Wavelet-based Scalable Coding of Still and Time-varying Stereoscopic Imagery

Sumit K. Nath

***Supervisor**: Dr. Eric Dubois*

School of Information Technology and Engineering (S.I.T.E),
Faculty of Graduate and Postdoctoral Studies,
University of Ottawa,
880 King Edward Avenue,
Ottawa ON - K1N 6N5
Canada

June 2004

Thesis submitted in partial fulfillment of requirements for the degree of Doctor of Philosophy.

Abstract

This thesis addresses the issue of encoding and decoding still and time-varying stereoscopic imagery. A review of current encoding techniques is undertaken, with special emphasis on algorithms having SNR and spatial scalability. A stereo image pair consists of two views of the same scene. Due to the redundant nature of both views, prediction-based techniques produce superior results when compared with independent encoding of both images. Some of the most widely used embedded still-image coding techniques rely on discrete wavelet transform (DWT)-based analysis. However, these schemes cannot be adapted in a straightforward manner to encode stereoscopic still-image pairs.

In this thesis, a novel DWT-based embedded stereoscopic still-image codec structure is proposed. This scheme preserves the progressive transmission capability of still-image coding algorithms, while suitably adapting to the nuances and special characteristics of stereoscopic imagery. A comparative study of variable-block and fixed-block disparity estimation is also undertaken. Partition artifacts result due to imperfect disparity compensation. Drawbacks in existing compensation techniques are discussed and a novel loop-filtering scheme is proposed. This is used to smooth disparity-compensated images before generating and subsequently encoding residual images. As seen from this thesis, this scheme improves on the performance of current techniques. In addition, the dyadic sampling structure of a 2-D DWT is exploited to obtain discrete levels of spatial-scalability and forms part of an embedded scheme for transmission of stereoscopic still-images at different spatial resolutions.

The proposed algorithm is suitably modified to encode time-varying stereoscopic imagery. Drawbacks of current moving-picture hierarchies are analyzed and a novel hierarchy is proposed that insures that a user has the flexibility to view a sequence either in monoscopic (default) or stereoscopic modes. Independent objective results, explaining SNR and spatial scalability features, are presented when encoding a few pictures of a stereoscopic moving image sequence. In addition, informal subjective results are presented when viewing encoded versions a time-varying sequence.

Acknowledgment

"*The secret to creativity is knowing how to hide your sources*"
Albert Einstein

I would like to graciously acknowledge the following individuals who have, directly and indirectly, helped me during the course of this research.

It is rightly said that the difference between ordinary and extraordinary is just a little *extra*. This sums up my work under the supervision of Dr. Eric Dubois. His patience and unrelenting zeal for perfection has helped me in raising the standard and outlook of the research work presented in this thesis. I would also like to thank Dr. James Walker at the University of Wisconsin, Eau Claire for useful discussions pertaining to the ASWDR algorithm. I would also like to acknowledge Dr. Tamás Frajka for the (FZ) results presented in Chapter 5. In addition, I would like thank Rahul Shukla at EPFL, Switzerland for the (RS) results presented in the same chapter. The display of stereoscopic moving image sequences, discussed in Chapter 7, was made possible by my colleague Xiaodong Huang at the VIVA Laboratory. I would also like to thank my other colleagues at the VIVA laboratory who have helped me in various miscellaneous aspects during the course of this research work. The "*angioMR*" stereo image pair has been provided by Dr. Thomas Langø at SINTEF Unimed, Trondheim, Norway. Source codes for the ASWDR algorithm was provided by Dr. James Walker .

In addition, this thesis would not have appeared in its current shape without the help and support of the following individuals. I would like to extend my gratitude to my good friend Mr. Andreas Moser for his help and support that made my arrival in Canada possible. I would also like to thank Dr. Jayanta Kumar Ray and Mr. Rakesh A. Nayak for their help and support during my stay in Calgary. I will always be indebted to my good friends Rohini and Katla Chandrasekhar for their support at various stages of this thesis. I would also like to extend the same feeling of indebtedness to my friend Ravi Sridhar Murthy . I would also like to my friend Mr. Kulbhushan

Kapoor and his extended family. They have been a surrogate family and have made my stay in Ottawa very worthwhile. In the same vein, I would like to thank Mr. Gérald Levert and his family. Last but not the least, I will always be indebted to my *Mama* and *Mami* in Calcutta, India for their emotional support throughout the course of this research.

