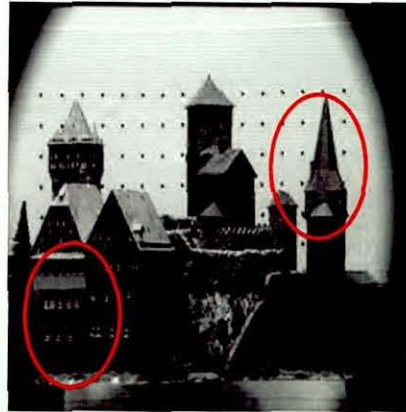
b   c

Figure 4.15 Reconstructed right 'Castle' image using bior9/7
(a) Comparison of rate-distortion graphs between benchmark, CODEC I,
CODEC II and CODEC III (b) Using CODEC I, 0.183 bpp, 26.52 dB,
(c) Using CODEC III, 0.183 bpp, 31.90 dB
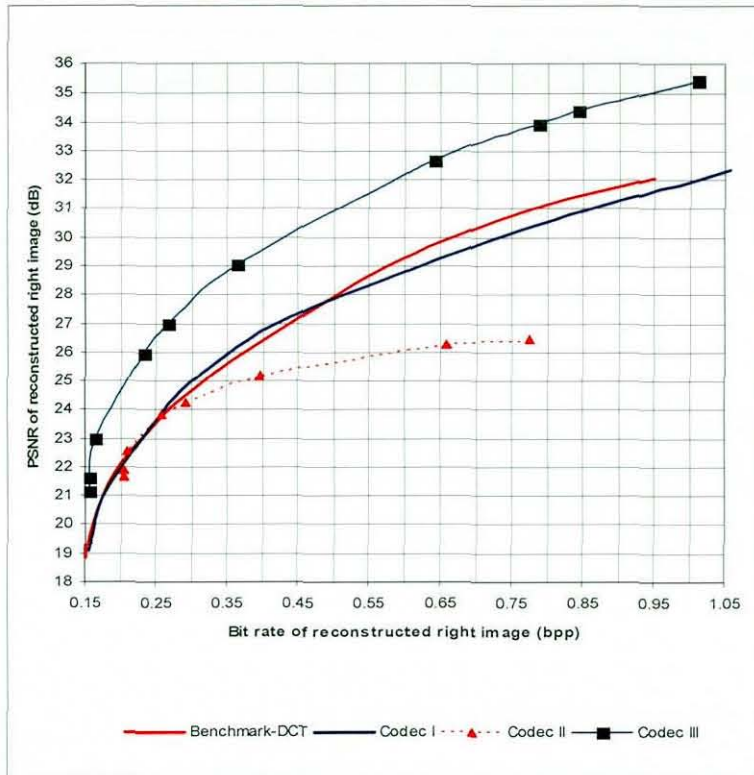
a
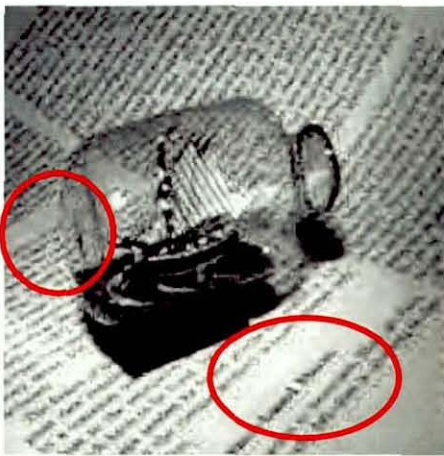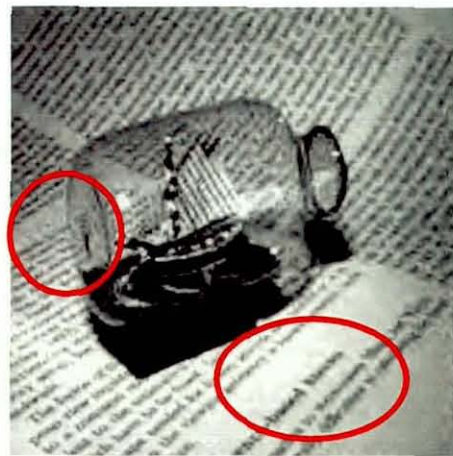
b   c

Figure 4.16 Reconstructed right 'Bottle' image using bior9/7
(a) Comparison of rate-distortion performance graphs between benchmark,
CODEC I, CODEC II and CODEC III (b) Using CODEC I, 0.2785 bpp, 24.45 dB,
(c) Using CODEC III, 0.2785 bpp, 27.25 dB

4.5.4 Conclusion

The large improvement of CODEC III over CODEC I, is mainly due to performing disparity estimation and compensation in the wavelet domain, instead of in pixel domain as in CODEC I. Due to the wavelet-block based disparity estimation and compensation technique adopted by CODEC III, coefficients of all ten subbands from the same spatial location are involved in the matching process, i.e. in the calculation of the prediction error. Note that these coefficients are also tied together. In CODEC II, the prediction error calculation is only based on the coefficients of the level-3 subband. The corresponding higher resolution coefficients are not taken into account in the matching process. Whilst matching with the help of only the low resolution detail of an image is bound to reduce spurious matching due to noise, it has an adverse effect on the compression performance as high resolution details would be badly matched. Therefore CODEC III performs better than CODEC II. Whilst the best approach could have been to do independent block based matching between each pair of corresponding subbands, this results in the requirement of a high bit budget for the coding of a large number of disparity values. Therefore the approach in CODEC III is an effective compromise. Thus in spite of the apparent eight-pixel jump in search of a best match, CODEC III outperforms all other CODECs.

5 PBDCPC and OBDC in wavelet-based coding of stereo image pairs

5.1 Overview

This chapter proposes four novel stereo image coding algorithms, which make use of the well known Pioneering Block-based Disparity Compensated Predictive Coding (PBDCPC) and Overlapped Block Disparity Compensation (OBDC) techniques, in the wavelet-based coding of stereo image pairs.

The PBDCPC was originally proposed in 1997 [51, 87] and was used for pixel domain disparity estimation and compensation in a stereo image CODEC that uses DCT as the base compression technology. PBDCPC is known for its ability to work under very low bit rate constraints, yet producing predicted images of acceptable image quality. The high compression gains that are possible, is due to the fact that it avoids the necessity of having to transmit the disparity vector field as overhead bits. Within the research context presented in this chapter, we show that the change of base technology from DCT to DWT further improves the low bit rate performance of a PBDCPC stereo image CODEC. However, due to the specific requirements of the PBDCPC coding architecture and the properties of DWT-based multiresolution decomposition [173,187], the above transfer of base technology is not straightforward and thus involves additional design and development of considerable novelty. Due to this reason the performance enhancement of a typical PBDCPC based CODEC obtainable by the base technology change from DCT to DWT, exceeds the straightforward performance enhancement one could expect due to a straightforward change of base technology alone. In our experiment, the original PBDCPC algorithm is used as a benchmark to show superiority of DWT-based over the DCT-based implementation.

The OBDC is better known as Overlapped Block Motion Compensation (OBMC) in video coding. It posses the ability to reduce blocking artifacts by linearly combining

the predictions generated using multiple motion vectors, which include the block motion vector as well as its neighbours [220]. The only paper found in literature that uses the above OBMC idea in stereo image coding is DCT-based OBDC stereo image compression by Woontack Woo in 1999 [220]. Within the research context of the proposed work we explore the possibility of applying OBDC in multiresolution DWT domain.

In the proposed CODEC VI, FSBM is performed only in the subbands of the highest level of decomposition. Based on a predefined threshold, the PEI blocks of these subbands are further refined by applying OBDC. Those PEI blocks to which OBDC have been applied are flagged by '1' (otherwise '0') in an OBDC indicator table. Subsequently considering inter-subband relationships, the disparity vectors and OBDC tables of the detail subbands of the highest level of decomposition are used to find and refine the corresponding PEI blocks in lower level subbands. The ultimate PEI, disparity vectors and the OBDC indicator tables are coded and transmitted for enable reconstruction at the decoder end.

In CODEC VII, we combine the ideas of PBDCPC and OBDC to develop a wavelet-based stereo image CODEC.

5.2   The benchmark - The pioneering block-based disparity compensated predictive coding

The DCT-based PBDCPC of [51, 87] is chosen as the main benchmark to evaluate the performance of CODECs proposed in this chapter, due to its excellent very low bit rate performance characteristics [213]. This CODEC uses the standard baseline-JPEG architecture, similar to that described in chapter 4. The only difference is that it uses *pioneering block search* (*PBS*) in the disparity estimation/compensation process. Therefore, in section 5.2.1 we only described the PBS technique and in section 5.2.2, two main requirements of the PBDCPC algorithm is discussed. In

addition to the above, we use CODEC III as a second benchmark since it provided the best performance out of CODECs proposed in chapter 4.

5.2.1 Pioneering block-based search (PBS)

Figure 5.1, illustrates the PBS procedure that was used in the original PBDCPC algorithm [51, 87]. As illustrated in figure 5.1, to encode a given block of pixels of image pixels (say size 8×8), R_{blk}, in the predicted (right) image R, the blocks PB_1 and PB_2, directly above and preceding R_{blk} respectively, are taken as *pioneering blocks* to search for a matching pair of blocks, within a selected windowed area of maximum likely of the reconstructed reference (left) image. The best matching pair is found using minimum Mean Squared Error (MSE) as the minimization criterion and using equal weights for the two blocks. Once the best matching pair PB_{1_BM} and PB_{2_BM} are found, the block $\hat{L}_{blk\_BM}$ is chosen as the best predictor, for R_{blk}. The above prediction strategy exploits the intra-frame redundancy of the two constituent images of the stereo image pair and their binocular redundancy, together. Next, the prediction error E_{blk} is calculated by $E_{blk} = R_{blk} - \hat{L}_{blk\_BM}$. Apart from being encoded and transmitted to the decoder end, E_{blk} has to be decoded and added to the predictor $\hat{L}_{blk\_BM}$, to produce the new preceding pioneering block $PB_2=$ ($\hat{E}_{blk} + \hat{L}_{blk\_BM}$), to be used for encoding the next predicted block. And the PBS process described above is repeated until all the blocks in the right image have been predicted.

At the decoder end, when decoding a given right image block, a search identical to that of the encoder end is performed based on its already decoded adjacent blocks as pioneering blocks and the reconstructed left frame as the reference. Thus as compared to traditional predictive coding algorithms, the PBDCPC algorithm enables reconstruction of the right image at the decoder end without recourse to the block disparity values, enabling a comparatively better compression performance. Note that at the boundaries of the image where one of the pioneering blocks (or both) is

unavailable, the above search procedure is replaced by a PBS with one block (or by direct prediction).
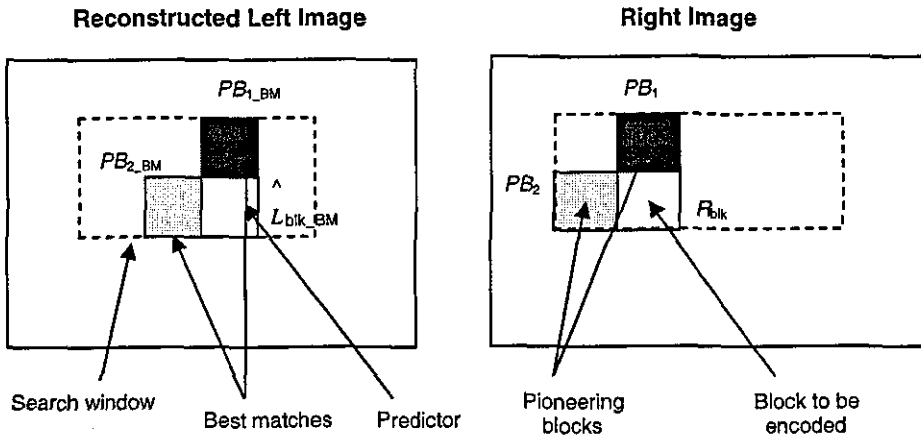
Figure 5.1 Pioneering block-based search in pixel domain

5.2.2 Two basic requirements of PBDCPC algorithm

Two specific criteria should be met in PBDCPC, to guarantee the reconstruction of the predicted image. Once the predictor, $\hat{L}_{blk\_BM}$, for the block to be encoded, R_{blk} is found, the predictor error E_{blk} is calculated as $E_{blk} = R_{blk} - \hat{L}_{blk\_BM}$. Note the use of $\hat{L}$ to represent the fact that the search is performed on a locally reconstructed left frame. This is the first specific requirement of the PBDCPC coder, as it guarantees that the procedure adopted at the encoder end is identical to that adopted at the decoder end. This is due to the fact that at the decoder end what is available is the reconstructed reference frame. In addition to meeting the above condition, a second specific condition needs to be met. When the prediction error E_{blk} is calculated at the encoder end, apart from it being encoded and transmitted to the decoder end, it has to be decoded and added to the predictor $\hat{L}_{blk\_BM}$ to produce the preceding pioneering block for the encoding of the next predicted block. Thus, always the pioneering block selection in the *predicted image is done on a 'so-far'*

reconstructed image. This satisfies the second condition in guaranteeing that the encoder and decoder prediction environments are identical.

Thus in the PBDCPC coder, the encoder and decoder work on identical reference frames and selects pioneering blocks from identical, 'so-far' reconstructed predicted frames. This enables the guaranteed reconstruction of the predicted image at the decoder end, without recourse to disparity information, which is the key feature of the PBDCPC scheme. We refer readers interested in the theoretical proof of the above mentioned quality guaranteeing procedure to [51].

5.3   Proposed wavelet-based implementation of the PBDCPC algorithm

In CODEC IV and CODEC V, the traditional disparity compensated predictive coding [81,161] is replaced by pioneering block-based predictive coding using DWT-based technology. Extending the pixel domain PBS procedure, summarized in section 5.2.1 to wavelet domain, to fully satisfy the second condition in guaranteeing the reconstruction of the predicted image, the prediction error blocks have to ultimately go through locally simulated lossy coding (transform coding/decoding and quantization/de-quantization) at the encoder end before being fed back to form the preceding pioneering block of the next block to be encoded.

Thus, a straightforward block-based, pixel domain PBDCPC to obtain the prediction error image in full, and the subsequent DWT-based coding of the same, is not a suitable DWT implementation of the PBDCPC algorithm. This is due to the fact that with such a scheme, one is unable to introduce the compression losses that would be introduced to the prediction error blocks, before they are fed back. As a solution to the above problem, within our present research context we propose a strategy based on subband domain PBDCPC in which, predictive error blocks are appropriately quantized and de-quantized before being fed back, to act as the preceding pioneering

block of the next iteration. This guarantees the reconstruction of the right image under all bit rates.

In addition to specific design criteria discussed in section 5.2.2, to allow a worthwhile comparison between the original DCT-based and proposed DWT-based implementations of PBDCPC, we have adapted a baseline-JPEG coding strategy based on wavelet-block coding [173] as in previous CODECs (see chapter 4), to code the reference image and the predictive error image (PEI). With this adaptation we aim to demonstrate that the coding gains obtained under the present research context is solely due to the application of PBDCPC in wavelet domain and is not due to the use of more efficient coding techniques to code wavelet transform coefficients as against the techniques used to code the DCT coefficients.

5.3.1 CODEC IV- Pioneering wavelet-block based disparity compensated predictive coding

The CODEC IV is similar to CODEC III (see chapter 4) in which disparity estimation and compensation is performed in wavelet-block domain. However in CODEC IV, disparity estimation is based on pioneering wavelet-block search technique, as against traditional disparity estimation in CODEC III.

Figure 5.3 illustrates a high-level block diagram of the proposed CODEC IV. The left and right images are first transformed into their respective wavelet-block domain as in CODEC III. Firstly, the original left image (reference) L_{ori} is directly transmitted similar to the manner they were coded and transmitted in CODECs proposed in chapter 4. A sample of the left image is simultaneously reconstructed at the encoder end, to be used in the pioneering wavelet-block based search. Subsequently, for each wavelet-block to be encoded WR_{blk} in the right image wavelet-blocks, pioneering wavelet-blocks WR_{PB_1} and WR_{PB_2}, above and preceding WR_{blk} respectively are used to search for the best matching wavelet-block in the reconstructed and re-transformed (to wavelet-block domain) left image, $\hat{WL}$. Due to the orientation of the two pioneering wavelet-blocks relative to the wavelet-block to be encoded, in encoding

wavelet-blocks in the first row of wavelet-blocks (other than the first block itself), only the preceding wavelet-block is used as the pioneering wavelet-block. A similar strategy is adapted to resolve the problem of encoding wavelet-blocks in the first column of wavelet-blocks where only the wavelet-block above the wavelet-block is used as the pioneering block. The wavelet-block at the top, left-hand corner of a subband is always coded directly, i.e. without any disparity compensation.

As in CODEC III, when using wavelet-blocks (each of size 8x8), one wavelet-block shift is equivalent to eight pixels shift in the pixel domain. Once the best match wavelet-block pair is found $\hat{WL}_{PB\_1\_BM}$ and $\hat{WL}_{PB\_2\_BM}$, the wavelet-block to the right of the best match is selected as the best match for the wavelet-block to be encoded and the corresponding prediction error wavelet-block is formed $E = WR_{blk} - \hat{WL}_{blk\_BM}$, which is subsequently added to the prediction error image.
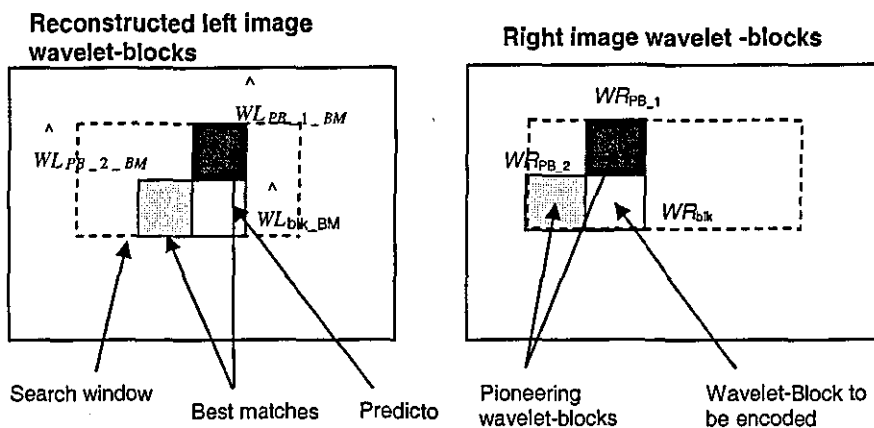
Figure 5.2 Pioneering block search in wavelet-block domain

At the decoder end, each right image wavelet-block is reconstructed using an identical procedure to that of at the encoder. As at the decoder the search is performed for the best match, in a reconstructed and re-transformed left image, at the encoder end it is essential that the search is performed in an identical reference image, thus the reason for using a reconstructed left frame for pioneering wavelet-block based search above. In addition to this, the preceding and wavelet-block above the wavelet-block to

decoded, which are used as the pioneering wavelet-blocks for the search of the best matching wavelet-block, are identical to its counterpart at the encoder end, due to the feedback loop shown in figure 5.3 of the encoder. These satisfy the two requirements mentioned in section 5.2.2 to guarantee the reconstruction of the predicted image without the need for the transmission of the disparity vectors.
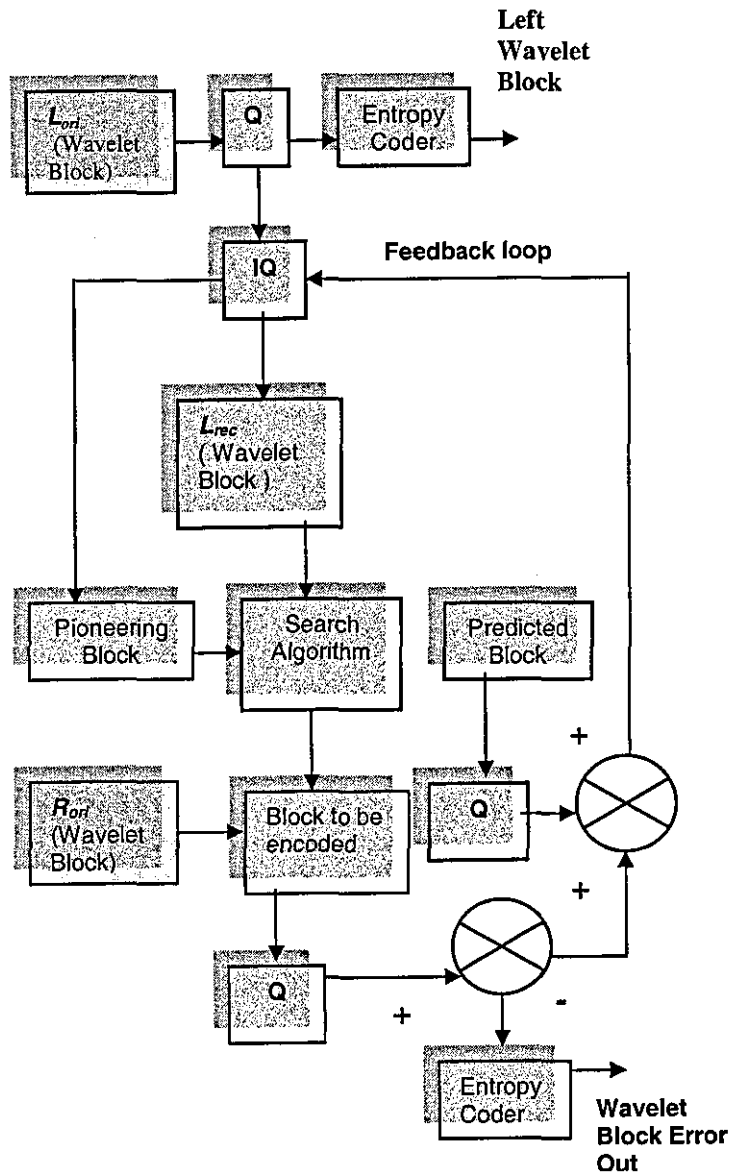
Figure 5.3 The proposed CODEC IV- The Encoder

5.3.2 CODEC V- Hierarchical pioneering block-based predictive coding in wavelet domain

In CODEC II we analysed that performing hierarchical disparity estimation in all subbands of the wavelet decomposition is highly inefficient. This is due to the need of a high proportion of overhead bits for disparity vector field coding. In CODEC V pioneering block-based predictive coding, which does not require disparity vector field coding, enables disparity estimation to be performed independently in each subband of the multiresolution, DWT domain.

Figure 5.4 illustrates the detailed block diagram of the encoder of CODEC V. The original left image (reference), L_{ori}, is independently coded using the baseline-JPEG like wavelet CODEC as in figure 5.5. After reconstruction of the left image at the decoder end, it provides a reference for the prediction of the right image. In addition to the above, the encoded reference image is locally decoded at the encoder end to produce a reconstructed left image L_{rec}, which is subsequently used as the reference in the wavelet-based PBS unit. Note that this satisfies the first requirement for a successful PBDCPC design. The original right image R_{ori} and reconstructed left image L_{rec} then undergo 3-level dyadic wavelet decomposition.

Within the wavelet-based PBS unit, for each corresponding subband pair (e.g. L_{LL3} and R_{LL3}, L_{HL1} and R_{HL1} etc.) a PBS is performed in the subband domain as described in section 5.2.1, to produce the corresponding subband of the Prediction Error Image (PEI). Note that we use a block size of 1×1, 2×2, and 4×4 for PBS at level-3, level-2 and level-1 subband, respectively. Once the prediction error image has been completely found (i.e. all of its subbands found), the baseline-JPEG like wavelet CODEC (as in previous CODECs) is used to transform it into wavelet-blocks and transmit after suitable quantization and entropy coding. Figure 5.4, also illustrates a prediction error (PE) block feedback loop, which converts the original domain prediction errors into the reconstructed domain by locally sending the PE through a quantization/de-quantization procedure. The error block, in its reconstructed state E_{rec}

is then added to the best predictor previously found from L_{rec} to form the block, which would act as the preceding (front) pioneering block for the next iteration of the PBS. Thus the purpose of the above mentioned error feedback loop is to satisfy the second requirement for a successful PBDCPC CODEC design as mention in section 5.2.2.
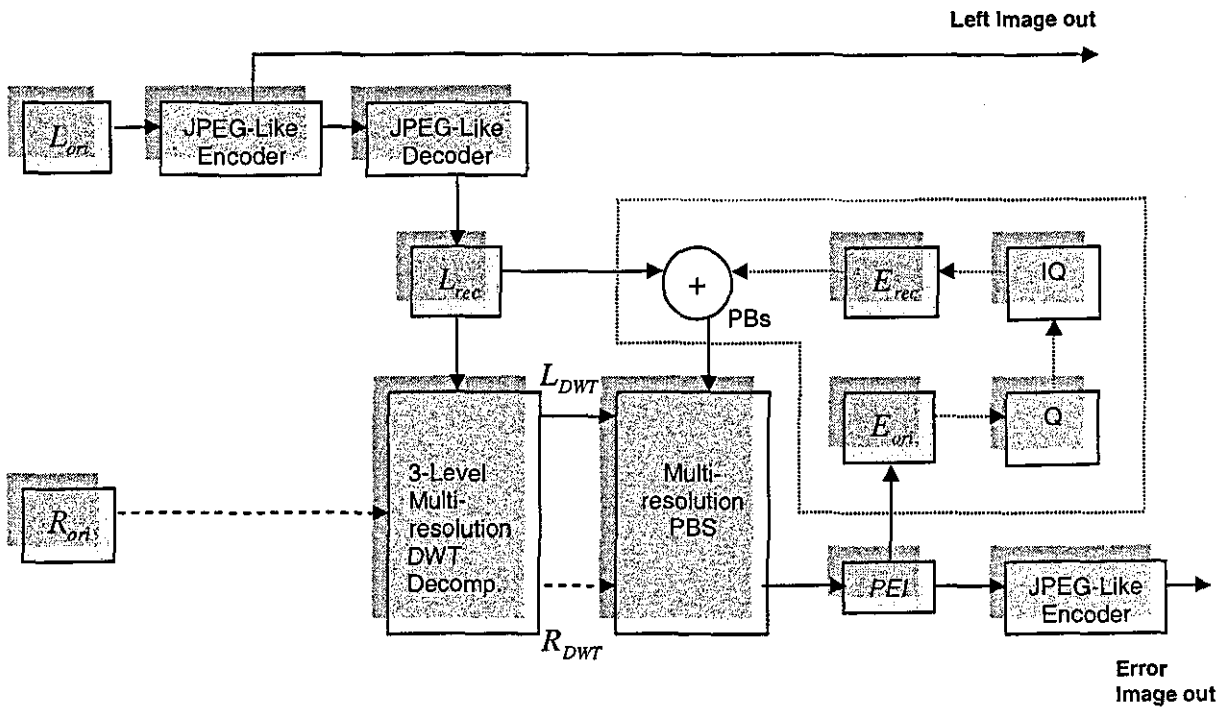
Figure 5.4 The proposed CODEC V- The Encoder
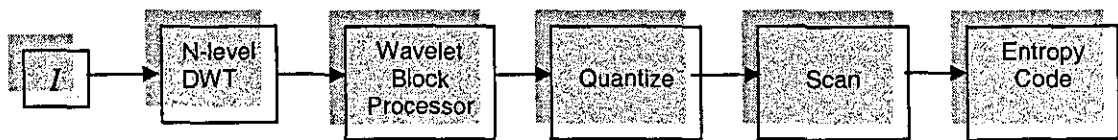
Figure 5.5 The JPEG-like wavelet-block encoder

The decoder of the proposed CODEC V works as the mirror image of the encoder. Due to the orientation of the two pioneering blocks relative to the block to be encoded, in encoding blocks in the first row of blocks (other than the first block itself), of a given subband, only the preceding block is used as the pioneering block.

A similar strategy in adapted to resolve the problem of encoding blocks in the first column of blocks where only the block above the block is used as the pioneering block. The block at the top, left-hand corner of a subband is always coded directly, i.e. without any disparity compensation.

5.3.3 Experimental results & analysis

As in previous CODECs, CODEC IV and V are implemented using purpose built MATLAB routines and tested on a popular set of test stereo image pairs. Parallel axis camera geometry has been used to acquire these images and we assume that pixels at the same spatial location to have the same brightness. Compactly supported and symmetric biorthogonal filters (bior 9/7) are used as in JPEG-2000.

In order to evaluate the performance of both CODECs, we compare their rate-distortion performance, against DCT-based pioneering block predictive coding algorithm of [51] and CODEC III (the best CODEC in chapter 4).The objective image quality is measured in terms of Peak Signal to Noise Ratio (PSNR), whereas the compression efficiency is measured in bits per pixels (bpp) as follows:

$$BR = \left[\frac{Tot\_Bits_{comp}}{N \times M} \right] \quad bpp$$

where, N and M represent the image dimensions and Tot_Bits_{comp} is the total number of bits required to encode the prediction errors only. For a fair comparison, the objective quality of the left image was set at equivalent levels for all CODECs for a given image pair.

In figure 5.6(a), figure 5.6(b) and figure 5.8(a), for stereo image pairs "SScastle","Bottle"and "Arch" respectively, the RD performance of CODEC IV and CODEC V, are compared against two benchmarks: the PBCDPC (no disparity vector field transmitted) and the previously proposed CODEC III (disparity vector field transmitted). The RD graphs indicate an improved perfomance by CODEC IV and

CODEC V, over the benchmarks, with both CODEC IV and CODEC V showing superiority over the DCT-based PBDCPC. This is mainly due to blocking artifacts present in the DCT-based technology, which particularly affect the low bit rate performance of the benchmark DCT-based PBDCPC. Note that blocking artifacts do not occur in DWT-based coding and thus the reference frame where PBS is performed would be smoother, resulting in it providing better predictions. Figures 5.6(a), 5.6(b) and 5.8(a) illustrate that at very low bit rates (0.12 bpp for SScastle, 0.1 bpp for Bottle and 0.11 for Arch), CODEC IV, CODEC V and DCT-based PBDCPC performance are almost the same. This is due to the fact that at very low bits rates only a small amount of bits are available for coefficient coding. Therefore only the lowest/coarsest resolution wavelet coefficients in CODEC IV & V and the lowest frequency coefficients (mostly the DC coefficients) of DCT-based PBDCPC would be coded, making the performance of CODECs comparable.

Figure 5.6(a), 5.6(b) and 5.8(a) also show that when using CODEC IV & V, for the SScastle image a PSNR improvement of up to 5 dB, for the Bottle image an improvement of up to 2 dB and for the Arch image an improvement of up to 6 dB respectively, can be obtained, over CODEC III. In addition figure 5.6(a), 5.6(b) and 5.8(a) , illustrate the ability of the pioneering block based techniques to operate at very low bit rates (0.15 bpp and lower) as compared to CODEC III. This is due to the fact that these schemes do not need the transmission of disparity vector fields. At various bit rates (05. bpp, 0.25 bpp and 0.15 bpp), Table 5.1 shows that the total bits (Tot bits) used to transmit the PEI and disparity vector field or PEI for CODEC III and the PEI in the benchmark (PBCDPC), CODEC IV and CODEC V. In CODEC IV and CODEC V, as all bits are available to be used for the PEI, higher PSNR of the reconstructed right image as compared to the CODEC III (where part of the total bits is used in transmitting the disparity vector field) could be expected.

Table 5.1 Comparison of total bits and PSNR of CODEC III and CODEC IV with the benchmark and CODEC II at various bit rates of reconstructed right 'Bottle' image

BRR (bpp)	Benchmark DCT-pioneer		CODEC III		CODEC IV		CODEC V	
	Tot bits	PSNR (dB)	Tot bits	PSNR (dB)	Tot bits	PSNR (dB)	Tot bits	PSNR (dB)
0.5	51934	28.42	51951	31.23	51901	31.45	51804	31.69
0.25	25804	23.92	25724	26.51	25699	27.23	25701	27.97
0.15	15660	21.18	16001	20.99	15366	23.45	15372	23.89

When comparing the performance of CODEC IV with that of CODEC V, a general improvement is illustrated in CODEC V. As illustrated in figure 5.6 and 5.8, for SScastle image, CODEC IV performs better than CODEC V in the range of 0.15 bpp to 0.25 bpp. For the Bottle image, CODEC V is better than CODEC IV at all bit rates. For the Arch image, CODEC IV performes better than CODEC V in the range 0.12 bpp to 0.18 bpp. However the PSNR difference is marginal, only up to 0.5 dB. This is due to the fact that in CODEC IV, the search for the best match is performed simultaneously within all subbands using wavevlet-blocks whereas in CODEC V, the search is done independently in each subband.

In term of CPU times, CODEC III takes about 3 seconds less than that of CODEC IV and CODEC V. However there is no significant difference in CPU times between CODEC IV and V (see Appendix I).

For further evaluation, figure 5.7, illustrates a comparison of the subjective image quality of the reconstructed predicted right 'Bottle' images obtained using the two benchmarks (PBDCPC and CODEC III), CODEC IV and CODEC V. At bit rate of 0.25 bpp, a PSNR difference of 3-4 dB between the benchmark PBDCPC DCT-based, CODEC IV and CODEC V is observed. For example, the marked areas in figure 5.7, highlight the difference in image quality between the CODECs and the benchmarks. Similar subjective image quality evaluation can be done for the reconstructed right Arch image in figure 5.8.

5.3.4 Conclusion

Two efficient wavelet-based implementations of PBDCPC stereo image compresion algorithms have been proposed. The proposed CODEC IV, performs pioneering block-based search simultaneously in all subbands using the wavelet-block concept and CODEC V further improved the performance of CODEC IV by performing the pioneering block-based search independently in each subband of the wavelet multiresolution domain. Both these CODECs successfully improved the performance over the benchmark CODECs (DCT-based) at all bit rates. This is mainly due to blocking artifacts present in the DCT-based technology which particularly affect the low bit rate performance of the benchmark DCT-based PBDCPC. Note that blocking artifacts do not occur in DWT-based coding and thus the reference frame where PBS is performed would be smoother, resulting in it providing better prediction. In addition to this, the proposed CODECs are able to operate at bit rates as low as 0.15 bpp, where other proposed CODECs (chapter 4) fail to operate due to the need a certain minimum bit budget for encoding disparity vector fields.

With both proposed CODECs adapting the same coding strategy as the benchmark DCT-based PBDCPC, it has been demonstrated above that the coding gains are obtainable. We have shown that these are not solely due to the transfer of technology from DCT to DWT, but also due to the application of PBDCPC in wavelet domain.
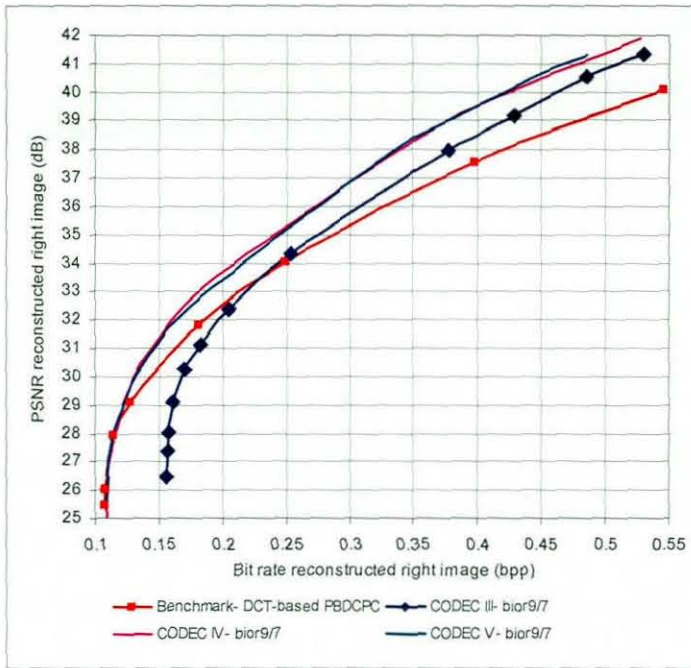
a
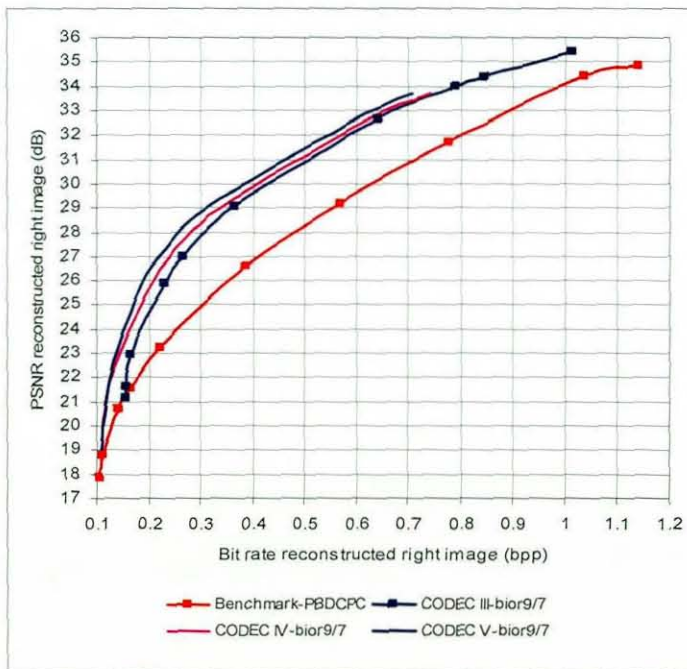
b

Figure 5.6 Rate-distortion performance graphs for
(a) SScastle, CODEC IV and CODEC V vs benchmark and CODEC III
(b) Bottle, CODEC IV and CODEC V vs benchmark and CODEC III
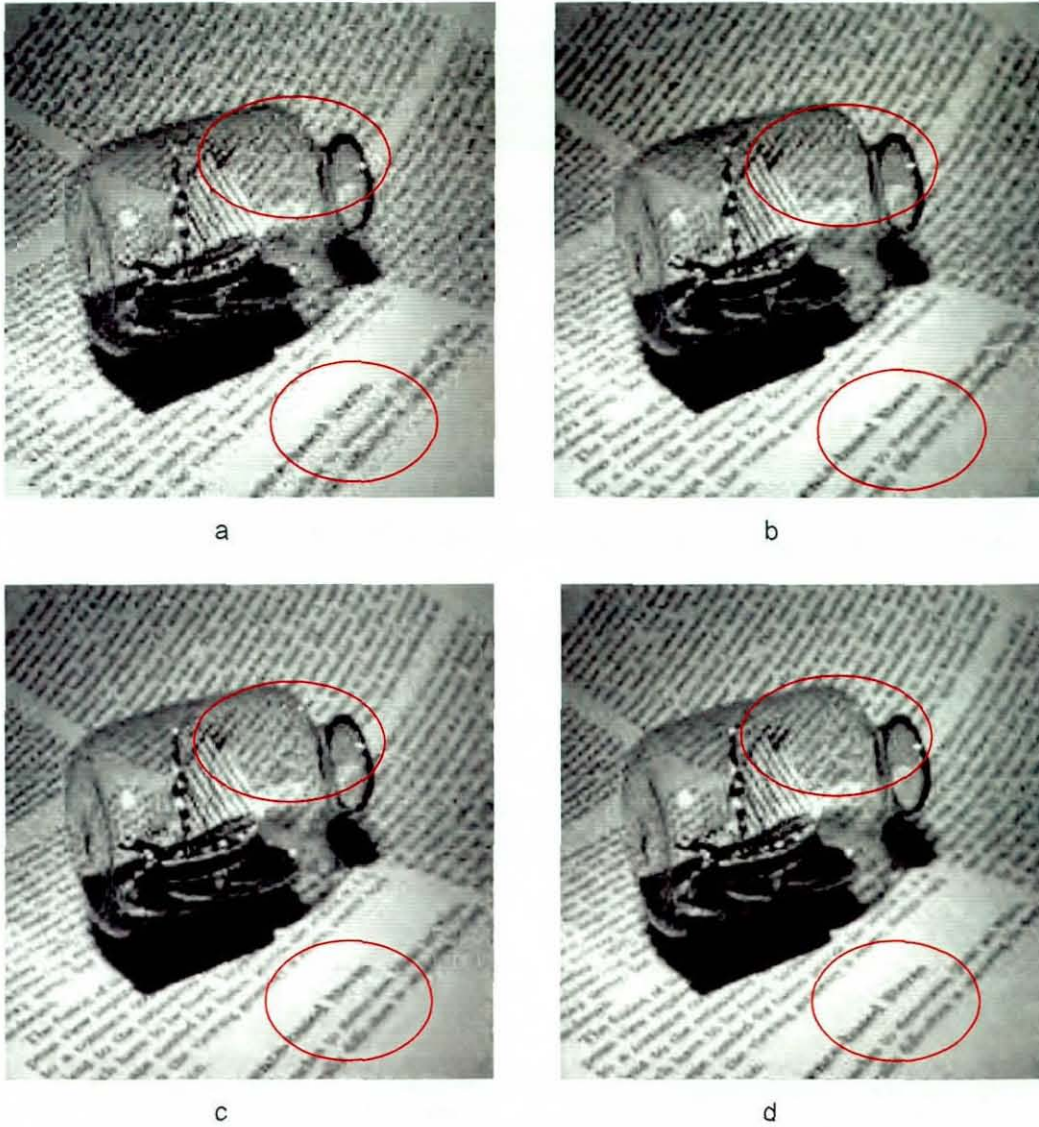
Figure 5.7 Subjective quality evaluation for 'Bottle' reconstructed right image at 0.25 bpp
(a) PBDCPC DCT-based at 23.92 dB (b) CODEC III at 26.51 dB (c) CODEC IV at 27.23 dB
(d) CODEC V at 27.97 dB
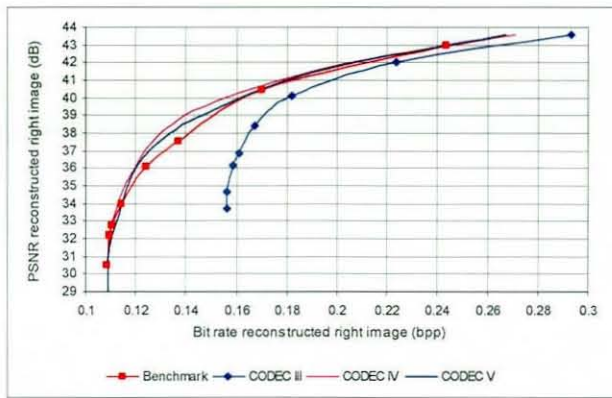
a
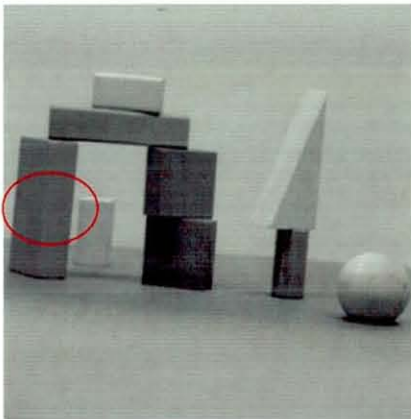
b

c
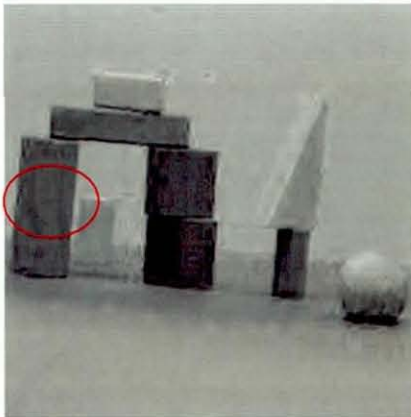
d
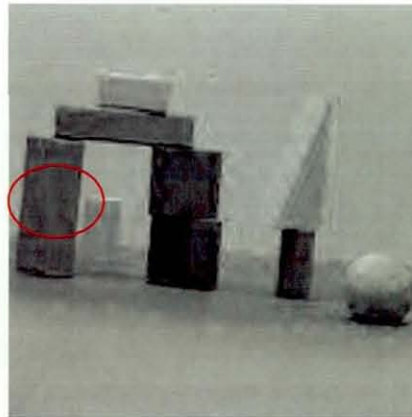
e

Figure 5.8 For reconstructed right image Arch, (a) Rate-distortion performance
graphs of the benchmark, CODEC IV and CODEC V (b) Original right image
(c) Benchmark, DCT-PBDCPC, 0.12 bpp, 35.35 dB (d) CODEC IV, 0.12 bpp, 36.31 dB
(e) CODEC V, 0.12 bpp, 35.96 dB

5.4 OBDC in wavelet-based stereo image coding

In this section, two wavelet-based stereo image coding algorithms, which uses Overlapped Block Disparity Compensation (OBDC) are proposed. The difference between the two proposed CODECs are as follows: In CODEC VI, neighbouring blocks are taken into consideration only in the selective DC process. In addition, bits have to be allocated for coding the OBDC tables and to transmit the disparity vectors of blocks in the lowest resolution subband for the reconstruction of predicted image at the decoder. Meanwhile, CODEC VII uses the PBDCPC framework and includes neighbouring blocks during DE/DC process. Further, in CODEC VII, OBDC is applied to all blocks in each subband of the DWT decomposition.

5.4.1 Design of CODEC VI

In this algorithm, FSBM is initially performed to search for the best match only in the highest decomposition level subbands (LL_3, LH_3, HL_3, HH_3), using non-overlapping blocks, and OBDC is later performed for disparity compensation.

At the encoder end, the left image and the right image undergo a 3-level dyadic DWT separately. For a block of each subband in level-3 (block size 2x2) of the right image using FSBM, a search is performed in the corresponding subband of the left image, within a maximum likelihood area. The best predictor is found by using the minimum MSE as the selection criterion.

For level-2 (each block of size 4x4) and level-1 (each block of size 8x8) detailed images (HL_2, LH_2, HH_2, HL_1, LH_1, HH_1), the predicted best matches are found by using the disparity vectors of the corresponding detailed images of level-3. For level-2, the best match is found by multiplying the disparity vector of level-3 by two while for level-1 the disparity value is obtained by multiplying the corresponding LL_3 disparity value by four. This results in the proposed algorithm, only having to transmit the disparity values that are in level-3 leading to a considerable decrease in the bit budget required for the transmission of the disparity vector field. Up to this stage, the

prediction errors are obtained for all the subbands. Then, blocks with MSE greater than a pre-specified threshold will be assigned a value '1' in a so-called, OBDC indicator table, which would indicate that OBDC needs to be performed for further refinement of the original prediction. An entry of '0' in the table means that OBDC need not be applied.

Figure 5.9, illustrates the OBDC procedure. The block is divided into four quadrants; 1, 2, 3 and 4. The prediction error of each quadrant is replaced by a new prediction error obtained by averaging its prediction error with that of its immediate neighbours. The new prediction errors are obtained as follows:
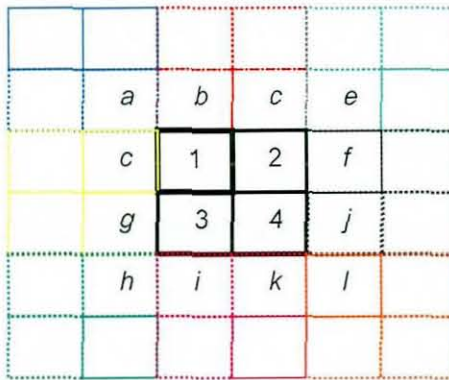
$$E_{1new} = (E_1 + E_a + E_b + E_c)/4$$
$$E_{2new} = (E_2 + E_d + E_e + E_f)/4$$
$$E_{3new} = (E_3 + E_g + E_h + E_i)/4$$
$$E_{4new} = (E_4 + E_j + E_k + E_l)/4$$

Figure 5.9 Block and its neighbouring blocks used in OBDC

To ensure that the new prediction reduces the energy level of the error block, its MSE is compared with the MSE of the old prediction error. If the MSE of the new prediction block is greater than the old MSE, then the old prediction error is used.

After OBDC has been applied on all blocks indicated by the OBDC indicator tables, the new prediction errors for each subband contains blocks predicted with and without OBDC. The new prediction errors are then used to form the new PEI. Finally each wavelet-block in the left image and the new PEI undergoes DWT JPEG-like encoding and are subsequently transmitted.

At the decoder end, by making the disparity vectors and the OBDC indicator tables available, the reverse process of that of the encoder is performed to obtain the reconstructed right image.

5.4.2 Design of CODEC VII

In CODEC VII, OBDC is used to enhance the performance of pioneering block-based predictive coding of CODEC V. CODEC VII operates in a manner similar CODEC V (refer to CODEC V block diagram) except that the blocks used for matching are *overlapping*.

In figure 5.10, to encode the shaded block (divide into four quadrant, labeled 1, 2, 3, 4), two *overlapping* pioneering blocks (the preceding and above block to be encoded), are used. They overlap by half the block to be encoded. Following a PBS search process as in CODEC V, after the best match is found, the block immediately below the upper block, but overlapped by half of the upper block (refer to shaded block in figure 5.7) is chosen as best match for the block to be encoded

The pioneering block selection in the predicted image is done on a 'so-far' reconstructed image and since its overlapped block, for example in figure 5.10, quadrant 1, 2 and 3 of a block, have already been reconstructed from the above and preceding blocks of the block to be encoded. Therefore, for quadrant 1, 2 and 3, average values are taken with the corresponding quadrant of the encoded block.

At the decoder, CODEC VII uses the PBDCPC coder framework, and satisfies the two PBDCPC conditions: using reconstructed reference frames at the encoder and selects overlapping pioneering blocks from identical, 'so-far' reconstructed predicted frames. This enables a similar process at the encoder and decoder ends enabling reconstruction of the predicted image without the need to transmit the disparity vectors.
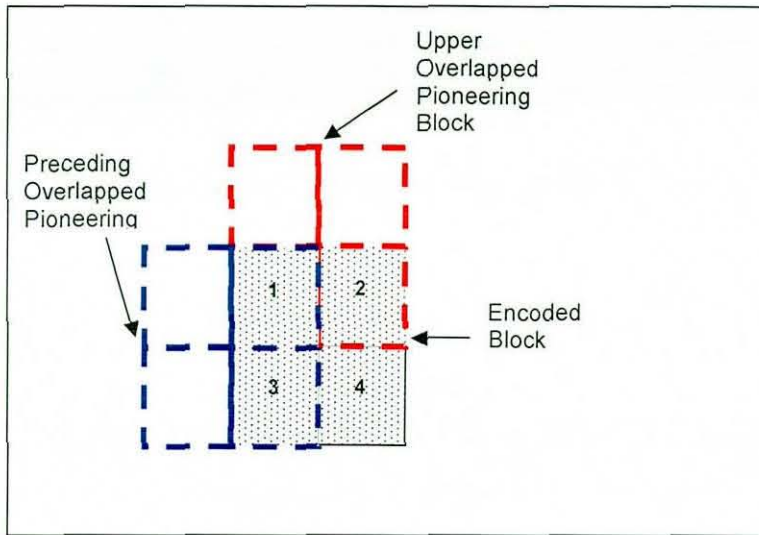
Figure 5.10 Illustration of overlapped pioneering blocks

5.4.3 Experiment results and analysis

The CODEC VI and CODEC VII are implemented using purpose built MATLAB routines and tested on a set of stereo image pairs as that used for testing the CODECs in previous section. Parallel axis camera geometry has been used to acquire these images and we assume that pixels at the same spatial location to have the same brightness. Compactly supported and symmetric biorthogonal filters (bior 9/7) are used as in JPEG-2000.

Both CODEC VI and CODEC VII , are compare to rate-distortion (RD) performance of CODEC V (the previous best). The objective image quality is measured in terms of Peak Signal to Noise Ratio (PSNR), whereas the compression efficiency is measured in bits per pixels (bpp) as follows:

$$BR = \left\lceil \frac{Tot\_Bits_{comp}}{N \times M} \right\rceil \quad bpp$$

where, N and M represent the image dimensions and Tot_Bits_{comp} is the total number of bits required to encode the prediction errors only. Except, CODEC VI, Tot_Bits_{comp},

includes the number of bits required to transmit the OBDC indicator tables and disparity vectors of the level-3 subbands. For a fair comparison, the objective quality of the left image was set at equivalent levels for all CODECs for a given image pair.

In figure 5.11 (a) and (b), the rate-distortion performance of CODEC VI and CODEC VII are compared with that of CODEC V for stereo image pairs, "Arch" and "Bottle", respectively. CODEC VI rate-distortion performance graphs are lower as compare to that of CODEC V and CODEC VII. The perfomance of CODEC VI is affected due to performing disparity estimation using FSBM without taking into consideration the affect of its neighbouring blocks. However note that neighbouring blocks are taken into consideration during the subsequent disparity compensation stage. Furthermore, part of the total bits are used in encoding of the disparity vectors and OBDC tables which made it unable to operate at a much lower bit rates (0.15 bpp as shown in figure 5.11).

Even though the rate-distortion performance graphs of CODEC VII are higher than those of CODEC VI, it shows no significant improvement when compared to CODEC V. One reason for the improvement of performance of CODEC VII over that of CODEC VI, is due to the fact that during DE and DC process, neighbouring blocks are taken into consideration. More importantly CODEC VII uses PBDCPC technique, which enables the use of bits that would otherwise bere used for disparity field coding to be used for coding the PEI, resulting in a better reconstructed right image. Meanwhile, CODEC VII shows no significant improvement over CODEC V, as by applying OBDC on every block of the subband degrades the disparity compensation efficiency particularly for those blocks, which can be compensated efficiently without OBDC. This is based on a claimes in [221] that OBDC is efficient only when high frequency components exists in the block. It was mentioned that OBDC is efficient in reducing the blocking artifacts [221]. However, CODEC VII is wavelet-based where blocking artifacts are non-existant and thus no significant compression gain can be obtained over CODEC V.
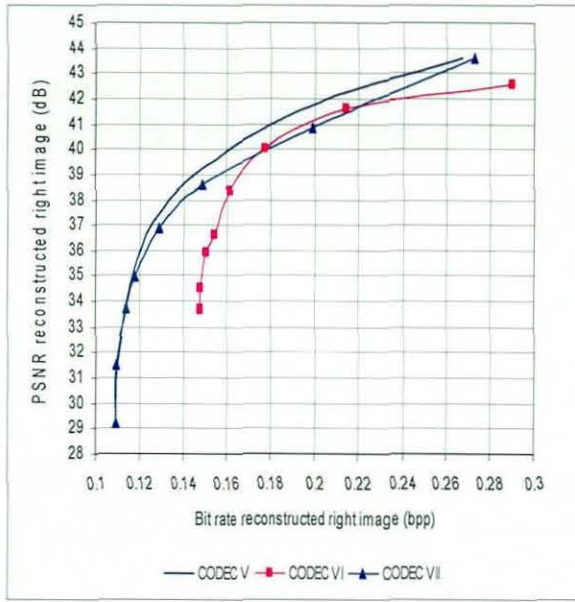
a
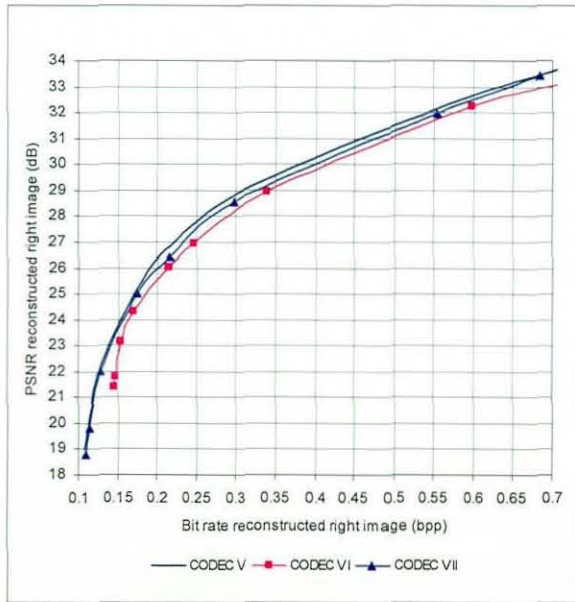
b

Figure 5.11 Rate-distortion performance graphs of
CODEC V, CODEC VI and CODEC VII for stereo image pairs
(a) Arch (b) Bottle

5.4.4 Conclusion

In both CODECs OBDC is implemented in multiresolution DWT domain. We have shown that CODEC VII, which uses PBDCPC in combination with OBDC, improves CODEC VI (does not use PBDCPC). However our experimental results showed that CODEC VII is unable to show improvement over CODEC V. In [120], it is found that performing DE/DC in wavelet domain is less effected by occlusion as the pixels in covered and uncovered areas are first decorrelated by spatial filtering (WT). Since the mismatching only occurs when occluded area contains edges, its likelihood is reduced in wavelet-domain DE/DC. Further, the wavelet-domain DE/DC, alleviates photometric distortion, due to its frequency selectivity (note that high-band coefficients used for block matching would be less affected than low-band ones). In addition , there are no blocking artifacts as compared to DCT-based coders. All above reasons contribute towards OBDC being ineffective in stereo image CODECs that performes DE/DC in wavelet-domain.

6   Stereoscopic extension of JPEG-2000

6.1 Overview

The CODECs proposed in Chapter 4, 5 and 6 do not posses scalability properties. In contrast, the new image compression standard, JPEG-2000 has resolution and fidelity scalability properties. Therefore, within the research context of this chapter we propose the design of a Rate Scalable Stereo Image CODEC (RASTER), by making use of the monocular compression technology present in JPEG-2000. We show that RASTER has the unique ability to preserve the image quality at binocular depth boundaries, which is an important requirement in the design of a stereo image CODEC. This chapter also discusses the functional and operational features of JPEG-2000 image coding standard and EBCOT coding.

6.2 JPEG-2000 background

JPEG, a DCT-based technology, became a still image compression standard in 1992 [62]. At that time, DCT-based transform techniques were well established as compared to wavelet analysis and wavelet coding, which were still very new technologies [200]. However, since the introduction of the DCT-based JPEG standard, the wavelet-based image coding techniques have developed significantly. The widespread application of wavelet transform based technologies to image compression originates from Shapiro's paper of Embedded Zero-tree Wavelet coding by Shapiro [185]. Shapiro's CODEC is effectively able to exploit the multiresolution properties of the wavelet transform, to produce efficient image compression.

In March 1997, when a new call for contributions were launched for the development of a new standard for the compression of still images (called JPEG-2000), of the many algorithms submitted, the Wavelet Trellis Coded Quantization (WTCQ) method

performed the best and was selected in November 1997, as the reference JPEG-2000 algorithm. In November 1998, the EBCOT algorithm [43] was presented to the working group by its developer, David Taubman, and was adopted as the method for encoding the wavelet coefficients. In March 1999, the MQ coder [43] was presented to the working group and was adopted as the arithmetic coder to be used in JPEG-2000. The standardization process, which is coordinated by the JTC1/SC29/WG1 of ISO/IEC, launched JPEG-2000 as a new International Standard (IS) in December 2000 [180]. JPEG-2000 provides an entirely new way of compressing images based on the wavelet transform, with rate-distortion performance superior than JPEG especially at low bit rates In addition, the WT which provide multiresolution decomposition of the image, enable new features such as resolution and fidelity scalability, region of interest, random access etc., which are important in present day applications [180].

6.2.1 Framework of JPEG-2000 coder

Input Image → Wavelet Transform → Quantization → Tier 1 Coding → Tier 2 Coding → *Code Stream*

Code Stream → Tier 2 Decoding → Tier 1 Decoding → Dequantization → Inv-Wavelet Transform → *Reconstructed Image*
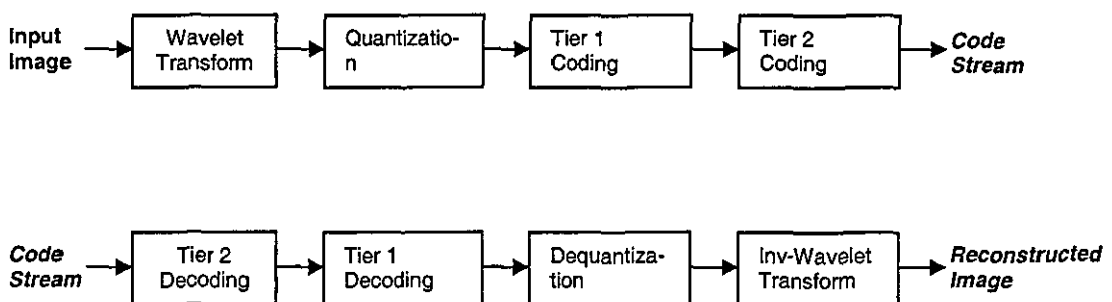
Figure 6.1 JPEG-2000 CODEC

For simplicity, we briefly described the JPEG-2000 algorithm with respect to a single tile of a single component (e.g., gray level) image. If the image being compressed is in color, it is divided into three components. Each component is compressed individually. If the image is very large (maximum image size in JPEG-2000 is $2^{32}-1$ by $2^{32}-1$), more than one tile is used. The purpose of tiling is to reduce the memory space. Each tile is compressed individually. Based on figure 6.1, the first

step is to compute a wavelet transform that results in subbands of wavelet coefficients. Two such transforms, an integer (reversible) and a floating-point (irreversible), are specified by the standard. The default reversible transform is implemented by means of the Le Gall 5-tap/3-tap filter [140]. The default irreversible transform is implemented by means of the biorthogonal Daubechies 9-tap/7-tap filter [146]. There are $L+1$ resolution levels of subbands, where L is a parameter determined by the encoder. In step two, the wavelet coefficients are quantized. The user specifying the target bit rate or quality level does this. A different quantizer is used for the coefficients of each subband, and each quantizer has only one parameter, its step size. Mathematically, the quantization process is defined as

$$V(x, y) = \lfloor |U(x, y)| / \Delta \rfloor \operatorname{sgn} U(x, y) \qquad (6.1)$$

where Δ is the quantizer step size, $U(x,y)$ is the input subband coefficient, and $V(x,y)$ denotes the output quantizer indices for the subband. The lower the bit rate, the coarser the wavelet coefficients have to be quantized. As for lossless compression, the step size is set equal to one ($\Delta = 1$). In this case, no quantization is done and all the coefficients remained unchanged. Step three (Tier 1 coding) uses the MQ coder (an encoder similar to the QM coder used in JBIG) to arithmetically encode the wavelet coefficients. The EBCOT algorithm [43] has been adopted for the encoding step. The principle of EBCOT is to divide each subband into blocks (termed code-blocks) that are coded individually. The code blocks are rectangular in shape with nominal width and height being an integer power of two and the product of the width and height cannot exceed 4096. Each code block is independently coded using a bit-plane coder. During the bit-plane coding technique, there are three coding passes per bit-plane. These coding passes are significance, refinement, and cleanup. The first coding pass, which is significant pass is used to convey significance and sign information of the each bit plane of each subband coefficients. The refinement pass, convey subsequent bits after the most significant bit of each sample. The final pass, which is the cleanup pass, send significance and sign information for those samples that have not yet been found to be significant and are predicted to remain insignificant during the processing of the current bit plane. The bit-plane encoding process generates a sequence of

symbols for each coding pass. The bits resulting from several code-blocks are group to become a packet and the packets are the components of the bit stream. The last step (Tier 2 coding) is to construct the bit stream. This step places the packets, as well as markers, in the bit stream. The markers can be used by the decoder to skip certain areas of the bit stream and to reach certain points quickly. Using markers, the decoder can, e.g., decode certain code-blocks before others, thereby displaying certain regions of the image before other regions. Another use of the markers is for the decoder to progressively decode the image in one of the several ways. The bit stream is organized in layers, where each layer is a natural way to achieve progressive image transmission and decomposition. The decoder block diagram is just the mirror image of the encoder as shown in figure 6.1. A more details description of the JPEG-2000 CODEC can be found in [7].

6.2.2 JPEG-2000 special features

As previously mention, JPEG-2000, besides having superior rate-distortion performance at low bit rates and subjective image quality, compared to JPEG, possesses features, which are demanded by state-of-the-art multimedia applications. These features are as follows:

- Support lossy and lossless compression of a single-component and multi-component imagery in a single framework.
- Scalability by fidelity (accuracy) or resolution (size).
- Region of interest coding, whereby difference parts of an image can be coded with differing fidelity.
- Random access to particular regions of an image without needing to decode the entire code stream.
- A flexible file format with provisions for specifying opacity information and image sequences.
- Good error resilience.

In this section we will briefly explain, how scalability by SNR and resolution are obtained in JPEG-2000 CODEC. Detailed explanations about other features in JPEG-

2000 can be found in these references [7, 43,180]. Scalability by SNR (accuracy) arises when the compressed bit stream contains elements, which can be discarded in order to obtain a lower quality (higher distortion) representation of the subband samples. This is referred to as distortion scalability. Ideally, the reduced quality representations obtained by discarding appropriate elements from a distortion scalable bit stream can be decoded to reconstruct the original image with a fidelity approaching that of an "optimal" coder, tailored to produce the same bit rate as the scaled bit stream.

Most practical means of achieving this goal involve some form of bit-plane coding, whereby the magnitude bits of the subband samples are coded one by one from most significant bit (MSB) to least significant bit (LSB). Discarding least significant bits is equivalent to coarser quantization of the original subband samples. The term "successive approximation" and "bit rate scalability" have also been used in connection with this type of scalability.

A second type, scalability by resolution (size) is one from which a reduced accuracy may be obtained simply by discarding unwanted subbands of compressed data. The lower accuracy representation should be identical to that which would have been obtained if the lower resolution image were compressed directly. The DWT's multiresolution properties arise from the fact that the LL_d is a reasonable low resolution representation of LL_{d-1}, with half the width and height. Here, the original image is interpreted as a LL_0 subband of highest resolution, while the lowest resolution is represented directly by the LL_D subband. The LL_d subband, $0 \leq d < D$, may be recovered from the subbands at levels $d+1$ through D by applying only $D-d$ stages of DWT synthesis. So long as each subband from DWT stage d, $0 < d \leq D$, is compressed without reference to information in any of the subbands from DWT stages d', $0 \leq d' < d$, we may convert a compressed image into a lower resolution compressed image, simply by discarding those subbands which are not required. The number of resolutions available in this way is $D+1$.

Although resolution scalability provides a crude mechanism for decreasing the bit rate and increasing distortion, this is not usually an efficient mechanism for trading distortion for compressed size. It has been observed that discarding subbands from a compressed bit stream generally produces lower resolution images with such small distortion (and large bit rate) as to be inappropriate for applications requiring significant compression. In order to produce a family of successively lower image resolutions with a consistent level of perceived or objective distortion (e.g., a consistent mean square error), the multiresolution transform should be combined with distortion scalable coding [43].

6.2.3 Independent Embedded Block Coding with Optimized Truncation [EBCOT]

The first step of EBCOT [43] is to use DWT to decompose the image pixels into spatial frequency subbands. Each subband is then partitioned into relatively small blocks, known as code-blocks. Each code-block is coded entirely independently, without reference to other blocks in the same or other subbands. Finally these bit streams are packed into so called quality layers in the codestream.

Ideally an embedded coder generates a bit stream such that every subsets of the stream can be decoded to reconstruct the original code-block with the SNR approaching that of an optimal coder, tailored to produce the same bit rate as the subset. The bit stream generated by the block coder may be truncated at any of a large number of points. The compressor is free to assign incremental contributions from each code-block to each quality layer in any desired fashion. The rate and distortion information, (R_i^n, D_i^n) for each candidate truncation point, n, in the embedded bit stream generated for code block B_i is collected and the information is used to assign code block contributions to quality layers in such a way that layers $1,2,3,...,\lambda$ contain a rate-distortion optimal representation of the image, for each λ. This strategy is applied to various distortion measures [43]. However, in the adoption of this technique to the JPEG-2000 standard, a slight modification exists in the fact that no

restriction is placed on the interpretation adopted by the compressor in its construction of quality layers.

A natural technique for embedded coding is provided by the well known bit-plane coding technique which could be described as follows: Let $\chi[n] \in \{1,-1\}$ denote the sign of a transform coefficient $s[n]$ at location, $n = [n_1, n_2]$, and let $v[n]$ denote the quantized magnitude, i.e., $v[n] = \left\lfloor \frac{|s[n]|}{\delta} \right\rfloor$ where δ is the step size of the relevant deadzone quantizer. The value of any individual samples $s[k]$, is represented by progressively coding its magnitude bits, from most to least significant. Let P denote the number of bit-planes required to represent all samples in the block, so that $v[n] < 2^P, \forall n$; then the embedded bit stream identifies the value of the p'th magnitude bit $v_p[k]$ for each $p = P-1,...,0$. The sign bit, $\chi[k]$, is coded immediately after the first significant bit position, $v_p[k] = 1$. As a result, truncation of the embedded representation is equivalent to discarding some number, p_k, of least significant magnitude bits from each sample, $s[k]$. This in turn is equivalent to employing a larger quantization step size, $2^{p_k} \delta$, for that sample. Thus for any given sample, the available quantization levels are quite coarsely space. However an important feature of the embedded coder used here, over plain bit-plane coding is the adaptive nature of the sequence in which bits from different samples are coded, which tends to encode the most valuable information (in the sense of reducing the distortion of the reconstructed image the most) as early as possible. The embedded block coder uses context modelling [43] to address both the ordering and the coding of these events.

The use of most valuable information as early as possible in the encoding process was the main motivation for using the JPEG-2000 (or EBCOT) technology for stereo image compression. When using the suppression theory in coding stereo image pairs in association with disparity compensated predictive coding, it is often necessary to code the reference frame directly at high bit rate and to only allocate the minimal amount of the bit budget to code the prediction error image (PEIs). Thus the PEIs are

prone to excessive loss in quality, which would adversely affect the overall quality of the reconstructed stereo image pair. The application of embedded block coding described above, would solve this problem by concentrating on the accurate coding of coefficients close to high activity areas (usually along object boundaries where binocular depth information is excessive) of the PEIs. The remaining low activity areas, being mostly zeros, would remain so, even under very high compression rates. Thus at the decoder end the disparity compensated procedure would be able to reconstruct a predicted image (right image) whose quality near high activity areas of the error image is determined by the efficiently coded PEIs and the quality near low activity areas would be largely determined by the reconstructed reference (left) image of superior quality.

6.3   RASTER: A JPEG-2000 stereo image Codec

This proposed RASTER CODEC, deviates from all earlier proposed CODECs in that it is based JPEG-2000 framework. Thus it is able to take advantage of new features in the JPEG-2000 which is not available in baseline-JPEG. The feature used in RASTER is scalability, in order to produce a rate scalable stereo image coding algorithm.

6.3.1 Introduction

To make use of the rate scalability provided by the JPEG-2000 in stereo image compression, one obvious way is to apply it independently in coding the stereo image pair. However, such a coder would be inefficient, as the stereo redundancy present within a stereo image pair is not exploited. Fortunately, a solution exists in the suppression theory of binocular vision [80], which is widely used in literature to address this issue and achieve image compression. In these methods one image of the stereo image pair (say left) retains the details of the scene, while the second image (right) retains the disparity information. Hence the second image can be highly compressed without affecting the depth information in the compressed stereo images

[80]. In this chapter we propose how the technology provided within the new JPEG-2000 standard could be used in association with the suppression theory of binocular vision to achieve scalable stereo image compression.

6.3.2 The RASTER Codec design

Figure 6.2 shows the block diagram of the proposed rate scalable stereo image CODEC. A careful observation of figure 6.2 leads us to the following:

- A JPEG-2000 encoder is used to encode the directly transmitted *left image* and the *prediction error image* (PEI).

- At the encoder end, disparity estimation and compensation is performed in comparison with a locally decoded reference image instead of the original reference image.

- Decoded reference frames at both encoder and decoder are locked at a given data rate *R1*.

- The decoder contains two JPEG-2000 decoders, operating at two data rates.

- The disparity vector (DV) field is encoded separately using a fixed length code.

In order to explain the above inclusions in more detail, let us explore some theoretical aspects of the coding process in more detail. Let d denote the disparity vector field. Let L be the reference image (left view). The predicted right image R_{pred} is obtained by rearranging the pixels in L relative to d. If D denotes this operation, the predicted right image can be written as,

$$R_{pred} = D(L, d) \qquad (6.2)$$

Thus, the prediction error image, E_{pred} for the right image, R is obtained as,

$$E_{pred} = R - R_{pred} \qquad (6.3)$$

At the decoder end the predicted right image, $\hat{R}_{pred}$, is obtained by,

$$\hat{R}_{pred} = D\left(\hat{L}, \hat{d}\right) \tag{6.4}$$

where, $\hat{L}$ and $\hat{d}$ are the decoded reference image and disparity vector field, respectively. Thus, the decoded right image can be obtained using the equation,

$$\hat{R} = \hat{R}_{pred} + \hat{E}_{pred} \tag{6.5}$$

where, $\hat{E}_{pred}$ is the decoded predictive error image.

The disparity vector field is losslessly encoded using DPCM coding [180]. This leads to the relationship, $d = \hat{d}$. Thus, if the same reference frame is maintained both at the encoder and decoder, i.e. if $L = \hat{L}$, comparing equations (6.2) and (6.4),

$$\Rightarrow R_{pred} = \hat{R}_{pred} \tag{6.6}$$

This results in the decoded PEI, $\hat{E}_{pred}$, being the only source of distortion in the reconstruction of the right image (equation (6.5)). Thus, in order to maximize the performance of the CODEC it is necessary to maintain the same reference frame at both encoder and decoder. In the proposed CODEC we achieve this by adding a prediction feed back loop in the encoder so that a decoded image is used as the reference image.

However in the proposed scalable CODEC, the decoded reference images have different distortions at different data rates. Hence it is impossible for the encoder to generate the exact reference images as in the decoder for all data rates. In order to solve this problem we use adaptive disparity compensation (ADC), a concept similar to adaptive motion compensation (AMC) [43]. In ADC, in order to maintain the same reference frame in the encoder and decoder, we introduce a feed back loop in the decoder, such that the decoded reference images at both ends could be locked to the same data rate, say R.

Assume that the target data rate is R_T, where $R_L \leq R_T \leq R_H$ and the disparity vector field is coded at a rate R_{DV} where $R_{DV} < R_L$. As R_{DV} is known both by the encoder and decoder, if we decode the embedded reference image bit stream at rate $R_L - R_{DV}$ $= R$ at both ends, the decoded reference frames would be identical, i.e. $\hat{L}$. It was mentioned above that at the decoder the embedded reference image bit stream is decoded at two data rates. The image decoded at $R_L - R_{DV}$ is used as the decoded reference image to find the predicted image, and the image decoded at rate $R_T - R_{DV}$ is added to the predicted frame to generate the final decoded right image. By following this strategy, the reference images at the encoder and decoder are maintained as identical, which leaves the decoded E_{pred} as the only source of distortion, thus eliminating error propagation over varying data rates.

Figure 6.2 The block diagram of the proposed RASTER

6.3.3 Experiments results & analysis

The Jasper JPEG-2000 implementation [43] was used as the base compression engine. Experiments were performed on a set of five stereo image pairs namely lamp, cans, texture, lab and packs, respectively containing an increasing amount of stereo data. All images were grayscale (value range 0- 255, 8 bits per pixel, uncompressed) and are of size 320×320 pixels. As the images have been obtained using parallel axis camera geometry, a search area of 8 pixels in the horizontal direction was used for all images in disparity compensation. The compression rate is measured in bits per pixel (bpp). The objective image quality is measured in terms of peak signal to noise ratio.

Table 6.1 illustrates the rate-distortion performance of the RASTER CODEC in comparison to the direct transmission of right image, at same bit rate. For all experiments the left image (reference image) is directly transmitted at twice the bit rate of right (predicted) image tabulated in the first column of table 6.1. As a block size of 8×8 and a search window size of 8 pixels (0-7) were used, $\log_2 8/(64) = 0.0469$ bpp is allocated for the fixed length coding of disparity vector field, DV. The remaining bits are used in coding the PEI. For example, when 0.4000 bpp has been allocated for the coding of right image, 0.8000 bpp is used for the coding of directly transmitted left image, 0.0469 bpp is used for the coding of disparity vector field and 0.3531 bpp is used for the coding of the PEI. Note that for a fair comparison the quality of the directly transmitted (if it was to) right image at this rate is measured at 0.4000 bpp rather than at 0.3531 bpp. The reason is that under such a situation, no DV field would have to be transmitted for the right image reconstruction. Note that for all the experiments, the right frame prediction at the encoder end and the decoding at the decoder end is done based on a left frame that is decoded at rate $R1=0.1000$ bpp.

The results in table 6.1, clearly indicates that the RASTER CODEC produces reconstructed images of better objective image quality (up to 3.7 dB excess PSNR) compared to the right images that could be obtained by direct compression using JPEG-2000 at the same bit rate. This observation is true for all right image bit rates

below 0.2 bpp. However, for bit rates above this, slight, subjectively unnoticeable image degradations are observed.

Table 6.1 – Rate-distortion performance comparison

Right BR (bpp)	Cans		Packs		Lamp		Lab		Texture	
	JPEG 2K(dB)	RASTER (dB)	JPEG 2K (dB)	RASTER (dB)	JPEG2 K (dB)	RASTER (dB)	JPEG 2K (dB)	RASTER (dB)	JPEG 2K (dB)	RASTER (dB)
0.4000	44.67	44.25	34.79	34.35	46.35	46.17	38.52	37.77	36.80	36.45
0.2000	39.67	39.71	30.08	30.35	42.73	42.62	33.41	33.50	32.26	32.79
0.1600	37.99	38.22	29.00	29.44	41.18	41.35	32.02	32.42	31.18	31.67
0.1333	36.77	37.52	27.76	28.64	39.95	40.42	31.01	31.50	30.37	30.82
0.1143	35.58	36.88	27.07	28.02	38.93	39.70	30.05	30.93	29.37	30.45
0.1000	34.58	36.07	26.50	27.75	38.07	39.10	29.24	30.49	28.70	29.88
0.0889	33.81	35.73	25.83	27.47	37.17	38.76	28.79	30.13	28.22	29.71
0.0800	33.32	35.51	25.42	27.22	36.30	38.49	28.32	29.81	27.69	29.48
0.0727	32.46	35.20	25.00	27.00	35.75	38.27	27.76	29.54	27.29	29.24
0.0667	31.73	35.00	24.65	26.83	35.07	37.73	27.20	29.39	27.01	29.15
0.0615	31.45	34.62	24.28	26.52	33.88	37.64	27.02	29.19	26.73	29.04

Figure 6.3 Rate-distortion performance graphs of 'Cans' and 'Packs'

Figure 6.3 shows the rate distortion graphs of the RASTER CODEC and JPEG-2000 coder in compressing the right image of the stereo image pairs 'Cans' and 'Packs'. It

clearly indicates that the RASTER CODEC provides better, reconstructed image quality at all right images bit rates above 0.2 bpp. Similar performance of the RASTER CODEC is observed for the other stereo images pairs used in the experiments.

Figure 6.4(a) and 6.4(b) illustrates the original left and right images of the stereo image pair 'Cans'. Figure 6.4(c) illustrates the error image that results from disparity compensated prediction. Figure 6.4(d) illustrates the reconstructed left image, when the original is coded at a bit rate of $2 \times 0.0615 = 0.123$ bpp (PSNR=36.22dB). Figure 6.4(e) and 6.4(f) compares the subjective image quality of the right images produced by the two coders (RASTER & JPEG-2000) at a right image bit rate of 0.0615 bpp. The image produced by RASTER CODEC (PSNR=34.62dB) shows less ringing artifacts around object boundaries as compared to the image formed by directly transmitted right image using JPEG-2000 (PSNR=31.45 dB). Visual comparison of figures 6.4(d), 6.4(e) and 6.4(f) shows that the RASTER CODEC provides a reconstructed image with visual image quality equal to the reconstructed left image (JPEG-2000 coded) at twice the bit rate, whereas the directly transmitted right image is of noticeably inferior image quality as compared to the image of figure 6.4(d).

The quality of the reconstructed right image of the RASTER CODEC is dependent on the quality of the locally decoded left image that is used at the encoder and decoder ends as a reference for disparity compensated prediction. In order to maintain rate scalability of the proposed coder within a larger range, it is essential that this reference image is coded at the lowest acceptable bit rate. However very low quality of this reference frame could result in the lowering of the reconstructed right image quality of the RASTER CODEC. At the same time, if this reference frame is coded at a higher bit rate, it would result in an increase of the lower bit rate limit of the CODEC.

Figure 6.5 illustrates the rate distortion performance of the RASTER CODEC under varying compression rates of this reference image. It is seen that the best performance is when the directly transmitted left frame is used, without any change, as the

reference frame. However at such a setting the scalability of the RASTER CODEC would suffer.
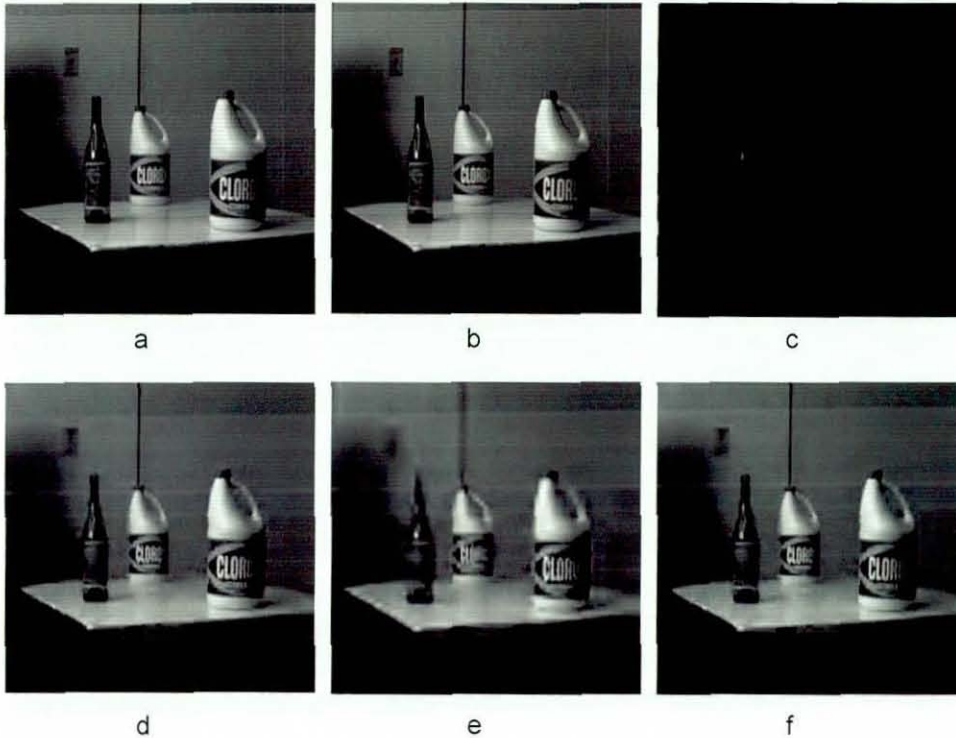
Figure 6.4 Subjective image quality illustrations for 'Cans'
(a) Original left image (b) Original right image (c) Error image (d) Reconstructed left image, JPEG-2000, at 0.123 bpp (e) Reconstructed right image, JPEG-2000, at 0.0615 bpp (f) Reconstructed right image, RASTER CODEC, at 0.0615 bpp

If scalability is not a requirement of the RASTER CODEC it is able to compress the right image at half the bit rate of the left image, still obtaining an image quality equivalent to that of the left image (see table 6.2). This is an interesting observation that shows that the proposed coder acts very competitively to most stereo image CODECs that have been proposed. It clearly shows that the RASTER CODEC efficiently codes the PEIs and uses information that it could from the reconstructed left frame to build areas where the prediction errors are negligible or zero.
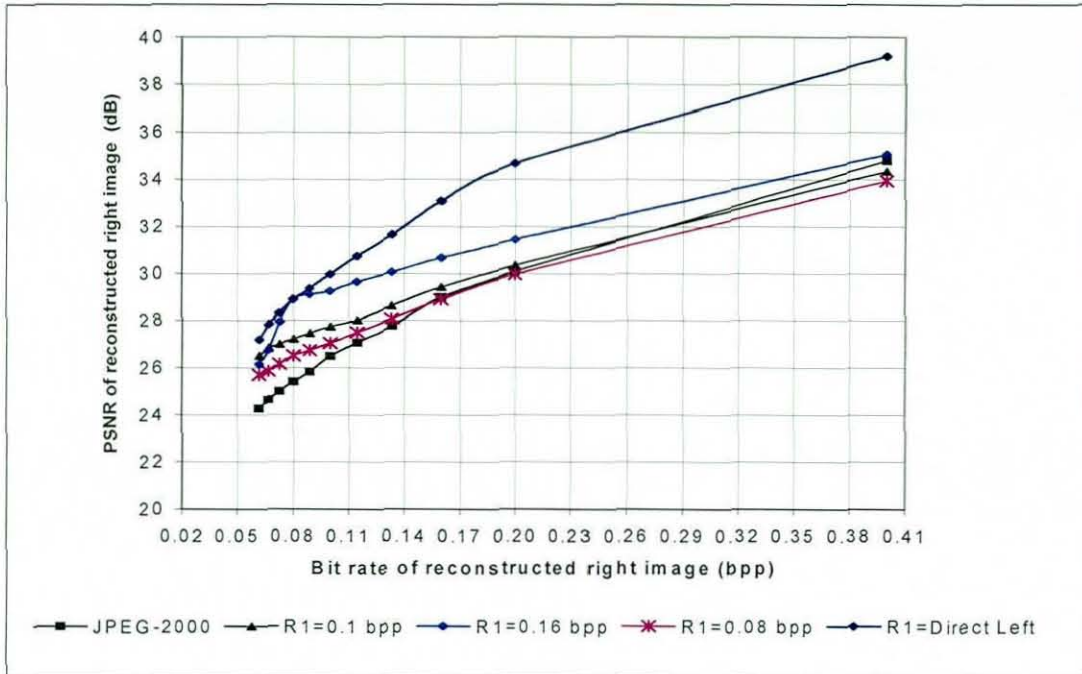
Figure 6.5 Rate-distortion performance of RASTER CODEC for stereo image pair 'Packs', under varying bit rates, R1, of the left image

Table 6.2 Rate distortion performance of RASTER with no rate scalability for image pair 'Packs'

Left Image Bit Rate (bpp)	Right Image Bit Rate (bpp)	Rec. Left Image PSNR (dB), JPEG2K	Rec. Right Image PSNR(dB), RASTER	Rec. Right Image PSNR(dB), JPEG2K
0.8000	0.4000	40.60	39.20	34.79
0.4000	0.2000	34.91	34.67	30.08
0.3200	0.1600	33.27	33.09	29.00
0.2666	0.1333	31.77	31.65	27.76
0.2286	0.1143	30.90	30.75	27.07
0.2000	0.1000	30.15	29.98	26.50
0.1778	0.0889	29.53	29.37	25.83
0.1600	0.0800	28.89	28.92	25.42
0.1454	0.0727	28.32	28.34	25.00
0.1334	0.0667	27.82	27.83	24.65
0.1230	0.0615	27.15	27.20	24.28

All the above experimental results clearly show that the RASTER CODEC shows rate scalability in stereo image compression and provides an efficient solution to many problems that are inherent to such CODECs. At low bit rates, the proposed CODEC is able to code the PEIs very efficiently. The prioritized coefficient coding technique adopted by JPEG-2000 acts well in the coding of disparity compensated images and

when combined with the disparity compensated stereo imaging techniques provides enhanced subjective and objective image quality in the reconstructed right images.

6.4 Conclusion

In this chapter, a rate scalable CODEC (RASTER) has been proposed for the efficient compression of stereo image pairs. The CODEC is designed based on the JPEG-2000 technology and the well known disparity compensated predictive coding technique of stereo image pairs. A JPEG-2000 CODEC is used as the base compression engine to code the directly transmitted reference image and the predictive error images. In order to maintain rate scalability adaptive disparity compensation (ADC) technique has been adopted. The experimental results have shown the proposed CODEC is able to achieve PSNR gains of up to 3.7 dB as compared to directly transmitting the right frame using JPEG-2000.

7 Conclusions & further research

This chapter summarizes the key results presented in Chapters 4, 5, and 6, draw conclusions and emphasizes the important contributions of this thesis. It also presents possible future directions of research to extend the functionality and efficiency of the proposed CODECs.

The thesis proposed seven, novel, DWT based stereo image CODECs. In CODEC I, DWT replaces DCT in the well-known DCTDPC stereo image compression algorithm of M.G. Perkins [155]. It was shown that performing disparity estimation and compensation in the pixel domain using non- overlapping 8x8 blocks of pixels, resulted in a predictive error image (PEI), which consists of artificially created block boundary discontinuities. These artificial discontinuities reduce the efficiency of subsequent DWT coding since they result in large wavelet coefficient values in the detail subbands. Consequently, CODEC I, only managed to record a marginal improvement (at most 1 dB) over the chosen benchmark.

To overcome the main drawback of CODEC I, for the remaining proposed CODECs, disparity estimation and compensation (DE/DC) are performed in the wavelet domain. Research has found that performing DE/DC in wavelet domain is less affected by occlusion as the pixels in covered and uncovered areas are decorrelated during DWT. Since the mismatching only occurs when the occluded area contains edges, its likelihood is reduced in wavelet-domain DE/DC. Further, the wavelet-domain DE/DC, alleviates photometric distortion, due to its frequency selectivity (note that high-band coefficients used for block matching would be less affected than low-band ones). However, the main problem in implementing DE/DC in multi-resolution subband domain is the high overhead bits required to transmit the resulting disparity vector field. To overcome this problem partially, in CODEC II, DE process is only done in the lowest resolution/ highest level subband of the multiresolution DWT domain. For the lower level subbands, the disparity vectors of its corresponding subbands in the highest level are used by

multiplying by a constant integer depending on the subband level. Although this reduced the overhead bits for disparity vectors, experimental results showed that propagating the disparities from the highest level subbands to the lower level subbands, significantly affect the performance of CODEC II. Therefore, by exploiting the inter-subband relationships, in CODEC III, disparity estimation was performed simultaneously in all subbands by the use of wavelet-blocks. Wavelet-blocks, which consist of wavelet coefficients belonging to the same spatial location from different subbands, allow efficient DE as compared to the DE strategy used in CODEC II without increasing the overhead bits required for disparity vector field coding. This was reflected in the significant improvement (up to 6 dB) obtained by CODEC III over CODEC II.

In CODEC IV, disparity estimation is done using wavelet-blocks as in CODEC III. However, due to the additional use of the pioneering wavelet-block based search technique, no disparity vector field is needed to reconstruct the right/predicted image at the decoder. Therefore, with extra bits available to transmit the PEI, the quality of the reconstructed right (predicted) image improved by up to 4 dB as compared to CODEC III. In addition, CODEC IV is able to operate at bit rates as low as 0.1 bpp, which is impossible with CODEC I, CODEC II and CODEC III. Note that in the latter CODECs a certain minimum bit budget is required for the lossless transmission of the disparity vector field. Pioneering block-based predictive coding allows disparity estimation to be performed independently on each subband, i.e. using wavelet coefficient blocks of the same subband in the coding process. CODEC V uses this idea to further improve the wavelet block based PBS of CODEC IV. Experiment results, recorded a 2dB PSNR improvement over CODEC IV. Also, both of these DWT-based PBDCPC CODECs successfully improved the performance over the DCT-based PBDCPC CODEC at all bit rates. It was shown that this is mainly due to blocking artifacts present in the DCT-based technology which particularly affect the low bit rate performance of the DCT-based PBDCPC. In comparison, blocking artifacts do not occur in DWT-based coding and thus

the reference frame where PBS is performed would be smoother, resulting in it providing a better prediction.