Contents

List of Figures

List of Tables

Acronyms

2DLS	2-D logarithmic search
3SS	3-step search
4SS	4-step search
AC	Arithmetic coding
ASWDR	Adaptively-scanned wavelet-difference-reduction
BPP	Bits-per-pixel
CA	Coding artifacts
CB	Color bleeding
CDF	Cohen-Daubechies-Feauveau
CONCOD	Conditional coder
CLDC	Closed loop disparity codec
DC	Disparity compensation
DCT	Discrete cosine transform
DDV	Displaced disparity vector
DE	Disparity estimation
DEMUX	Demultiplexer
DFD	Displaced frame difference
DWT	Discrete wavelet transform
EBCOT	Embedded block coding with optimized truncation
EOS	End-of-scan
EPNR	Edge preserving noise reduction
EZW	Embedded zerotree wavelet
FB	Fixed-block
FS	Full search
FZ	Frajka and Zeger's algorithm
GOP	Group of pictures
HBDE	Hierarchical (fixed or variable) block-based disparity estimation
HBME	Hierarchical (fixed or variable) block-based motion estimation
HDTV	High definition television
HVS	Human visual system
JPEG	Joint photographic experts group
LF	Loop filter
MAD	Mean-absolute-difference
MC	Motion compensation
ME	Motion estimation
MGE	Multigrid embedding
MPEG	Moving pictures experts group
MSE	Mean-squared error
MUX	Multiplexer
MV	Motion vector
MVP	Multiview profile

ND_1	Proposed algorithm with "CDF-9/7" filters and with loop-filtering
ND_2	Proposed algorithm with "Odegard-9/7" filters and with loop-filtering
ND_3	Proposed algorithm with "Cooklet-17/11" filters and with loop-filtering
ND_{nf}	Proposed algorithm with "CDF-9/7" filters and without loop-filtering
OBDC	Overlapped-block disparity compensation
OBMC	Overlapped-block motion compensation
OLDC	Open loop disparity codec
PR	Perfect-reconstruction
PSNR	Peak signal-to-noise ratio
RGB	Red-Green-Blue color space
QoS	Quality of service
QTMap	Quadtree Map
R-D	Rate-distortion
RS	Refinement search
SAD	Sum-of-absolute-difference
SDTV	Standard definition television
SDV	Standard disparity vector
SNR	Signal-to-noise ratio
SPIHT	Set partitioning in hierarchical trees
RS	Shukla and Radha's algorithm
VB	Variable block
VQEG	Video quality experts group
WDR	Wavelet-difference-reduction
YCbCr	Luminance-Chrominance color-space

Notations

Image

$\underline{\boldsymbol{I}}$	Color image with three, two-dimensional, gray-scale components
$\boldsymbol{I}_{(.)}$	2-D gray-scale image
$\widehat{\boldsymbol{I}}_{(.)}$	Reconstructed 2-D gray-scale image
$[M, N]$	Dimensions of a 2-D image
4:4:4	Unsampled image in RGB domain
4:2:0	Sub-sampled image in YCbCR domain
$\mathcal{K}_{(.)}$	Overall bitrate of Y, Cb or Cr components