In CODEC VI and CODEC VII, the main focus was on designing a CODEC that would further improve on the performance of CODEC V. Overlapped block disparity compensation (OBDC), which is known to improve the performance of DCT based stereo image CODECs, was tested in the DWT domain in CODEC VI and CODEC VII. Experiments showed that significant improvements using OBDC could be expected only in the presence of blocking artifacts. Since in DWT-domain there are no blocking artifacts, performance analysis of CODEC VI and CODEC VII showed no significant improvement over CODEC V.

All the CODECs (I - VII) are designed based on the building blocks of a baseline-JPEG CODEC, which make them more adoptable to be used with software, hardware modules that use baseline-JPEG coding architecture. This could be particularly important at a time, image and video coding technology is undergoing a steady change from DCT-based coding schemes to DWT- based coding schemes.

However, all proposed CODECs in this thesis can easily be converted to an alternative architecture, which use DWT coefficient encoding techniques more suitable for wavelets coefficients such as the scalable EZW [185], SPHIT [16] or EBCOT [43]. By using these techniques, the multiresolution coding features of DWT can be further exploited and functionality can be further extended. Thus, in the last part of the research JPEG-2000 technology was used to take advantage of the rich features offered by it, i.e., the scalability feature of JPEG-2000 was used to produce a rate scalable stereo image compression algorithm. Rate scalability of the stereo image CODEC was made possible, through a feed back loop introduced at both the encoder and decoder, so that the decoded reference image at both ends could be locked at the same rate. Experiments results showed gains of up to 3.7 dB as compared to directly transmitting the right image using JPEG-2000.

Further improvements, in terms of algorithmic and computational efficiency are also possible by using faster search techniques. For example, one could explore the possibility of using the search information for already reconstructed neighbouring blocks when searching for the best match for the current block. For example, in CODEC V, the re-use of search information of lower resolution subbands for the search operations of higher subbands will expediate the overall search operation .

For colour stereo image compression, it can be approximately regarded as compression of multiple component YC_bC_r (transformation of the RGB components) of stereo image pairs. The DE/DC process can be performed between each component YC_bC_r of the stereo image pairs independently. However to reduce the overhead bits the disparity is obtained by taking the average disparity of the three components.

The DWT-based stereo image coding techniques proposed in this thesis could further be developed to cater for stereoscopic video coding. Of particular interest would be subjectively optimised CODECs, which carefully selects and prioritise the coding of visually important regions/DWT coefficients, in a stereoscopic scene.

Fractal image compression has been known to be very effective in exploiting the self-similarity in the spatial image domain. Thus, the fractal coding technique can be extended in the wavelet domain which can be viewed as a technique for block prediction from lower resolution subbands to the higher resolution ones. Therefore, the research ideas extensively tested within this thesis could help in an attempt to use fractal in stereo image/video coding.

References

[1] ISO/IEC JTC 1/SC 29/WG 1, ISO/IEC FCD 15444-1: Information technology- JPEG 2000 Image Coding System: Core coding system, International Organization for Standardization, ISO/IEC JTC1 SC29/WG1. (www.jpeg.org/FCD15444-1.html)

[2] A. Luo and H. Burkhardt (1995), "An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions", International Journal of Computer Vision, vol.15, no. 3, pp.171-188, July 1995.

[3] A. Redert, E. Hendricks, et al. (1999), "Correspondence estimation in image pairs", IEEE Signal Processing Magazine, vol.16, pp. 29-46, May 1999.

[4] A.K. Louis, P. Maab, et al. (1997), Wavelets Theory and Applications, John Wiley & Sons.

[5] A.R. Calderbank, I. Daubechies, et al. (1998), "Wavelet transforms that map integers to integers", Applied and Computational Harmonic Analysis, vol.5, no.3, pp. 332-369, July 1998.

[6] Aaron F. Bobick and S. S. Intille (1999), "Large occlusion stereo", International Journal of Computer Vision, vol. 33, no. 3, pp. 181-200.

[7] Adams, M. D. (2001), Jasper Software Reference Manual Version 1.500.4, pp 1-21, December 2001.

[8] Adrian Munteanu, Jan Cornelis, et al. (1998), "Wavelet lossy and lossless image compression technique-Use of the Lifting scheme", IWSSIP 1998, Zagreb, Croatia, pp. 12-19.

[9] Adrian Munteanu, Jan Cornelis, et al. (1999), "Wavelet image compression-The Quadtree coding approach", IEEE Signal on Information Technology in Biomedicine, vol. 3, no.3, pp. 176-185, September 1999.

[10] Ahmed H. Tewfik, Deepen Sinha, et al. (1992), "On the optimal choice of a wavelet for signal representation", IEEE Transactions on Information Theory vol.38, no.2, pp.747-765, March 1992.

[11] Aldo Morales and S. Agili (2003), "Implementing the SPIHT algorithm in MATLAB", ASEE/WFEO International Colloquium, American Society for Engineering Education, pp.1-10, June 2003.

[12] Ali N. Akansu and R. A. Haddad (1996), Fundamentals and Optimal Designs of Subband and Wavelet Transforms. Subband and Wavelet Transforms: Design and Applications, Kluwer Academic Publishers, pp. 33-81.

[13] Amir Averbuch, Danny Lazar, et al. (1996), "Image compression using wavelet transform and multiresolution decomposition", IEEE Transactions on Image Processing, vol. 5, no.1, pp. 4-15, January 1996.

[14] Amir Said and W. A. Pearlman (1996), "An image multiresolution representation for lossless and lossy compression", IEEE Transactions on Image Processing, vol.5, pp.1303-1310, September 1996.

[15] Amir Said and W.A. Pearlman (1993), "Image compression using the spatial-orientation tree", IEEE International Symposium Circuits and Systems, Chicago, IL.

[16] Amir Said and W.A. Pearlman (1996), "A new, fast, and efficient image codec based on set partitioning in hierarchical trees", IEEE Transactions on Circuits and Systems for Video Technology, vol.6, no.3, pp. 243-250, June 1996.

[17] Andrew Woods, Tom Docherty, et al. (1993), "Image distortions in stereoscopic video systems", Stereoscopic Displays and Applications IV, San Jose, USA, vol. 1915, February 1993.

[18] Aria Nosratinia and M. T. Orchard (1996), "Optimal unified approach to warping and overlapped block motion estimation in video coding", Proceedings SPIE VCIP '96, Florida, USA, March 1996.

[19] Athanassios Skodras, C. Charilaos, et al. (2001), "The JPEG 2000 still image compression standard", IEEE Signal Processing Magazine, vol. 18, no. 5, pp.36-58, September 2001.

[20] Battaglia, G. J. (1996), "Mean Square Error", AMP Journal of Technology, vol. 5, pp. 31-36, June 1996.

[21] Bernd Girod, Frank Hartung, et al. (1996), Subband Image Coding. Subband and Wavelet Transforms: Design and Applications, Kluwer Academic Publishers, pp. 213:249.

[22] Bierling, M. (1988), "Displacement estimation by hierarchical block matching", SPIE conference on VCIP '88, vol. 1001, pp.942-951.

[23] Cambridge Autostereo 3D Project.
(http://www.cl.cam.ac.uk/Research/Rainbow/projests)

[24]   C. Christopoulos, Joel Askelof, et al. (2000), "Efficient methods for encoding regions of interest in the upcoming JPEG 2000 still image coding standard", IEEE Signal Processing Letters, vol. 7, no.9, pp. 247-249, September 2000.

[25]   C. Christopoulos, A. Skodras, et al. (2000). "The JPEG 2000 still image coding system: An overview." IEEE Trans. on Consumer Electronics, vol. 46, no.4, pp.1103-1127, November 2000.

[26]   C. Lawrence Zitnick and T. Kanade (2000), "A cooperative algorithm for stereo matching and occlusion detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, no. 7, pp. 675-684, July 2000.

[27]   Carlos Vazquez, Janusz Konrad, et al. (2000), "Wavelet-based reconstruction of irregularly-sampled images: application to stereo imaging", International Conference on Image Processing (ICIP'00), Vancouver, Canada, pp.1-4, September 2000.

[28]   Cathlyn Y. Wen and R. J. Beaton (1996), "Subjective image quality evaluation of image compression techniques", The Human Factors and Ergonomics Society 40th Annual Meeting.

[29]   Chae Whan Lim, Nam Chul Kim, et al. (2000), "Rate-distortion based image segmentation using recursive merging", IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, no. 7, pp. 1121-1134, October 2000.

[30]   Chang Po-Rong and M.-J. Wu (2000), "A wavelet multiresolution compression technique for 3D stereoscopic image sequence based on mixed-resolution psychophysical experiments", Signal Processing: Image Communication, vol. 15, pp. 705-727.

[31]   Cheung Auyeung and J. Kosmach (1992), "Overlapped block motion compensation", Proceedings SPIE: VCIP'92, vol.1818, pp.561-572, November 1992.

[32]   Chia-Wen Lin, Er-Yin Fei, et al. (1998), "Hierarchical disparity estimation using spatial correlation", IEEE Transactions on Consumer Electronics, vol. 44, no. 3, pp. 630-637, August 1998.

[33]   Chia-Yuan Teng and D. L. Neuhoff (1996), "Quadtree-guided wavelet image coding, IEEE Data Compression, Snowbird, Utah, USA, pp. 406-415, March 1996.

[34]   Chun-Jen Tsai and A. K. Katsaggelos (1999), "Dense disparity estimation with a divide-and-conquer disparity space image technique." IEEE Transactions on Multimedia, vol. 1, no.1, pp. 18-29, March 1999.

[35]   Silvia Manolache Cirstea (2003), "3D-Object space reconstruction from planar recorded data of 3D-Integral images", Journal of VLSI Signal Processing, vol. 35, pp. 5-18.

[36]   D. Drascic and J. Grodski (1993), "Defence teleoperation and stereoscopic video", Proceedings SPIE Stereoscopic Displays and Applications IV, San Jose, California, vol. 1915, pp. 58-69, February 1993.

[37]   D. Milovanovic, Z. Bojkovic, et al. (2002), "Development of wavelet techniques in JPEG2000 image coding standard", Proceedings of the 2nd IASTED International Conference on Visualization, Imaging, and Image Processing, Malaga, Spain, pp. 794-797, September 2002.

[38]   D.S. Cruz, T. Ebrahimi, et al. (1999), "Region of interest coding in JPEG2000 for interactive client/server applications", IEEE Third Workshop on Multimedia Signal Processing, pp. 389-394, September 1999.

[39]   Daniel Scharstein and R. Szeliski (2002), "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", International Journal of Computer Vision, vol. 47, no. 1, pp. 7-42.

[40]   Daubechies, I. (1992). Ten lectures on wavelets, CBMS-NSF regional conference series in applied mathematics. Philadelphia, Pa, Society for Industrial and Applied Mathematics.

[41]   David Taubman, Erik Ordentlich, et al. (2002), "Embedded block coding in JPEG 2000", Signal Processing: Image Communication, vol.17, pp. 49-72.

[42]   David Taubman and A. Zakhor (1994), "Multirate 3-D subband coding of video", IEEE Transactions on Image Processing, vol. 3, pp. 572-588, June 1994.

[43]   David Taubman (2000), "High performance scalable image compression with EBCOT", IEEE Transactions on Image Processing, vol. 9, no. 7, pp. 1158-1170, July 2000.

[44]   Detlev Marpe and H. L. Cycon (2002), "High-performance wavelet-based video coding using variable block-size motion compensation and adaptive arithmetic coding", International Conference on Signal and Image Processing (SIP'02), IASTED, pp. 404-409, August 2002.

[45]   Diego Santa-Cruz and T. Ebrahimi (2000), "A study of JPEG 2000 still image coding versus others standards", The X European Signal Processing Conference, Tampere, Finland, vol. 2, pp. 673-676, September 2000.

[46]   Dimitrios Tzovaras, Nikos Grammalidis, et al. (1997), "Object-based coding of stereo image sequence using joint 3-D motion/disparity compensation",

IEEE Transactions on Circuits and Systems for Video Technology, vol. 7, no.2, pp. 312-327, August 1997.

[47]   Dirck Schilling and P. Cosman (1999), "Edge-enhanced image coding for low bit rates", IEEE International Conference on Image Processing (ICIP'99), vol. 3, pp. 747-751.

[48]   Dirck Schilling and P. Cosman (2001), "Feature-preserving image coding for very low bit rates", Data Compression Conference 2001, Snowbird, USA, IEEE Computer Society, pp. 103-112, March 2001.

[49]   Edward Aboufadel and S. Schlicker (1999), Discovering Wavelets, John Wiley & Sons, Inc.

[50]   Eran Edirisinghe and J. Jiang (1999), "A block-object hybrid approach towards stereo image compression", Proceedings of the International Conference on Information, Communication and Signal Processing (ICICS '99), Singapore, December 1999.

[51]   Eran Edirisinghe and J. Jiang (1999), "Towards eliminating disparity field coding of stereo image pairs", Proceedings International Conference on Information, Communications and Signal Processing (ICICS' 99), Singapore, December 1999.

[52]   Eran Edirisinghe (2000), "Stereo image compression approach via pixel value variation trend adaptive subspace projection", Proceedings of SPIE Photonics East Conference, Boston, MA, USA, November 2000.

[53]   Eran Anusha Edirisinghe, "Data compression of stereo images & videos", PhD Thesis, Loughborough University, UK, October 1996.

[54]   Eui-Sung Kang, Toshiba Tanaka, et al. (1998), "A multi-threshold embedded zerotree wavelet coder", Proceedings of ITC/CSCC '98, Sokcho, Japan, vol. 1, pp. 117-120, July 1998.

[55]   Fowler, J. E. (2000), "Adaptive vector quantization for efficient zerotree-based coding of video with non stationary statistics", IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, pp. 1478-1488, December 2000.

[56]   Francoys Labonte, Chon Tam Le Dinh, et al. (1999), "Spatiotemporal spectral coding of stereo image sequences", IEEE Transaction on Circuits and Systems for Video Technology, vol. 9, no.1, pp. 144-155, February 1999.

[57]   Frederic Dufaux and M. Kunt (1992), "Multigrid block matching motion estimation with an adaptive local mesh refinement", VCIP'92, vol. 1818, pp. 97-109

[58] G. Davis and A. Nosratinia (1998), "Wavelet-based image coding: An overview", Applied and Computational Control, Signals, and Circuits, vol. 1, no.1, pp. 205-269.

[59] G. Van der Auwera, Adrian Munteanu, et al. (1998), "A new technique for motion estimation and compensation of the wavelet details images", European Signal Processing Conference, EUSIPCO'98, Island of Rhodes, Greece, pp. 913-996.

[60] G. Van der Auwera, A. Munteanu, et al. (1998), "Video coding based on motion estimation in the wavelet detail images", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'98), Seattle, USA.

[61] G. Van Meerbergen, M. Vergauwen, et al. (2002), "A hierarchical symmetric stereo algorithm using dynamic programming", International Journal of Computer Vision , vol. 47, pp. 275-285.

[62] G.K.Wallace (1992), "The JPEG still picture compression standard", Communications of the ACM, vol.34, no.4, pp. 30-44, April 1991.

[63] Garcia-Perez, M. A. (1992), "The perceived image: Efficient modelling of visual inhomogeneity", Spatial Vision, vol. 6, no.2, pp. 89-99.

[64] Gilbert Strang and T. Nguyen (1996), Wavelets and Filter Banks, Wellesley-Cambridge Press.

[65] Gloria Menegaz and J. P. Thiran (2002), "Lossy to lossless object-based coding of 3-D MRI data", IEEE Transactions on Image Processing, vol.11, no. 10, pp.1053-1061.

[66] Gormish, M. J. (1999), "JPEG 2000: Worth the Wait?".
      http://rii.ricoh.com/~gormish/pdf/mwscs99-j2k-worthwait.pdf

[67] Goyal, V. K. (2001), "Theoretical foundations of transform coding", IEEE Signal Processing Magazine, vol.18, pp. 9-21, September 2001.

[68] Graham J. Woodgate, David Ezra, et al. (1998), "Autostereoscopic 3D display systems with observer tracking", Signal Processing: Image Communication, vol. 14, pp. 131-145.

[69] Guiwei Xing, Jin Li, et al. (2001), "Arbitrarily shaped video-object coding by wavelet", IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no.10, pp. 1135-1139, October 2001.

[70] H. Yamaguchi and e. al. (1989), "Stereoscopic images disparity for predictive coding", IEEE Proceedings ASSP '89, pp.1976-1979.

[71] Halle, M. (1997), "Autostereoscopic displays and computer graphics", Computer Graphic, ACM SIGGRAPH, vol. 31, no. 2, pp. 58-62.

[72] Haluk Aydinoglu and M. H. H. III (1998), "Stereo image coding: A projection approach", IEEE Transactions on Image Processing, vol. 7, no.4, pp. 506-516, April 1998.

[73] Harry Veron, D.A. Southard, et al. (1990), "Stereoscopic displays for terrain database visualization", SPIE Proceedings, Stereoscopic Displays and Applications, vol. 1256, pp. 124-135.

[74] Hiroshi Watanabe and S. Singhal (1991), "Windowed motion compensation", SPIE conference on VCIP '91, vol. 1605, pp. 582-589.

[75] Holschneider, M. (1995), Wavelets: An Analysis Tool, Oxford University Press.

[76] Hongen Liao, Nobuhiko Hata, et al. (2004), "Image-guidance for cardiac surgery using dynamic autostereoscopic display system", Proceedings 2004 IEEE International Symposium on Biomedical Imaging, Arlington, Va, USA, pp. 265-268, April 2004.

[77] Hosam Khalil, Amir F. Atiya, et al.(1999), "Three-dimensional video compression using subband/wavelet transform with lower buffering requirements", IEEE Transaction on Image Processing, vol. 8, no. 6, pp. 762-773, June 1999.

[78] Hwang-Seok Oh and H.-K. Lee (2000), "Block-matching algorithm based on an adaptive reduction of the search area for motion estimation", Real-Time Imaging, vol. 6, pp. 407-414.

[79] Hyuk Choi and T. Kim (2000), "Blocking-artifacts reduction in block-coded images using wavelet-based subband decomposition", IEEE Transactions on Circuit and Systems for Video Technology , vol.10, no.5, pp. 801-805, August 2000.

[80] I. Dinstein and e. al. (1989), "Compression of stereo images and the evaluation of its effects on 3-D perception", SPIE Applications of Digital Image Processing XII, vol. 1153, pp. 522-530.

[81] I. Dinstein and e. al. (1988), "On stereo image coding", 9th International Conference on Pattern Recognition, pp. 357-359.

[82] I. Sexton and P. Surman (1999), "Stereoscopic and autostereoscopic display systems", IEEE Signal Processing, vol. 16, pp. 85-99, May 1999.

[83] Ian P. Howard and B. J. Rogers (2002), Seeing in Depth: Depth Perception, Ontario, Canada, I. Porteous.

[84] Ian P. Howard and B. J. Rogers (2002), Seeing in Depth: Basic Mechanisms, Ontario, Canada, I. Porteous.

[85] III, W. J. P. (1999), MATLAB for Engineering Applications, WCB/McGraw-Hill.

[86] Ingemar J. Cox, Sunita L. Hingorani, et al. (1996), "A maximum likelihood stereo algorithm", Computer Vision and Image Understanding, vol. 63, no.3, pp. 542-567, May 1996.

[87] J. Jiang, E.A. Edirisinghe, et al. (1997), "A novel predictive coding algorithm for 3-D image compression", IEEE Transactions on Consumer Electronics, vol. 43, no. 3, pp. 430-437, August 1997.

[88] J. Jiang, Eran Edirisinghe, et al. (1997), "Algorithm for compression of stereo image pairs", Electronics Letters, vol. 33, no.12, pp. 1034-1035, June 1997.

[89] J. Li and S. Lei (1999), "An embedded still image coder with rate-distortion optimization", IEEE Transactions on Image Processing, vol. 8, pp. 913-924, July 1999.

[90] J. Xu, Z. Xiong, et al. (2001), "Three-dimensional embedded subband coding with optimized truncation (3-D ESCOT)", Journal Applied Computational Harmonic Analysis, vol. 10, pp. 290-315, May 2001.

[91] James S. Walker and T. Q. Nguyen (2000), "Adaptive scanning methods for wavelet difference reduction in lossy image compression", ICIP 2000, Vancouver, Canada.

[92] Jelena Kovacevic and M. Vetterli (2001), "Transform coding: Past, present and future", IEEE Signal Processing Magazine, vol. 18, pp. 1, September 2001.

[93] Jeonghun Yang, Hyuk Choi, et al. (2000), "Noise estimation for blocking artifacts reduction in DCT coded images", IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, no.7, pp. 116-1120, October 2000.

[94] Jian Sun, Nan-Ning Zheng, et al. (2003), "Stereo matching using belief propagation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no.7, pp. 787-800, July 2003.

[95]   Jin Liu and D. Przewozny (1998), "Stereo image segmentation using hybrid analysis technique", Workshop on Non-Linear Model Based Image Analysis, Glasgow, UK, pp. 1-4.

[96]   Jizheng Xu, Zixiang Xiong, et al. (2000), "High performance wavelet-based stereo image coding", IEEE International Symposium on Circuits and Systems, Scottsdale, Arizona, USA, May 2000.

[97]   Joel Jung, Marc Antonini, et al. (2003), "Optimal decoder for block-transform based video coders", IEEE Transactions on Multimedia, vol. 5, no. 2, pp. 1-16, June 2003.

[98]   Johanson, M. (2001), "Stereoscopic video transmission over the Internet", Proceedings of the Second IEEE Workshop on Internet Applications (WIAPP'01), San Jose, California, USA.

[99]   John G. Proakis and D. G. Manolakis (1992), Digital Signal Processing: Principles, Algorithms, and Applications, Macmillan Publishing Company.

[100]   John Y.A. Wang and E. H. Adelson (1994), "Representing moving images with layers", IEEE Transactions on Image Processing Special Issue, vol. 3, no.5, pp. 625- 638, May 1994.

[101]   Jon K. Rogers and P. C. Cosman (1998), "Wavelet zerotree image compression with packetization", IEEE Signal Processing Letters, vol. 5, no.5, pp. 105-107, May 1998.

[102]   Jonathan K. Su and R. M. Mersereau (1997), "Non interactive rate constrained motion estimation for OBMC", Proceedings IEEE ICSP '97, pp.   33-36, October 1997.

[103]   Jones, P.W, et al. (1995), "Comparative study of wavelet and DCT decompositions with equivalent quantization and encoding strategies for medical images", Proceedings of the Society for Photographic Instrumentation Engineers: Medical Imaging, Bellingham, WA, vol. 2431, pp. 571-582.

[104]   Jong-Nam Kim and T.-S. Choi (1998), "A fast three-step search algorithm with minimum checking points using unimodal error surface estimation", IEEE Transactions on Consumer Electronics, vol. 44, no.3, pp. 638-648, August 1998.

[105]   K. Shen and E. J. Delp (1999), "Wavelet based rate scalable video compression", IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, no.1, pp. 105-122, February 1999.

[106]   Kalman Cinkler and A. Mertins (1996), "Coding of digital video with the edge-sensitive discrete wavelet transform", Proceedings of IEEE International

Conference on Image Processing, ICIP'96, Lausanne, Switzerland, September 1996.

[107]   Kanade, T. (1995), "Development of a video rate stereo machine", Proceedings of International Robotics and Systems Conference (IROS 95), Pittsburgh, PA, August 1996.

[108]   Karsten Muhlmann, Dennis Maier, et al. (2002), "Calculating dense disparity maps from color stereo images, an efficient implementation", International Journal of Computer Vision, vol. 47, pp.79-88.

[109]   Ken Perlin, Salvatore Paxia, et al. (2000), "An autostereoscopic display", Proceedings of the 27th Annual Conference on Computer Graphic and Interactive Technique, pp. 319-326.

[110]   Konrad, J. (1999), "Enhancement of viewer comfort in stereoscopic viewing: parallax adjustment", Proceedings of SPIE/IST symposium on electronic imaging, stereoscopic displays and virtual reality system, pp. 179-190.

[111]   Kyung-Nam Park, Gun-Woo Lee, et al. (2002), "Blocking artifacts reduction in block-coded image using interpolation and SAF based on edge map", The 2002 International Technical Conference on Circuits/Systems, Computers and Communications, Phuket, Thailand, pp. 1107-1011, July 2002.

[112]   Kyung-tae Kim, Mel Siegel, et al. (1998), "Synthesis of a high-resolution 3D-stereoscopic image pair from a high-resolution monoscopic image and a low-resolution depth map", Proceedings SPIE Stereoscopic Display and Virtual Reality Systems V, San Jose, VA, USA, pp. 76-86, January 1998.

[113]   L. Gooding, M.E. Miller, et al. (1991), "The effect of viewing distance and disparity on perceived depth", SPIE: Stereoscopic Displays and Application II, vol. 1457, pp. 259-266.

[114]   L. Liu and C. M. Schor (1994), "The spatial properties of binocular suppression zone", Vision Research, vol. 34, no.7, pp. 937-947.

[115]   L.B.Stelmach, W.J. Tam, et al. (2000), "Human perception of mismatched stereoscopic 3D inputs", ICIP'00, Vancouver, BC, Canada.

[116]   Lacotte, B. (1995), "Elimination of keystone and crosstalk effects in stereoscopic video", Technical report 95-31. Quebec, INRS Telecommunications.

[117]   Lee, T. S. (1996), "Image representation using 2D Gabor wavelets", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no.10, pp. 1-13, October 1996.

[118] Lew B. Stelmach and W. J. Tam (1998), "Stereoscopic image coding: Effects of disparate image-quality in left- and right-eye views", Signal Processing: Image Communication, vol. 14, pp.111-117.

[119] Li, X. (2003), "On exploiting geometric constraint of image wavelet coefficients", IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 11, pp. 1378-1387, November 2003.

[120] Li, Xin (2004), "Scalable video compression via over complete motion compensated wavelet coding", Signal Processing: Image Communication, vol. 19, pp. 637-651.

[121] Lin Luo, Feng Wu, et al. (2004), "Advanced motion threading for 3D wavelet video coding", Signal Processing: Image Communication, vol.19, pp. 601-616.

[122] Lionel Oisel, Etienne Memin, et al. (2003), "One-dimensional dense disparity estimation for three-dimensional reconstruction", IEEE Transactions on Image Processing, vol. 12, no. 9, September 2003.

[123] M. Bilal Ahmad and T.-S. Choi (2001), "Edge detection-based block motion estimation", Electronics Letters, vol. 37, no. 1, pp.17-18, January 2001.

[124] M. Charrier, D.S. Cruz, et al. (1999), "JPEG2000, the next millennium compression standard for still images", IEEE International Conference on Multimedia Computing and Systems, vol. 1, pp. 1-3, June 1999.

[125] M. Pommeray, Jean-Claude Kastelik, et al. (2003), "Image crosstalk reduction in stereoscopic laser-based display systems", Electronic Imaging, vol.12, no.4 pp. 689-697, October 2003.

[126] M. Vetterli and C. Herley (1992), "Wavelets and filter banks: Theory and design", IEEE Transactions on Signal Processing, vol. 40, no. 9, pp.2207-2232, September 1992.

[127] M.D.Adams (2000), "JasPer project homepage".
(http://www.ece.ubc.ca/~mdadams/jasper)

[128] M.E.Lukacs (1986), "Predictive coding of multi-viewpoint images sets", Proceedings on IEEE ICASSP '86, Tokyo, Japan, vol. 1, pp. 521-524, April 1986.

[129] M.G. Strintzis and S. Malassiotis (1999), "Object-based coding of stereoscopic and 3D image sequence", IEEE Signal Processing Magazine, pp.14-28, May 1999.

[130] M.K. Mandal, E. Chan, et al. (1996), "Multiresolution motion estimation techniques for video compression", Optical Engineering, vol.35, no.1, pp. 128-136.

[131] M.W. Siegel, V.S. Grinberg, et al. (1995), "Software for 3D-TV and 3D-stereoscopic workstations", International Workshops on Stereoscopic and Three- Dimensional Imaging 95, Thessaloniki, Greece.

[132] Majani, E. (1994), "Biorthogonal wavelets for image compression", SPIE Symposium on VCIP'94, Chicago, IL, USA, vol. 2308, pp. 478-488.

[133] Mallat, S. G. (1989), "A theory for multiresolution signal decomposition: The wavelet representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 674-693, July 1989.

[134] Mallat, S. G. (1998), A wavelet tour of signal processing, San Diego, Academic Press.

[135] Malvar, H. S. (1999), "Fast progressive wavelet coding", IEEE DCC 99, Snowbird, Utah, USA, pp. 336-344.

[136] Marco Accame, Francesco G.B. De Natale, et al. (1995), "Hierarchical block matching for disparity estimation in stereo sequences", Proceedings of IEEE ICIP'95, Washington, USA, vol. 2, pp. 374-377, October 1995.

[137] Marcus J. Nadenau, Julien Reichel, et al. (2003), "Wavelet-based color image compression: Exploiting the contrast sensitivity function", IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no.1, pp. 58-70, January 2003.

[138] Mark S. Moellenhoff and M. W. Maier (1998), "Transform coding of stereo image residuals", IEEE Transactions on Image Processing, vol. 7, no.6, pp. 804-812, June 1998.

[139] Mark S. Moellenhoff and M. W. Maier (1998), "Characteristics of disparity-compensated stereo image pair residuals", Signal Processing: Image Communication, vol. 14, pp. 55-69.

[140] Masatoshi Okutomi, Yasuhiro Katayama, et al. (2002), "A simple stereo algorithm to recover precise object boundaries and smooth surfaces", International Journal of Computer Vision, vol. 47, pp. 261-273.

[141] Mel Siegel and S. Nagata (2000), "Just enough reality: Comfortable 3D viewing via microstereopsis", IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, no. 3, pp. 387-396, April 2000.

[142] Mel Siegel, Sriram Sethuraman, et al. (1997), "Compression and interpolation of 3D-stereoscopic and multi-view video", SPIE IS&T Conference, San Jose, California, vol. 3012, pp. 227-238, February 1997.

[143] Meriem Mazri, Amar Aggoun (2003), "Compression of 3D integral images using wavelet decomposition", Proceedings of SPIE- The International Society for Optical Engineering, vol. 5150, pp. 1181-1192.

[144] Michael M. Chang, A. Murat Tekalp, et al. (1997), "Simultaneous motion estimation and segmentation", IEEE Transactions on Image Processing, vol.6, no.9, pp.1326-1333, September 1997.

[145] Michael T. Orchard and G. J. Sullivan (1994), "Overlapped block motion compensation: An estimation-theoretic approach", IEEE Transactions on Image Processing, vol. 3, no.5, pp. 693-699, September 1994.

[146] Michael Unser and T. Blu (2003), "Mathematical properties of the JPEG2000 wavelet filters", IEEE Transactions on Image Processing, vol.12, no.9, pp. 1080-1090, September 2003.

[147] Michael W. Marcellin, Margaret A. Lepley, et al. (2002), "An overview of quantization in JPEG 2000", Signal Processing: Image Communication, vol. 17, pp. 73-84.

[148] Michel Misiti, Yves Misiti, et al. (1996), Wavelet Toolbox for use with MATLAB, The MathWorks Inc.

[149] Mi-Hyun Kim and K.-H. Sohn (1999), "Edge-preserving directional regularization technique for disparity estimation of stereoscopic images", IEEE Transactions on Consumer Electronics, vol. 45, no.3, pp. 804-811, August 1999.

[150] Minglun Gong and Y.-H. Yang (2002), "Genetic-based stereo algorithm and disparity map evaluation", International Journal of Computer Vision, vol. 47, pp.63-77.

[151] Mohamed Ben Slima, Janusz Konrad, et al. (1997), "Improvement of stereo disparity estimation through balanced filtering: The sliding-block approach", IEEE Transactions on Circuits and Systems for Video Technology, vol. 7, no.6, pp.913-919, December 1997.

[152] Mulcahy, C. (1997), "Image compression using the Haar wavelet transform", Spelman College Science & Mathematics Journal,vol.1,no.1,pp.22-31, April 1997.

[153] Nam Chul Kim, Ick Hoon Jang, et al. (1998), "Reduction of blocking artifacts in block-coded images using wavelet transform", IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, no.3, pp. 253-257, June 1998.

[154] Nicholas J. Wade and M. T. Swanston (2001), Visual Perception, Taylor & Francis Inc.

[155] Nikolaos V. Boulgouris and M. G. Strintzis (2002), "A Family of wavelet-based stereo image coders", IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no.10, pp. 898-902, October 2002.

[156] Nikos Grammalidis and M. G. Strintzis (1998), "Disparity and occlusion estimations in multiocular systems and their coding for the communications of multiview image sequences", IEEE Transactions on Circuits and System for Video Technology, vol. 8, no. 3, pp. 328-344, June 1998.

[157] Pamela C. Cosman, Robert M. Gray, et al. (1996), "Vector quantization of image subbands: A survey", IEEE Transactions on Image Processing, vol. 5, no. 2, pp.202-225, February 1996.

[158] Pan, H.-P. (1996), "General stereo image matching using symmetric complex wavelets", SPIE Proceedings: Wavelet Applications in Signal and Image Processing IV, Denver, USA, vol. 2825, pp. 1-24, August 1996.

[159] Pastoor, S. (1993), "Human factors of 3D displays in advanced image communications", Displays, vol. 14, pp. 150-157.

[160] Paul Angel and C. Morris (2000), "Analyzing the Mallat wavelet transform to delineate contour and textural features", Computer Vision and Image Understanding, vol. 80, pp. 267-288.

[161] Perkins, M. G. (1992), "Data compression of stereo pairs", IEEE Transactions on Communication, vol. 40, pp. 684-696, April 1992.

[162] Peter Schelkens, Adrian Munteanu, et al. (2003), "Wavelet Coding of Volumetric Medical Datasets", IEEE Transactions on Medical Imaging, vol. 22, no. 3, pp.441-458, March 2003.

[163] Phil Surman, Ian Sexton, et al. (2003), "Beyond 3D television: The multi-modal, multi-view TV system of the future", 12th International Symposium Advanced Display Technologies, Korlev, Russia, pp. 208-210, August 2003.

[164] Ping An , Zhaoyang Zhang, et al. (2001), "Theory and experiment analysis of disparity for stereoscopic image pairs", Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, vol. pp. 68-71, May 2001.

[165] Po-Cheng Wu and L.-G. Chen (2001), "An efficient architecture for two-dimensional discrete wavelet transform", IEEE Transactions on Circuits and System for Video Technology, vol. 11, no.4, pp. 536-545, April 2001.

[166] Po-Cheng Wu and L.-G. Chen (1998), "A block shifting method for reduction of blocking effects in subband/wavelet image coding", IEEE Transactions on Consumer Electronics, vol. 44, no. 1, pp. 170-177, February 1998.

[167] Q. Jiang, J.J. Lee, et al. (1999), "A wavelet based stereo image coding algorithm", Proceedings of ICASSP '99, vol. 6, pp. 3157-3160, March 1999.

[168] Rahul Shukla and H. Radha (2003), "Disparity segmentation based stereo image coding", IEEE International Conference on Image Processing, Barcelona, Spain, September 2003.

[169] Rajesh Rajagopal, Ephraim Feig, et al. (1998), "Motion optimization of ordered blocks for overlapped block motion compensation", IEEE Transactions on Circuit and Systems for Video Technology, vol. 8, no.2, pp.119-123, April 1998.

[170] Reginald L. Lagendijk, Frank Bosveld, et al. (1996), Subband Video Coding. Subband and Wavelet Transforms: Design and Application, Kluwer Academic Publishers, pp. 251-280.

[171] Reinhard Borner, Bernd Duckstein, et al. (2000), "A family of single-user autostereoscopic displays with head-tracking capabilities", IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, no.2, pp. 234-243, March 2000.

[172] Renato Pajarola and P. Widmayer (2000), "An image compression method for spatial search", IEEE Transactions on Image Processing, vol. 9, no. 3, pp. 357-365.

[173] Ricardo de Queiroz, C.K. Choi, et al. (1997), "Wavelet transforms in a JPEG-Like image coder", IEEE Transactions on Circuits and Systems for Video Technology , vol.7, no.2, pp. 419-424, April 1997.

[174] Robert Sekuler and R. Blake (2002), Perception, McGraw-Hill.

[175] Ronald A. Devore, Bjorn Jawerth, et al. (1992), "Image compression through wavelet transform coding", IEEE Transactions on Information Theory, vol. 38, no. 2, pp.719-746, March 1992.

[176] Ronald A. Devore and B. J. Lucier (1992), Wavelet. Acta Numerica. A. Iserles, Cambridge University Press, pp. 1-56.

[177] S. Pastoor and M. Wopking (1997), "3-D displays: A review of current technologies", Displays, vol.17, pp.100-110.

[178] S.Vassiliadis, E.A. Hakkennes, et al. (1998), "The sum-absolute-difference motion estimation accelerator", 24th EUROMICRO'98, vol. 2, pp. 559-566.

[179] Saha, S. (2001), "Image compression- from DCT to wavelet: A review", ACM Crossroad Student Magazine, January 2001. (www.acm.org/crossroad/xrds6-3/sahaimgcoding.html)

[180] Salomon, D. (2000), Data Compression: The Complete Reference, Springer-Verlag.

[181] Sami Brandt and J. Heikkonen (2001), "Multi-resolution matching of uncalibrated images utilizing epipolar geometry and its uncertainty, IEEE International Conference on Image Processing (ICIP 2001), Thessaloniki, Greece.

[182] Sang Hwa Lee, Yasuaki Kanatsugu, et al. (2001), "Object segmentation in stereo image using cooperative line field in stochastic diffusion", International Conference on Image Processing (ICIP'01), Thessaloniki, Greece.

[183] Sang-Hoon Seo and M. R. Azimi-Sadjadi (2001), "A 2-D filtering scheme for stereo image compression using sequential orthogonal subspace updating", IEEE Transactions on Circuits and System for Video Technology, vol. 11, no.1, pp. 52-66, January 2001.

[184] Savage, N. (2003), "Reaching New Depth", SPIE's OE, vol. 3, pp. 37-39, October 2003.

[185] Shapiro, J. M. (1993), "Embedded image coding using zerotrees of wavelet coefficients", IEEE Transactions on Signal Processing, vol. 41, no.12, pp. 3445-3462.

[186] Simon Baker, Terence Sim, et al. (2003), "When is the shape of a scene unique given its light-field: A fundamental theorem of 3D vision?", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no.1, pp. 100-109, January 2003.

[187] Sotiris Malassiotis and M. G. Strintzis (1997), "Object-based coding of stereo image sequences using three-dimensional models", IEEE Transactions on Circuits and Systems for Video Technology, vol. 7, no. 6, pp. 892-905, December 1997.

[188] Sriram Sethuraman, Angel G. Jordan, et al. (1994), "Multiresolution based hierarchical disparity estimation for stereo image pair compression",

Proceedings of the Symposium on Application of Subbands and Wavelets, Newark, New Jersey, USA, pp. 1-7.

[189] Sriram Sethuraman, M.W.Siegel, et al. (1995), "A multiresolution region based segmentation scheme for stereoscopic image compression", SPIE, Digital Video Compression, vol. 2419.

[190] Stanley Coren, Lawrence M. Ward, et al. (1999), Sensation and Perception, Harcourt Brace College Publishers.

[191] Stathis Panis, Robert Kutka, et al. (1997), "Reduction of block artifacts by selective removal and reconstruction of the block borders", Proceedings 1997 Picture Coding Symposium, Berlin, Germany, pp. 705-708.

[192] Stephane G. Mallat and W. L. Hwang (1992), "Singularity detection and processing with wavelets", IEEE Transactions on Information Theory, vol. 38, no.2, pp. 617-643, March 1992.

[193] Steven Dewitte and J. Cornelis (1997), "Lossless integer wavelet transform", IEEE Signal Processing Letters, vol. 4, no.6, pp.158-160.

[194] Strang, G. (1994), "Wavelets", American Scientist, vol. 83, pp. 250-255, April 1994.

[195] Takeshi Maemura, Masahide Kaneko, et al. (1999), "Compression and representation of 3-D images", IEICE Transactions on Information & System, vol. E82-D, pp. 558-567, March 1999.

[196] Tamas Frajka and K. Zeger (2003), "Residual image coding for stereo image compression", Optical Engineering, vol.42, no.1, pp. 1-8, January 2003.

[197] Tieh-Yuh Chen, Alan Conrad Bovik, et al. (1999), "Stereoscopic ranging by matching image modulations", IEEE Transactions on Image Processing, vol. 8, no. 6, pp.785-796, June 1999.

[198] Tzovaras Dimitrios, Nikos Grammalidis, et al. (1998), "Disparity field and depth map coding for multiview 3D image generation", Signal Processing: Image Communication, vol.11, pp. 205-230.

[199] Usery, E. L. (2003), "Autostreoscopy - Three-dimensional visualization solution or myth?", UCGIS Workshop: Visualization White Papers.

[200] Usevitch, B. E. (2001), "A tutorial on modern lossy wavelet image compression: Foundations of JPEG 2000", IEEE Signal Processing Magazine, vol. 18, pp. 22-35, September 2001.

[201] V.Baskaran and K. Konstantinides (1997), Image and Video Compression Standards: Algorithms and Applications, Kluwer Academic Publishers.

[202] Vetterli, M. (2001), "Wavelets, approximation, and compression", IEEE Signal Processing Magazine, vol. 18, pp. 59-73, September 2001.

[203] Vetterli, M. (1995), Wavelets and subband coding, Engelwood Cliffs, N.J, Prentice Hall.

[204] Victor S. Grinberg, Gregg Podnar, et al. (1994), "Geometry of binocular imaging", Proceedings Stereoscopic Displays and Virtual Reality System, San Jose, California, USA, vol. 2177, pp. 56-65, February 1994.

[205] W. -H. Kim and S.-W. Ra (1999), "Fast disparity estimation using geometric properties and selective sample decimation for stereoscopic image coding", IEEE Transactions on Consumer Electronics, vol. 45, no.1, pp. 203-209, February 1999.

[206] W. LI and E. Salari (1993), "Efficient coding method for stereo image pairs", SPIE, vol. 2094, pp. 1470-1476.

[207] W.B. Pennebaker and J. L. Mitcell (1993), JPEG: Still Image Data Compression Standard, Van Nostrand Reinhold.

[208] W.D. Reynolds and R. V. Kenyon (1996), "The wavelet transform and the suppression theory of binocular vision for stereo image compression", 3rd IEEE International Conference on Image Processing, Lausanne, Switzerland, pp. 16-19, September 1996.

[209] Wa James Tam and L. B. Stelmach (1997), "Perceived image quality of MPEG-2 stereoscopic sequences", Electronic Imaging: Science and Technology, San Jose, California, vol. 3016, pp. 296-301, February 1997.

[210] Walker, J. S. (2000), "Lossy image codec based on adaptively scanned wavelet difference reduction", Optical Engineering, vol. 39, no.7, pp. 1891-1897.

[211] Walker, J. S. (1999), A Primer on Wavelets and their Scientific Applications, Chapman & Hall/CRC.

[212] Wen-Jyi Hwang and Y.-C. Huang (2000), "Multistage motion vector quantization for video coding", Optical Engineering, vol. 39, pp.1994-2002, July 2000.

[213] Wilburn, B.S, "A predictive for stereo image compression". (http://ise.stanford.edu/class/ee392/demos/wilburn/).

[214] Woo Woontack and A. Ortega (1996), "Stereo image compression based on the disparity compensation using the MRF model", SPIE VCIP'96, vol. 2727, pp. 28-41.

[215] Woo Woontack and A. Ortega (1999), "Optimal blockwise dependent quantization for stereo image coding", IEEE Transactions on Circuit and Systems for Video Technology, vol. 9, no. 6, pp. 861-868, September 1999.

[216] Woo Woontack and A. Ortega (1998), "Dependent quantization for stereo image coding", SPIE VCIP'98, vol. 3309, pp. 902-913.

[217] Woo Woontack and A. Ortega (1997), "Stereo image compression based on the disparity field segmentation", SPIE VCIP'97, vol. 3024, pp. 391-402.

[218] Woo Woontack, "Rate-distortion based dependent coding for stereo images and video: Disparity estimation and dependent bit allocation", PhD Thesis, University of Southern California, USA, December 1998.

[219] Woods, A. (2000), "Stereoscopic presentations: Taking the difficulty out of 3D", The 6th International Workshop on 3-D Imaging Media Technology, Seoul, Korea, pp. 1-6, July 2000.

[220] Woontack Woo and A. Ortega (2000), "Overlapped block disparity compensation with adaptive windows for stereo image coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, no. 2, pp. 194-200, March 2000.

[221] Woontack Woo and A. Ortega (1999), "Modified overlapped block disparity compensation for stereo image coding", SPIE VCIP'99, San Jose, Ca, USA, vol. 3653, pp. 570-581.

[222] Yiannis Andreopoulos, Andrian Munteanu, et al. (2002), "Wavelet-based fully-scalable video coding with in-band prediction", 3rd IEEE Benelux Signal Processing Symposium, Leuven, Belgium, pp. 1-4, March 2002.

[223] Ying Luo and Rabab K. Ward (2003), "Removing the blocking artifacts of block-based DCT compressed images", IEEE Transaction on Image Processing, vol. 12, no. 7, pp. 838-842, July 2003.

[224] Yong Suk-Kum, Jun-Jae Lee, et al. (1997), "Stereo matching algorithm based on modified wavelet decomposition process", Pattern Recognition, vol. 30, no. 6, pp. 929-952.

[225] Yoshifumi Kitamura, Takashige Konishi, et al. (2001), "Interactive stereoscopic display for three or more users", Proceedings 28th Annual

Conference on Computer Graphic and Interactive Technique, ACM SIGGRAPH, pp. 231-239.

[226] Z. Xiong, K. Ramchandran, et al. (1999), "A comparative study of DCT and wavelet based image coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, pp. 692-695, August 1999.

[227] Zhang, L. (2001), "Hierarchical block-based disparity estimation using mean absolute difference and dynamic programming", International Workshop on Very Low Bit Rate Video Coding, Athen, Greece.

[228] Zhaohui Cai, K.R. Subramaniam, et al. (2001), "Modifications on morphology-based image coding", Electronics Letters, vol. 37, no.7, pp. 421-422, March 2001.

[229] Zhou Wang, Alan C. Bovik, et al. (2002), "Why is image quality assessment so difficult?", IEEE International Conference on Acoustics, Speech, & Signal Processing (ICASSP'02).

[230] Zhou Wang, Alan C. Bovik, et al. (2004), "Image quality assessment: From error measurement to structural similarity", IEEE Transactions on Image Processing 13.

[231] Zixiang Xiong, Kannan Ramchandran, et al. (1997). "Space-frequency quantization for wavelet image coding." IEEE Transactions on Image Processing, vol. 6, no. 5, pp. 677-693.

APPENDIX I

A graph plotting CPU time (y-axis, from 2 to 6) against PSNR reconstructed right image (dB) (x-axis, from 15 to 40), with three curves labeled CODEC III, CODEC IV, and CODEC V.
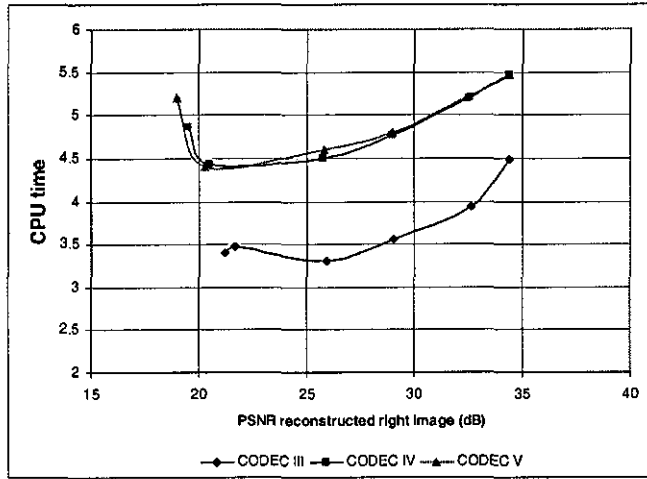
APPENDIX II

The following original contributions have been made. The resulting conference and journal papers are included in Appendix I, and are referenced as A1,A2,...,A8.

A1. **Nayan, M.Y.**, Edirisinghe, E.A. and Bez, H.E., "Two Novel Wavelet-block Based Stereo Image Compression Algorithms", *Proceedings of the Second IASTED International Conference: Visualisation, Imaging and Image Processing*, ACTA Press, 2002, pp. 251-256, ISBN 0-88986-345-3.

A2. **Nayan, M.Y.**, Edirisinghe, E.A. and Bez, H.E., "Baseline JPEG-Like DWT CODEC for Disparity Compensated Residual Coding of Stereo Images", *Proceedings of Eurographics UK*, IEEE Computer Society Press, De Montfort University, Leicester, UK, 2002, pp. 67-74, ISBN: 0-7695-1518-5.

A3. **Nayan, M.Y.**, Edirisinghe, E.A. and Bez, H.E., "Wavelet-based Stereo Image Pair Coding with Pioneering Block Search", *Recent Trends in Multimedia Information Processing*, Word Scientific, Proceedings of the 9th International Workshop on Systems, Signals and Image Processing, 2002, pp. 239-243, ISBN: 981-238-243-7.

A4. Edirisinghe, E.A., **Nayan, M.Y.** and Bez, H.E., "A Pioneering Block Based Stereo Image CODEC in Wavelet Domain", *Proceedings of the SPIE International Conference on Stereoscopic Displays and Applications XIV*, 5006, January 2003, pp.399-406, ISBN: 0-8194-4806-0

A5. Rajkumar, G., **Nayan, M.Y.** and Edirisinghe, E.A., "RASTER: A JPEG-2000 Stereo Image CODEC", *Proc. of the IASTED International Conference on Visualization, Imaging & Image Processing*, Spain, September 2003, pp. 233-238, ISBN: 0-88989-382-2.

A6. **Nayan.M.Y.**, Edirisinghe, E.A. and Bez, H.E., "Transform Domain Overlapped Block Disparity Compensation in Wavelet Coding of Stereo Image Pairs" , *Proceedings of the 2003 joint conference of the 4th International Conference on Information, Communications and Signal Processing and the 4th Pacific-Rim Conference on Multimedia* , IB2.1 , Singapore, December 2003, ISBN: 0-7803-8186-6 .

A7. Edirisinghe, E.A., **Nayan, M.Y.** and Bez, H.E., "A Wavelet Implementation of the Pioneering Block-Based Disparity Compensated Predictive Coding Algorithm for Stereo Image Pair Compression", *International Journal on Signal Processing: Image Communication*, 19(1), January 2004, pp. 37-46, ISSN: 0923-5965,

[WWW] Available from: http://authors.elsevier.com/sd/article/S0923596503001085.

A8. **Nayan, M.Y.**, Edirisinghe, E.A., Bez, H.E., "A Family of Wavelet Transform Based Stereo Image Compression Algorithms", Accepted for publication in Int. Journal. of Computers & Applications.

TWO NOVEL WAVELET-BLOCK BASED STEREO
IMAGE COMPRESSION ALGORITHMS

M.Y.Nayan, E.A.Edirisinghe, H.E.Bez
Department of Computer Science, Loughborough University, UK

Abstract

In this paper we propose two novel stereo image
coding techniques, which use an architecture similar to
that of a Discrete Cosine Transform (DCT) based
baseline JPEG-CODEC [1], but effectively replaces the
DCT technology by the more recently popularized
Discrete Wavelet Transform (DWT) technology. Both
techniques perform disparity compensation in the
wavelet-block transform domain but differ in the actual
strategies used for this purpose. We show that the
proposed CODECs have superior rate distortion and
subjective image quality performance as compared to
DCT based stereo image compression techniques [2,3]
and DWT-JPEG-Like coder proposed by us in [16]. As
compared to these benchmark techniques, we report
peak-signal-to-noise-ratio (PSNR) gains of up to 5 dB
for the proposed CODECs.

Key Words: Stereo Image Compression, Discrete
Wavelet Transforms, JPEG, Disparity Compensation

1. Introduction

The recent advances in auto-stereoscopic display
technology [4-6] have enabled users to experience
stereoscopic vision without the aid of special eyeglasses
or helmet-mounted display kits that often result in
noticeable user discomfort. Currently these new
developments are driving stereo imaging into further
heights by widening its scope to cover a more diverse
application area that includes CAD/CAM, remote
surveillance, medical imaging, tele-medicine, tele-
robotics, HDTV, entertainment and virtual reality.

A stereo image pair consists of two separate views
of a three-dimensional scene captured and/or recorded
by two cameras, one corresponding to the left eye of the
human visual system (left image) and the other to the
right (right image). Unfortunately the presence of two
images as against one image that is sufficient to
describe a two-dimensional scene brings upon a
fundamental problem of data compression. One
straightforward solution to this problem is to use
standard image compression techniques for the
independent coding of the two constituent images.
However in stereo imaging, the presence of inter-frame
redundancy between the stereo image pair and the usual
intra-frame redundancy that is exploited well by
monocular image compression schemes could be

positively used in obtaining further compression. The
most common method used in literature is fixed size
block-based, disparity compensated predictive coding
[2,3], which effectively results in a Predictive Error
(PE) image, that is usually coded using the compression
technique used to independently compress the reference
(say left) image. The coded reference and PE images
are transmitted to the decoder, along with the disparity
vector field, which is used at the decoder to reconstruct
the predicted image based on the reconstructed
reference and PE images. However, in most such
attempts the basic compression engine that is used to
compress the reference and PE images, is the DCT
based baseline-JPEG standard [1]. As DCT based
compression schemes suffer from homogeneous
blocking artifacts at low bit rates, such schemes are not
suitable for low bit rate stereo image coding. However,
they have the advantage of having a simple, well-
established coding architecture, which may be
beneficial in certain application areas.

To cater for the continual expansion of multimedia
and Internet applications, image compression
technologies have to continually improve and evolve in
functionality and efficiency. Worldwide efforts in this
direction have recently led to the standardization of
DWT based JPEG-2000 [7], which is expected to
replace the popular DCT based image compression
technology offered by the baseline-JPEG standard. The
dyadic DWT has versatile time-frequency localization
due to pyramid-like multi-resolution decomposition
architecture and offers improved rate-distortion
performance characteristics as compared to DCT.
Pioneered by the new JPEG-2000 standard, DWT
technology is currently destined to lead the research and
developments efforts in image and video coding fields
at least for the next decade. Therefore it is essential that
efforts be taken to explore the possibility of using DWT
technology in stereo imaging. Few attempts have
already been made in this direction [8-10] within the
stereo image coding research community. These
methods directly attempt to adapt established
monocular DWT based image compression techniques
to stereo imaging.

In this paper we extend our work of [16] and
propose two, novel, wavelet-block based stereo image
coding techniques. As against disparity compensation
in the pixel domain [16], in CODEC-1 we show that
disparity compensation in the wavelet-block domain
produces improved rate distortion performance. In
CODEC-2 we improve this idea further by adapting to

pioneering wavelet-block based prediction, where we modify and apply our DCT based pioneering block based technique of [14,15] in wavelet-block domain. Due to the avoidance of transmitting the disparity vectors under this scheme (CODEC-1), we show that it inherits some special properties in addition to its improved rate-distortion performance as compared to the other similar techniques discussed.

The rest of the paper is arranged as follows. In section 2 we introduce dyadic wavelet decomposition of an image and the transformation of resulting DWT coefficient image to its so-called, *wavelet-block domain*. Section 3 introduces the reader to the two proposed algorithms. Section 4 provides experimental results and analyses the effectiveness of the proposed coders under varying bit rate settings. Finally Section 5 concludes with a foresight to the future research in this direction.

2. Wavelet Decomposition of Stereo Image Pair and Formation of Wavelet Blocks

In applying DWT, the wavelet coefficients are generated by independently applying a cascade of two-channel filter banks to the two constituent images of the stereopair. Figure 1(a) represents the general, three-level wavelet decomposition of a two-dimensional image.

The idea behind *wavelet blocks* is to group the DWT coefficients into blocks as illustrated in figure.1, so that the grouping is similar to that used by a DCT based subband-coding procedure (common spatial locations, different subbands).
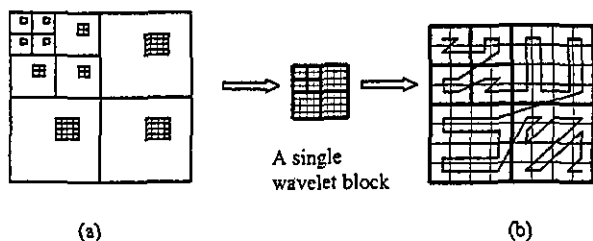

Figure 1. Formation of a single wavelet block from a 3-level decomposed image (a) decomposed image (b) scan order per single wavelet block

The final result is the transformation of the input image into a re-organized structure of original image size, but consisting of a two-dimensional array of non-overlapping wavelet blocks, each representing different subband elements from the location corresponding to the block. In other words, for a S-level DWT, blocks of $2^S \times 2^S$ samples are constructed. Each block is subsequently scanned into a vector in order to be processed by the remaining parts of the baseline JPEG-like coding procedure.

In using DWT filters of more than two taps long, the convolution process results in the formation of additional coefficient values that needs wavelet-block representation in a manner similar to that described above. For an accurate reconstruction of the transformed image, these additional wavelet blocks also need encoding and transmission along the lines of normal wavelet blocks.

3 Proposed Stereo Image Compression Algorithms

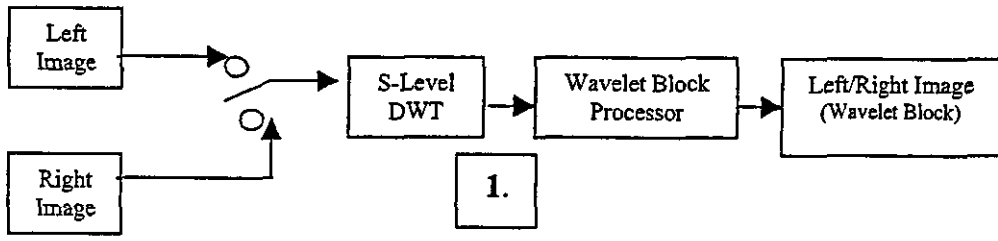### 3.1 CODEC-1: Disparity compensated search in wavlet-block domain

In this algorithm the search for the best match of a given wavelet block of the reference image (left) is done on the right image (the image to be predicted) in wavelet-block domain.

Figure 2 represents the complete block diagram of the proposed CODEC-I. At the encoder end, the left image (reference image) and the right image (predicted image) undergo a 3-level dyadic DWT separately and are subsequently converted into their wavelet-block images (WL and WL respectively) as described in section 2. For each block in WR a search is performed in WL for best matching block, within a horizontally oriented windowed area of maximum likelihood on WL. Note that due to the nature of the wavelet-block domain, the above search is performed with 8 pixel jumps, considered as one unit block shift. The best predictor is found using the Mean Squared Error (MSE) as the selection criteria. The prediction errors of all the non-overlapping right frame wavelet blocks are then used to form the Prediction Error (PE) image, which would be in wavelet-block domain.

Finally, each wavelet block in WL and PE undergo scanning as illustrated in fig. 1(b) as against the zig-zag scanning procedure adapted by baseline-JPEG [1]. This modification of scanning procedure is essential as in the wavelet block representation, for a given block, there are coefficients that not only belong to different subbands, but also multiple coefficients which belong to the same subband.

After the modified scanning procedure discussed above, the ordered coefficients undergo scalar quantization. We use the strategy adapted in [13] to determine a suitable quantization table, that provides results equivalent to the 64-entry, uniform, fixed quantization table used by the baseline-JPEG standard. After the quantization, the quantized coefficients are subjected to entropy coding (runlength/Huffman coding) as in baseline-JPEG.

The decoder operations are essentially the inverse processes of the encoder processes and are clearly illustrated in figure 2.

Figure 2. CODEC I - ENCODER/DECODER

3.2 CODEC-2: Pioneering block search in the wavelet-block domain.

The basic aim of this technique [14, 15] is to avoid the transmission of disparity vectors. The left and right images are first transformed into their respective wavelet-block domains as discussed in section-2.

At the encoder end, for each wavelet-block to be encoded in the right image, a pioneering block, which is the block preceding it, is used to search for the best matching wavelet-block in the reconstructed and re-transformed (to wavelet-block domain) left image. Once the best match is found the block to the right of the best match is selected as the best match for the block to be encoded and the corresponding prediction error block is formed, which is subsequently added to the prediction error image. The detailed block diagram of the encoder of CODEC-2 is shown in figure 3.

Figure 3. CODEC-2: Encoder

At the decoder end, each right image wavelet-block is reconstructed using an identical procedure to that of at the encoder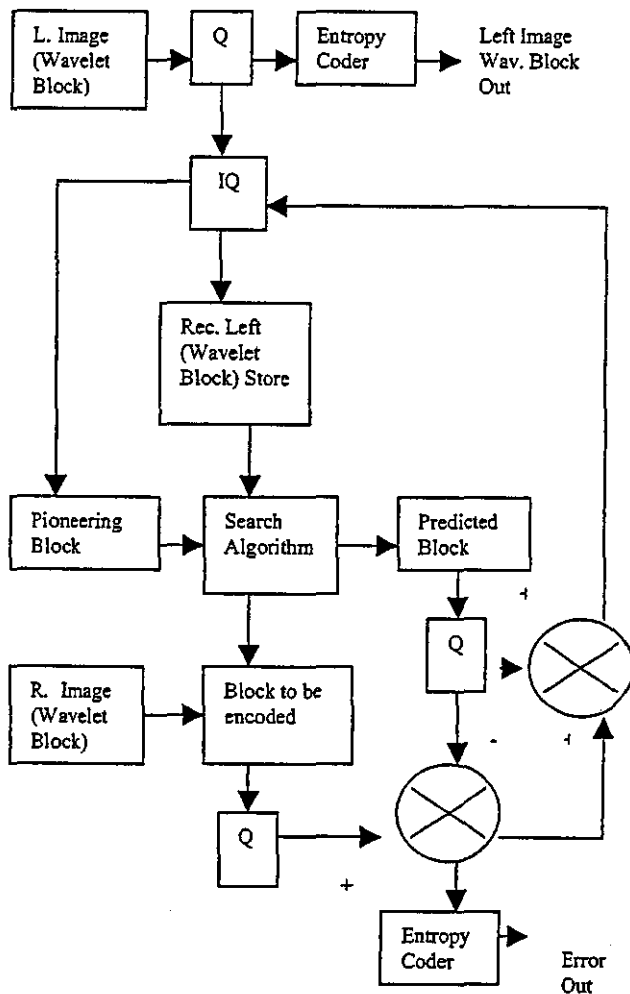. As at the encoder the search is performed for the best match, in a reconstructed and re-transformed left image, at the decoder the search is performed in an identical reference image. In addition

to this, the preceding wavelet block to the block to be decoded, which is used as the pioneering block for the search of the best matching block, is identical to its counterpart at the encoder end, due to the feedback loop shown in figure 3. These guarantee the reconstruction of the predicted image without the need for the transmission of the disparity vectors. Readers interested in more detail of this search technique are referred to our previous papers [14,15]. The savings of the bit-budget that is obtained due to the non-necessity of disparity vector field transmission is expected to improve the CODEC-2's rate distortion performance by considerable margins.

4. Experimental Results & Analysis

The proposed CODECs were implemented using purpose built MATLAB routines and were tested on a set of six commonly used test stereo image pairs representing indoor, outdoor, natural and synthetic scenes. All test stereo image pairs were of size 512×512. Parallel axis camera geometry has been used to acquire these images. Thus for both CODECs a windowed search area representing a disparity range of 0-7 block shifts was used. For CODEC-1, the disparity vector field was coded using a fixed length code, requiring three bits per block. A Daubachies compactly supported orthornormal filter (db7) was used as the wavelet transform.

In order to evaluate the performance of proposed CODECs, we compare their rate-distortion performances, against that of a similar, disparity compensated predictive coding based stereo image coder [2,3] that uses DCT (i.e. baseline- JPEG) and our previously proposed coder, DWT-JPEG-like [16]. The objective image quality is measured in terms of Peak Signal to Noise Ratio (PSNR), whereas the compression efficiency is measured in bits per pixels (bpp). The compression efficiency (or Bit Rate) is measured in terms of bits per pixel (bpp) as follows:

$$BR = \left[\frac{Tot\_Bits_{comp}}{N \times M} \right] \quad bpp$$

where, N and M represent the image dimensions and Tot_Bits_{comp} is the total number of bits required to represent a given image in its compressed format. For benchmark, DWT-JPEG-like and CODEC-1, Tot_Bits_{comp} includes the bits that are required to code the disparity vector field, in addition to the bit requirement to encode and transmit the predictive error image. However, for CODEC-2, Tot_Bits_{comp} represents only the bit requirement to encode the predictive error image. For a fair comparison, the objective quality of the left image

(reference) was set at equivalent levels (41.6 dB) for all CODECs.

For the stereo image pair 'SsCastle', figure 4, compares the subjective image quality of the reconstructed right images when using the proposed and benchmark techniques. Figure 4 shows that at a bit rate of 0.1555 bpp, the reconstructed right images obtained using CODEC-1 and CODEC-2 have better subjective and objective image qualities as compared to that of 4(a) and 4(b). Particularly, the subjective image quality of CODEC-2 outperforms ass others due to the fact that the savings that are obtained by not having to transmit the disparity vector field is used in more accurately coding the prediction error frame. Even though a more accurate subjective quality measurement should involve stereo viewing of the two images (i.e. both left and right) previous research has shown that preservation of subjective image quality of constituent images (i.e. left and right monocular images) near object boundaries, where stereo depth cues are most likely to be of higher magnitude is a vital requirement for comfortable stereo viewing. Thus it could be inferred that the proposed methods would perform subjectively better under stereoscopic viewing conditions.
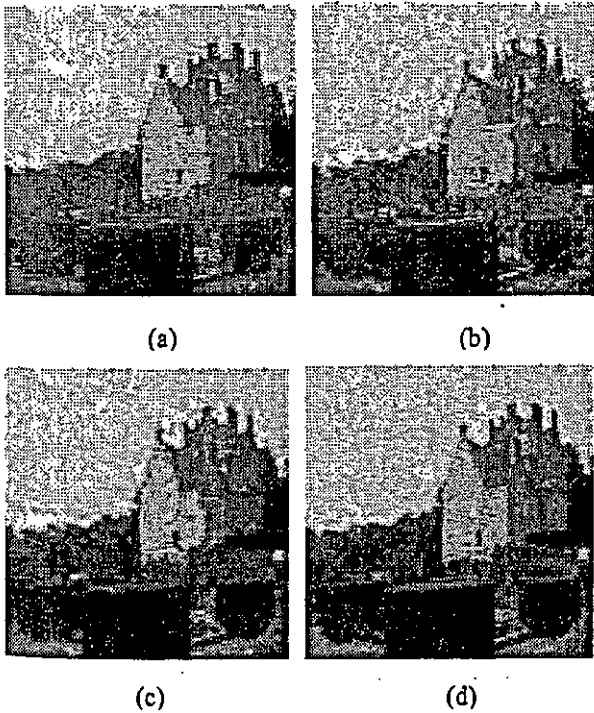

(a)   (b)

(c)   (d)

Figure 4. Subjective quality evaluation for Sscastle: Reconstructed right image of (a) DCT-JPEG, 0.1555 bpp, 26.39 dB (b) DWT-JPEG [16], 0.1555 bpp, 25.87 dB (c) CODEC-1, 0.1555 bpp, 27.18 dB (d) CODEC-2, 0.1555 bpp, 31.4 dB

For further evaluation of the performance of the proposed CODECs, in figure 5, we provide the rate distortion performance graphs for the proposed and benchmark techniques. Figure 5 clearly illustrates the

superiority of the proposed CODECs. Especially CODEC-1 has the ability to perform at bit rates lover that 0.15 bpp, due to the fact that no bits need to be allocated to the disparity vector fields. CODEC-2 shows objective image quality gains of up to 5 dB for most bit rates, whereas CODEC-2 performs relatively better as against the benchmarks at bit rates above 0.4 bpp.
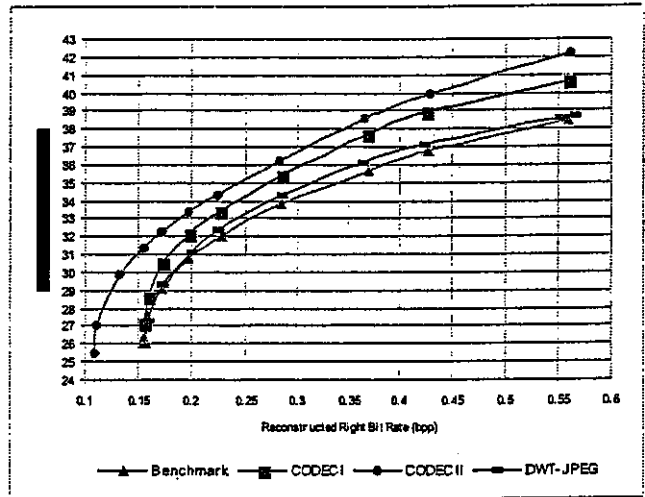
Figure 5. Rate distortion performance graphs

In addition to the above performance improvements achieved by the proposed CODECs, they also provide a simple architecture for DWT based coding of stereo images. As the design is based on the building blocks of a JPEG CODEC, much use could be made of the software, hardware modules that are already available for baseline-JPEG coding in a software/hardware implementation of the proposed CODECs. This could be particularly important at a time image and video coding technology is undergoing a steady change, from DCT based coding schemes to DWT based coding schemes.

5. Conclusions

In this paper we have proposed two, novel, DWT based stereo image CODECs, which use a DCT, based baseline-JPEG like architecture. We have shown that the proposed CODECs have improved rate-distortion performance and subjective image quality performances as compared to the popular DCT based, disparity compensated predictive coding methods [2,3] of stereo images and a similar baseline JPEG-like CODEC proposed by us previously [16]. The proposed CODECs perform better at all bit rates, with PSNR gains of up to 5 dB. In addition to this, the proposed CODECs have the advantage of having a simple architecture that doesn't need image buffering, and also are also able to . produce reconstructed images with less blocking and ringing artifacts.

At present research is being carried out to apply the proposed CODECs and their coding principles in coding stereo video sequences.

Acknowledgements

References

[1] W.B.Pennebaker, J.L.Mitchell, JPEG: Still Image Compression Standard, NY: Van Nostrand Reinhold, 1993.

[2] M.G.Perkins, Data Compression of Stereopairs, *IEE Trans. on Comm.*, 40(4), 1992, 684-696.

[3] I.Dinstein, M.G.Kim, A.Hanik, Compression of Stereo Images Using Subsampling and Transform Coding, *Journal of Optical Engineering*, 30, 1991, 1359-1364.

[4] M.McCormick, Integral 3D imaging for broadcast, *Proc. 2nd Int. Display Workshop*, Hamamatsu, 1995, 77-80.

[5] E.Schulze, Synthesis of moving holographic stereograms with high-resolution spatial light modulators, *Proc. of SPIE 2406*, 1995, 124-126.

[6] D.Takemori, K.Kanatani, S.Kishimoto, S.Yoshii, H.Kanayama, 3-D Display with large double lenticular lens screens, *SID 95 Digest*, 1995, 55-58.

[7] *ISO/IEC JTC 1/SC 29/WG 1, ISO/IEC FCD 15444-1*: Information technology – JPEG-2000 Image Coding System:
Core coding system, 2000, www.jpeg.org/FCD15444-1.htm

[8] S.Sethuramam, M.W.Siegel, A.G.Jordan, A multi-resolution framework for stereoscopic image sequence compression, *Proceedings of the ICIP-95, 2*, 1995, 361-365.

[9] E.A.Edirisinghe, Rate Scalable Stereo Image Coding Using EZW Algorithm, *Proc. of the IEEE Int. Conf. on Information,*
Communications and Signal Proc., Singapore, Oct. 2001

[10] N.V.Boulgouris, M.G.Strintzis, Embedded Coding of Stereo Images, *Proc. of the IEEE Int. Conf. on Image Proc., ICIP 2000*, June 2000.

[11] S.A.Martucci, I.Sodagar, T.Chiang, Y.Zhang, A Zerotree Wavelet Video Coder, *IEEE Trans. on CSVT, 7*(1), 1997, 109-1997.

[12] S.A.Martucci, I.Sodagar, Zerotree Entropy Coding of Wavelet Coefficients for Very Low Bit Rate Video Coding, *Proc. of 1996 IEEE Int. Conf. on Image Proc.*, Lausanne, Switzerland, Sept. 1996.

[13] R. de Queiroz, C.K. Choi, Y. Huh, K.R. Rao, Wavelet Transforms in a JPEG – Like Image Coder, *IEEE Trans. on CSVT, 7*(2), 1997, 419 -424.

[14] J.Jiang, E.A.Edirisinghe, H.Schroder, A Novel Predictive Coding Algorithm for 3-D Image Compression, IEEE Trans. on Cons. Elec., 43(3), 1997, 430-437.

[15] J.Jiang, E.A.Edirisinghe, H.Schroder, Algorithm for the Compression of Stereo Image Pairs, *IEE Electronic Letters, 33*(12), 1997, 1034-1035.

[16] M.Y.Nayan,E.A.Edirisinghe,H.E.Bez, Baseline JPEG-Like DWT CODEC for Disparity Compensated Residual Coding of Stereo Images, *Proc. of Eurographic 2002*, Leicester, UK, 2002, 67-74.

Baseline JPEG-Like DWT CODEC for
Disparity Compensated Residual Coding of Stereo Images

M.Y.Nayan, E.A.Edirisinghe, H.E.Bez

Department of Computer Science, Loughborough University, UK

M.Y.Nayan@lboro.ac.uk

Abstract

In this paper we propose a novel stereo image coding technique, which uses an architecture similar to that of a Discrete Cosine Transform (DCT) based baseline JPEG-CODEC [1], but effectively replaces the DCT technology by the more recently popularized Discrete Wavelet Transform (DWT) technology. We show that as a result of this hybrid design, which combines the advantage of two popular technologies, the proposed CODEC has improved rate distortion and subjective image quality performance as compared to DCT based stereo image compression techniques [2,3]. In particular, at very low bit rates (0.15 bpp), we report peak-signal-to-noise-ratio (PSNR) gains of up to 3.66 dB, whereas at higher bit rates we report gains in the order of 1 dB.

1. Introduction

The recent advances in auto-stereoscopic display technology [4-6] have enabled users to experience stereoscopic vision without the aid of special eyeglasses or helmet-mounted display kits that often result in noticeable user discomfort. Currently these new developments are driving stereo imaging into further heights by widening its scope to cover a more diverse application area that includes CAD/CAM, remote surveillance, medical imaging, tele-medicine, tele-robotics, HDTV, entertainment and virtual reality.

A stereo image pair consists of two separate views of a three-dimensional scene captured and/or recorded by two cameras, one corresponding to the left eye of the human visual system (left image) and the other to the right (right image). Unfortunately the presence of two images as against one image that is sufficient to describe a two-dimensional scene, brings upon a fundamental problem of data compression. One straightforward solution to this problem is to use standard image compression techniques for the independent coding of the two constituent images. However in stereo imaging, the presence of inter-frame redundancy between the stereo image pair and the usual intra-frame redundancy that is exploited well by monocular image compression schemes, could be positively used in obtaining further compression. The most common method used in literature is fixed size block-based, disparity compensated predictive coding [2,3], which effectively results in a Predictive Error (PE) image, that is usually coded using the compression technique used to independently compress the reference (say left) image. The coded reference and PE images are transmitted to the decoder, along with the disparity vector field, which is used at the decoder to reconstruct the predicted image based on the reconstructed reference and PE images. However, in most such attempts the basic compression engine that is used to compress the reference and PE images, is the DCT based baseline-JPEG standard [1]. As DCT based compression schemes suffer from homogeneous blocking artifacts at low bit rates, such schemes are not suitable for low bit rate stereo image coding. However, they have the advantage of having a simple, well-established coding architecture, which may be beneficial in certain application areas.

To cater for the continual expansion of multimedia and Internet applications, image compression technologies have to continually improve and evolve in functionality and efficiency. Worldwide efforts in this direction have recently led to the standardization of DWT based JPEG-2000 [7], which is expected to replace the popular DCT based image compression technology offered by the baseline-JPEG standard. The dyadic DWT has versatile time-frequency localization due to a pyramid-like multi-resolution decomposition architecture and offers improved rate-distortion performance characteristics as compared to DCT. Pioneered by the new JPEG-2000 standard, DWT technology is currently destined to lead the research and developments efforts in image and video coding fields at least for the next decade. Therefore it is essential that efforts be taken to explore the possibility of using DWT technology in stereo imaging. Few attempts have already been made in this direction [8-10] within the stereo image coding research community. These methods directly attempt to adapt established monocular DWT based image compression techniques to stereo imaging.

In this paper we propose the use of *wavelet blocks*, which is the fundamental concept behind DWT based video coding algorithms that are based on the so-called 'zero-trees' [11,12] concept. However in contrast to such attempts made by the video coding community, we propose to preserve and use the simple coding architecture of baseline-JPEG [1], in order to get the best of both, i.e. simplicity and the superior rate-distortion efficiency in coding. By adapting a DWT coding architecture that uses the building blocks of baseline-JPEG [13], the proposed scheme avoids the necessity of buffering the images and multiple passes and thus avoids very complex processing.

The rest of the paper is arranged as follows. In section 2 we introduce dyadic wavelet decomposition of an image and the formation of wavelet blocks using this decomposition. Section 3 contains a comprehensive insight into the proposed coding algorithm. Section 4 provides experimental results and analyses the effectiveness of the proposed coder under varying bit rate settings. Section 5 concludes with a foresight to the future research in this direction.

2. Wavelet Decomposition & Formation of Wavelet Blocks

In applying DWT, the coefficients are generated by applying a cascade of two-channel filter banks to the image. Figure 1(a) represents a three-level wavelet decomposition of an image. Usually in DWT subband coding procedures, the coefficients are grouped to subband oriented groups (common subband, different spatial location) whereas in DCT subband coding procedures, they are grouped into blocks (common spatial locations, different subbands). The idea behind wavelet blocks is to group the DWT coefficients into blocks as illustrated in figure.1, so that the grouping is similar to that used by a DCT based subband coding procedure.

As illustrated in fig.1, the final result is the transformation of the input image into a re-organized structure of original image size, but consisting of a two-dimensional array of non-overlapping wavelet blocks, each representing different subband elements from the location corresponding to the block. In other words, for a S-level DWT, blocks of $2^S \times 2^S$ samples are constructed. Each block is subsequently scanned into a vector in order to be processed by the remaining parts of the baseline JPEG-like coding procedure.

In using DWT filters of more than two taps long, the convolution process results in the formation of additional coefficient values that needs wavelet-block representation in a manner similar to that described above. For an accurate reconstruction of the transformed image, these additional wavelet blocks also need encoding and transmission along the lines of normal wavelet blocks.

3. Proposed Wavelet Based Stereo Image CODEC

Figure 2 represents the block diagram of the proposed CODEC. Note that it is assumed here that the disparity values are suitably coded and transmitted losslessly to the decoder end [2].

At the encoder end, the right image is divided into non-overlapping 8×8 pixel blocks. For each of these blocks a search is performed within a windowed area of maximum likelihood on the original left image (considered as reference image) to find the best predictor. Assuming that parallel axis camera geometry is used to obtain the stereo image pair, one could limit the above mentioned search window in only a horizontal direction, spanning to the left of the corresponding block (of the block to be encoded) in the left image [2].
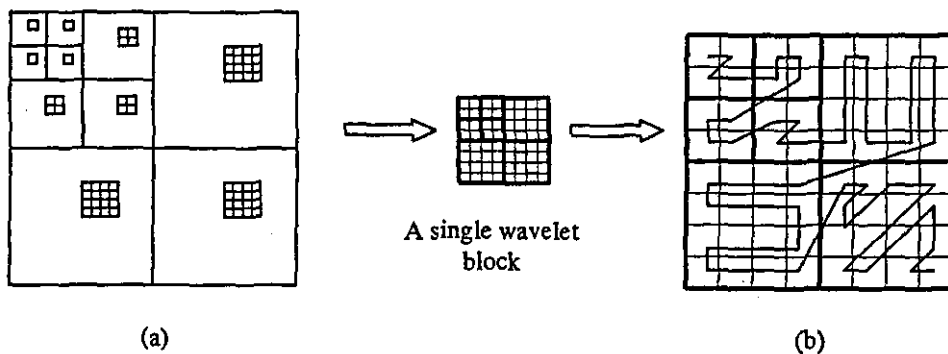

(a)   (b)

Figure 1. Formation of a single wavelet block from a 3-level decomposed image (a) decomposed image (b) scan order per single wavelet block

In the above search procedure the best predictor is found using the Mean Squared Error (MSE) as the selection criteria. The prediction errors of all the non-overlapping right frame blocks are then used to form the Prediction Error (PE) image. The left image, L and the PE image then undergo 3-level dyadic DWT separately and are subsequently converted into their wavelet block representations, by following the procedure described in detail in section 2. Each wavelet block of the two representations then undergo scanning as illustrated in fig. 1(b) as against the zig-zag scanning procedure adapted by baseline-JPEG [1]. This modification of · scanning procedure is necessary as in the wavelet block representation, for a given block, there are coefficients that not only belong to different subbands, but also multiple coefficients which belong to the same subband.

After the modified scanning procedure discussed above, the ordered coefficients undergo scalar quantization. We use the strategy adapted in [13] to determine a suitable quantization table, that provides results equivalent to the 64-entry, uniform, fixed quantization table used by the baseline-JPEG standard (Fig 3).

For easy reference we provide a brief description of this Quantizer design as follows: Let the block size be $N \times N$ $(N = 2^S)$ and the step size be Δ_{ij} for $0 \le (i, j) \le N - 1$.
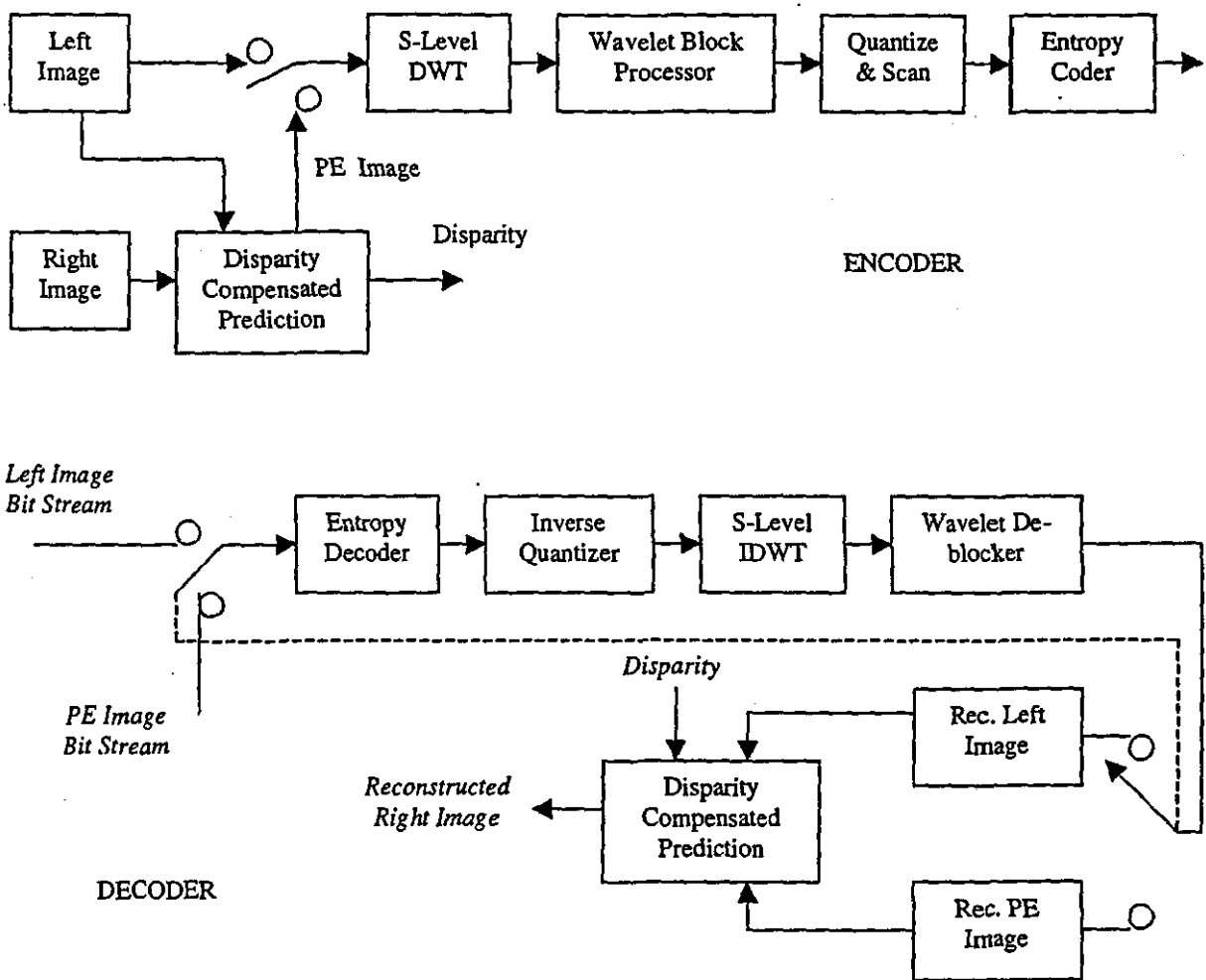

Figure 2. Proposed Wavelet Based Stereo Image CODEC

The step size used in the proposed scheme is then calculated as,

$$\Delta_{ij} = \frac{A}{H\left(\frac{\sqrt{i^2 + j^2}}{N}\right)}$$

where A is a scaling parameter that is used to control the bit rate. The model for $H(x)$ is defined as,

$$H(x) = \left(a_0 + a_1 x + a_2 x^2 + a_3 x^3\right) e^{\left(a_4 x + a_5 x^2\right)}$$

The constant parameters $a_0, a_1 ..., a_5$ are then optimized for $S=3$ (3-levels of decomposition) in such a way as to minimize the error between the achieved table and the luminance quantization table provided by baseline JPEG (Fig.3). The parameters thus found are,

$a_0 = 0.794$, $a_1 = -1.639$, $a_3 = 0.614$, $a_3 = 0.470$,
$a_4 = 4.277$ and $a_5 = -4.892$.

Due to the use of DWT in its block based representation, in a wavelet block there will be several coefficients in the same band and the model developed for the DCT will have different step sizes for different coefficients in a block. In order to solve this problem the Quantizer steps obtained above are averaged, on a per-subband basis and all coefficients in a particular subband are replaced by this average value.

$$Q_{DWT} = \begin{bmatrix} 8 & 7 & 8 & 8 & 34 & 34 & 34 & 34 \\ 7 & 7 & 8 & 8 & 34 & 34 & 34 & 34 \\ 8 & 8 & 12 & 12 & 34 & 34 & 34 & 34 \\ 8 & 8 & 12 & 12 & 34 & 34 & 34 & 34 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \end{bmatrix}$$

$$Q_{JPG} = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

Figure 3. Comparison of Quantizer Tables

Figure 3 compares the Quantizer table obtained with A=6.7 using the strategy described above, with the standard luminance quantizer table of baseline-JPEG.

After the quantization, the quantized coefficients are subjected to entropy coding (runlength/Huffman coding) as in baseline-JPEG. The decoder operations are essentially the inverse processes of the encoder processes and are clearly illustrated in figure 2.

4. Experimental Results & Analysis

The proposed CODEC was implemented using purpose built MATLAB routines and was tested on a set of six commonly used test stereo image pairs representing indoor, outdoor, natural and synthetic scenes. All test stereo image pairs were of size 512×512. Parallel axis camera geometry has been used to acquire these images. Thus a windowed search area representing a disparity range of 0-7 was used. The disparity vector field was coded using a fixed length code, requiring three bits per block. A Daubachies compactly supported orthornormal filter (db7) was used as the wavelet transform.

In order to evaluate the performance of the proposed coder, in figure 4, we compare its rate-distortion performance, against that of a similar, disparity compensated predictive coding based stereo image coder [2] which uses DCT (i.e. baseline-JPEG) as the basic compression engine. This algorithm has been widely used within the stereo image coding research community as a representative benchmark for evaluating novel block based stereo image coding algorithms [16-18]. The objective image quality is measured in terms of Peak Signal to Noise Ratio (PSNR) between the original right image and reconstructed predictive (right) image. The compression efficiency (or Bit Rate) is measured in terms of bits per pixel (bpp) as follows:

$$BR = \left[\frac{Tot\_Bits_{comp}}{N \times M}\right] \quad bpp$$

where, N and M represent the image dimensions and Tot_Bits_{comp} is the total number of bits required to represent a given image in its compressed format. For both benchmark and proposed algorithms, Tot_Bits_{comp} includes the bits that are required to transmit the disparity vector field, in addition to the bit requirement to encode and transmit the predictive error image. Further, for a fair comparison, the objective quality of the left image (reference) was set at equivalent levels (39.39 dB) for both CODECs.
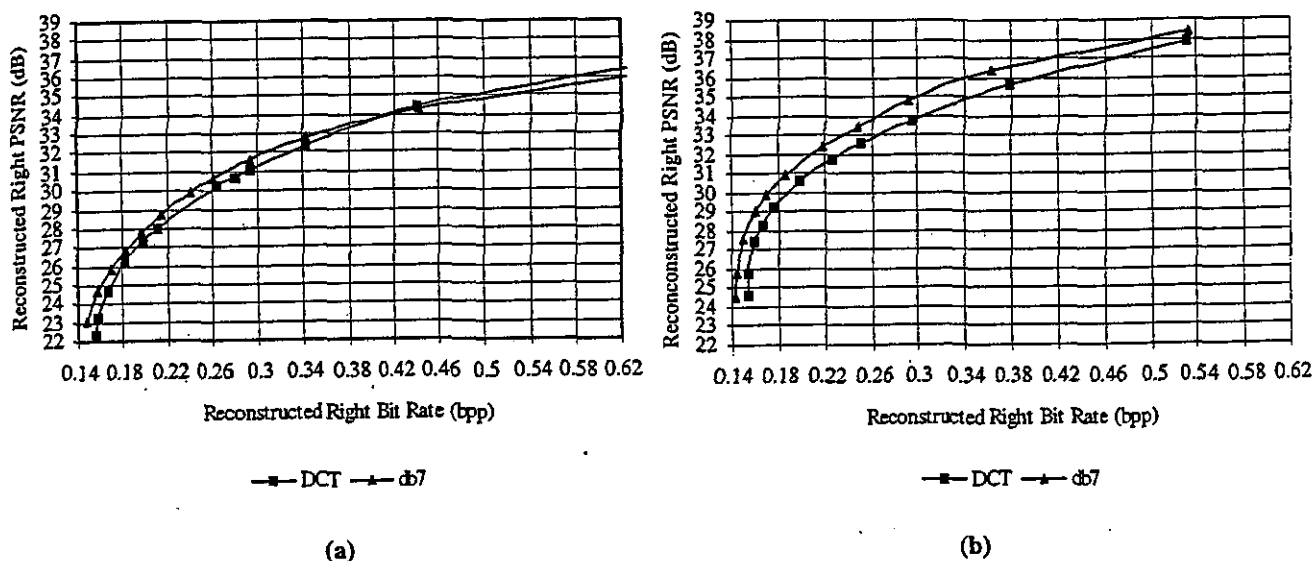
Figure 4. Rate distortion performance for image pair (a) Castle (b) SSCastle

The rate distortion performance graphs of fig. 4 indicate the improvements achieved by the proposed CODEC at all right image bit rates lower than 0.4 bpp for both images. For the SSCastle image, the proposed CODEC provides enhanced performance for right image bit rates of even above 0.5 bpp. Further, for this image, at 0.16 bpp, a PSNR enhancement of 3.66 dB is indicated. Experiments on other test image pairs showed similar performance characteristics.