Wavelets

$\boldsymbol{c_m}$	1-D input signal
$\boldsymbol{c_{m-1}}$	Approximate coefficients of $\boldsymbol{c_m}$
$\boldsymbol{d_{m-1}}$	Detail coefficients of $\boldsymbol{c_m}$
$\tilde{\boldsymbol{H}}(\boldsymbol{z})$	z-transform of 1-D analysis low-pass filter coefficients
$\tilde{\boldsymbol{G}}(\boldsymbol{z})$	z-transform of 1-D analysis high-pass filter coefficients
$\boldsymbol{H}(\boldsymbol{z})$	z-transform of 1-D synthesis low-pass filter coefficients
$\boldsymbol{G}(\boldsymbol{z})$	z-transform of 1-D synthesis high-pass filter coefficients
DWT_k	k-level 2-D separable forward discrete wavelet transform
DWT_k^{-1}	k-level 2-D separable inverse discrete wavelet transform
$\boldsymbol{c}_{0x}$	All low-pass subband
$\boldsymbol{d}_{1i}$	High-low (HL) subband at scale-i
$\boldsymbol{d}_{2i}$	Low-high (LH) subband at scale-i
$\boldsymbol{d}_{3i}$	High-high (HH) subband at scale-i
$\tilde{\boldsymbol{P}}(z)$	Polyphase components of analysis wavelet filters
$\boldsymbol{P}(z)$	Polyphase components of synthesis wavelet filters
$S_i(z), T_i(z)$	Lifting step polynomials

ASWDR

T	Maximum threshold used during ASWDR encoding
w	Wavelet-transformed image coefficient
γ	Maximum value of all wavelet-transformed image coefficients
$\boldsymbol{ICS}$	List containing all coefficients of an image
$\boldsymbol{SCS}$	List containing significant coefficients
$\boldsymbol{TPS}$	List containing scan-updated coefficients
`pos.`	relative positions of significant coefficients
R	Rfinement bit
C	EOS value
C_{sec}	Length of secondary list at end of current dominant scan

D_i	i^{th} dominant scan
S_i	i^{th} refinement scan
$\mathcal{E}(\mathcal{K}_i)$	All steps of an ASWDR encoding scheme, at bit-rate $\mathcal{K}_i$, excluding a forward DWT
$\mathcal{D}(\mathcal{R}_i)$	All steps of an ASWDR decoding scheme, at bit-rate $\mathcal{R}_i$, excluding an inverse DWT

Estimation

$\boldsymbol{B_i}$	i^{th} block from target image
$\boldsymbol{v_i}$	Actual displacement vector representing either motion vector (MV) or disparity vectors (DDV, SDV)
$\widehat{\boldsymbol{v}}_{\boldsymbol{i}}$	Estimated displacement vector
$\boldsymbol{d(x, v_i)}$	Displaced frame difference
$\widehat{\boldsymbol{v}}_{\boldsymbol{i}}^{\boldsymbol{k}}$	Estimated displacement vector at scale k
$\widehat{\boldsymbol{\Delta}}_{\boldsymbol{i}}^{\boldsymbol{k}}$	Estimated refinement vector at scale k
$\widehat{\boldsymbol{v}}_{\boldsymbol{i}}^{\boldsymbol{k-1}}$	Estimated displacement vector at scale $(k-1)$
V_t	Homogeneity threshold during quadtree partitioning

Compensation

X_n	Signal to be quantized
$\bar{U}_n$	Signal subtracted from X_n
e_n	Unquantized error signal
$\widehat{X}_n$	Quantized signal
$\widehat{U}_n$	Quantized version of $\bar{U}_n$
$\widehat{e}_n$	Quantized error signal
Q	Quantization operator
$E[.]$	Expected value of a random variable
$\mathcal{E}$	Distortion due to quantization process

Loop Filter

$[n_1, n_2]$	Co-ordinates of image being filtered
$f[n_1, n_2]$	Input image being filtered
$[\alpha, \beta]$	Additional parameters dependent on $[n_1, n_2]$
λ	Smoothing parameter
$g[n_1, n_2]$	Filtered output image