For the stereo image pair 'Castle', figure 5 compares the subjective image quality of the reconstructed images (at two different bit rates) when using the proposed and benchmark techniques. It illustrates the improved subjective image quality achieved in the predicted image by the proposed CODEC, specifically at very low bit rates. At a bit rate of 0.16 bpp, the reconstructed right image using the benchmark algorithm shows clear shape distortions along the boundaries of most objects, such as the trees at the bottom center of the image, the triangular shaped roofs and the windows in the buildings. However when the proposed technique is used, the overall quality degradations within the same areas are much less and the shapes of the objects are more clearly perceptible. Even though a more accurate subjective quality measurement should involve stereo viewing of the two images (i.e. both left and right) previous research has shown [19] that preservation of subjective image quality of constituent images (i.e. left and right monocular images) near object boundaries, where stereo depth cues are most likely to be of higher magnitude is a vital requirement for comfortable

stereo viewing. Thus it could be inferred that the proposed method would perform subjectively better under stereoscopic viewing conditions.

In order to further evaluate the subjective quality improvements achievable by the proposed algorithm, in figure 6, we compare the reconstructed right images obtained by the two techniques when the SSCastle image is used. It shows subjective quality improvements near object boundaries similar to that illustrated in figure 5 and discussed above.

In addition to the above performance improvements achieved by the proposed CODEC, it also provides a simple architecture for DWT based coding of stereo images. As the design is based on the building blocks of a baseline-JPEG CODEC, much use could be made of the software, hardware modules that are already available for baseline-JPEG coding in a software/hardware implementation of the proposed CODEC. This could be particularly important at a time image and video coding technology is undergoing a change, from DCT based coding schemes to DWT based coding schemes.

In measuring the objective image quality of the reconstructed images we have used the peak-signal-to-noise ratio as the evaluation metric. Similar performances improvements were indicated when comparison metrics such as MSE (mean squared error) and MAD (mean absolute difference) were used, indicating that the proposed wavelet based technique results in the improvement of image quality, on average, over the entire area of the predicted image.
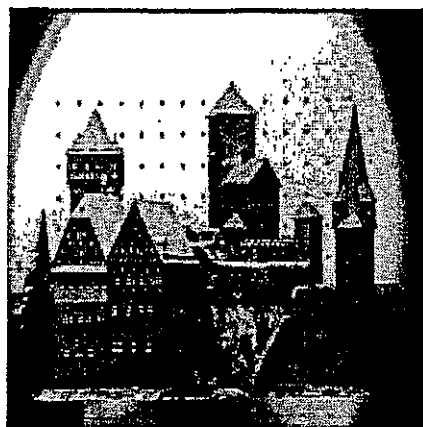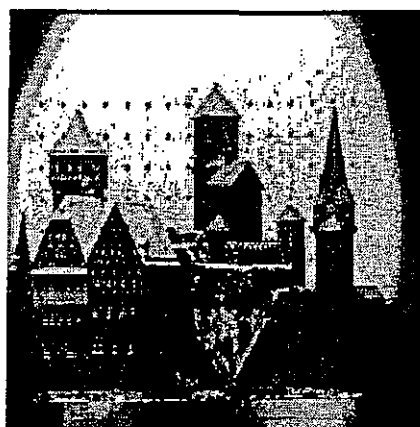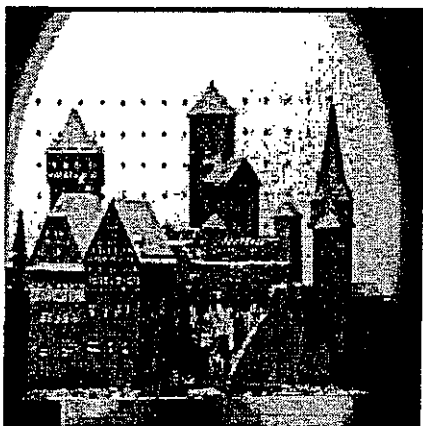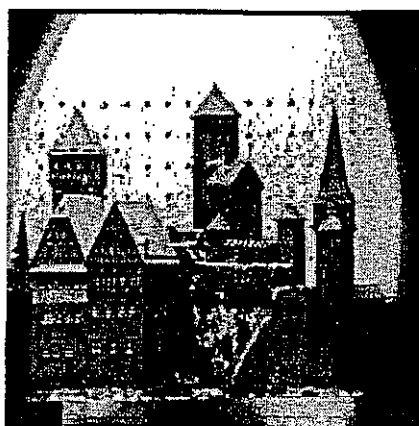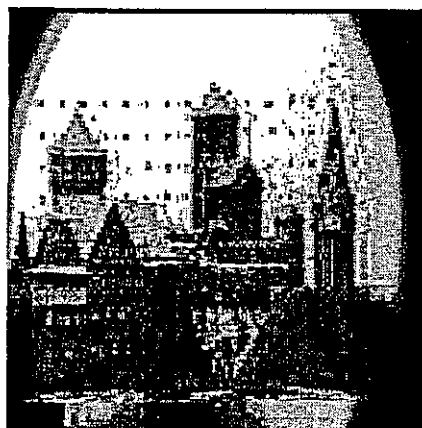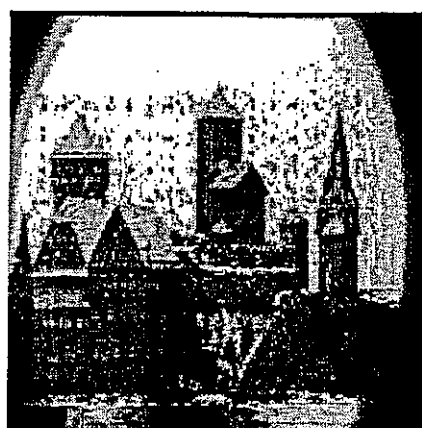
Figure 5. Subjective quality evaluation for Castle (a) original left image
(b) original right image (c) benchmark reconstructed right image, 0.34 bpp, 32.29 dB
(d) proposed coder reconstructed right image, 0.34 bpp, 32.79 dB
(e) benchmark reconstructed right image, 0.16 bpp, 23.31 dB
(f) proposed coder reconstructed right image, 0.16 bpp, 24.81 dB

(a)

(b)
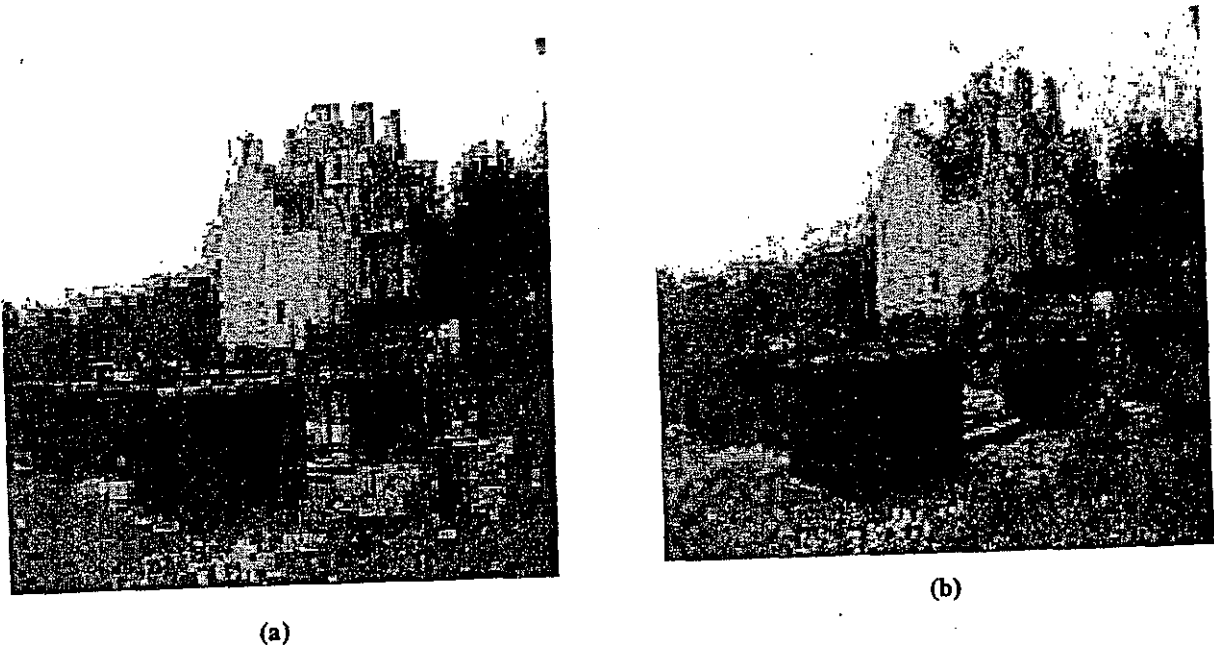
Figure 6. Subjective quality evaluation for SSCastle
(a) benchmark reconstructed right image, 0.1542 bpp, 24.66 dB
(b) proposed coder reconstructed right image, 0.1540 bpp, 28.32 dB

5. Conclusions

In this paper we have proposed a DWT based stereo image CODEC, which uses a DCT based, baseline-JPEG like architecture. We have shown that the proposed CODEC better rate-distortion performance and subjective image quality as compared to the popular DCT based, disparity compensated predictive coding of stereo images. The proposed CODEC performs best at low bit rates, specifically providing PSNR gains of up to 3.66 dB at bit rates of 0.16 bpp. In addition to this, the proposed CODEC has the advantage of having a simple architecture that doesn't need image buffering, and also provides reconstructed images with less blocking and ringing artifacts.

At present further research is being carried out to evaluate the subjective quality improvements of the proposed technique based on stereoscopic viewing conditions. Further research is proposed in the use of different wavelet basis functions, adaptive coefficient scanning and quantization and in applying the proposed technique for object based coding of stereo image pairs. The ideas presented above could also be extended and applied in coding stereo video sequences.

Further research has been successfully carried out to extend the use of the proposed CODEC in pioneering block based disparity compensated predictive coding, introduced by us in [14,15], in order to remove the necessity of transmitting the disparity value as overhead information. This has shown to further improve the rate distortion performance.

6. References

[1]. W.B.Pennebaker and J.L.Mitchell, "JPEG: Still Image Compression Standard", Van Nostrand Reinhold, New York, 1993.

[2]. M.G.Perkins, "Data Compression of Stereopairs", IEE Trans. on Comm., Vol.40, No.4, April 1992, pp..684-696.

[3]. I.Dinstein, M.G.Kim and A.Hanik, "Compression of Stereo Images Using Subsampling and Transform Coding", Journal of Optical Engineering, Vol.30, Sept. 1991,pp. 1359-1364.

[4]. M/McCormick, "Integral 3D imaging for broadcast", Proc. 2nd Int. Display Workshop, Hamamatsu, 1995, pp. 77-80.

[5]. E.Schulze, "Synthesis of moving holographic stereograms with high-resolution spatial light modulators", Proc. SPIE 2406, 1995, pp. 124-126.

[6]. D.Takemori, K.Kanatani, S.Kishimoto, S.Yoshii, H.Kanayama, "3-D Display with large double lenticular lens screens", SID 95 Digest, 1995, pp. 55-58.

[7]. ISO/IEC JTC 1/SC 29/WG 1, ISO/IEC FCD 15444-1: Information technology – JPEG-2000 Image Coding System: Core coding system, March 2000, www.jpeg.org/FCD15444-1.htm

[8]. S.Sethuramam, M.W.Siegel, A.G.Jordan, "A multi-resolution framework for stereoscopic image sequence compression', Proceedings of the ICIP-95, Vol.2, pp. 361-365.

[9]. E.A.Edirisinghe, "Rate Scalable Stereo Image Coding Using EZW Algorithm", Proc. of the IEEE Int. Conf. on Information, Communications and Signal Proc., ICICS 2001, Singapore, Oct. 2001, ISBN: 981-04-5149-0.

[10]. N.V.Boulgouris, M.G.Strintzis, "Embedded Coding of Stereo Images", Proc. of the IEEE Int. Conf. on Image Proc., ICIP 2000, June 2000.

[11]. S.A.Martucci, I.Sodagar, T.Chiang, Y.Zhang, "A Zerotree Wavelet Video Coder", IEEE Trans. on CSVT, Vol.7, No.1, February 1997, pp. 109-1997.

[12]. S.A.Martucci, I.Sodagar, "Zerotree Entropy Coding of Wavelet Coefficients for Very Low Bit Rate Video Coding", Proc. of 1996 IEEE Int. Conf. on Image Proc., Lausanne, Switzerland, Sept. 1996.

[13].R. de Queiroz, C.K. Choi, Y. Huh, K.R. Rao, "Wavelet Transforms in a JPEG – Like Image Coder," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, No. 2, pp 419 -424, April 1997.

[14].J.Jiang, E.A.Edirisinghe, H.Schroder, "A Novel Predictive Coding Algorithm for 3-D Image Compression", IEEE Trans. on Cons. Elec., Vol.43, N0.3, pp. 430-437, August 1997.

[15]. J.Jiang, E.A.Edirisinghe, H.Schroder, "Algorithm for the Compression of Stereo Image Pairs", IEE Electronic Letters, Vol.33, No.12, pp. 1034-1035, June 1997.

[16]. J.Jiang, E.A.Edirisinghe, "A Hybrid Approach for Low Bit-Rate Coding of Stereo Images", IEEE Trans. On Image Proc., Vol. 11, No. 2, pp 123-134, February 2002.

[17]. H.Aydinouglu, H.Hayes, "Stereo image coding: A projection approach," IEEE Trans. On Image Proc., Vol.7, pp. 506-516, April 1998.

[18]. F.Labonte, C.T.Le Dinh, J.Faubert, P.Cohen, "Spatiotemporal Spectral Coding of Stereo Image Sequences, " IEEE Trans. On CSVT, Vol. 9, pp. 144-155, February 1999.

[19]. I.Dinstein et. al., "Compression of stereo images and the evaluation of its effects on 3-D perception", Proc. of Inst. Elect. Eng., 1989.

WAVELET BASED STEREO IMAGE PAIR CODING WITH PIONEERING BLOCK SEARCH

M.Y.NAYAN, E..A.EDIRISINGHE AND H.E.BEZ

Department of Computer Science,
Loughborough University, UK
E-mail: M.Y.Nayan@lboro.co.uk

In this paper, we propose two novel CODECs which use pioneering block search in the wavelet domain to improve the performance of disparity-based stereo image pair compression techniques. The proposed CODEC-1, does the pioneering block search simultaneously for all subband images using wavelet blocks and the proposed CODEC-2, adapt the pioneering block search separately for each subband image. As a result of effectively exploiting both intra and inter frame redundancy, in both CODECs, coding of the disparity vector field is avoided. We show that the proposed CODECs have superior rate-distortion performance and subjective image quality, compared to Discrete Cosine Transform (DCT) based pioneering block technique proposed by us in [1]. As compared to this benchmark, for the proposed CODECs we report peak-signal-to-noise-ratio (PSNR) gains of up to 3 dB.

1. Introduction

In [1] we proposed a pioneering block based disparity compensated predictive coding scheme in DCT domain for stereo image compression, which eliminates the need of coding the disparity vector field but is yet able to reconstruct the predicted image. However a drawback of this technique is the blocking effect at low bit rates, which is an inherent problem in any DCT based coding algorithm. The blocking effects cause poor image quality especially at the intensity and object edges, which are visually important features of the stereo image pair that any compression technique should try to preserve.

To overcome the above problem, we propose two novel CODECs to improve the performance of our earlier work [1]. Both CODECs replace the DCT based compression engine by a DWT based engine, which has versatile time-frequency localization characteristics due to a pyramid-like multiresolution decomposition architecture, thus improving the rate – distortion performance as compared to the use of DCT. The proposed CODECs also follow the current trend in research & development lead by the standardization of JPEG-2000.

2. Overview of Pioneering Block Search.

With reference to our earlier work [1], pioneering block based predictive coding

algorithm is described as follows. The block preceding the block to be encoded is selected as a pioneering block to search a windowed area in the reconstructed left reference frame. When the best match is found the block immediately to the right is taken as the best predictor to the block to be encoded. The encoded prediction errors after transmission are reconstructed back at the encoder end and used with the predicted block to reconstruct each encoded block which in then used as the pioneering block, in the next iteration. This ensures that both images involved in all operations of the pioneering block-based search and prediction will be the same at both ends, which in turn will guarantee the accurate reconstruction of the right image. At the decoder end, with the availability of the reconstructed left frame and the block preceding the block to be decoded, a similar search can be performed at the decoder end to find the best match from the reconstructed left frame. The corresponding error block is added to this best match to obtain the decoded right frame block. Following the same procedure, the right frame would be reconstructed in full on a block-by-block basis without any disparity information.

3.   Proposed CODEC-1 & CODEC-2

3.1. *Pioneering block search using wavelet-blocks: CODEC-1*

At the encoder end, the left image (reference image) and the right image (predicted image) undergo a 3-level dyadic DWT separately and the subband images are subsequently grouped into wavelet blocks such that each block consists of wavelet coefficients from a common spatial location but from different subbands [3-4].

Adapting the technique described in section 2, by replacing each block by a wavelet-block, able the pioneering block search to be performed simultaneously for different subband images. At the decoder, each right image wavelet-block is reconstructed using an identical procedure to that of the encoder

3.2 *Pioneering block search in subband images: CODEC-2*

As in CODEC-1, after the left (reference image) and the right (predicted) images undergo a 3-level dyadic DWT separately, for each image, 10 subband images are formed, which consists of 1 approximation image and 3 detail images (horizontal, vertical and diagonal detail images) at level 3, followed by 3 detail images at level 2 and finally 3 detail images at level 1.

For the left image, subband images are converted into wavelet blocks, and are quantized, entropy coded and transmitted. The subband images of the left image are reconstructed locally at the encoder end to be used as the reference for pioneering block search & prediction. This is to ensure that both encoder and

decoder use the same subband images in predicting and searching for the best match for the subband images of the right image.

With the ten reconstructed subband images of the left image and ten subband images of the right image, a pioneering block search is performed as described in section 2, separately between each corresponding subband image pairs. At the end of the pioneering block search process, ten prediction error subband images are obtained. Subsequently these prediction errors are converted into wavelet blocks, quantized, entropy coded and transmitted.

At the decoder end, the wavelet blocks of the prediction errors and the left image are entropy decoded, de-quantized and are put back into their corresponding subband images. From the ten reconstructed subband images of the left image and ten subband images of the prediction errors, a pioneering block search technique to that of the encoder end is performed at the decoder end. By adding the corresponding prediction error block to the best match in the reconstructed left subband image, the right subband image is reconstructed. This process is repeated until all ten subband images of the right image have been reconstructed. The Inverse Discrete Wavelet Transform (IDWT) is then applied to the reconstructed subband image of the left and right images to reconstruct the left and the right images. Applying the pioneering block separately to each subband image is expected to improve the rate distortion performance by a considerable margin.

4. Experiment Results & Analysis

Both CODECs were implemented using purpose built MATLAB routines and were tested on a set of five commonly used test stereo image pairs representing indoor, outdoor, natural and computer rendered scenes. All test stereo image pairs were of size 512x512. Parallel axis camera geometry has been used to acquire these images, and pixels at the same spatial location are assumed to have the same brightness. Thus for both CODECs a windowed search area representing a disparity range of 0-7 blocks shifts in the horizontal direction only, was used. A Daubachies compactly supported orthornormal filter (db7) was used as the wavelet transform. In order to evaluate the performance of proposed CODECs, we compared their rate-distortion performances against that of DCT based pioneering block based predictive coding algorithm [1] previously proposed by us. The objective image quality is measured in terms of Peak Signal to Noise Ratio (PSNR), whereas the compression efficiency is measured in bits per pixels (bpp) The compression efficiency (or Bit Rate, BR) is measured as follows:

$$BR = \left[\frac{Tot\_Bits_{comp}}{N \times M} \right] \quad bpp$$

where, N and M represent the image dimensions and Tot_Bits_{comp} is the total number of bits required to represent a given image in its compressed format that consists of only the bit requirement to encode the prediction errors.

For the stereo image pair 'Castle', figure 2, compares the subjective image quality of the reconstructed right images obtained when using the proposed CODEC-1, CODEC-2 and the benchmark techniques. Figure 2 shows that at a bit rate of 0.1558 bpp, the reconstructed right images obtained using CODEC-1 and CODEC-2 have better subjective and objective image qualities as compared to that of Figure 2(a) especially around the edges of the castle.

a   .   b   c .
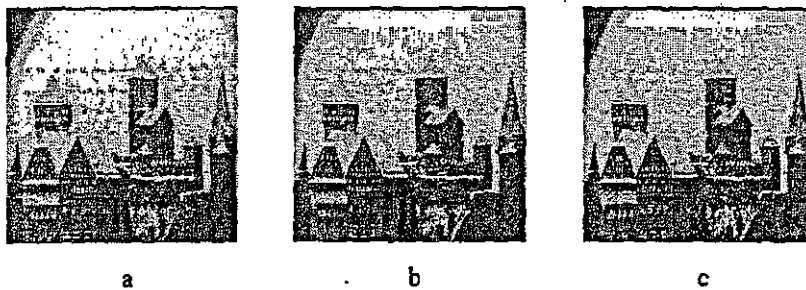
Figure 2. Subjective quality evaluation for Castle reconstructed right image
(a) Benchmark, 0.1558 bpp, 25.94 dB   (b) CODEC-1, 0.1558 bpp, 28.93 dB
(c) CODEC-2, 0.1498 bpp, 29.03 dB

For further evaluation of the performance of the proposed CODECs, in figure 3, we compare the rate-distortion performance graphs for the proposed CODECs and the benchmark. It clearly illustrates the superiority of the proposed CODECs. Both proposed CODECs show improvements as compared to the DCT based implementation of pioneering block search. A gain of up to 3 dB, as observed in figure 3.

5. Conclusion

In this paper, we have proposed two, novel wavelet based stereo image pair compression algorithms, which improve the performance of our previously proposed DCT based, pioneering block based stereo image compression. The proposed CODECs perform better at all bit rates, producing PSNR gains of up to

3 dB. At present research is being carried out to apply the proposed CODECs and their coding principles in coding stereo video sequences.
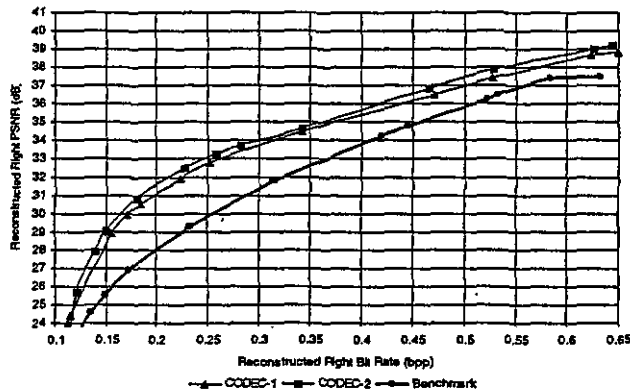
Figure 3. Rate distortion performance graphs

References
[1] J. Jiang, E.A. Edirisinghe, H.Schroder, "A Novel Predictive Coding Algorithm for 3-D Image Compression", IEEE Trans. On Cons. Elec., Vol. 43, No. 3, pp. 430-437, August 1997.
[2] M.G. Perkins, "Data Compression of Stereopairs", IEE Trans. On Comm., Vol 40, No. 4, pp. 684-696, April 1992.
[3] R. de Queiroz, C.K. Choi, Y. Huh, K.R. Rao, "Wavelet Transforms in a JPEG-Like Image Coder," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, No. 2, pp. 419-424, April 1997.
[4] M.Y.Nayan, E.A. Edirisinghe, H.E. Bez, " Baseline JPEG-Like DWT CODEC for Disparity Compensated Residual Coding of Stereo Images", Proceeding of the 20th Eurographics UK Conference 2002, June 2002, pp. 67-74, ISBN 0-7695-1518-5.
[5] E.A. Edirisinghe, "Rate Scalable Stereo Image Coding Using EZW Algorithm", Proceeding of the IEEE Int. Conf. On Information, Communications and Signal Proc., ICICS Oct 2001.
[6] N.V. Boulgouris, M.G. Strintzis, "Embedded Coding of Stereo Image", Proceeding of the IEEE Int. Conf. On Image Processing , ICIP 2000, June 2000.
[7] ISO/IEC JTC 1/SC 29/WG 1, ISO/IEC FCD 15444-1: Information Technology- JPEG-2000 Image Coding System: Core Coding System, March 2000, www.jpeg.org/FCD15444-1htm.
[8] S. Sethuraman, M.W. Seigel, A.G. Jordan, " a multiresolution framework for stereoscopic image sequence compression", Proceeding of the ICIP-95, Vol. 2, pp. 361-365.

Pioneering Block Based Stereo Image CODEC in Wavelet Domain

E.A.Edirisinghe [*], M.Y.Nayan, H.E.Bez

Department of Computer Science, Loughborough University, UK

ABSTRACT

In this paper we propose the wavelet domain implementation of our original pioneering block based stereo image compression algorithm and compare its performance with traditional, DCT based and state-of-the art, DWT based stereo image compression algorithms. Due to the special requirements of the pioneering block based CODEC and the properties of DWT based multi-resolution decomposition, the implementation of the original algorithm in the wavelet domain is not straightforward and thus provides knowledge and understanding of significant novelty. Experiments were performed on a set of eight stereo image pairs representing, natural, synthetic, in-door and out-door images. We show that for the same bit rates, objective quality gains of up to 5 dB (PSNR) are obtained as compared to the benchmark algorithms. One significant property of the proposed CODEC is its ability to produce reconstructed right images of up to 25dB at right image bit rates as low as 0.1 bpp. Significant gains in subjective image quality are also obtained as compared to benchmark methods.

Indexing terms: Stereo image compression, wavelet-blocks, predictive coding and disparity compensation.

1. INTRODUCTION

The recent advancement of auto-stereoscopic display technologies are currently driving stereo imaging into applications domains such as digital television, computer simulations and Internet technologies that have been so far dominated by the state-of-the-art monocular image and video coding techniques. Thus at present, efficient algorithms that are capable of coding and transmitting (storing) stereo images and image sequences are attracting significant research efforts from the stereo imaging research community. To this extent various coding algorithms based on block, object and region based predictive coding techniques have been reported in literature [3-15].

In parallel to the abovementioned shift of technology from monocular to binocular vision capability, the popular, base compression technology that was used in almost all image and video coding standards (baseline-JPEG, MPEGs and H26x), Discrete Cosine Transforms [1], are at present being steadily replaced by DWT based technologies that are capable of providing improved rate-distortion performance and additional functionality [2,13-20]. The culmination of research in this direction resulted in the recent standardization of JPEG-2000 [2] and wavelet domain implementations of MPEG-4 video coding standard. As most stereo image and sequence compression techniques that have been proposed in literature are based on DCT based technologies [3-12], currently there exists a trend in transfer of stereo imaging technology to DWT based techniques. To this effect few wavelet based stereo image compression algorithms have been reported in literature [13-15] during the past few years. Most of these algorithms are a straightforward application of the popular wavelet based monocular image compression algorithms to stereo image compression.

Following the abovementioned trends in technology and research directions, within the present research context, we propose a wavelet-based implementation for the PBDCPC algorithm, which was originally implemented in [5,6] using DCTs as the base technology. The PBDCPC algorithm is known for its ability to work under very low bit rate constraints [6], still producing predicted images of acceptable image quality. This is mainly due to the fact that it avoids the necessity of having to transmit the so-called, disparity vector field as overhead bits. We show that the change of base technology from DCT to DWT further improves its low bit rate performance. However, due to the specific requirements of the PBDCPC coding structure and properties of DWT based multi-resolution decomposition [13,19], the above transfer of base technology is not straightforward and thus involves additional design and development of considerable novelty. The original PBDCPC algorithm is summarized in section 2 for quick reference. Readers interested in its detailed design and behavior are advised to refer to the original work reported in [5,6].

For clarity of presentation, the rest of the paper is organized as follows. Apart from section 1, which is a general introduction to the context of the proposed research, section 2, summarizes the principles and the strict requirements behind the effective implementation of the PBDCPC algorithm. Section 3, discuses the DWT based multi-resolution decomposition of a stereo image pair and the formation of wavelet-block based structure, which forms the foundation for the proposed implementation. Section 4 discusses PBDCPC algorithm's implementation in wavelet domain by first discussing a Multi-Resolution Pioneering Block Search (MRPBS) procedure and subsequently by re-structuring the resulting Predictive Error (PE) image and the reference image within a wavelet-block structure [19,20] for entropy coding and transmission. In section 5, we compare the performance of the proposed CODEC with that of traditional DCT based DCPC schemes [3,4] and a more recently published DWT based DCPC scheme [20] and prove the novel algorithm's improved rate-distortion performance and improved range of performance. Finally in section 6, we conclude with suggestions for further improvements of the proposed algorithms and its possible use in stereo image sequence compression.

2. PIONEERING BLOCK BASED SEARCH (PBS)

Figure 1, illustrates the *pioneering block based search* (PBS) procedure that is used in PBDCPC algorithm [6].
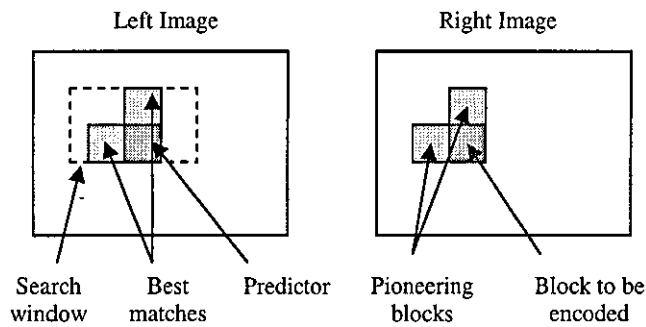
Figure 1. Pioneering block based search procedure

As illustrated in fig. 1, to encode a given block of image pixels (say size 8×8) in the predicted (say right) image, the block preceding it and the block directly above it are taken as *pioneering blocks* to search for a pair of blocks, within a selected windowed area of maximum likely of the locally reconstructed reference (left) image. The best matching pair is found using mean squared error as the minimization criterion and using equal weights for the two blocks. Once the best matching pair is found, the block directly below the upper block is chosen as the predictor for the block to be encoded. Note that at the boundaries of the image where one of the pioneering blocks (or both) is unavailable, the above search procedure is replaced by a PBS with one block (or by direct prediction). Once the predictor, $\hat{L}$ for the block to be encoded, R is found, the prediction error, E is calculated as, $E = R - \hat{L}$. Note the use of $\hat{L}$ to represent the fact that the search is performed on a locally reconstructed left frame. This is the first essential requirement of the PBDCPC coder, as it guarantees that the procedure adopted at the encoder end is identical to that adopted at the decoder end. This is due to the fact that at the decoder end what is available is the reconstructed reference frame. In addition to meeting the above condition, a second condition needs to be met. When the prediction error E is calculated at the encoder end, while it is being encoded and transmitted to the decoder end, the same coded error has to be decoded, and added to the predictor, $\hat{L}$, to produce the preceding pioneering block for the encoding of the next predicted frame block. Thus, always the pioneering block selection in the predicted image is done on a 'so-far' reconstructed image. This satisfies the second condition in guaranteeing that the encoder and decoder processes are identical. Thus in the PBDCPC coder the encoder and decoder work on identical reference frames and selects pioneering blocks from identical, 'so-far' reconstructed predicted frames. This enables the guaranteed reconstruction of the predicted image at the decoder end, without recourse to disparity information. We refer readers interested in the theoretical proof of the above-mentioned quality guaranteeing procedure to [21].

Following the pixel domain PBS procedure, summarized above, the actual prediction (i.e. calculation of prediction error, E) could be either performed in the pixel domain itself, or in the transformed domain. However, in order to satisfy the second condition in guaranteeing the reconstruction of the predicted image, the prediction errors have to go through transform coding/decoding and quantization/de-quantization at the encoder end, thus making it an integral part of the PBDCPC algorithm. Therefore in any attempt to change the base transform technology of PBDCPC algorithm to DWT the above-mentioned requirement takes precedence. Thus, a straightforward block based, pixel domain disparity compensated prediction to obtain the prediction error image in complete, and the subsequent DWT based coding of the same, is not a suitable DWT implementation of the PBDCPC algorithm. This is due to the fact that with such a scheme one is unable to introduce the compression losses that would be introduced to the prediction errors, before they are 'fed-back' to form the pioneering blocks for the next block to be encoded. Thus within our present research context we propose a strategy based on the formation of a wavelet-block coding structure, which removes the major obstacle faced by the abovementioned straightforward strategy.

3. WAVELET DECOMPOSITION OF A STEREO IMAGE PAIR & THE FORMATION OF WAVELET BLOCKS

The first stage of the proposed DWT implementation of the PBDCPC algorithm is the independent wavelet decomposition of the stereo image pair into N levels, forming $3N + 1$ sub-bands. The coefficients of these sub-bands are generated by applying a cascade of two-channel filter banks to the image. Figure 2 illustrates this decomposition structure for a given image, with N=3. We name the lowest resolution band, LL3.

As described above the common practice in DWT subband coding procedures, is to group the coefficients into subband-oriented groups. In other words, coefficients of a given subband, comes from different spatial locations of the image. However in block based, DCT subband coding procedures, the coefficients are grouped into blocks, i.e. they are from common spatial locations, but different subbands. Since the adaptation of a DCT-like (baseline-JPEG like) coding structure would enable the proper design of the PBDCPC algorithm in the wavelet domain, we propose the transformation of the decomposition structure of figure 2 to wavelet-blocks as depicted in figure 3. The idea behind wavelet blocks is to group the DWT coefficients into blocks, so that the grouping is similar to that used by a DCT based subband coding procedure.
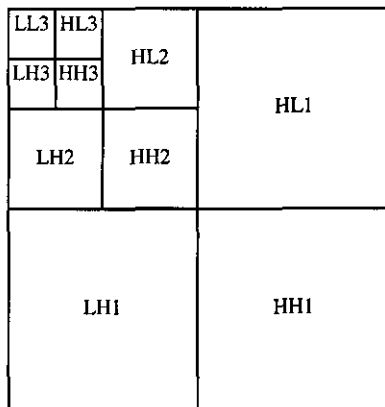

Figure 2. Three level wavelet decomposition of an image
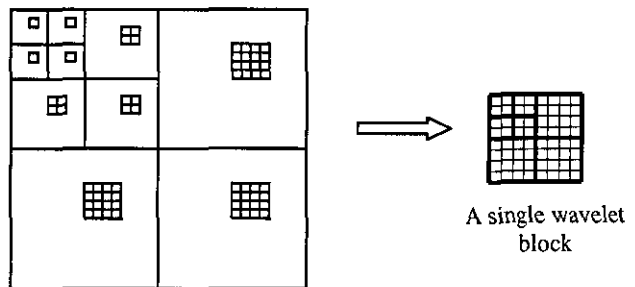
Figure 3. Formation of a single wavelet block from a 3-level decomposed image

As illustrated in fig.3, the final result is the transformation of the input image into a re-organized structure of original image size, but consisting of a two-dimensional array of non-overlapping wavelet blocks, each representing different

subband elements from the location corresponding to the block. In other words, for a S-level DWT, blocks of $2^S \times 2^S$ samples are constructed.

4. WAVELET BASED IMPLEMENTATION OF THE PBDCPC ALGORITHM

Figure 4 illustrates a high-level block diagram of the proposed CODEC. The original left image (reference) L_{ori}, is independently coded using a JPEG-like wavelet CODEC proposed in [19]. After reconstruction at the decoder end, it provides a reference for the prediction of the right image. In addition to the above, the encoded reference image is locally decoded at the encoder end to produce a reconstructed left image L_{rec}, which is subsequently used as the reference in the wavelet based PBS unit. Note that by adapting this strategy we satisfy the first condition of guaranteeing the reconstruction of the predicted image, under the pioneering block based coding technique. The original right image R_{ori} and reconstructed right image L_{rec} then undergo N level dyadic wavelet decomposition as described in section 3. For our experiments we have used $N=3$ and we follow the convention depicted in figure 2, in naming the ten resultant subbands, with the lowest subband being named as *LL3*.
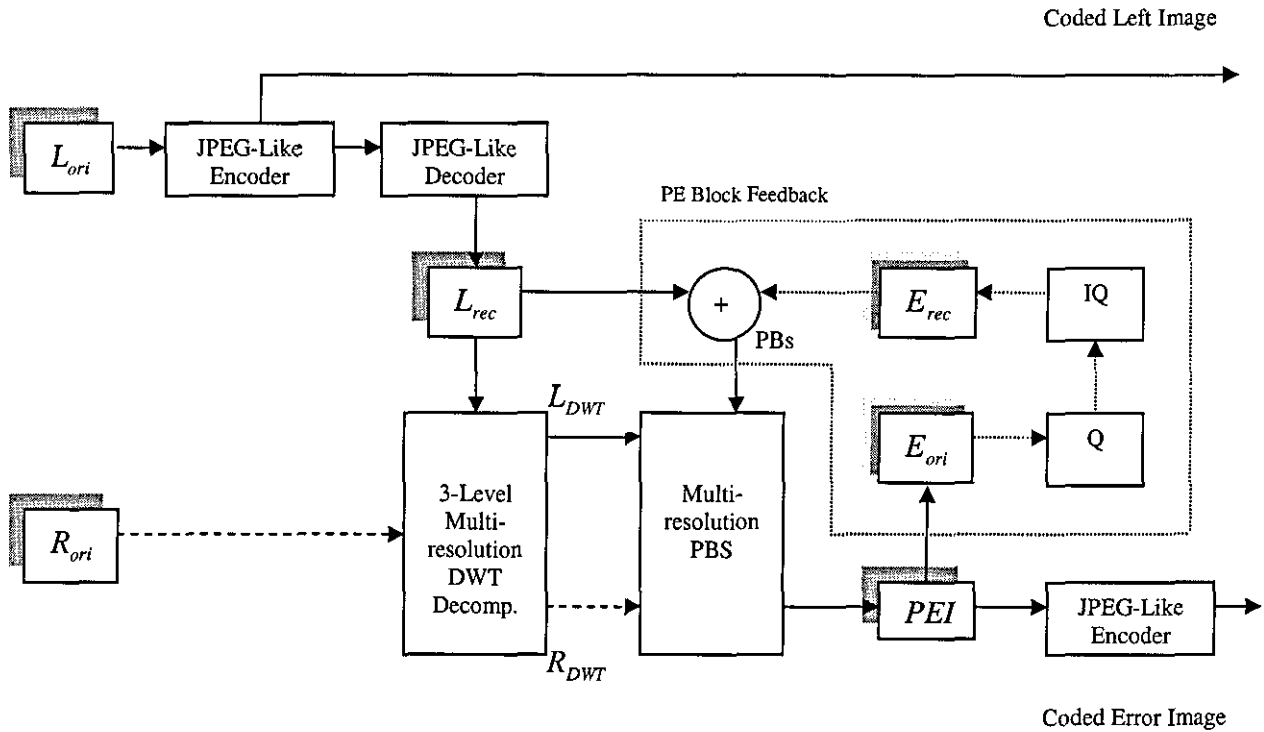
Figure 4. Proposed Encoder

Within the wavelet based PBS unit, for each corresponding subband pair (e.g. L_{LL3} and R_{LL3}, L_{HL1} and R_{HL1}) a PBS is performed as described in section 2, to produce the corresponding subband of the Prediction Error Image (PEI). Once the prediction error image has been completely found (i.e. all of its subbands found), a JPEG-like CODEC (see figure 5, for details refer [19]) is used to transform them into wavelet blocks and transmit after suitable quantization and entropy coding. Fig. 4 also illustrates a prediction error (PE) block feedback loop, which converts the original domain prediction errors into the reconstructed domain by locally sending the PE through a quantization/de-quantization procedure. The error block, in its reconstructed state, E_{rec} is then added to the best predictor previously found from L_{rec} to form the block, which would act as the preceding (front) pioneering block for the next iteration of the PBS. Thus the purpose of

the above mentioned error feedback loop is to satisfy the second requirement for a successful PBDCPC CODEC design (see section 2).

The decoder of the proposed CODEC works in a manner similar to the encoder. Due to the orientation of the two pioneering blocks relative to the block to be encoded, in encoding blocks in the first row of blocks (other than the first block itself), of a given subband, only the block above is used as the pioneering block. A similar strategy in adapted to resolve the problem of encoding blocks in the first column of blocks. The block at the top, left-hand corner of a sub-band is always coded directly, i.e. without any disparity compensation.

5. EXPERIMENTAL RESULTS & ANALYSIS

The proposed CODEC was implemented using purpose built MATLAB routines and was tested on a set of eight commonly used test stereo image pairs representing indoor, outdoor, natural and synthetic scenes. All test stereo image pairs were of size 512×512. As parallel axis camera geometry has been used to acquire these images we have limited the PBS along a horizontal direction only. For both CODECs, a windowed search area representing a disparity range of 0-7 pixel shifts was used. For the benchmark CODEC, the disparity vector field was coded using a fixed length code, requiring three bits per block. A Daubachies compactly supported orthonormal filter (db7) was used as the wavelet transform.

In order to evaluate the performance of the proposed CODEC, in figure 5, we compare its performance against that of the well-known benchmark algorithm of Perkin's (benchmark-1), that uses DCT as the basic compression engine, and that of our previously proposed CODEC of [20] (benchmark-3), which is essentially a wavelet-block implementation of benchmark-1.
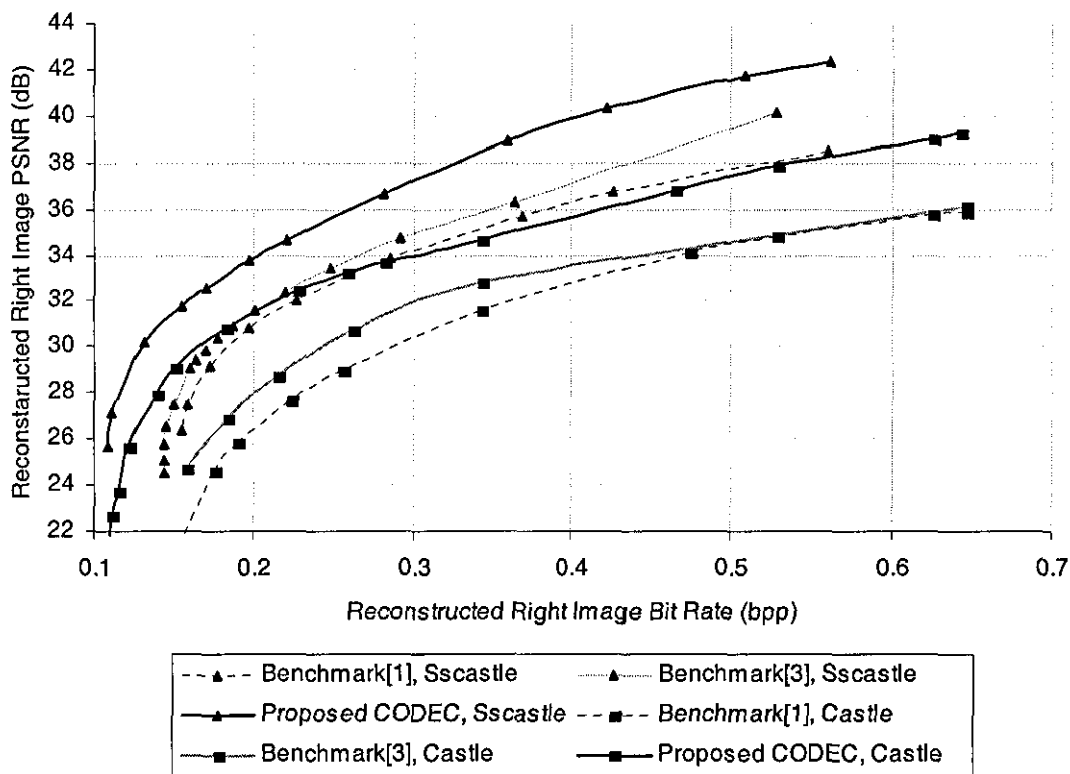
Figure 5. Comparison of rate-distortion performance graphs

The objective image quality is measured in terms of Peak Signal to Noise Ratio (PSNR), whereas the compression efficiency is measured in bits per pixels (bpp). The compression efficiency (or Bit Rate) is measured as follows:

$$BR = \left\lceil \frac{Tot\_Bits_{comp}}{N \times M} \right\rceil \quad bpp$$

where, N and M represent the image dimensions and Tot_Bits_{comp} is the total number of bits required to represent a given image in its compressed format. For benchmark-1 and benchmark-3, Tot_Bits_{comp} include the bits that are required to code the disparity vector field, in addition to the bit requirement to encode and transmit the predictive error image. However, for the proposed CODEC, Tot_Bits_{comp} represents only the bit requirement to encode the predictive error image. For a fair comparison, for a given stereo image pair, the objective quality of the left image (reference) was set at equivalent levels for all three CODECs.

As illustrated in fig. 5, the rate-distortion graphs show that the proposed PBDCPC CODEC outperforms the benchmark algorithms by considerable margins both at medium and low bit rates. At a reconstructed right image bit rate of 0.2 bpp, for Castle image, the PBDCPC CODEC produces a PSNR gain of approximately 3.5 dB as against benchmark-3 and 5 dB as against bechmark-1. For Sscastle image improvements of up to 3 dB are indicated at this bit rate. At higher bit rates, improvements in the range of 2-4 dB are indicated. Further analysis of the results in fig. 5 indicates the ability of the proposed PBDCBC CODEC to perform at bit rates as low as 0.11 bpp. In contrast, the benchmark CODECs are unable to perform at bit rates lower than 0.145 bpp. This is due to the fact that both benchmark CODECs require an allocation of a certain fixed bit budget (0.047 bpp in our experiments) for coding the resultant disparity vector fields.

(a)   (b)
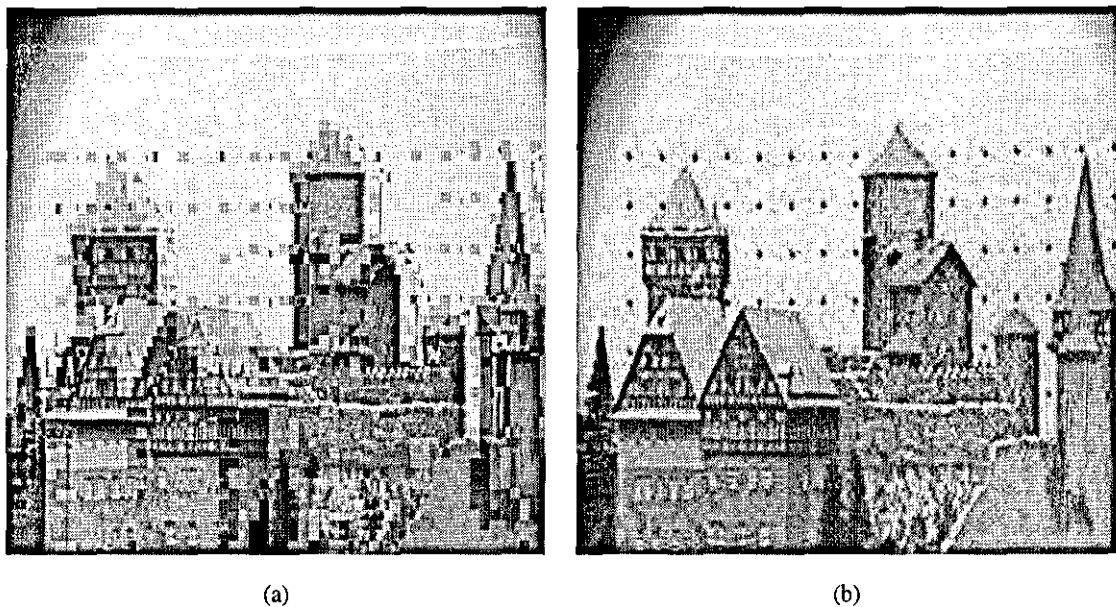
Figure 6.   Subjective quality evaluation for Castle reconstructed right image (a)
Benchmark-1, 0.1550 bpp, 21.29 dB   (b) Proposed CODEC, 0.1498 bpp, 29.03 dB

Figure 6 compares the subjective image quality performance of the proposed CODEC against the DCT based Benchmark-1 at a bit rate of 0.15 bpp. It clearly illustrates that in the proposed CODEC the image quality near object boundaries are much better as compared to that of Benchmark-1.

In predictive coding of stereo image pairs, the predictive errors are highest near object boundaries, which represent areas of binocular disparity. Therefore the prediction error frames have highest values in these regions. Thus when these frames are coded under constrained bit budgets, the maximum loss is introduced in object boundary areas. This is particularly true at bit rates lower than 0.25 bpp for most images. Even though the proposed CODEC does not result in a more accurate prediction, the bit budget that could be allocated to coding the prediction errors is higher as compared to benchmark techniques 1 and 3 due to the fact that no bits have to be allocated for the coding of disparity vector fields. Particularly at very low bit rates, the bits required for the lossless coding of disparity vector fields is significantly higher than what is required for coding the prediction error fields. Thus the proposed CODEC's relative performance as compared to the benchmark techniques improves with the decrease of bit rate. A closer analysis of the fig. 6(a) indicates that the subjective quality degradation near object boundaries is of a *blocky* nature and in fig 6(b) it is of *ringing* nature. This is due to the inherent characteristics of DCT and DWT based compression. However, due to the usage of the pioneering block technique, in figure 6(b) the otherwise expected significant ringing effects have been largely eliminated.

6. CONCLUSION & FUTURE

In this paper we have proposed an efficient DWT implementation of our previously proposed DCT based pioneering block based stereo image compression algorithm. We have shown that due to the specific requirements of the design of the CODEC to guarantee the reconstruction of predicted images, without the use of disparity information, the DWT implementation is not a straightforward conversion of the base technology from DCT to DWT. The novel DWT based PBDCPC scheme results in improved, objective rate-distortion performance and a subjective quality performance. PSNR improvements of up to 5.5 dB are reported for transmission bit rates of 0.15 bpp. In addition to this the proposed CODEC is able to operate at bit rates as low as 0.1 bpp, where other CODECs fail to operate due to the need of a certain minimum bid budget for encoding disparity vector fields.

Further improvements to the proposed technique, in terms of algorithmic and computational efficiency is possible by using faster multi-resolution search techniques [13] that avoid searches in all sub-bands but would rather effectively re-use search information at lower resolution sub-bands within the search operations of higher resolution sub-bands. We are currently in the process of expanding the ideas proposed in this paper to wavelet based stereoscopic sequence coding.

REFERENCES

1. W.B.Pennebaker, J.L.Mitchell, "JPEG: Still Image Compression Standard", Van Nostrand Reinhold, NY, 1993.
2. ISO/IEC JTC 1/SC 29/WG 1, ISO/IEC FCD 15444-1: Information technology – JPEG-2000 Image Coding System: Core coding system, March 2000, www.jpeg.org/FCD15444-1.htm
3. M.G.Perkins, "Data Compression of Stereopairs", IEE Trans. on Comm., Vol.40, No.4, pp. 684-696, April 1992.
4. I.Dinstein, M.G.Kim, A.Hanik, "Compression of Stereo Images Using Subsampling and Transform Coding", Journal of Optical Engineering, Vol.30, pp. 1359-1364, Sept. 1991.
5. J.Jiang, E.A.Edirisinghe, H.Schroder, "A Novel Predictive Coding Algorithm for 3-D Image Compression", IEEE Trans. on Cons. Elec., Vol.43, No.3, pp. 430-437, August 1997.
6. J.Jiang, E.A.Edirisinghe, H.Schroder, "Algorithm for the Compression of Stereo Image Pairs", IEE Electronic Letters, Vol.33, No.12, pp. 1034-1035, June 1997.
7. W.Woo, A.Ortega, "Optimal Blockwise Dependent Quantization for Stereo Image Coding", IEEE Trans. on CSVT, Vol. 9, No.6, pp. 861-867, Sep. 1999.
8. W.Woo, A.Ortega, "Overlapped Block Disparity Compensation with Adaptive Windows for Stereo Image Coding, "IEEE Trans. on CSVT, Vol. 10, No.2, pp.194-200, March 2000.

9. N.D.Doulamis, A.D.Doulamis, Y.S.Avrithis, K.S.Ntalianis, S.D.Kollias, "Efficient Summarization of Stereoscopic Video Sequences", IEEE Trans. on CSVT, Vol. **10**, No.4, pp. 501-517, June 2000.
10. R.Wang, Y.Wang, "Multiview Video Sewquence Analysis, Compression and Virtual Viewpoinr Synthesis", IEEE Trans. on CSVT, Vol. **10**, No.3, pp. 397-410, April 2000.
11. N.Grammalidis, D.Beletsiotis, M.G.Strintzis, "Sprite Generation and Coding in Multiview Image Sequences", IEEE Trans. on CSVT, Vol. **10**, No.2, pp. 302-311, March 2000.
12. J.Jiang, E.A.Edirisinghe "A Hybrid Scheme for Low Bit-Rate Coding of Stereo Images", IEEE Transactions on Image Processing, Vol. **11**, Issue. **2**, pp. 123-134, Feb. 2002.
13. S.Sethuramam, M.W.Siegel, A.G.Jordan, "A multi-resolution framework for stereoscopic image sequence compression', Proceedings of the ICIP-95, Vol.2, pp. 361-365, 1995.
14. N.V.Boulgouris, M.G.Strintzis, "Embedded Coding of Stereo Images", Proc. of the IEEE Int. Conf. on Image Proc., ICIP 2000, June 2000.
15. E.A.Edirisinghe, "Rate Scalable Stereo Image Coding Using EZW Algorithm", Proc. of the IEEE Int. Conf. on Information, Communications and Signal Proc., ICICS 2001, Singapore, Oct. 2001.
16. S.A.Martucci, I.Sodagar, T.Chiang, Y.Zhang, "A Zerotree Wavelet Video Coder", IEEE Trans. on CSVT, Vol.7, No.1, pp. 109-1997, February 1997.
17. K.Shen, E.J.Delp, "Wavelet Based Rate Scalable Video Compression, "IEEE Trans on CSVT, Vol.9, No. **1**, pp. 109-122, Feb. 1999.
18. S.A.Martucci, I.Sodagar, "Zerotree Entropy Coding of Wavelet Coefficients for Very Low Bit Rate Video Coding", Proc. of 1996 IEEE Int. Conf. on Image Proc., Lausanne, Switzerland, Sept. 1996.
19. R. de Queiroz, C.K. Choi, Y. Huh, K.R. Rao, "Wavelet Transforms in a JPEG – Like Image Coder," IEEE Transactions on Circuits and Systems for Video Technology, Vol. **7**, No. **2**, pp 419 -424, April 1997.
20. M.Y.Nayan,E.A.Edirisinghe,H.E.Bez, "Baseline JPEG-Like DWT CODEC for Disparity Compensated Residual Coding of Stereo Images", Proc. of Eurographic 2002, UK, pp. 67-74, June 2002.
21. Edirisinghe, E.A., Jiang, J., "Towards Eliminating Disparity Field Coding of Stereo Image Pairs", Proc. of the IEEE Int. Conf. on Information, Communication & Signal Processing, December 1999.

* Correspondence: E.A.Edirisinghe@lboro.ac.uk; Telephone: +44 (0)1509 228234; Fax: +44 (0)1509 211586

RASTER: A JPEG 2000 STEREO IMAGE CODEC

G.RAJKUMAR*, M.Y.NAYAN**, E.A.EDIRISINGHE**

*School of Medicine, Imperial College, UK,
**Department of Computer Science, Loughborough University, UK

ABSTRACT

With the increasing need of multimedia technologies, image compression requires higher performance as well as new functionality. To address this need in the specific area of still image coding, a new standard, JPEG-2000 has been designed and standardized [6]. JPEG-2000 provides additional functionality that its predecessors were not able to address efficiently or address at all. In this paper we attempt to extend the use of the technology provided by JPEG-2000 in solving many unresolved problems in stereo image coding. We show that it is possible to design a *Rate* Scalable *Stereo* image CODEC (RASTER), which has the unique ability of preserving the quality at binocular depth boundaries, an important requirement in the design of a stereo image CODEC. It is shown that RASTER provides PSNR gains of up to 3.7 dB at very low bit rates and provides reconstructed images of superior subjective image quality as compared to the direct transmitted images of equal bit rate.

KEY WORDS

Stereo image compression, JPEG-2000, Rate scalability

1. INTRODUCTION

To make use of the rate scalability provided by the JPEG-2000 in stereo image compression, one obvious way is to apply it independently in coding the stereo image pair. However, such a coder would be inefficient, as the stereo redundancy present within a stereo image pair is not exploited. Fortunately, a solution exists [in the suppression theory of binocular vision [14], which is widely used in literature [1-5] to address this issue and achieve image compression. In these methods one image of the stereo image pair (say left) retains the details of the scene, while the second image (right) retains the disparity information. Hence the second image can be highly compressed without affecting the depth information in the compressed stereo images [14]. In this paper we propose how the technology provided within the new JPEG-2000 standard could be used in association with the suppression theory of binocular vision to achieve scalable stereo image compression.

For clarity of presentation we divide this paper into five sections. Section-2 describes the basics of disparity compensated predictive coding of stereo image pairs and some associated problems in coding the so-called predictive error frames. Section-3 proposes the novel scalable stereo image compression framework and section-4 provides experimental results and an analysis. Finally section-5 concludes with an insight to further improvements and possible applications.

2 DISPARITY COMPENSATED PREDICTIVE CODING

The idea of disparity compensated predictive coding is similar to the idea of motion compensated predictive coding, used in video coding. The image to be predicted (say the right image) is blocked into fixed size (usually 8×8), non-overlapping blocks. In encoding these individual blocks, a search is made within a certain maximum likelihood area within the reference frame for a *best match* under a given matching criteria, such as Mean Squared Error (MSE). This best match is subsequently selected as the predicted block for the block to be encoded and is used to replace it, in a so-called *disparity predicted* image. A *prediction error image* (PEI) is obtained by taking the pixel domain difference between the image to be predicted and the disparity predicted image. The displacement to the best matching block from the block to be encoded is obtained as the *disparity value* of the block to be encoded. The values thus obtained for all blocks in the predicted image are used to form the so-called *disparity vector* (DV) *field* of the stereo image pair.

The PEIs are usually encoded using either block-based transforms such as DCT [1-3], or non block-based coding such as subband coding or the wavelet transforms [5]. The major problem with the DCT based transform-coding algorithm is the existence of the visually unpleasant blocking artifacts, especially at low data rates. This problem can be eliminated using the wavelet transforms, which are usually obtained over the whole image. In literature, wavelet transform approaches have been used for stereo image compression [5,15]. However these algorithms are not scalable. If we use wavelet based rate scalable algorithms such as EZW [7], SPIHT [8], EBCOT [9], to compress the reference image and the PEI, rate scalable stereo image compression can be achieved.

The use of most valuable information as early as possible in the encoding process in scalable coding schemes

described above was the main motivation behind the use of the JPEG-2000 (or EBCOT) technology for stereo image compression within our present research context. When using the suppression theory for coding stereo image pairs in association with disparity compensated predictive coding, it is often necessary to code the reference frame directly at a high bit rate (higher quality) and to only allocate the minimal amount of the bit budget to code the prediction error images (PEIs). Thus the PEIs are prone to excessive loss in quality, which would adversely affect the overall quality of reconstructed stereo image pair. The application of embedded block coding described above, would solve this problem by concentrating on the accurate coding of coefficients (thus pixels) close to high activity areas (usually along object boundaries where binocular depth information is excessive) of the PEIs. The remaining low activity areas, being mostly zeros, would remain so, even under very high compression rates. Thus at the decoder end the disparity compensation procedure would be able to reconstruct a predicted image (right image) whose quality near high activity areas of the error image is determined by the efficiently coded PEIs and the quality near low activity areas would be largely determined by the reconstructed reference (left) image of superior quality.

In the following section we introduce the reader to the design and implementation of the RASTER CODEC.

3. THE RASTER CODEC DESIGN

Figure-1 shows the block diagram of the proposed rate scalable stereo image CODEC. A careful observation of fig. 1 leads us to the following:

1. A JPEG-2000 encoder is used to encode the directly transmitted *left image* and the *prediction error image* (PEI).
2. At the encoder end, disparity estimation and compensation is performed in comparison with a locally decoded reference image instead of the original reference image.
3. Decoded reference frames at both encoder and decoder are locked at a given data rate R1.
4. The decoder contains two JPEG-2000 decoders, operating at two data rates.
5. The disparity vector (DV) field is encoded using a fixed length code separately.

In order to explain the above inclusions in more detail; let us explore some theoretical aspects of the coding process in more detail.

Let d denote the disparity vector field obtained by the method described in section 2.1. Let L be the reference image (left view). The predicted right image, R_{pred} is obtained by rearranging the pixels in L relative to d. If D

denotes this operation, the predicted right image can be written as,

$$R_{pred} = D(L, d) \tag{1}$$

Thus, the prediction error image, E_{pred} for the right image, R is obtained as,

$$E_{pred} = R - R_{pred} \tag{2}$$

At the decoder end the predicted right image, $\hat{R}_{pred}$, is obtained by,

$$\hat{R}_{pred} = D(\hat{L}, \hat{d}) \tag{3}$$

where, $\hat{L}$ and $\hat{d}$ are the decoded reference image and disparity vector field, respectively. Thus, the decoded right image can be obtained using the equation,

$$\hat{R} = \hat{R}_{pred} + \hat{E}_{pred} \tag{4}$$

where, $\hat{E}_{pred}$ is the decoded predictive error image.

The disparity vector field is losslessly encoded using DPCM coding [1]. This leads to the relationship, $d = \hat{d}$. Thus, if the same reference frame is maintained both at the encoder and decoder, i.e. if $L = \hat{L}$, comparing equations (1) and (3),

$$\Rightarrow R_{pred} = \hat{R}_{pred} \tag{5}$$

This results in the decoded PEI, $\hat{E}_{pred}$, being the only source of distortion in the reconstruction of the right image (equation (4)). Thus, in order to maximize the performance of the CODEC it is necessary to maintain the same reference frame at both encoder and decoder. In the proposed CODEC we achieve this by adding a prediction feedback loop in the encoder so that a decoded image is used as the reference image.

However in the proposed scalable CODEC, the decoded reference images have different distortions at different data rates. Hence it is impossible for the encoder to generate the exact reference images as in the decoder for all data rates. In order to solve this problem we use adaptive disparity compensation (ADC), a concept similar to adaptive motion compensation (AMC) [9]. In ADC, in order to maintain the same reference frame in the encoder and decoder, we introduce a feedback loop in the decoder, such that the decoded reference images at both ends could be locked to the same data rate, say R.

234

Assume that the target data rate is R_T, where $R_L \leq R_T \leq R_H$ and the disparity vector field is coded at a rate R_{DV} where $R_{DV} < R_L$. As R_{DV} is known both by the encoder and decoder, if we decode the embedded reference image bit stream at rate $R_L - R_{DV}$ (=R) at both ends, the decoded reference frames would be identical, i.e. $\hat{L}$. It was mentioned above that at the decoder the embedded reference image bit stream is decoded at two data rates. The image decoded at $R_L - R_{DV}$ is used as the decoded reference image to find the predicted image, and the image decoded at rate $R_T - R_{DV}$ is added to the predicted frame to generate the final decoded right image. By following this strategy, the reference images at the encoder and decoder are maintained as identical, which leaves the decoded E_{pred} as the only source of distortion, thus eliminating error propagation over varying data rates.

255, 8 bits per pixel, uncompressed) and are of size 320×320 pixels. As the images have been obtained using parallel axis camera geometry, a search area of 8 pixels in the horizontal direction was used for all images in disparity compensation. The compression rate is measured in bits per pixel (bpp). The objective image quality is measured in terms of peak signal to noise ratio.

Table.1 illustrates the rate distortion performance of the RASTER CODEC in comparison to the direct transmission of right image, at same bit rate. For all experiments the left image (reference image) is directly transmitted at twice the bit rate of right (predicted) image tabulated in the first column of Table.1. As a block size of 8×8 and a search window size of 8 pixels (0-7) were used, $\log_2 8/(64) = 0.0469$ bpp is allocated for the fixed length coding of disparity vector field, DV. The remaining bits are used in coding the PEI. For example, when 0.4000 bpp has been allocated for the coding of right image, 0.8000 bpp is used for the coding of directly transmitted left image, 0.0469 bpp is used for the coding

Figure 1. Block Diagram of the Proposed Rate Scalable Stereo Image CODEC

4. RESULTS & ANALYSIS

The above CODEC was implemented on a UNIX platform using C language. The JasPer JPEG-2000 implementation (13) was used as the base compression engine. Experiments were performed on a set of five stereo image pairs namely lamp, cans, texture, lab and packs, respectively containing an increasing amount of stereo data. All images were grayscale (value range 0-

of disparity vector field and 0.3531 bpp is used for the coding of the PEI. Note that for a fair comparison the quality of the directly transmitted (if it was to) right image at this rate is measured at 0.4000 bpp rather than at 0.3531 bpp. The reason is that under such a situation, no DV field would have to be transmitted for the right image reconstruction. Note that for all the experiments, the right frame prediction at the encoder end and the decoding at the decoder end is done based on a left frame that is decoded at rate R1=0.1000 bpp. (see Section 3.0).

Table 1 – Rate Distortion Performance Comparison

Right BR (bpp)	Cans		Packs		Lamp		Lab		Texture	
	JPEG 2K(dB)	RASTER (dB)	JPEG 2K (dB)	RASTER (dB)	JPEG2K (dB)	RASTER (dB)	JPEG 2K (dB)	RASTER (dB)	JPEG 2K (dB)	RASTER (dB)
0.4000	44.67	44.25	34.79	34.35	46.35	46.17	38.52	37.77	36.80	36.45
0.2000	39.67	39.71	30.08	30.35	42.73	42.62	33.41	33.50	32.26	32.79
0.1600	37.99	38.22	29.00	29.44	41.18	41.35	32.02	32.42	31.18	31.67
0.1333	36.77	37.52	27.76	28.64	39.95	40.42	31.01	31.50	30.37	30.82
0.1143	35.58	36.88	27.07	28.02	38.93	39.70	30.05	30.93	29.37	30.45
0.1000	34.58	36.07	26.50	27.75	38.07	39.10	29.24	30.49	28.70	29.88
0.0889	33.81	35.73	25.83	27.47	37.17	38.76	28.79	30.13	28.22	29.71
0.0800	33.32	35.51	25.42	27.22	36.30	38.49	28.32	29.81	27.69	29.48
0.0727	32.46	35.20	25.00	27.00	35.75	38.27	27.76	29.54	27.29	29.24
0.0667	31.73	35.00	24.65	26.83	35.07	37.73	27.20	29.39	27.01	29.15
0.0615	31.45	34.62	24.28	26.52	33.88	37.64	27.02	29.19	26.73	29.04

The results in table.1 clearly indicates that the RASTER CODEC produces reconstructed images of better objective image quality (up to 3.7 dB excess PSNR) compared to the right images that could be obtained by direct compression using JPEG-2000 at the same bit rate. This observation is true for all right image bit rates below 0.2 bpp. However, for bit rates above this, slight, subjectively unnoticeable image degradations are observed.

Figure.2 shows the rate-distortion graphs of the RASTER CODEC and JPEG-2000 coder in compressing the right image of the stereo image pair 'Cans'. It clearly indicates that the RASTER CODEC provides better, reconstructed image quality at all right image bit rates above 0.2 bpp.
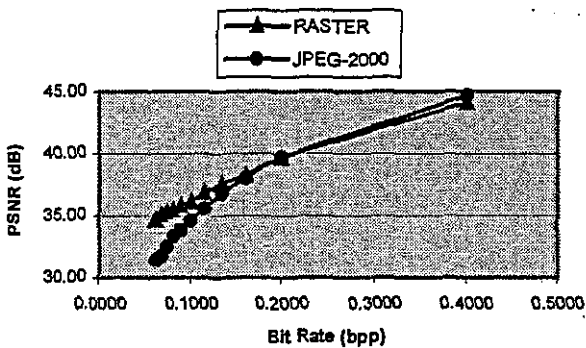

Figure 2. Rate-Distortion Performance for 'Cans'.

Figure 3(a) and 3(b) illustrates the original left and right images of the stereo image pair 'Cans'. Figure 3(c) illustrates the error image that results from disparity compensated prediction. Figure 3(d) illustrates the reconstructed left image, when the original is coded at a bit rate of 2×0.0615=0.123 bpp (PSNR=36.22dB). Figure 3(e) and 3(f) compares the subjective image quality of the right images produced by the two coders (RASTER & JPEG-2000) at a right image bit rate of 0.0615 bpp. The image produced by RASTER CODEC (PSNR=34.62dB) shows less ringing artifacts around object boundaries as compared to the image formed by directly transmitted right image using JPEG-2000 (PSNR=31.45 dB). Visual comparison of figures 3(d), 3(e) and 3(f) shows that the RASTER CODEC provides a reconstructed image with visual image quality equal to the reconstructed left image (JPEG-2000 coded) at twice the bit rate, whereas the directly transmitted right image is of noticeably inferior image quality as compared to the image of fig. 3(d).

The quality of the reconstructed right image of the RASTER CODEC is dependent on the quality of the locally decoded left image that is used at the encoder and decoder ends as a reference for disparity compensated prediction. In order to maintain rate scalability of the proposed coder within a larger range, it is essential that this reference image is coded at the lowest acceptable bit rate (see Section 3). However very low quality of this reference frame could result in the lowering of the reconstructed right image quality of the RASTER CODEC. At the same time, if this reference frame is coded at a higher bit rate, it would result in an increase of the lower bit rate limit of the CODEC.

Figure 4 illustrates the rate distortion performance of the RASTER CODEC under varying compression rates of this reference image. It is seen that the best performance is when the directly transmitted left frame is used, without any change, as the reference frame. However at such a setting the scalability of the RASTER CODEC would suffer.
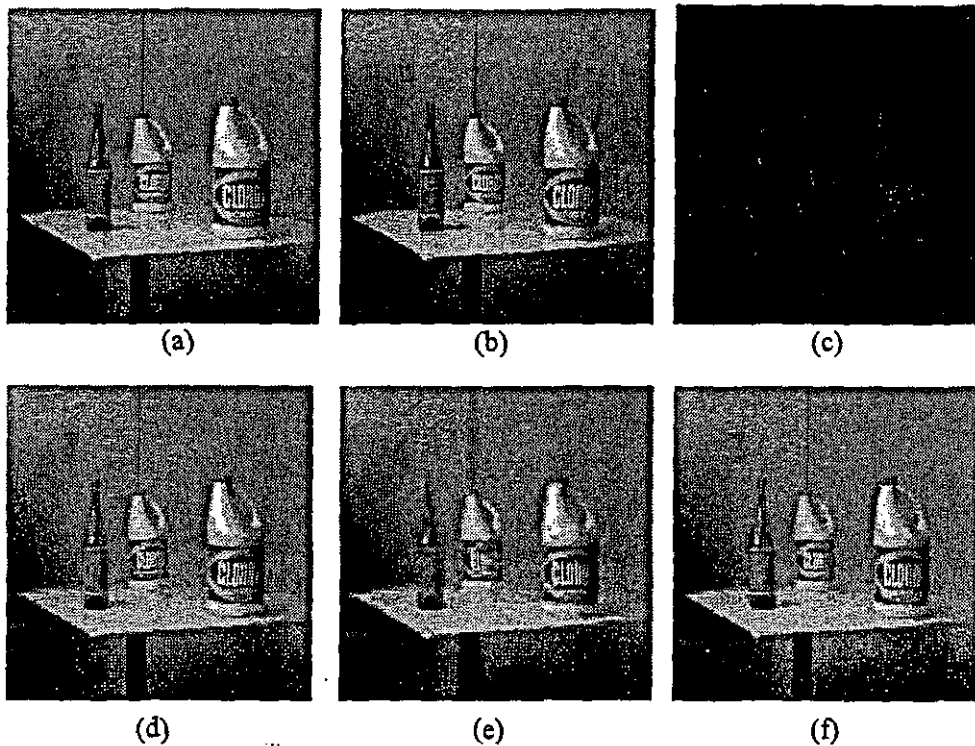
Figure 3. Subjective Image Quality Illustration for 'Cans' (a) Original left image (b) Original right image (c) Error image (d) Reconstructed left image, JPEG-2000, at 0.123 bpp (e) Reconstructed right image, JPEG-2000, at 0.0615 bpp (f) Reconstructed right image RASTER CODEC, at 0.0615 bpp.
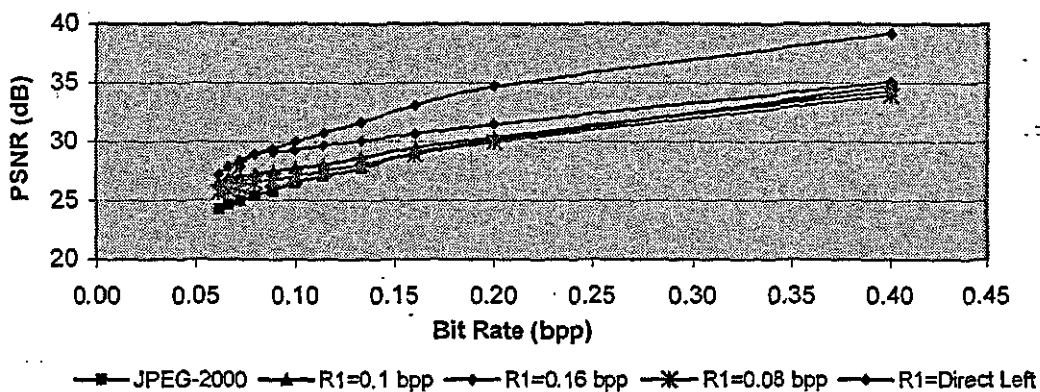

Figure 4. Rate-distortion performance of RASTER CODEC for stereo image pair 'Packs', under varying bit rates, R1 of left reference frame.

If scalability is not a requirement of the RASTER CODEC it is able to compress the right image at half the bit rate of the left image, still obtaining an image quality equivalent to that of the left image (see table.2). This is an interesting observation that shows that the proposed coder acts very competitively to most stereo image CODECS that have been proposed in literature in the past. It clearly shows that the RASTER CODEC efficiently codes the PEIs and uses information that it could obtain from the reconstructed left frame to build areas where the prediction errors are negligible or zero.

The above experimental results clearly show that the RASTER CODEC shows rate scalability in stereo image compression and provides an efficient solution to many problems that are inherent to such CODECS. At low bit rates, the proposed CODEC is able to code the PEIs very efficiently. The prioritized coefficient coding technique adopted by JPEG-2000 acts well in the coding of disparity compensated images and when combined with the disparity compensated stereo imaging techniques provides enhanced subjective and objective image quality in the reconstructed right images.

Left Image Bit Rate (bpp)	Right Image Bit Rate (bpp)	Rec. Left Image PSNR (dB), JPEG2K	Rec. Right Image PSNR(dB), RASTER	Rec. Right Image PSNR(dB), JPEG2K
0.8000	0.4000	40.60	39.20	34.79
0.4000	0.2000	34.91	34.67	30.08
0.3200	0.1600	33.27	33.09	29.00
0.2666	0.1333	31.77	31.65	27.76
0.2286	0.1143	30.90	30.75	27.07
0.2000	0.1000	30.15	29.98	26.50
0.1778	0.0889	29.53	29.37	25.83
0.1600	0.0800	28.89	28.92	25.42
0.1454	0.0727	28.32	28.34	25.00
0.1334	0.0667	27.82	27.83	24.65
0.1230	0.0615	27.15	27.20	24.28

In coding the PEIs we have adopted a strategy whereby we map the maximum positive (or negative) value to +255 (or –255) and accordingly use the same linear mapping to adjust the values of all other (negative & positive) error values. This was found to be necessary in order to maximize the efficiency of using a JPEG-2000 CODEC for the coding of PEIs. We have also experimented on an adaptive gain based preprocessing technique for PEI coding [10], whereby smaller predictive errors are shifted upwards in magnitude relative to the larger errors to give extra priority in coding them. However the minimal gains obtainable under careful parameter settings was not found to give any useful rate-distortion performance enhancement to the proposed CODEC.

5. CONCLUSIONS

We have proposed a rate scalable CODEC (RASTER) for the efficient compression of stereo image pairs. The CODEC is designed based on the JPEG-2000 technology and the well-known disparity compensated predictive coding techniques of stereo image pairs. A JPEG-2000 CODEC is used as the base compression engine to code the directly transmitted reference image and the predictive error images. In order to maintain rate scalability an adaptive disparity compensation (ADC) technique has been adopted. We have shown that the proposed CODEC is able to achieve PSNR gains of up to 3.7 dB as compared to directly transmitting the right frame using JPEG-2000. In anticipation of the standardization of moving JPEG-2000 [14] we are currently in the process of further extending this idea to stereo video sequence coding.

REFERENCES

[1] M.G.Perkins, "Data Compression of Stereopairs", IEEE Transactions on Communications, Vol. 40, No. 4, pp 684-696, April 1990.

[2] J.Jiang, E.A.Edirisinghe, H.Schroder, "Algorithm for compression of stereo image pairs", IEE Electronic Letters, Vol. 33, No. 12, pp. 1034-1035, June 1997.

[3] E.A.Edirisinghe, J.Jiang, "A Block-Object Hybrid Approach Towards Stereo Image Compression", Proc. of the Int. Conf. on Inf., Commun. & Signal Processing, ICICS'99, Singapore, December 1999.

[4] E.A.Edirisinghe, J.Jiang, "Stereo Image Compression Approach via Pixel Value Variation Trend Adaptive Subspace Projection", Proceedings of SPIE Photonics East Conference, Boston, MA, USA, Nov. 2000.

[5] S.Sethuramam, M.W.Siegel, A.G.Jordan, "A Multiresolution Region Based Segmentation Scheme for Stereoscopic Image Compression", SPIE Digital Video Comp., Alg. and Tech., Vol. 2419, 1995.

[6] ISO/IEC JTC 1/SC 29/WG 1, ISO/IEC FCD 15444-1: Information technology – JPEG 2000 Image Coding System: Core coding system, Mar. 2000. http//www.jpeg.org/FCD15444-1.htm.

[7] J.M.Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients", IEEE Trans. on Signal Processing, Vol. 41, No. 12, December 1993.

[8] A.Said, W.A.Pearlman, "A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees", IEEE Trans. on CSVT, Vol. 6, No. 3, pp. 243-250, June 1996.

[9] D.Taubman, "High Performance Scalable Image Compression with EBCOT", IEEE Trans. on Image Proc., Vol 9, Jul 2000, pp. 1158-1170.

[10] K. Shen, E.J.Delp, "Wavelet Based Rate Scalable Video Compression", IEE Trans. On CSVT, Vol. 9, No. 1, pp 105-122, February 1999.

[11] Q.Wang, M.Ghanbari, "Scalable Coding of Very High Resolution Video Using the Virtual Zerotree", IEEE Trans. On CSVT, Vol.7, No.5, pp. 719 – 727, October 1997.

[12] http://www.ece.ubc.ca/~mdadams/jasper/jasper.html

[13] http://www.jpeg/org/public/fcd 15444-3.pdf.

[14] Dinstein, M.G.Kim, J.Tzelgov, A.Henik, "Compression of Stereo Images and the Evaluation of Its Effects on 3-D Perception", SPIE Apps. of Digital Image Proc. XII, Vol. 1153, pp. 522-530, 1989.

Transform Domain Overlapped Block Disparity Compensation in Wavelet Coding of Stereo Image Pairs

M.Y.Nayan, E.A.Edirisinghe, H.E.Bez
Department of Computer Science,
Loughborough University, UK

Abstract

In this paper we propose a stereo image CODEC, that uses overlapped block disparity estimation/compensation in multiresolution wavelet transform domain. Overlapped block matching is known to be able to reduce blocking artifacts by linearly combining multiple blocks provided by the disparity vectors of a block and its neighbours. This capability of OBM and discrete wavelet transform's (DWT) versatile time-frequency localization is expected to improve rate distortion performance. With the propose CODEC gains of up to 1 dB at low bit rates as compared to benchmark [1] used in the paper.

1. Introduction

Advances in telecommunication and multimedia technologies have reached a stage where there are new demands of more realistic three-dimensional (3-D) display technologies. Stereoscopic images/sequences provide a promising means to stimulate 3-D perception as it can artificially imitate the human visual system in perceiving 3-D. The recent breakthrough of autostereoscopic display systems which enable users to view stereo images/sequences without having to wear special eyeglasses or helmet-mounted display kits has resulted in an increased popularity of stereo imaging in diverse application domains such as computer games, tele-robotics, surgery, virtual environment, etc.

A stereo image pair consists of two separate views of a three-dimensional scene captured and/or recorded by two cameras, one corresponding to the left eye of the human visual system (left image) and the other to the right eye (right image). However, independent transmission/storage of stereo image pairs require a two fold increased in the bandwidth/storage space requirement. Exploiting the inter and intra redundancies between the stereo image pairs is the key factor in stereo image compression in order to avoid transmitting/storing two images (left and right views) independently. Based on this, many pixel, block and object based stereo image compression algorithms have been proposed [1-7,9-10].Majority of these algorithms are based on Discrete Cosine Transform (DCT) technology and uses Fixed Size Block Matching (FSBM) in disparity estimation/compensation. Unfortunately using FSBM with DCT results in annoying blocking artifacts at low bit rates. This is partly due to FSBM assuming constant disparity for the entire block and the use of fixed length basis functions in DCT. In stereo imaging applications such artifacts cause viewer discomfort and create a major problem. Overlapped block matching (OBM), which reduces blocking artifacts and Wavelet Transform that uses variable length basis functions, provide a good solution to this problem.

OBM was originally used in video coding and has receives much interest within the stereo image compression research community since its introduction in [7]. OBM method reduces blocking artifacts by using multiple disparity vectors to estimate a block and linearly combining the prediction, based on the disparity of the block and its neighbouring blocks, as opposed to FSBM which depends only on the disparity of the block itself.

Meanwhile the use of wavelet transforms in new still image coding standard, JPEG-2000 and video coding standard, MPEG-4 which are capable of providing improved rate-distortion performance and functionality, has motivated us using OBM in wavelet based coding of stereo image pairs [8-10].

We propose two novel CODECs in wavelet transform domain for stereo image coding. CODEC-1 performs multiresolution hierarchical disparity estimation/compensation in wavelet domain to produce improved rate distortion performance. In CODEC-2, we further improve this algorithm by adding OBM to CODEC-1.

The rest of the paper is arranged as follows. In section 2 we provide the difference of multiresolution hierarchical disparity estimation of the benchmark and our proposed CODECs. Section 3 describes in detail the two algorithms of the proposed CODECs. Section 4 provides experimental results and compares the effectiveness of the proposed CODECs with regard to the benchmark. Finally, Section 5 concludes with a foresight to our future research in this direction.

2. Multiresolution Hierarchical Disparity Estimation & Compensation

The Discrete Wavelet Transform (DWT) decomposition based on [12] forming a pyramid structure is an efficient way of representing information of an image at progressively lower resolutions. The pyramid structure enables efficient multiresolution based disparity estimation/compensation to be performed due to availability of significance features such as edges that are crucial in stereo perception at a multitude of resolutions levels [1].

The left image (reference image) and the right image (predicted image) undergo a 3-level DWT decomposition where the coefficients are generated by applying a cascade of two-channel filter banks to each image. At each level there are four subimages that are one-fourth the size of the image at

the level below. One subimage is obtained by lowpass filtering in both directions (rows and columns) and is referred to as the coarse image (LL); it carries most of the energy present in the original image and

possess intensity distribution similar to that of the original image. The remaining three subimages are detail images, which contain high frequency details in either horizontal (LH), vertical (HL), or both (HH) directions. Only the coarse image at each level is subjected to further decomposition to obtain the next level of subimages.

At the end of the 3-level DWT decomposition there are 10 subimages for each image consisting of one coarse image (LL3), three horizontal detail images (LH1, LH2, LH3), three vertical detail images (HL1, HL2, HL3) and three diagonal detail images (HH1, HH2, HH3) as shown in figure 1.
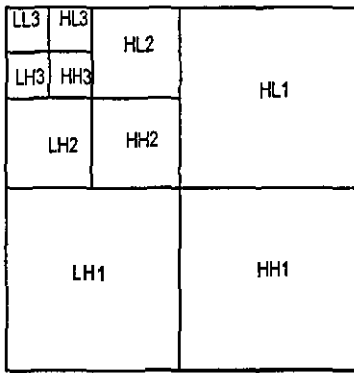
Figure 1. Three level wavelet decomposition of an image

2.1 Multiresolution Hierarchical Disparity Estimation – The benchmark technique

In the benchmark technique [1], disparity estimation is initially performed on the coarsest sub-images, LL3, of the stereo image pair. These disparity values are subsequently propagated to the LL2 level as an initial estimate of the disparity. A search is now performed at LL2 level, within a search space of double the size used in LL3 level, for a further refinement of this disparity value. This value is finally propagated as the initial estimate for the LL1 level. At the first level of decomposition (LL1 in out case), the disparity value refinement is continued for a final time as discussed above. Finally the three detailed subimages, HL1, LH1 and HH1 uses the refined disparity value thus obtained from LL1 as the final estimate of disparity. Figure-2 illustrates this process. The right image will be subsequently estimated from the four sub-images of the highest resolution level.

2.2 Multiresolution Hierarchical Disparity Estimation - Proposed

In the proposed CODECs, disparity estimation is performed separately on the coarse image and the three detail images of

the lowest resolution level, as in figure 3.

The disparity vectors of detail images of the lowest resolution level are multiplied by 2^n, n=1,2... and used as disparity vector of next higher resolution level for corresponding detail images.

In the propose CODECs; the right image is estimated using all subimages of each level as opposed to the benchmark which uses only the level-1 subimage. This results in a better performance of the propose CODECs over the benchmark, as shown by our experiment result described in section 4.
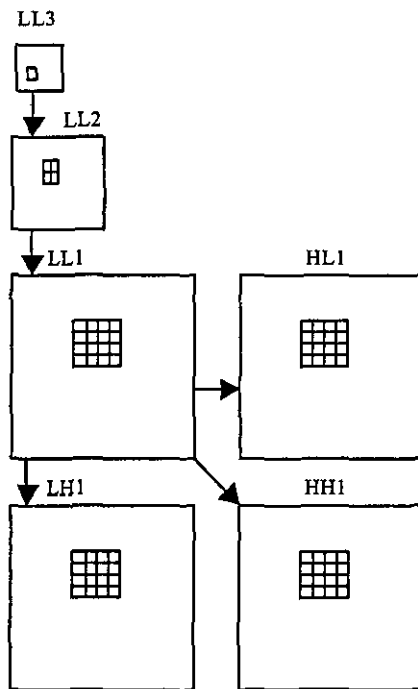

Figure 2. Hierarchical disparity estimation – The benchmark technique

3. The Proposed CODECs

3.1 CODEC-1

In this algorithm, the search for the best match of a given block of each subimage in level-3 of the reference image (left) is done independently with the corresponding subimage in level-3 of the predicted image (right image).

Figure 4 represents the block diagram of the proposed CODEC-1. At the encoder end, the left image and the right image undergo a 3-level dyadic DWT separately. For a block of each subimage in level-3 of the left image a search is performed in the corresponding subimage of the right image, within a maximum likelihood area on the subimage. The best predictor is found by using the Mean Squared Error (MSE) as the selection criterion.

For level-2 and level-1 detailed images (HL, LH, HH), the predicted best matches are found by using the disparity vectors of the corresponding detailed images of level-3. For

level-2, the best match is found by multiplying the disparity vector of level-3 by two while for level-1 the disparity value is obtained by multiplying the corresponding LL3 disparity value by four. This results in the proposed algorithm, only having to transmit the disparity values that are in level 3 leading to a considerable decrease in the bit budget required for the transmission of the disparity vector field.

The prediction errors of all the ten subimages are then used to form the Prediction Error Image (PEI), which would initially be in the form of a 3-level decomposed image. This PEI is subsequently converted to the wavelet-block domain as explained in our previous paper [9-11]. Finally, each wavelet block in the left and the PEI undergo a JPEG-like encoder with modified scanning and quantization. [9-11].

At the decoder end, the reverse process of the encoder takes place until the right image is reconstructed.

To ensure that the new prediction reduces the energy level of the error block, its MSE is compared with the MSE of the old prediction error. If the MSE of the new prediction block is greater than the old MSE, then the old prediction error is used.

After OBM has been applied on all blocks indicated by the OBM indicator table, the new prediction errors for each subimage contains blocks predicted with and without OBM. The new prediction errors are then used to form the new PEI. Finally each wavelet block in the left image and the new PEI undergoes JPEG-like encoding as described in our work [9-11] and transmitted.

At the decoder end, by making the disparity vectors and the OBM indicator table available, the reverse process of that of the encoder is performed to obtain the reconstructed right image.
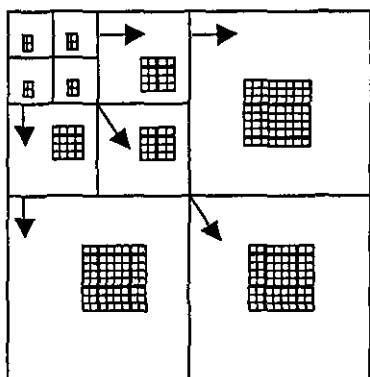

Figure 3. Hierarchical disparity estimation – Proposed CODEC

Figure 4. CODEC-1- Encoder

3.2 CODEC-2

The design of CODEC-2 is the same as CODEC-1 up to the stage where prediction errors are obtained for all the subimages. Blocks with MSE greater than a pre-specified threshold will be assigned a value 1 in a so-called, OBM indicator table, indicate that OBM needs to be performed for further refinement. An entry of 0 in the table means that OBM no applied.

Figure 5 illustrates the OBM procedure. The block is divided into four quadrants; 1, 2, 3 and 4. The prediction error of each quadrant is replaced by a new prediction error obtained by averaging its prediction error with that of its immediate neighbours. The new prediction errors are obtained as follows:

a	b	d	e
c	1	2	f
g	3	4	j
h	i	k	l

Figure 5. Block and it neighbouring blocks used in OBM

$$1_{new} = (1+a+b+c)/4$$
$$2_{new} = (2+d+e+f)/4$$
$$3_{new} = (3+g+h+i)/4$$
$$4_{new} = (4+j+k+l)/4$$

4. Experimental Results & Analysis

Both CODECs were implemented using purpose built MATLAB routines. In order to test the effectiveness of the proposed algorithms, we have simulated its performance on two pairs of stereo images; a synthesized scene, Room, and a natural scene, Aqua. The size of the stereo image pairs is 128x128. Parallel axis geometry has been used to acquire these images. The disparity vector field and OBM indicator were coded using fixed length code. Different Daubachies compactly supported orthornormal filters; db4 (for level-1), db2 (for level-2) and db1 (for level-1) were used as the wavelet transforms. Using shorter length filters for lower resolution sub-images (as above) have shown to be effective in improving the subjective and objective performance of wavelet based image CODECs.

The objective image quality is measured in terms of Peak Signal to Noise Ratio (PSNR). The compression efficiency (or Bit Rate) is measured in terms of bits per pixel (bpp) as follows:

$$BR = \left[\frac{Tot\_Bits_{comp}}{NxM} \right] bpp$$

where, N and M represent the image dimensions and Tot_Bits_{comp} is the total number of bits required to represent a given image in its compressed format. For the benchmark and CODEC-1, Tot_Bits_{comp} includes the bits that are required to code the disparity vector field, in addition to the bits requirement to encode and transmit the predictive error image. For CODEC-2, it also includes the bits required to code the OBM indicator. For a fair comparison, the objective quality of the left image (reference) was set at equivalent level (44.61 dB) for all CODECs.

For the stereo image pair 'Aqua' figure 6, compares the subjective image quality of the reconstructed right images when using the proposed and benchmark techniques.

Figure 6 shows that at a bit rate of about 0.1572 bpp, the reconstructed right images obtained using CODEC-1 and CODEC-2 have better subjective and objective quality as compared to that of Figure 6(b). This is particularly evident at the upper left edges of the stone and the small fish in the upper middle part of the image.

For further evaluation of the performance of the proposed CODECs, in figure 7, we provide the rate distortion performance graphs for the proposed and benchmark techniques. Figure 7 illustrates small improvement of CODEC-2 over CODEC-1 since with wavelet based techniques there are ringing artifacts as opposed to blocking artefacts in DCT based techniques. Thus OBM, which is mainly used to reducing the blocking artefacts has only a slight impact on the performance of the CODEC-2.

CODEC-1 and CODEC-2 show objective image quality gains of up to 1 dB over the benchmark. This is mainly due to the difference in the hierarchical disparity estimation between the proposed and the benchmark techniques.
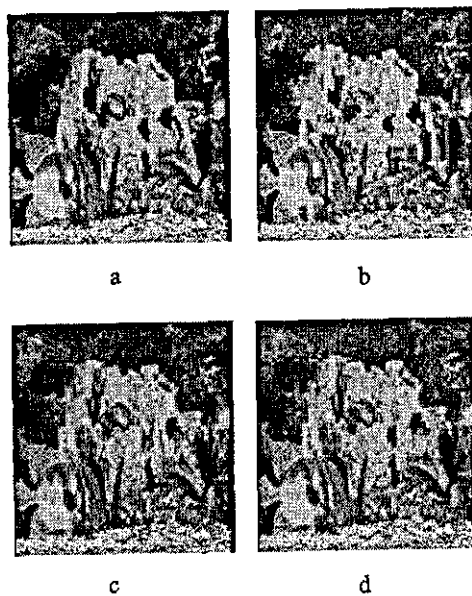

a   b

c   d

Figure 6. Subjective quality evaluation for Aqua:
(a) Original right image
(b) Reconstructed right image, benchmark, at 0.1574 bpp, PSNR 21.07 dB
(c) Reconstructed right image, CODEC-1 At 0.1575 bpp, PSNR 21.78 dB
(d) Reconstructed right image, CODEC-2 At 0.1572 bpp, PSNR 22 db

5. Conclusion

We have proposed two DWT based stereo image pair compression algorithms that performs multiresolution hierarchical disparity estimation and overlapped block matching in wavelet transform domain. The proposed CODECs have produced better performance than the benchmark with PSNR gains of up to 1 dB. These CODECs also have better subjective performance.

Fixed Length coding was used to code the disparity vector field and the OBM indicator. Further experiments showed that disparity vector field is quite smooth and OBM indicator uses value of either 0 or 1. Thus replacing Fixed Length coding with Differential Pulse Code Modulation (DPCM) would result in a more efficient coding of the above. The bits saved could be used to improve the subjective and objective quality of the predicted image.

We are currently in the process of replacing the JPEG-like CODEC used within our present research context, with a JPEG-2000 CODEC. This is expected to improve the results further.
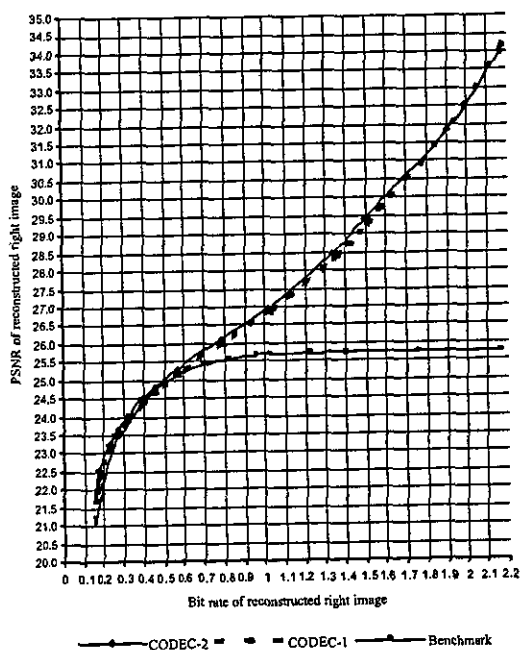
Figure 7. Rate-distortion performance
graphs

References

[1] Sriram Sethuraman, Angel G. Jordan, M.W. Siegel, "Multiresolution based hierarchical disparity estimation for stereo image pair compression," Proc. of the Symposium on Application of subbands and wavelets, March 1994.

[2] M.G.Perkins, "Data Compression of Stereopairs", IEEE Transactions on Communications, Vol. 40, No. 4, pp. 684-696, April 1990.

[3] J.Jiang, E.A.Edirisinghe, H.Schroder, "Algorithm for compression of stereo image pairs", IEE Electronic Letters, Vol. 33, No. 12, pp. 1034-1035, June 1997.

[4] E.A.Edirisinghe, J.Jiang, "A Block-Object Hybrid Approach Towards Stereo Image Compression", Proc. of the Int. Conf. On Inf., Commun. & Signal Processing, ICICS '99, Singapore, December 1999.

[5] E.A.Edirisinghe, J.Jiang, "Stereo Image Compression Approach via Pixel Value Variation Trend Adaptive Subspace Projection", Proceedings of SPIE Photonics East Conference, Boston, MA, USA, November 2000.

[6] Dinstein, M.G.Kim, J.Tzelgov, A.Henik, "Compression of Stereo Images and the Evaluation of Its Effects on 3-D Perception", SPIE Applications of Digital Image Processing XII, Vol. 1153, pp. 522-530, 1989.

[7] Woontack Woo, Antonio Ortega, "Overlapped Block Disparity Compensation with Adaptive Windows for Stereo Image Coding," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 2, pp. 194-200, March 2000.

[8] ISO/IEC JTC 1/SC 29/WG 1, ISO/IEC FCD 15444-1: Information Technology – JPEG 2000 Image Coding System: Core coding system, Mar. 2000. http://www.jpeg.org /FCD15444-1.htm

[9] M.Y.Nayan, E.A.Edirisinghe, H.E.Bez, "Baseline JPEG-Like CODEC for Disparity Compensated Residual Coding of Stereo Images," 20th Eurographics UK Conference, pp. 67-73, Leicester, UK, June 2002.

[10] M.Y.Nayan, E.A.Edirisinghe, H.E.Bez, "Two Novel Wavelet-block based Stereo Image Compression Algorithms," Proc. of the Second IASTED Intl. Conference Visualization, Imaging and Image Processing, pp. 251-256, Malaga, Spain, September 2002.

[11] R. de Queiroz, C.K.Choi, Y.Huh, K.R.Rao, "Wavelet Transforms in a JPEG-Like Image Coder," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, No.2, pp. 419-424, April 1997.

[12] David Salomon, "Data Compression Complete Reference," Springer-Verlag, 2000, Chapter 5, pp.457-567.

ELSEVIER

A wavelet implementation of the pioneering block-based disparity compensated predictive coding algorithm for stereo image pair compression

E.A. Edirisinghe*, M.Y. Nayan, H.E. Bez

Department of Computer Science, Loughborough University, Loughborough, Leicestershire LE11 3TU, UK

Abstract

In this paper, we propose a novel, discrete wavelet transform (DWT) domain implementation of our previously proposed, pioneering block-based disparity compensated predictive coding algorithm for stereo image compression. Under the present research context we perform predictive coding in the form of pioneering block search in the sub-band domain. The resulting transform domain predictive error image is subsequently converted to a so-called wavelet-block representation, before being quantized and entropy coded by a JPEG-like CODEC. We show that the proposed novel implementation is able to effectively transfer the inherent advantages of DWT-based image coding technology to efficient stereo image pair compression. At equivalent bit rates, the proposed algorithm achieves peak signal to noise ratio gains of up to 5.5 dB, for reconstructed predicted images, as compared to traditional and state of the art DCT and DWT-based predictive coding algorithms.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Stereo image compression; Discrete wavelet transform; Predictive coding; Disparity compensation; Pioneering blocks; Wavelet blocks

1. Introduction

The recent advancement of auto-stereoscopic display technologies are currently driving stereo imaging into applications domains such as digital television, computer simulations and Internet technologies that have been so far dominated by the state-of-the-art monocular image and video coding techniques. Thus at present, efficient algorithms that are capable of coding and transmitting (storing) stereo images and image sequences are attracting significant research efforts from the stereo imaging research community. To this extent various coding algorithms based on block, object- and region-based predictive coding techniques have been reported in literature [1–4,6,8–10,15,17,19–21].

In parallel to the above-mentioned shift of technology from monocular to binocular vision capability, the popular, base compression technology that was used in almost all image and video coding standards (baseline-JPEG, MPEGs and

---

*Corresponding author. Tel.: +44-1509-228234; fax: +44-1509-211586.

E-mail addresses: e.a.edirisinghe@lboro.ac.uk (E.A. Edirisinghe), m.y.nayan@lboro.ac.uk (M.Y. Nayan), h.e.bez@lboro.ac.uk (H.E. Bez).

H26x), discrete cosine transforms [14], are at present being steadily replaced by discrete wavelet transform (DWT)-based technologies that are capable of providing improved rate-distortion performance and additional functionality [1,4,7, 11–13,16–18]. The culmination of research in this direction resulted in the recent standardization of JPEG-2000 [7] and wavelet domain implementations of MPEG-4 video coding standard. As most stereo image and sequence compression techniques that have been proposed in literature are based on DCT-based technologies [2,3,6,8–10,15,19–21], currently there exists a trend in transfer of stereo imaging technology to DWT-based techniques. To this effect few wavelet-based stereo image compression algorithms have been reported in literature [1,4,17] during the past few years. Most of these algorithms are a straightforward application of the popular wavelet-based monocular image compression algorithms to stereo image compression.

Following the above-mentioned trends in technology and research directions, within the present research context, we propose a wavelet-based implementation for the pioneering block-based disparity compensated predictive coding (PBDCPC) algorithm, which was originally implemented in [9,10] using DCTs as the base technology. The PBDCPC algorithm is known for its ability to work under very low bit rate constraints [9], still producing predicted images of acceptable image quality. The high compression gains that are possible are due to the fact that it avoids the necessity of having to transmit the so-called, disparity vector field as overhead bits. Within our present research context we show that the change of base technology from DCT to DWT further improves its low bit rate performance. However, due to the specific requirements of the PBDCPC coding architecture and the properties of DWT-based multi-resolution decomposition [16,17], the above transfer of base technology is not straightforward and thus involves additional design and development of considerable novelty. The original PBDCPC algorithm is summarized in Section 2 for quick reference. Readers interested in its detailed design and behavior are advised to refer to the original work reported in [9,10].

For clarity of presentation, the rest of the paper is organized as follows. Apart from this section, which is a general introduction to the context of the proposed research, Section 2, summarizes the principles and the strict requirements behind the effective implementation of the PBDCPC algorithm. Section 3, discuses the DWT-based multi-resolution decomposition of a stereo image pair and the formation of wavelet-block-based structure, which forms the foundation for the proposed implementation. Section 4 discusses PBDCPC algorithm's implementation in wavelet domain by first discussing a multi-resolution pioneering block search (MRPBS) procedure and subsequently by re-structuring the resulting predictive error (PE) image and the reference image within a wavelet-block structure [13,16] for entropy coding and transmission. In Section 5, we compare the performance of the proposed CODEC with that of traditional DCT-based DCPC schemes [2,15] and a more recently published DWT based DCPC scheme [13] and prove the novel algorithm's improved rate-distortion performance and improved range of performance. Finally in Section 6, we conclude with suggestions for further improvements of the proposed algorithms and its possible use in stereo image sequence compression.

2. Pioneering block-based search and its special requirements

In predictive coding of stereo image pairs we assume that the reference, left image of the stereo image pair is coded independently and that the right image is predictive coded based on the reference image. Under this coding strategy it is possible to code the predictive image at a much lower bit rate as compared to the reference frame, still producing a reconstructed right image of quality equivalent to that of the reference frame, which enables good stereoscopic viewing capability. The PBDCPC algorithm follows this basic coding architecture.

Fig. 1 illustrates the *pioneering block-based search* (PBS) procedure that was used in the original PBDCPC algorithm [9].
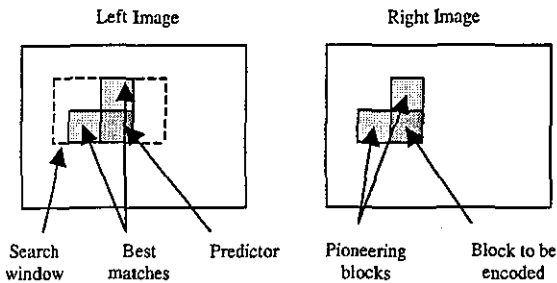
Left Image   Right Image

Search   Best   Predictor   Pioneering   Block to be
window   matches   blocks   encoded

Fig. 1. Pioneering block-based search procedure.

As illustrated in Fig. 1, to encode a given block of image pixels (say size 8×8) in the predicted (right) image, the block preceding it and the block directly above it are taken as *pioneering blocks* to search for a matching pair of blocks, within a selected windowed area of maximum likely of the reference (left) image. The best matching pair is found using mean squared error as the minimization criterion and using equal weights for the two blocks. Once the best matching pair is found, the block directly below the upper block is chosen as the predictor for the block to be encoded. The above prediction strategy exploits the intra-frame redundancy of the two constituent images of the stereo image pair and their inter-frame redundancy, together. At the decoder end, when decoding a given right image block, a search identical to that of the encoder end is performed based on it's already decoded adjacent blocks as pioneering blocks and the reconstructed left frame as the reference. Thus as compared to traditional predictive coding algorithms, the PBDCPC algorithm enables reconstruction of the right image at the decoder end without recourse to the block-disparity values, enabling a comparatively better compression performance. Note that at the boundaries of the image where one of the pioneering blocks (or both) is unavailable, the above search procedure is replaced by a PBS with one block (or by direct prediction).

Two specific criteria should be met in PBDCPC, to guarantee the reconstruction of the predicted image. Once the predictor, $\hat{L}$ for the block to be encoded, R is found, the prediction error (PE), E is calculated as, $E = R - \hat{L}$. Note the use of $\hat{L}$ to represent the fact that the search is performed on a locally reconstructed left frame. This is the first specific requirement of the PBDCPC coder, as it guarantees that the procedure adopted at the encoder end is identical to that adopted at the decoder end. This is due to the fact that at the decoder end what is available is the reconstructed reference frame. In addition to meeting the above condition, a second specific condition needs to be met. When the PE E is calculated at the encoder end, apart from it being encoded and transmitted to the decoder end, it has to be decoded and added to the predictor, $\hat{L}$, to produce the preceding pioneering block for the encoding of the next predicted frame block. Thus, always the pioneering block selection in the predicted image is done on a 'so-far' reconstructed image. This satisfies the second condition in guaranteeing that the encoder and decoder prediction environments are identical. Thus in the PBDCPC coder the encoder and decoder work on identical reference frames and selects pioneering blocks from identical, 'so-far' reconstructed predicted frames. This enables the guaranteed reconstruction of the predicted image at the decoder end, without recourse to disparity information, which is the key feature of the PBDCPC scheme. We refer readers interested in the theoretical proof of the above-mentioned quality guaranteeing procedure to [5].

Following the pixel domain PBS procedure, summarized above, to fully satisfy the second condition in guaranteeing the reconstruction of the predicted image, the PE blocks have to ultimately go through locally simulated lossy coding (transform coding/decoding and quantization/de-quantization) at the encoder end before being fed back to form the preceding pioneering block of the next block to be encoded. Thus, a straightforward block based, pixel domain PBDCPC to obtain the PE image in full, and the subsequent DWT-based coding of the same, is not a suitable DWT implementation of the PBDCPC algorithm. This is due to the fact that with such a scheme one is unable to introduce the compression losses that would be introduced to the PE blocks, before they are 'fed-back'. As a solution to the above problem, within our present research context we propose a strategy based on sub-band domain PBDCPC in which PE blocks are appropriately quantized and

de-quantized before being fed back to act as the preceding pioneering block of the next iteration. This guarantees the reconstruction of the right image under all bit rates.

In addition to specific design criteria discussed above, to allow a worthwhile comparison between the original DCT based and proposed DWT based implementations of PBDCPC, we have adapted a JPEG-like (DCT-based) coding strategy based on wavelet-block coding [16], to code the reference image and the predictive error image (PEI). With this adaptation we aim to demonstrate that the coding gains obtained under the present research context is solely due to the application of PBDCPC in sub-band domain and is not due to the use of more efficient coding techniques to code wavelet transform coefficients as against the techniques used to code the DCT coefficients. The JPEG-like coding strategy proposed in [16], which provides the basis for our design above, guarantees this. It also provides the additional advantage of using a popular, simple coding architecture in quantizing and entropy coding wavelet transform coefficients. Section 3 describes the sub-band decomposition of a stereo image pair and the conversion of such a decomposed structure to the so-called wavelet-block representation.

3. Wavelet decomposition of a stereo image pair and formation of wavelet blocks

The first stage of the proposed DWT implementation of the PBDCPC algorithm is the independent wavelet decomposition of the stereo image pair into N-levels, forming $3N + 1$ subbands. The coefficients of these sub-bands are generated by applying a cascade of two-channel filter banks to the image. Fig. 2 illustrates this decomposition structure for a given image, with $N = 3$. We name the lowest resolution band, LL3.

As described above the common practice in DWT subband coding procedures, is to group the coefficients into subband-oriented groups. In other words, coefficients of a given subband, comes from different spatial locations of the image. However in block based, DCT subband coding procedures, the coefficients are grouped into blocks, i.e. they

are from common spatial locations, but different subbands. Since the adaptation of a DCT-like (baseline-JPEG like) coding structure would enable fair comparison of the DWT-based PBDCPC algorithm with it's original DCT-based counterpart (see Section 2), we propose the transformation of the decomposition structure of Fig. 2 to wavelet-blocks as depicted in Fig. 3, prior to coding the reference frame and PE frame that results from the proposed CODEC (see Section 4). The idea behind wavelet blocks is to group the DWT coefficients into blocks, so that the grouping is similar to that used by a DCT-based subband coding procedure.

As illustrated in Fig. 3, the final result is the transformation of the input image into a re-organized structure of original image size, but consisting of a two-dimensional array of non-overlapping wavelet blocks, each representing different subband elements from the location
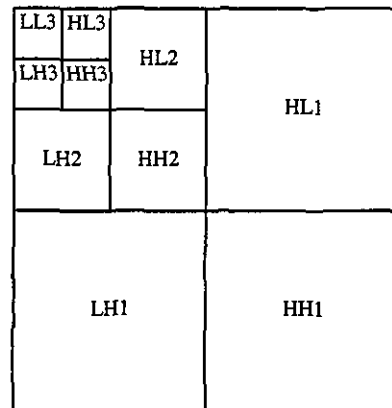

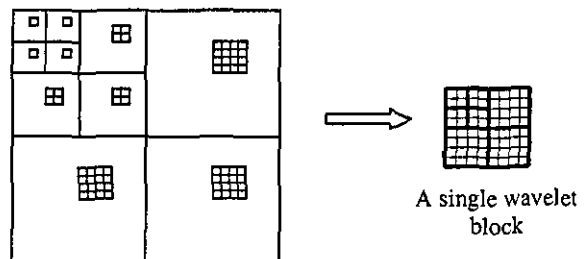Fig. 2. Three-level wavelet decomposition of an image.

Fig. 3. Formation of a single wavelet block from a three-level decomposed image.

corresponding to the block. In other words, for a S-level DWT, blocks of $2^S \times 2^S$ samples are constructed.

The following section describes the detail design of the proposed CODEC.

4. Wavelet-based implementation of the PBDCPC algorithm

Fig. 4 illustrates a high-level block diagram of the proposed CODEC. The original left image (reference) L_{ori}, is independently coded using the JPEG-like wavelet CODEC (see Fig. 5) proposed in [16]. After reconstruction at the decoder end, it provides a reference for the prediction of the right image. In addition to the above, the encoded reference image is locally decoded at the encoder end to produce a reconstructed left image L_{rec}, which is subsequently used as the reference in the wavelet-based PBS unit. Note that this satisfies the first requirement for a successful PBDCPC design. The original right image R_{ori} and reconstructed left

image L_{rec} then undergo N-level dyadic wavelet decomposition as described in Section 3. For our experiments we have used $N = 3$ and we follow the convention depicted in Fig. 2, in naming the ten resultant subbands, with the lowest subband being named as $LL3$.

Within the wavelet-based PBS unit, for each corresponding subband pair (e.g. L_{LL3} and R_{LL3}, L_{HL1} and R_{HL1}, etc.) a PBS is performed in the subband domain as described in Section 2, to produce the corresponding subband of the prediction error image (PEI). Note that we use a block size of 2×2 for PBS at the lowest sub-band level, whereas at the highest sub-band level we used a block size of 8×8. Once the PEI has been completely found (i.e. all of its subbands found), the JPEG-like CODEC [16] is used to transform it into wavelet blocks as described in Section 3 and transmit after suitable quantization and entropy coding. Fig. 4 also illustrates a PE block feedback loop, which converts the original domain PEs into the reconstructed domain by locally sending the PE through a quantization/de-quantization
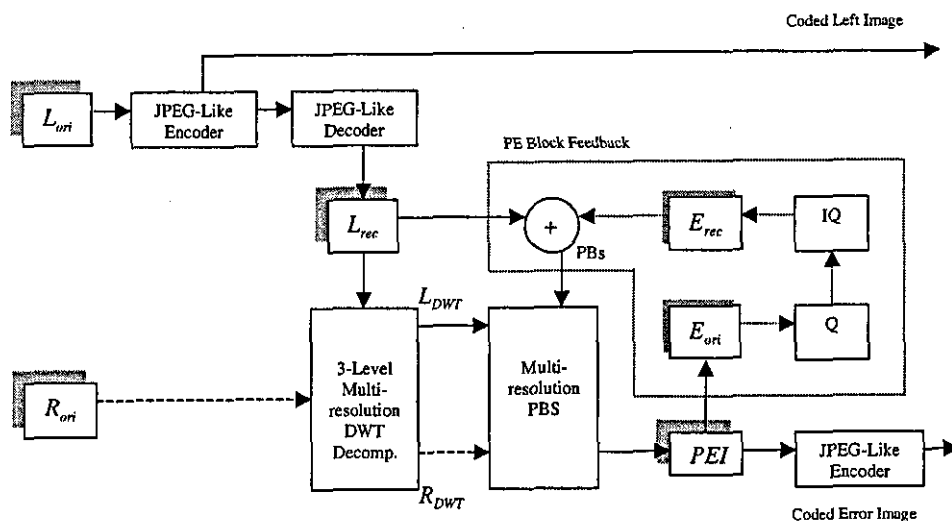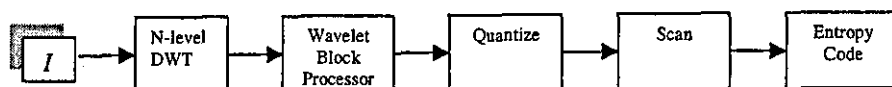

Fig. 4. Proposed encoder.

Fig. 5. JPEG-like wavelet block encoder.

$$Q_{DWT} = \begin{bmatrix} 8 & 7 & 8 & 8 & 34 & 34 & 34 & 34 \\ 7 & 7 & 8 & 8 & 34 & 34 & 34 & 34 \\ 8 & 8 & 12 & 12 & 34 & 34 & 34 & 34 \\ 8 & 8 & 12 & 12 & 34 & 34 & 34 & 34 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \end{bmatrix}$$

Fig. 6. The quantizer table.

procedure. The error block, in its reconstructed state, E_{rec} is then added to the best predictor previously found from L_{rec} to form the block, which would act as the preceding (front) pioneering block for the next iteration of the PBS. Thus the purpose of the above-mentioned error feedback loop is to satisfy the second requirement for a successful PBDCPC CODEC design (see Section 2).

Within the JPEG-like coding strategy adopted for coding wavelet-blocks, of both left image and PEI, we have used the quantizer depicted in Fig. 6. Thus all coefficients in band HH3 are quantized using 7 as the scalar quantizer and all coefficients in band HH1 are quantized using 55 as the quantization step size. This simple quantizer step allocation simplifies the above error-block feedback procedure considerably by providing a single factor for quantizing all coefficients of a PE block at each sub-band level.

The decoder of the proposed CODEC works in a manner similar to the encoder. Due to the orientation of the two pioneering blocks relative to the block to be encoded, in encoding blocks in the first row of blocks (other than the first block itself), of a given subband, only the block above is used as the pioneering block. A similar strategy in adapted to resolve the problem of encoding blocks in the first column of blocks. The block at the top, left-hand corner of a sub-band is always coded directly, i.e. without any disparity compensation.

5. Experimental results and analysis

The proposed CODEC was implemented using purpose built MATLAB routines and was tested on a set of eight commonly used test stereo image pairs representing indoor, outdoor, natural and synthetic scenes. All test stereo image pairs were of size 512×512. In order to evaluate the performance of the proposed CODEC, we compare its performance against that of the well-known benchmark algorithm of Perkin's (benchmark-1), that uses DCT as the basic compression engine, our previously proposed CODEC of Nayan et al. [13] (benchmark-2), which is essentially a wavelet-block-based implementation of benchmark-1 and our original DCT-based PBDCPC algorithm of Jiang et al. [9,10] (benchmark-3). As parallel axis camera geometry has been used to acquire these images we have limited the PBS along a horizontal direction only. For all four CODECs, a windowed search area representing a disparity range of 0–7 pixel shifts was used. For the benchmark-1 and benchmark-2 CODECs, the disparity vector field was coded using a fixed length code, requiring three bits per block. A Daubachies compactly supported orthonormal filter (db7) was used as the wavelet transform.

The objective image quality is measured in terms of peak signal to noise ratio (PSNR), whereas the compression efficiency is measured in bits per pixels (bpp). The compression efficiency (or bit rate) is measured as follows:

$$BR = \left[\frac{Tot\_Bits_{comp}}{N \times M} \right] bpp,$$

where N and M represent the image dimensions and Tot_Bits_{comp} is the total number of bits required to represent a given image in its compressed format. For benchmark-1 and benchmark-2, Tot_Bits_{comp} include the bits that are required to code the disparity vector field, in addition to the bit requirement to encode and transmit the PEI. However, for the proposed CODEC and benchmrk-3, Tot_Bits_{comp} represents only the bit requirement to encode the PEI. For a fair comparison, for a given stereo image pair, the objective quality of the left image (reference) was set at equivalent levels for all four CODECs.

As illustrated in Fig. 7, the rate-distortion graphs show that the proposed PBDCPC CODEC outperforms the benchmark-1 and benchmark-2 algorithms by considerable margins both at
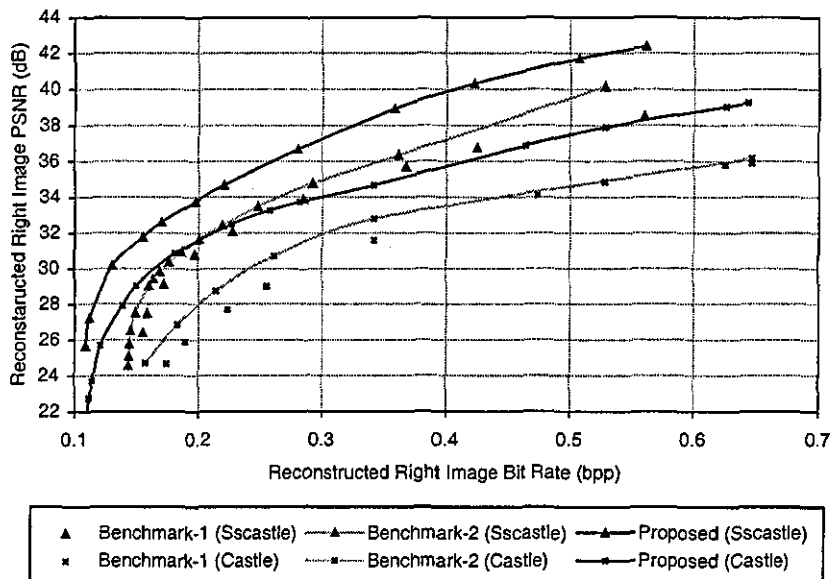
Fig. 7. Comparison of rate-distortion performance graphs.

medium and low bit rates. At a reconstructed right image bit rate of 0.2 bpp, for Castle image, the proposed CODEC produces a PSNR gain of approximately 3.5 dB as against benchmark-2 and 5 dB as against bechmark-1. For Sscastle image improvements of up to 3 dB are indicated at this bit rate. At higher bit rates, improvements in the range of 2–4 dB are indicated. Given the fact that the only difference between benchmark-1 and benchmark-2 is that in benchmark-2, the DCT-based coding has been replaced by a DWT-based coding technique, which uses a quantization and entropy coding stage that performs equivalently to that of the benchmark-1, all improvements in the rate-distortion performance could be accounted towards the transfer of base technology from DCT to DWT. Note that the coding PSNR gains obtained using this replacement does not exceed 2.5 dB. Arguing in similar lines, as the only difference between the benchmark-2 and the proposed CODEC is the replacement of the traditional disparity compensation by PBDCPC, the outstanding PSNR gains of the range 2–7 dB shown against benchmark-2 CODEC, could be accounted to the non-transmission of the disparity vector field. The gains are mostly significant at very low bit rates (7 dB at 0.14 bpp for Sscastle),

where, with the scheme of benchmark-2, the majority of the bit budget would be used to transmit the disparity vector field. With the proposed CODEC the complete bit budget could be allocated to the coding of the PEs.

Further analysis of the results in Fig. 7 indicates the ability of the proposed PBDCBC CODEC to perform at bit rates as low as 0.11 bpp. In contrast, the benchmark-1 and benchmark-2 CODECs are unable to perform at bit rates lower than 0.145 bpp. This is due to the fact that both of these benchmark CODECs require an allocation of a certain fixed bit budget (0.047 bpp in our experiments) for coding the resultant disparity vector fields.

Fig. 8 compares the subjective image quality performance of the proposed CODEC against the benchmark-1 CODEC at a bit rate of 0.15 bpp. It clearly illustrates that in the proposed CODEC the image quality near object boundaries are much better as compared to that of benchmark-1. In predictive coding of stereo image pairs, the PEs are highest near object boundaries, which represent areas of binocular disparity. Therefore the PE frames have highest values in these regions. Thus when these frames are coded under constrained bit budgets, the maximum loss is introduced in object
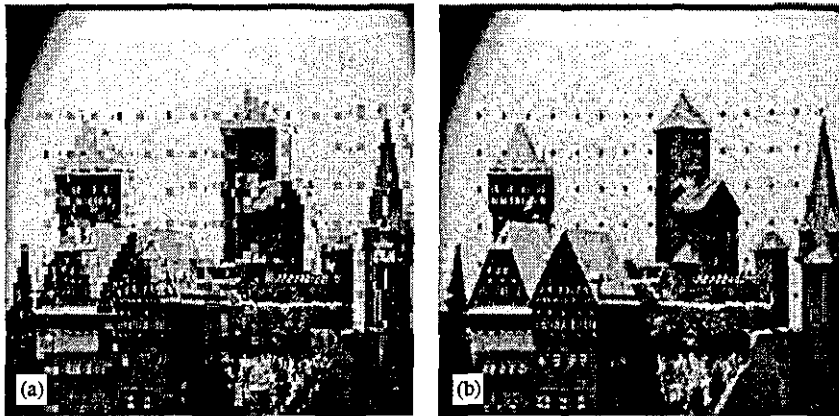
Fig. 8. Subjective quality evaluation for castle reconstructed right image: (a) Benchmark-1, 0.1550 bpp, 21.29 dB; and (b) proposed CODEC, 0.1498 bpp, 29.03 dB.

boundary areas. This is particularly true at bit rates lower than 0.25 bpp for most images. Even though the proposed CODEC does not result in a more accurate prediction, the bit budget that could be allocated to coding the PEs is higher as compared to benchmark techniques 1 and 2 due to the fact that no bits have to be allocated for the coding of disparity vector fields. Particularly at very low bit rates, the bits required for the lossless coding of disparity vector fields is significantly higher than what is required for coding the PE fields. Thus the proposed CODEC's relative performance as compared to the benchmark CODECs 1 and 2 improves with the decrease of bit rate. A closer analysis of the Fig. 8(a) indicates that the subjective quality degradation near object boundaries is of a *blocky* nature and in Fig. 8(b) it is of *ringing* nature. This is due to the inherent characteristics of DCT- and DWT-based compression. However, due to the usage of the pioneering block technique, in Fig. 8(b) the otherwise expected significant ringing effects have been largely eliminated.

In order to further analyze the performance of the proposed DWT implementation of the PBDCPC algorithm, in Fig. 9, we compare its rate-distortion performance with that of benchmark-3, the original DCT implementation [9,10] of the PBDCPC algorithm. It illustrates the fact that the performance of the proposed CODEC supersedes the performance of the benchmark-3

CODEC, particularly at medium and high bit rates (approx. 2–4.5 dB gain). Note that the only architectural difference between the two CODECs is that the proposed CODEC uses DWT technology as the base compression technique as against the benchmark-3 using DCT technology. However, the results in Fig. 7 have previously shown that the PSNR gains obtainable under such a change of base technology is limited to about 1–2.5 dB for the test images under consideration. Therefore the additional 1–2 dB gain that is obtained with the proposed CODEC as against the benchmark-3 CODEC should be due to the inherent advantage in using PBS in the subband domain as against its use in pixel domain. In Section 4 it was shown that in the design of the PBDCPC algorithm, two special conditions need to be satisfied. The first is that the search for the best match should be done on a reconstructed reference frame. The second condition was that the selection of the two pioneering blocks of a block to be encoded should be done on a "so far" reconstructed right image. Both these conditions mean that the selections for the pioneering blocks from the "so far" reconstructed right image (*selection image*-1) and their best matching blocks from the reference image (*selection image*-2) are done on lossy *selection* images. However, in the case of using DCT as the base compression technique, the two, pixel domain selection images would consist of artificial block boundaries due to
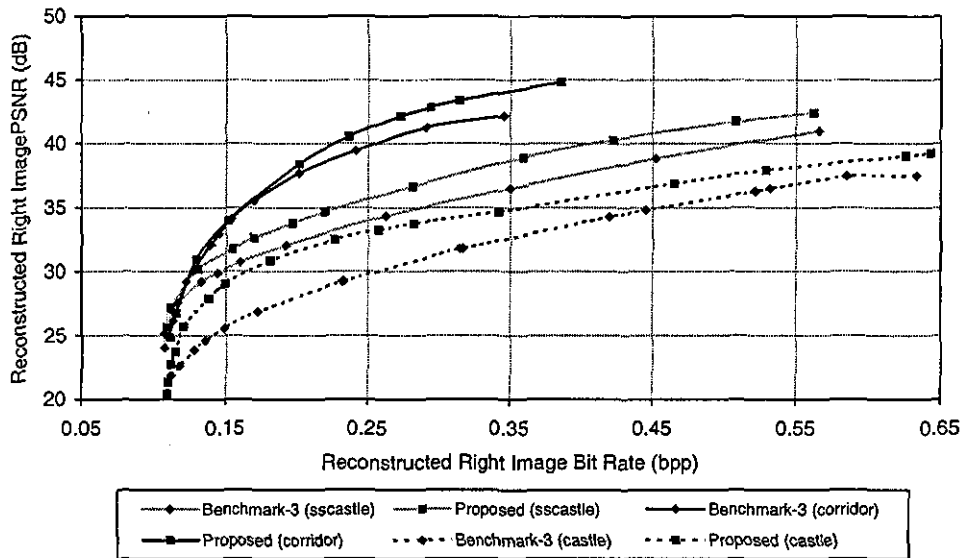
Fig. 9. Comparison of DCT-based vs. DWT-based PBDCPC Algorithms.

the presence of blocking artifacts inherent to DCT-based compression schemes. This would result in a significant reduction in redundancy between adjacent blocks at very low bit rates of coding. The corresponding effect in DWT-based schemes is ringing. However in the proposed DWT implementation, the two selection images are in the transform domain and as a result would not contain any ringing artifacts. However due to the introduction of coding losses in the form of uniform scalar quantization/de-quantization, the redundancy between adjacent blocks would still be marginally reduced. Therefore, as the accuracy in prediction of PBS schemes depends on the presence of redundancy between adjacent blocks in the constituent images of a stereo image pair, PBS in the DWT domain (i.e. in the proposed CODEC) would be more efficient compared to the PBS in the pixel domain (in the original CODEC). This is particularly true at very low bit rates of coding.

6. Conclusion and future

In this paper, we have proposed an efficient DWT implementation of our previously proposed DCT-based pioneering block-based stereo image compression algorithm. We have shown that due to the specific requirements of the design of the CODEC to guarantee the reconstruction of predicted images, without the use of disparity information, the DWT implementation is not a straightforward conversion of the base technology from DCT to DWT. The novel DWT-based PBDCPC scheme results in improved, objective rate-distortion performance and a subjective quality performance. PSNR improvements of up to 5.5 dB are reported for transmission bit rates of 0.15 bpp. In addition to this the proposed CODEC is able to operate at bit rates as low as 0.1 bpp, where other CODECs fail to operate due to the need of a certain minimum bid budget for encoding disparity vector fields.

Further improvements to the proposed technique, in terms of algorithmic and computational efficiency is possible by using faster multi-resolution search techniques [17] that avoid searches in all sub-bands but would rather effectively re-use search information at lower resolution sub-bands within the search operations of higher resolution sub-bands. We are currently in the process of expanding the ideas proposed in this paper to wavelet-based stereoscopic sequence coding.

References

[1] N.V. Boulgouris, M.G. Strintzis, Embedded coding of stereo images, Proceedings of the IEEE International Conference on Image Processing, ICIP 2000, Vancouver, Canada, June 2000.

[2] I. Dinstein, M.G. Kim, A. Hanik, Compression of stereo images using subsampling and transform coding, J. Opt. Eng. 30 (September 1991) 1359–1364.

[3] N.D. Doulamis, A.D. Doulamis, Y.S. Avrithis, K.S. Ntalianis, S.D. Kollias, Efficient summarization of stereoscopic video sequences, IEEE Trans. CSVT 10 (4) (June 2000) 501–517.

[4] E.A. Edirisinghe, Rate scalable stereo image coding using EZW algorithm, Proceedings of the IEEE International Conference on Information, Communications and Signal Processing, ICICS 2001, Singapore, October 2001, ISBN: 981-04-5149-0.

[5] E.A. Edirisinghe, J. Jiang, Towards eliminating disparity field coding of stereo image pairs, Proceedings of the IEEE International Conference on Information, Communication & Signal Processing, Singapore, December 1999.

[6] N. Grammalidis, D. Beletsiotis, M.G. Strintzis, Sprite generation and coding in multiview image sequences, IEEE Trans. CSVT 10 (2) (March 2000) 302–311.

[7] ISO/IEC JTC 1/SC 29/WG 1, ISO/IEC FCD 15444-1: Information Technology—JPEG-2000 Image Coding System: Core Coding System, March 2000, www.jpeg.org/ FCD15444-1.htm.

[8] J. Jiang, E.A. Edirisinghe, A hybrid scheme for low bit-rate coding of stereo images, IEEE Trans. Image Process. 11 (2) (February 2002) 123–134.

[9] J. Jiang, E.A. Edirisinghe, H. Schroder, Algorithm for the compression of stereo image pairs, IEE Electron. Lett. 33 (12) (June 1997) 1034–1035.

[10] J. Jiang, E.A. Edirisinghe, H. Schroder, A novel predictive coding algorithm for 3-D image compression, IEEE Trans. Cons. Elec. 43 (3) (August 1997) 430–437.

[11] S.A. Martucci, I. Sodagar, Zerotree entropy coding of wavelet coefficients for very low bit rate video coding, Proceedings of the 1996 IEEE International Conference on Image Processing, Lausanne, Switzerland, September 1996.

[12] S.A. Martucci, I. Sodagar, T. Chiang, Y. Zhang, A zerotree wavelet video coder, IEEE Trans. CSVT 7 (1) (February 1997) 109–118.

[13] M.Y. Nayan, E.A. Edirisinghe, H.E. Bez, Baseline JPEG-Like DWT CODEC for disparity compensated residual coding of stereo images, Proceedings of the Eurographic 2002, UK, June 2002, pp. 67–74.

[14] W.B. Pennebaker, J.L. Mitchell, JPEG: Still Image Compression Standard, Van Nostrand Reinhold, New York, 1993.

[15] M.G. Perkins, Data compression of stereo pairs, IEE Trans. Commun. 40 (4) (April 1992) 684–696.

[16] R. de Queiroz, C.K. Choi, Y. Huh, K.R. Rao, Wavelet transforms in a JPEG–Like image coder, IEEE Trans. Circuits Syst. Video Technol. 7 (2) (April 1997) 419–424.

[17] S. Sethuramam, M.W. Siegel, A.G. Jordan, A multi-resolution framework for stereoscopic image sequence compression, Proceedings of the ICIP-95, Washington, USA, Vol. 2, pp. 361–365.

[18] K. Shen, E.J. Delp, Wavelet based rate scalable video compression, IEEE Trans. CSVT 9 (1) (February 1999) 109–122.

[19] R. Wang, Y. Wang, Multiview video sequence analysis, compression and virtual viewpoint synthesis, IEEE Trans. CSVT 10 (3) (April 2000) 397–410.

[20] W. Woo, A. Ortega, Optimal blockwise dependent quantization for stereo image coding, IEEE Trans. CSVT 9 (6) (September 1999) 861–867.

[21] W. Woo, A. Ortega, Overlapped block disparity compensation with adaptive windows for stereo image coding, IEEE Trans. CSVT 10 (2) (March 2000) 194–200.

A Family of Wavelet Transform Based Stereo Image Compression Algorithms

M.Y.Nayan, E.A.Edirisinghe, H.E.Bez

Department of Computer Science, Loughborough University, UK

Abstract: In this paper we propose four, wavelet-block based stereo image compression algorithms that use a simple coding architecture similar to that of a DCT based baseline-JPEG CODEC. Due to this unique design, the proposed CODECs are able to transfer the inherent advantages of DWT-based technology to stereo image compression, without the usual extra algorithmic and implementation complexity that otherwise needs to be introduced. The four CODECs differ in the manner in which disparity compensated predictive coding is performed in predicting one image from its reference image. We provide experimental results to evaluate the detailed performance of the proposed CODECs and compare them with the performance of two well-known predictive coding based benchmark algorithms for stereo image compressions. We report PSNR gains of up to 7 dB at very low bit rates, for the CODEC with the best performance.

1. Introduction

Recent increase in the demand for more realistic viewing media, in conjunction with the advances in auto-stereoscopic display technology, are at present driving the use of stereo imaging into diverse areas of applications such as, remote surveillance, medical imaging, HDTV, virtual reality, tele-medicine, tele-robotic and CAD/CAM. Therefore research into the design and development of more efficient, functional, stereoscopic imaging systems, is attracting considerable attention from the image and video coding research community, at present [1-19].

In capturing a 3D scene, one simulates the human's eyesight effect by independently capturing two, two-dimensional images of the scene, through two horizontally separated perspectives. These two images are labeled as left image (left eye view) and right image (right eye view). If the above images were to be separately transmitted with the intention of reconstructing the 3D scene at the receiver end,

using binocular fusion, it would require double the bandwidth otherwise required for the transmission of a monocular image. Therefore, an essential requirement of a stereoscopic imaging system is the efficient compression of stereo image pairs. Although a series of JPEG [20,21] and MPEG standards represent worldwide efforts for developing image and video compression technology, they are not optimal for the compression of stereo image pairs. This is due to the fact that in stereo image pairs, for optimal compression, the presence of inter-frame redundancy and the usual intra-frame redundancy should both be exploited effectively. Unfortunately the direct, independent use of monocular image compression schemes [20,21], are unable to exploit inter-frame redundancy. Thus specific coding algorithms are required for coding stereo image pairs.

The most popular stereoscopic image-coding scheme is predictive coding. Under this scheme, the left image is chosen as the reference image and is compressed using some well-established monocular image compression algorithm/standard. Then, given the full knowledge of the left image, the right image (predicted image) is predictive coded. The most common method used in literature is fixed size block-based, disparity compensated predictive coding [2,3], which results in a Predictive Error (PE) image, that is usually coded using the compression engine used to independently compress the reference (left) image. The coded reference and PE images are transmitted to the decoder, along with the disparity vector field, which is used at the decoder to reconstruct the predicted image based on the reconstructed reference and PE image. However, in most such attempts the basic compression engine that is used to compress the reference and PE images, is the DCT-based baseline-JPEG standard [20]. As DCT based compression schemes suffer from homogeneous blocking artifacts at low bit rates, such schemes are not suitable for low bit rate stereo image coding. However, they have the advantage of having a simple, well-established coding architecture, which may be beneficial in certain application areas.

To cater for the continual expansion of multimedia and Internet applications, image compression technologies have to continually improve and evolve in functionality and efficiency. Worldwide efforts in this direction have recently lead to the standardization of DWT based JPEG-2000 [21], which would soon replace DCT-based image compression technology offered by the baseline-JPEG standard. The dyadic DWT has versatile time-scale localization characteristics [25,26] due to a pyramid-like multi-resolution decomposition structure and offers superior rate-distortion performance characteristics as compared to DCT. Lead by the ideas of JPEG-2000, DWT technology is currently destined to lead the research and developments efforts in image and video coding fields at least for the next decade. Therefore it is essential that every effort be taken to explore the possibility of using

DWT technology in stereo imaging, which is widely expected to be the imaging technology of future. Few attempts have already been made in this direction [6,16,17,19] within the stereo image coding research community.

In this paper we propose four DWT-based stereo image compression algorithms that preserve the simple block based coding architecture of DCT based baseline-JPEG standard, by arranging the wavelet coefficients in so-called wavelet blocks, prior to further processing [24]. Thus the proposed algorithms are able to retain the best of both approaches, i.e. simplicity of baseline-JPEG architecture and the superior rate-distortion efficiency of DWT based coding. Further, by adapting this hybrid coding architecture, the proposed CODECs avoid the necessity of buffering the images and multiple passes and thus avoid complex processing. The four CODECs, namely CODEC-1, CODEC-2,CODEC-3 and CODEC-4, while adopting the Wavelet-Block structure to rearrange the DWT coefficients and for subsequent coding, differs in terms of the method that is used for disparity compensation and prediction. In CODEC-1, the traditional block based disparity compensated predictive coding [2,3] is done in the pixel domain. In CODEC-2, the same is performed in the wavelet-block domain. In CODEC-3, pioneering block based disparity compensated predictive coding [9,10] in the wavelet-block domain is used. In CODEC-4, we perform, multi-resolution pioneering block based disparity compensated predictive coding. Table. 1 summarizes the Disparity Compensation and Prediction Coding (DCPC) techniques used by the four CODECs (see Appendix I for search procedure used in the four CODECs).

Name of CODEC	Method of DCPC
CODEC-1	In pixel domain, with traditional DCPC
CODEC-2	In wavelet-block transform domain, with traditional DCPC
CODEC-3	In wavelet-block transform domain, with pioneering block based DCPC
CODEC-4	In multiresolution DWT domain, with pioneering block based DCPC

Table-1: Summary of DCPC schemes used in the proposed CODECs

The rest of the paper is arranged as follows. In section 2 we introduce dyadic wavelet decomposition of an image and the formation of wavelet-blocks. Section 3 contains a comprehensive insight into the proposed coding algorithms. Section 4 provides experimental results and analyses the effectiveness of the proposed CODECs under varying bit rate settings. Section 5 concludes with a foresight to the future research in this direction.

3

2. Wavelet Decomposition & The Formation of Wavelet Blocks

In applying DWT, the coefficients are generated by applying a cascade of two-channel filter banks to the image. Each image is decomposed into a collection of $3N+1$ subband images, where N is the number of levels. Figure 1 represents a wavelet decomposition of an image, with $N=3$. Specifically, a 1-dimensional low pass filter (L) and a 1-dimensional high pass filter (H), each based on a mother wavelet [25,26], are applied horizontally and vertically to the original image, and the outputs are subsampled by a factor of 2, to create 4 subband images. A first DWT stage (level 1) decomposes the image into four subbands, denoted by LL1 (low-resolution version of original image), HL1 (vertical features of the image), LH1 (horizontal features of the image) and HH1 (diagonal features of the image). The next DWT stage decomposes this LL1 subband into four more subbands, denoted LL2, LH2, HL2 and HH2. The process continues for some number of stages, N, producing a total of $3N+1$ subbands whose samples represent the original image.
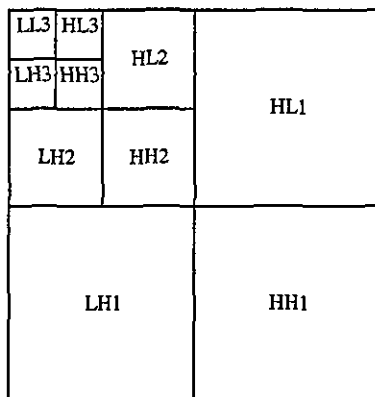
Figure 1. Three level wavelet decomposition

Usually in DWT subband coding, the coefficients are grouped to subband oriented groups (common subband, different spatial location) whereas in DCT subband coding, they are grouped into blocks (common spatial locations, different subbands). The idea behind *wavelet-blocks* is to group the DWT coefficients into blocks as illustrated in figure 2, so that the grouping is similar to that used by a DCT-based subband-coding.

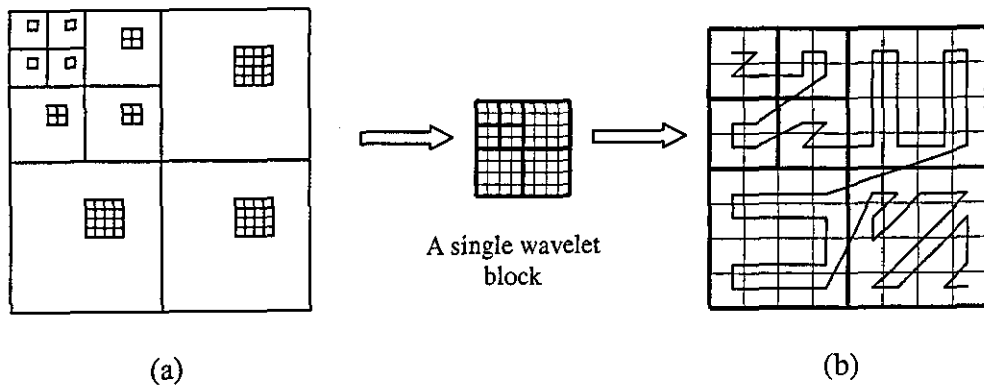(a)   (b)

Figure 2. Formation of a single wavelet block from a 3-level decomposed image
(a) decomposed image (b) scan order per single wavelet block

The final result is the transformation of the input image into a re-organized structure of original image size, but consisting of a two-dimensional array of non-overlapping wavelet-blocks, each representing different subband elements from the location corresponding to the block. In other words, for a N-level DWT, blocks of $2^N \times 2^N$ samples are constructed. The above conversion is carried out in all proposed CODECs. The following section introduces the reader to the specific details of these CODECs.

3. Proposed Wavelet-Based Stereo Image CODECs

3a. CODEC-1

Figure 3 represents the block diagram of the proposed CODEC-1. Note that it is assumed here that the disparity values are suitably coded using a fixed length code and transmitted losslessly to the decoder end [21].

At the encoder end, the right image is divided into non-overlapping 8×8 pixel blocks. For each of these blocks a search is performed within a windowed area of maximum likelihood on the original left image (considered as reference image) to find the best predictor. Assuming that parallel axis camera geometry is used to obtain the stereo image pair under consideration, one could limit the above mentioned search window in only a horizontal direction, spanning to the left of the corresponding block (of the block to be encoded) in the left image [3]. The best predictor is found using the Mean Squared Error (MSE) as the selection criteria. The prediction errors of all the non-overlapping right frame blocks are then used to form the Prediction Error (PE) image.
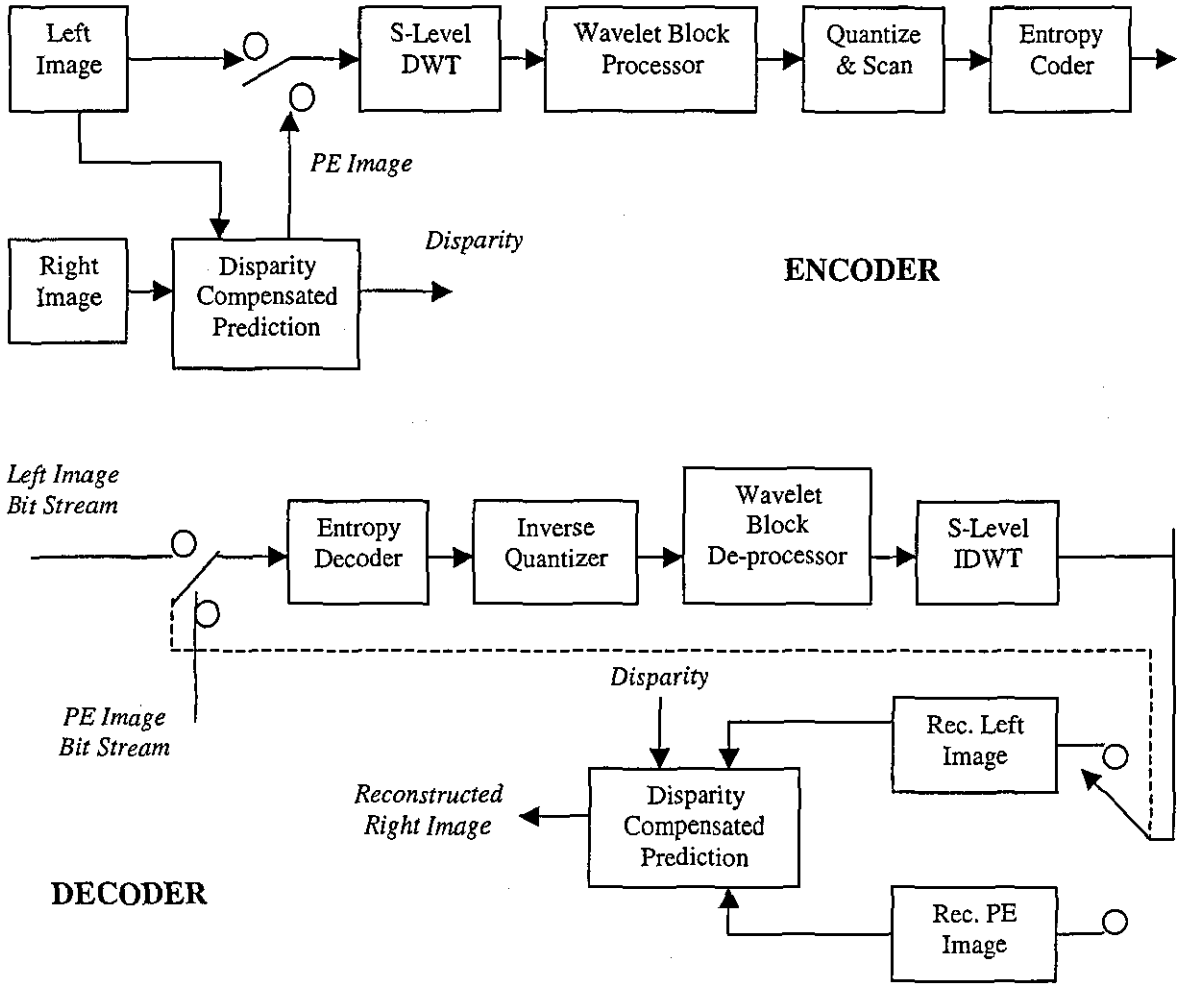
5

Figure 3. Proposed CODEC-1

The left image, L and the PE image then undergo 3-level dyadic DWT separately and are subsequently converted into their wavelet block representations, by following the procedure described in detail in Section 2. Each wavelet block of the two representations then undergo scanning as illustrated in Figure 1(b) as against the zig-zag scanning procedure adapted by baseline-JPEG [20]. This modification of scanning procedure is essential as in the wavelet block representation, for a given block; there are coefficients that not only belong to different subbands, but also multiple coefficients which belong to the same subband [24].

After the modified scanning procedure discussed above, the ordered coefficients undergo scalar quantization. We use the strategy adapted in [24] to determine a suitable quantization table, that provides results equivalent to the 64-entry, uniform, fixed quantization table used by the baseline-

6

JPEG standard. After the quantization, the quantized coefficients are subjected to entropy coding (runlength/Huffman coding) as in baseline-JPEG. The decoder operations are essentially the inverse processes of the encoder processes and are illustrated separately in figure 3.
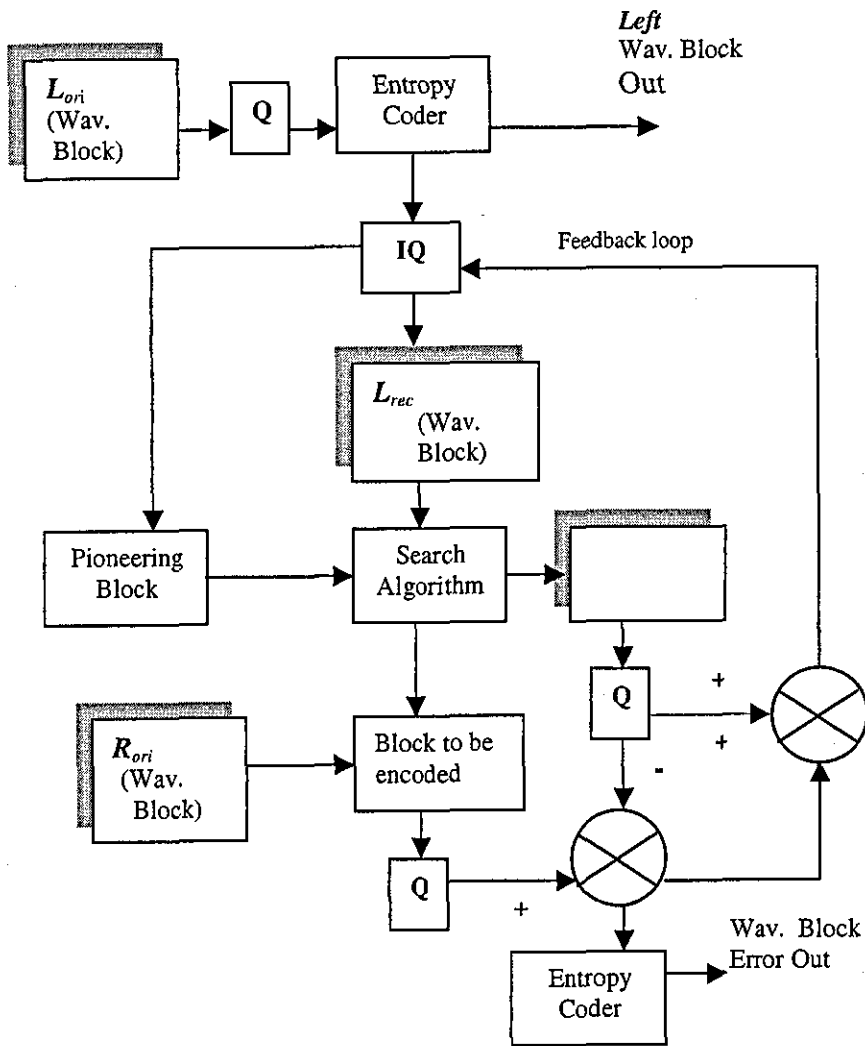
3b. CODEC-2

In CODEC-2, the disparity compensated predictive coding is done in transform, i.e. wavelet-block domain, rather than in the pixel domain as in CODEC-1. Thus at the encoder end, the left image (reference image) and the right image (predicted image) undergo a 3-level dyadic DWT separately and are subsequently converted into wavelet-block images (say WL and WR respectively) as described in Section 2. For each wavelet-block in WR a search is performed in WL for best matching block, within a horizontally oriented windowed area of maximum likelihood on WL. Note that due to the nature of the wavelet-block domain, the above search is performed with 8 pixels jumps, considered as one unit shift. The best predictor is found using the MSE as the selection criteria. The prediction errors of all the non-overlapping right frame wavelet blocks are then used to form the Prediction Error (PE) image, which would be obtained in wavelet-block domain. Finally, each wavelet block in WL and PE undergo modified scanning, scalar quantization and entropy coding as described for CODEC-1 (section 3b). The decoder operations are essentially the inverse processes of the encoder processes.

3c. CODEC-3

In CODEC-3, the traditional disparity compensated predictive coding [2,3] is replaced by pioneering block based predictive coding, originally proposed by us in [9,10] for DCT based disparity compensated predictive coding. Pioneering block based predictive coding is used to avoid the necessity of transmitting the disparity vector field altogether, thus achieving better compression gains.

Figure 4 illustrates the block diagram of the Encoder of CODEC-3. The left and right images are first transformed into their respective wavelet-block domains as discussed in Section 2. Firstly, the left image (reference) is directly transmitted as discussed in CODEC-1. A sample of the left image is simultaneously reconstructed at the encoder end, to be used in the pioneering block based search. Subsequently, for each wavelet-block to be encoded in the right image, a *pioneering wavelet-block*, which is the wavelet-block preceding it, is used to search for the best matching wavelet-block in the reconstructed and re-transformed (to wavelet-block domain) left image. Once the best match is found

7

the block to the right of the best match is selected as the best match for the wavelet-block to be encoded and the corresponding prediction error wavelet-block is formed, which is subsequently added to the prediction error image.

Figure 4. CODEC-3: Encoder

At the decoder end, each right image wavelet-block is reconstructed using an identical procedure to that of at the encoder. As at the decoder the search is performed for the best match, in a reconstructed and re-transformed left image, at the encoder end it is essential that the search is performed in an identical reference image, thus the reason for using a reconstructed left frame for pioneering block based search above. In addition to this, the preceding wavelet block to the block to be decoded, which is used as the pioneering block for the search of the best matching block, is identical to its counterpart

at the encoder end, due to the feedback loop shown in Figure 4, i.e., of the encoder. These two requirements guarantee the reconstruction of the predicted image without the need for the transmission of the disparity vectors. Readers interested in more detail of this search technique can refer to Appendix I in this paper. The savings of the bit-budget that is obtained due to the non-necessity of disparity vector field transmission is expected to improve the CODEC-3's rate distortion performance over both CODEC-1 and CODEC-2.

3d. CODEC-4

In CODEC-4, the pioneering block based predictive coding is performed in the multiresolution, DWT domain, instead of the wavelet-block domain. Figure 5, illustrates the detailed block diagram of the Encoder of CODEC-4.
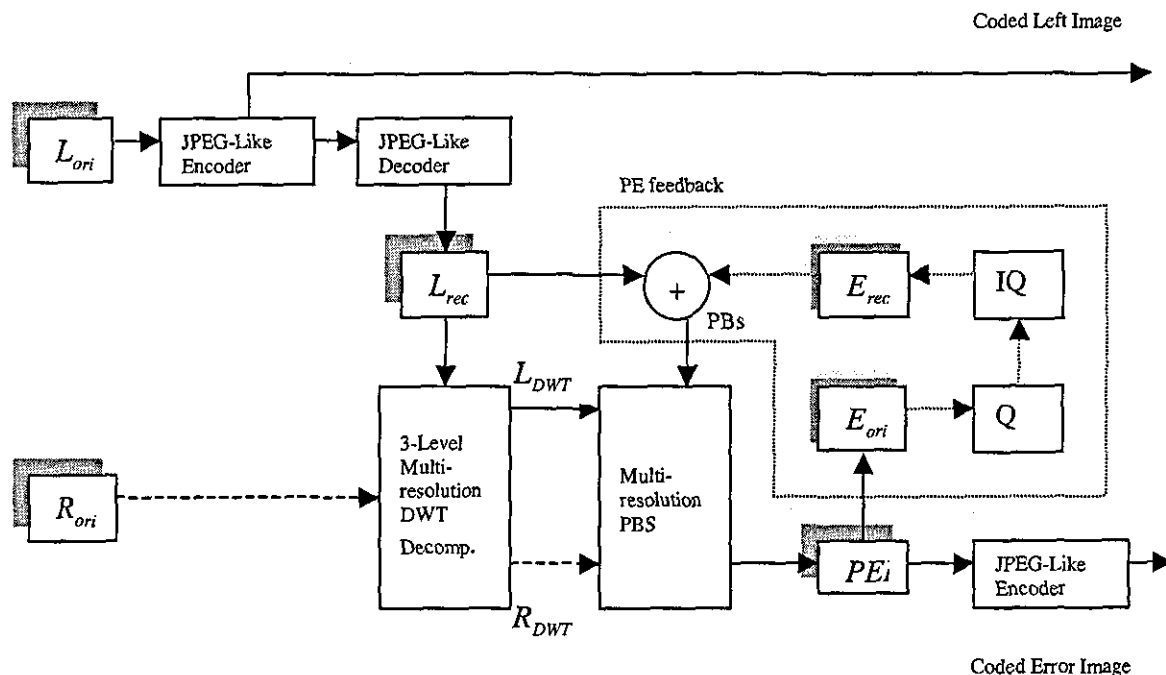

Figure 5. CODEC-4: The Encoder

The original left image (reference) L_{ori}, is independently coded using the baseline-JPEG like coding procedure discussed in sections 2 and 3b [24]. After reconstruction at the decoder end, it provides a reference for the prediction of the right image. In addition to the above, the encoded reference image is locally decoded at the encoder end to produce a reconstructed left image L_{rec}, which is subsequently used as the reference image in the wavelet based pioneering block search (PBS) unit.

9

Note that by adapting this strategy we satisfy the first condition of guaranteeing the reconstruction of the predicted image, under the pioneering block based coding technique [9]. The original right image R_{ori} and reconstructed right image L_{rec} then undergo N level dyadic wavelet decomposition as described in section 2. For our experiments we have used $N=3$ and we follow the convention illustrated in figure 1, in naming the ten resultant subbands, with the lowest resolution subband being named as *LL3*.

Within the wavelet based PBS unit, for each corresponding subband pair (e.g. L_{LL3} and R_{LL3}, L_{HL1} and R_{HL1}) a PBS is performed as described in section 3c, to produce the corresponding subband of the Prediction Error Image (PEI). Generally, for the PBS search associating a sub-band pair at level r, we use a pioneering, coefficient-block size of $2^{N-r} \times 2^{N-r}$. Once the prediction error image has been completely found (i.e. all of its subbands found), the JPEG-like CODEC [24] is used to transform them into wavelet blocks (see Section 2) and transmit after suitable quantization and entropy coding. Fig. 5 also illustrates a prediction error (PE) block feedback loop, which converts the original domain prediction errors into the reconstructed domain by locally sending the PE through a quantization/de-quantization procedure. The error block, in its reconstructed state, E_{rec} is then added to the best predictor previously found from L_{rec} to form the block, which would act as the preceding (front) pioneering block for the next iteration of the PBS. Thus the purpose of the above mentioned error feedback loop is to satisfy the second requirement for a successful pioneering block based CODEC design [11].

Within the JPEG-like coding strategy adopted for coding wavelet-blocks, of both left image and predictive error image, we have used the quantizer depicted in figure 10. Thus all coefficients in LL3 bands are quantized using 8 as the scalar quantizer and all coefficients in band HH3 are quantized using 55 as the quantization step size. This simple quantizer step allocation simplifies the above error-block feedback procedure considerably by providing a single factor for quantizing all coefficients at each subband level. The decoder of the proposed CODEC works in a manner similar to the encoder.

4. Experimental Results & Analysis

The proposed CODECs were implemented using purpose built MATLAB routines and were tested on a set of six commonly used test stereo image pairs representing indoor, outdoor, natural and computer

rendered scenes. All test stereo image pairs were of size 512×512. Parallel axis camera geometry has been used to acquire these images and we assume that pixels at the same spatial location to have the same brightness. Thus a horizontally spread windowed search area representing a disparity range of 0-7 pixels was used. For CODEC-1 and CODEC-2, the resulting disparity vector fields were coded using a fixed length code, requiring three bits per block. Daubechies' compactly supported orthornormal filters [25] have been used to reduce the filtering complexity. It has been observed [26] that there is no appreciable improvement in using filter orders above six. Hence in our experiments, we have chosen DAUB7 filters for all CODECs.
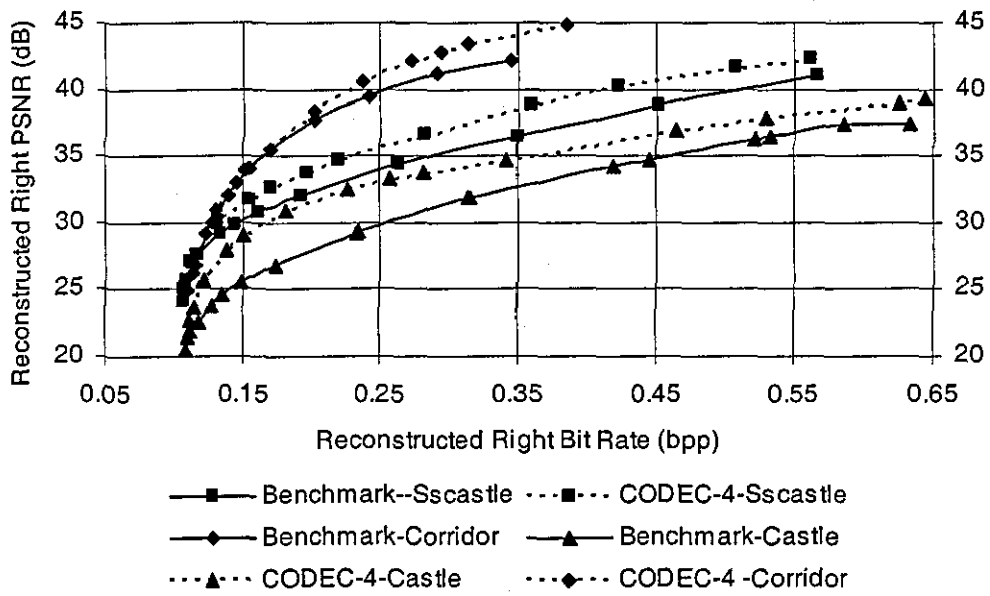
In order to evaluate the performance of the proposed CODECs, in figure 6, we compare their rate-distortion performance, against the disparity compensated predictive coding algorithm of [5] which uses DCT as the basic compression engine. Authors in literature to compare their proposed stereo image coding commonly use this benchmark. But the CODECs using, DWT based pioneering block predictive coding, i.e. CODEC-3 and CODEC-4, we use our recently DCT based pioneering block predictive coding algorithm proposed in [9,10], as a benchmark for comparison. The objective image quality is measured in terms of Peak Signal to Noise Ratio (PSNR), whereas the compression efficiency is measured in bits per pixels (bpp) as follows:

$$BR = \left[\frac{Tot\_Bits_{comp}}{N \times M} \right] \quad bpp$$

where, N and M represent the image dimensions and Tot_Bits_{comp} is the total number of bits required to represent a given image in its compressed format. For CODEC-1 and CODEC-2, Tot_Bits_{comp} consist of the number of bits required to encode the prediction errors and disparity vector fields. As for CODEC-3 and CODEC-4, Tot_Bits_{comp} exclude the number of bits required to encode the disparity vector fields. For a fair comparison, the objective quality of the left image (reference) was set at equivalent levels for all CODECs for a given image pair.
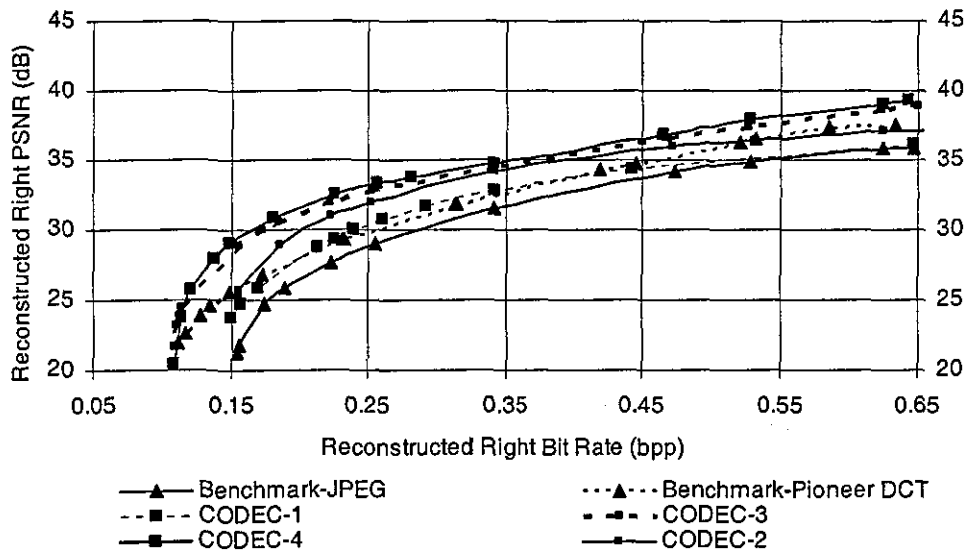
(a)

(b)

Figure 6 Rate-Distortion Performance Comparison Graphs
(a) For 'Castle' image: All four proposed CODECs vs. two benchmarks.
(b) For three stereo image pairs: CODEC-4 vs. DCT based pioneering block predictive
    coding.

12

In figure 6(a), for the stereo image pair "castle", we compare the rate-distortion performance of all four proposed CODECs, against that of the benchmarks, namely, disparity compensated predictive coding algorithm of [5] and DCT based pioneering block predictive coding algorithm proposed by us in [9,10]. The rate-distortion graphs clearly indicate an improved performance by all proposed CODECs, over the benchmarks, with the CODEC-4 showing the highest improvement among the proposed. Figure 6(a) also illustrates the ability of the pioneering block based techniques to operate at lower bit rates as compared to other algorithms. This is due to the fact that these schemes do not need the transmission of disparity vector fields.

To further evaluate the performance of CODEC-4 against that of the DCT based pioneering block predictive coding algorithm, in figure 6(b), we compare the rate-distortion graphs obtained by the two CODECs when coding stereo image pairs, 'Castle', 'Sscastle' and 'Corridor'. It further proves that CODEC-4 performs better as compared to the benchmark. In addition the graphs show that with the increase of stereoscopic data within the image pair, at low bit rates, the relative performance of the CODEC-4 against the benchmark, improves. This is illustrated by the fact that below 0.35 bpp, in "Castle" image, where the stereoscopic information is maximum (i.e. out of the three test image sets), the improvement of subjective quality obtained is higher than as obtained for the "corridor" image pair, with much less stereoscopic information. This relative improvement is not apparent at bit rates above 0.35 bpp.

When comparing the performance of CODEC-4 with that of CODEC-3, an improvement is illustrated in CODEC-4. This is justifiable, as CODEC-4 would result in a better overall prediction as compared to CODEC-3. This is due to the fact that in CODEC-3, the search for the best match is performed under integer wavelet-block shifts whereas in CODEC-4, the search is on a multi-resolution domain, with the possibility of single coefficient shifts in all sub-bands. In contrast to this observation, CODEC-2, with only the possibility of integer wavelet-block shifts in disparity compensation and prediction performs better than CODEC-1, which has the possibility of a integer pixel shifts in search. Further experiments revealed that this is due to the accuracy of matching in transform domain as compared to pixel domain, outweighing the disadvantages of integer wavelet-block shifts in disparity compensation and prediction.
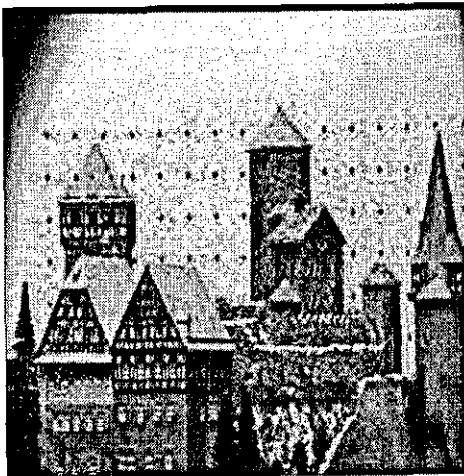
For further evaluation, in figure 7, we illustrate a comparison of the subjective image quality of the reconstructed predicted images obtained by using the two benchmark algorithms and proposed CODECs, 2 and 4.
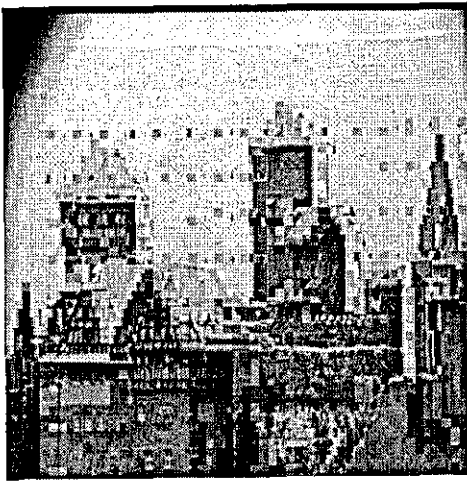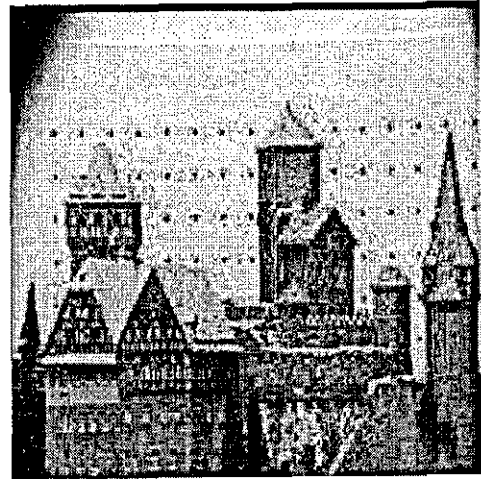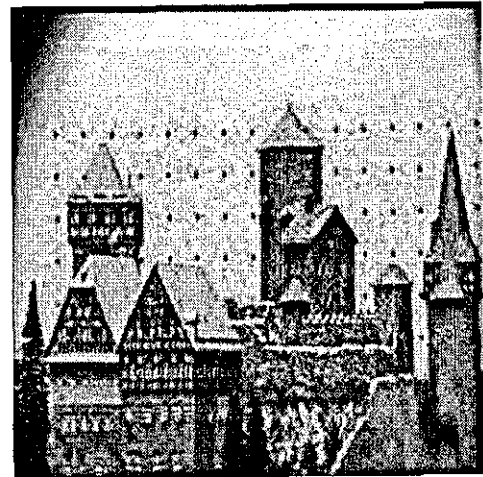
Figure 7 Subjective Quality Evaluation for Reconstructed Right Image of "Castle" Image.
(a) Benchmark: Traditional Predictive Coding (DCT based), BR = 0.155 bpp., PSNR = 21.29 dB
(b) Benchmark: Pioneering Block Based Predictive Coding (DCT based), BR = 0.1548 bpp.,
   PSNR = 25.88 dB   (c) CODEC-2, BR = 0.1558 bpp, PSNR = 25.12 dB (d) CODEC-4,
   BR = 0.1498 bpp, PSNR = 29.03 dB

Figure 7 clearly illustrates the superior image quality obtainable by using DWT based compression schemes as compared to DCT based compression schemes at a right image bit rate of approximately 0.155 bpp. This is shown by the fact that figures 7(c) and 7(d) has less visual artifacts as compared to figures 7(a) and 7(b). Further the illustrations in figure 7, proves the subjective quality improvements obtainable by the pioneering block based prediction scheme as against the traditional direct, block

14

based prediction scheme. This is proved by the fact that the reconstructed right images in figures, 7(b) and 7(d), illustrates improved subjective quality as compared to their traditional predictive coding based counterparts, illustrated by figures 7(a) and 7(c) respectively. Summarizing the above one could conclude that CODEC-4, which uses both DWT compression and pioneering block based predictive coding, gives best subjective image quality.

Even though a more accurate subjective quality measurement should involve viewing both images, previous research has shown that preservation of subjective image quality of left and right image near object boundaries, where depth cues are most likely to be of higher magnitude is a vital requirement for comfortable stereoscopic viewing. Thus it could be inferred that the proposed methods would perform subjectively better under stereoscopic viewing conditions.

5. Conclusion

We have proposed four, wavelet block based stereo image compression algorithms that use a simple coding architecture similar to that of a DCT based baseline-JPEG CODEC. We have shown that due to this unique design the proposed CODECs are able to transfer the inherent advantages of DWT based technology to stereo image compression, without the usual extra algorithmic and implementation complexity that otherwise needs to be introduced. We report PSNR gains of up to 7 dB at very low bit rates, for the CODEC-4, which uses DWT based multiresolution, pioneering block predictive coding as its basic compression algorithm. All other CODECs, performing less efficiently as compared to CODEC-4, still provides improvements over the performance of benchmark CODECs at all bit-rates. We have also shown that the proposed CODECs also provide reconstructed images of better subjective image quality.

At present we are extending the ideas presented in this paper to stereoscopic video sequence compression. Further, the use of zerotree entropy coding [22,23] in association with the ideas presented above is also being investigated.

Appendix I

1. Block-Based Disparity Compensation
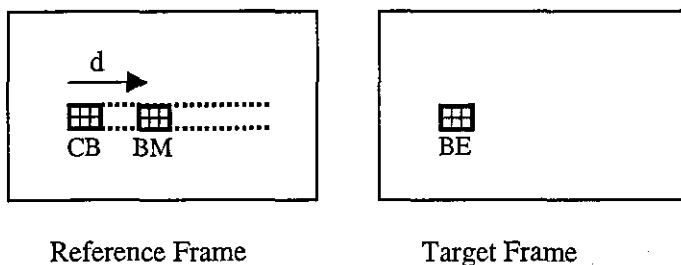

Reference Frame   Target Frame

Figure 8. Block-based search procedure in pixel
domain

Assuming parallel axes geometry, the search for best match is within predefined uni-directional horizontal search window as in figure 8. With the reference frame (left image) and target frame (right image), the searching for the block to be encoded (BE), starts from the corresponding block (CB) in the same position as BE in the reference frame and horizontally shifting by a pixel to the left of CB until the best match (BM) is found. BM is the best match if mean squared error between BE and BM is the minimum compared with other blocks in the search window. The number of pixels shift from CB to BM is d, and called the disparity. If disparity compensation is in wavelet block domain, similar procedure as explained above for pixel domain except that the shift will be block shift (i.e., 1 wavelet block shift in the wavelet domain = 1 pixel shift in the pixel domain).

2. Block-Based Pioneering Block Technique

Figure 9, illustrates the pioneering block-based search in the wavelet block domain which is similar to pioneering block-based search procedure [9,10] in pixel domain except in wavelet block domain the shift is per wavelet block whereas in pixel domain the shift is per pixel. To encode a given wavelet block (WB) of wavelet coefficients (say size 8x8) in the predicted (right) wavelet block, the wavelet block preceding (PB_{R1}) it and the wavelet block directly above (PB_{R2}) it are taken as pioneering blocks to search for a matching pair of wavelet blocks, within a selected window area of maximum likely of the reference (left) wavelet block (starting with PB_{L1} and PB_{L2}.). The best matching pair is found using means squared error as minimization criterion and using equal weights for the two

16

wavelet blocks. Once the best matching pair is found, the wavelet block directly below (BM) the upper block is chosen as the best match for WB.

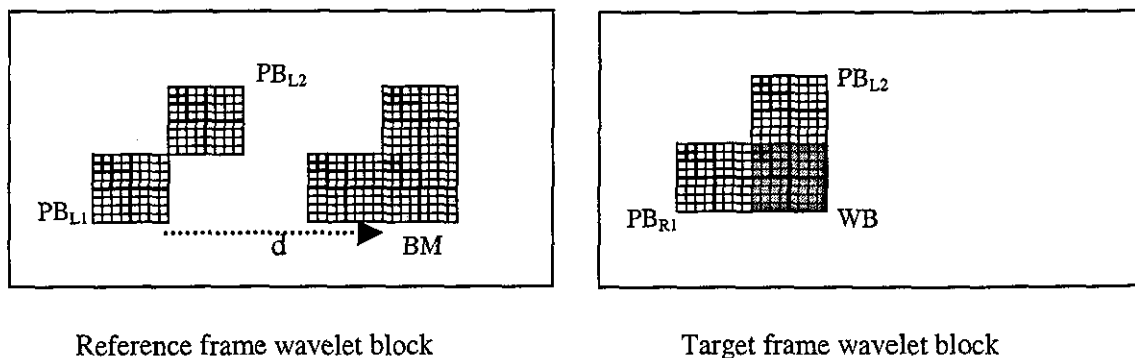Reference frame wavelet block   Target frame wavelet block

Figure 9. Pioneering block-based search
procedure in wavelet domain

References

[1]. Dinstein et.al. 'Compression of stereo images and the evaluation of its effects on 3-D perception' Proceedings of the IEE Applications of Digital Image Processing XII, 1989.

[2]. I.Dinstein, M.G.Kim, A.Hanik, "Compression of Stereo Images Using Sub sampling and Transform Coding", Journal of Optical Engineering, Vol.30, pp. 1359-1364, Sept. 1991.

[3]. M.G.Perkins, "Data Compression of Stereopairs", IEE Trans. on Comm., Vol.40, No.4, pp. 684-696, April 1992.

[4]. J.L. Dugelay, D. Pele 'Motion and disparity analysis of a stereoscopic sequence, application to 3DTV coding', EUSIPCO'92, October 1992, pp 1295-1298.

[5]. N. Grammalidis, S. Malassiotis, D. Tzovaras, M.G. Strintzis, '3D motion estimation and compensation for stereoscopic image sequence coding' Proceedings of 4th Euro Workshop on 3DTV, Rome, 1993.

[6]. S.Sethuramam, M.W.Siegel, A.G.Jordan, "A multi-resolution framework for stereoscopic image sequence compression', Proceedings of the ICIP-95, Vol.2, pp. 361-365.

[7]. D.V. Papadimitriou, T.J. Dennis, '3D parameter estimation from stereo image sequences for model-based image coding' Signal Processing, Image Communications, Vol7, 1995, pp471-487.

[8]. B. Choquet, J.L. Dugelay, D. Pele, 'A coding scheme for stereoscopic television seuqnces based on motion estimation-compensation using a 3D approach', IEE Conference in Image Processing and its Applications (IPA), July 1995, pp 188-192.

[9]. J.Jiang, E.A.Edirisinghe, H.Schroder, "A Novel Predictive Coding Algorithm for 3-D Image Compression", IEEE Trans. on Cons. Elec., Vol.43, N0.3, pp. 430-437, August 1997.

[10].J.Jiang, E.A.Edirisinghe, H.Schroder, "Algorithm for the Compression of Stereo Image Pairs", IEE Electronic Letters, Vol.33, No.12, pp. 1034-1035, June 1997.

[11].D. Tzovaras, N. Grammalidis, M.G. Strintzis, 'Object-based coding of stereo image sequences using joint 3D motion/disparity compensation', IEEE Transactions on Circuits and Systems for Video Technology, Vol 7, No 2, April 1997, pp 312-327.

[12].M.S. Moellenhoff and M.W. Maier 'Transform coding of stereo image residuals', IEEE Transactions on Image Processing, Vol 7, No 6, 1998, pp 804-812.

[13].H. Aydinoglu, H. Hayes, 'Stereo image coding: A projection approach' IEEE Transactions on Image Processing, Vol 7, No 4, 1998, pp 506-516.

[14].W. Woo and A. Ortega 'Optimal blockwise dependent quantization for stereo image coding', IEEE Transactions on Circuits and Systems for Video Technology, Vol 9, No 6, 1999, pp861-867.

[15].O. Woo and A. Ortega 'Overlapped block disparity compensation with adaptive windows for stereo image coding' IEEE Transactions on Circuits and Systems for Video Technology, 10: (2) March, 2000 pp194-200.

[16].N.V.Boulgouris, M.G.Strintzis, "Embedded Coding of Stereo Images", Proc. of the IEEE Int. Conf. on Image Proc., ICIP 2000, June 2000.

[17].E.A.Edirisinghe, "Rate Scalable Stereo Image Coding Using EZW Algorithm", Proc. of the IEEE Int. Conf. on Information, Communications and Signal Proc., ICICS 2001, Singapore, Oct. 2001, ISBN: 981-04-5149-0.

[18].J.Jiang, E.A.Edirisinghe "A Hybrid Scheme for Low Bit-Rate Coding of Stereo Images", IEEE Transactions on Image Processing, Volume 11, Issue. 2, pp 123-134, February 2002.

[19].M.Y.Nayan, E.A.Edirisinghe, H.E.Bez, "Baseline JPEG-Like DWT CODEC for Disparity Compensated Residual Coding of Stereo Images", Proceeding of the 20th Eurographics UK Conference 2002, pp. 67-74, June 2002.

[20].W.B.Pennebaker, J.L.Mitchell, "JPEG: Still Image Compression Standard", NY: Van Nostrand Reinhold, 1993.

[21].ISO/IEC JTC 1/SC 29/WG 1, ISO/IEC FCD 15444-1: Information technology – JPEG-2000 Image Coding System: Core coding system, March 2000, www.jpeg.org/FCD15444-1.html

[22].S.A.Martucci, I.Sodagar, "Zerotree Entropy Coding of Wavelet Coefficients for Very Low Bit Rate Video Coding", Proc. of 1996 IEEE Int. Conf. on Image Proc., Lausanne, Switzerland, Sept. 1996.

[23].S.A.Martucci, I.Sodagar, T.Chiang, Y.Zhang, "A Zerotree Wavelet Video Coder", IEEE Trans. on CSVT, Vol.7, No.1, February 1997, pp. 109-1997.

[24].R. de Queiroz, C.K. Choi, Y. Huh, K.R. Rao, "Wavelet Transforms in a JPEG – Like Image Coder," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, No. 2, pp 419 -424, April 1997.

[25].Ingrid Daubechies, 'Orthornormal bases of compactly supported wavelets', Communications of Pure and Applied Mathematics, Vol. 41, pp. 909-996, 1988.

[26].W.R. Zettler, et al, 'Application of compactly supported wavelets to image compression', SPIE Vol. 1244, Image Processing Algorithms and Techniques, pp. 150-160, 1990.