Chapter 1

Problem Definition and Thesis Scope

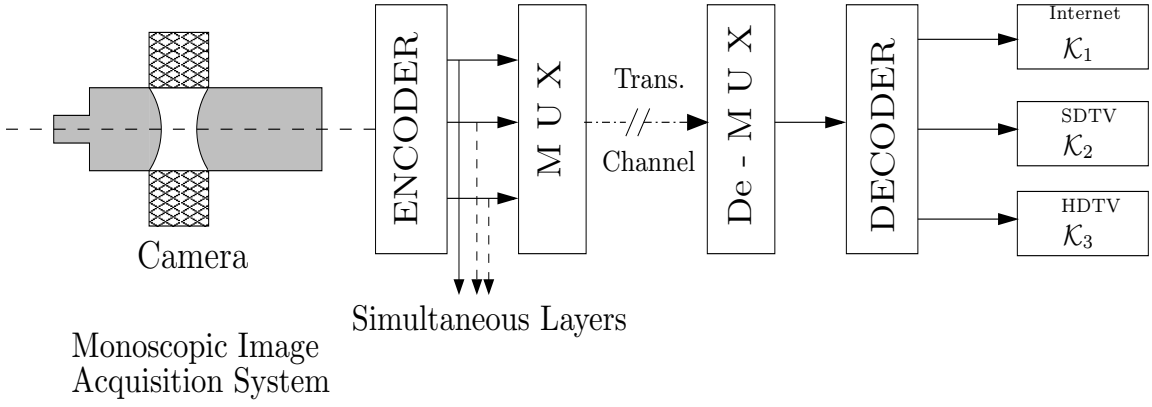
1.1 Background information

THE advent of high definition television (HDTV) systems has revolutionized the way in which we view visual information. It is envisaged that some future television systems will be able to display *stereoscopic imagery*. This is different from *monoscopic imagery* as it has two views of visual information. This mimics the binocular nature of the human visual system (HVS). It is also envisaged that future telemedicine applications (e.g., remote surgeries) will require transmission of stereoscopic images. Due to the binocular composition of stereoscopic imagery, *depth* from the scene being imaged can be perceived by the HVS.

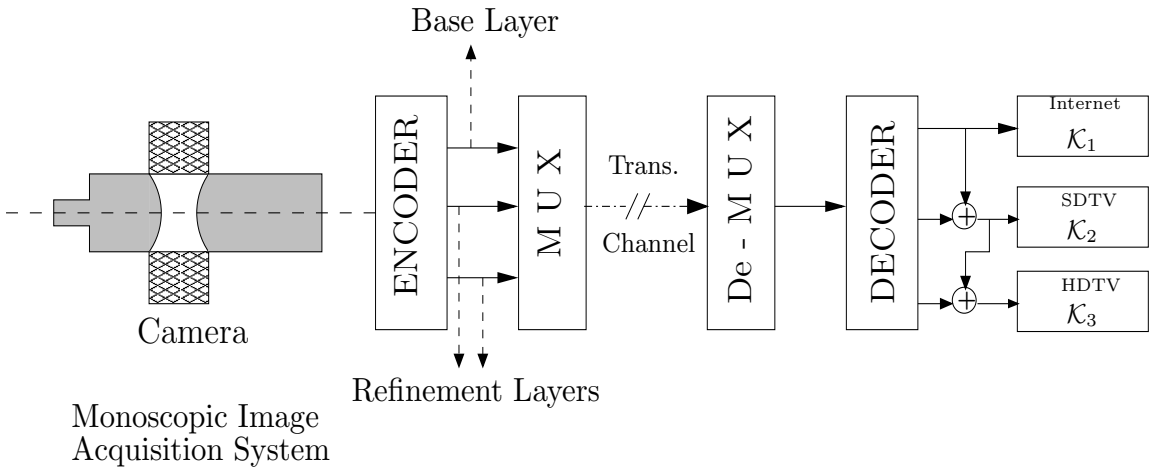
The advent of high speed networks and high density digital versatile discs (DVD) have greatly affected the means by which stereoscopic imagery can be transmitted or stored. As gargantuan amounts of data are involved, compressing them would indeed improve the performance of such networks or storage devices. Recently there has been tremendous growth of HDTV systems and Internet based broadcasting (otherwise known as *webcasting*). Transmission and delivery of stereoscopic moving-image content to consumers having these systems, in addition to traditional standard definition television (SDTV) systems is a challenging problem. This leads to the concept of *scalable trans-*

mission.

This can be defined as a simultaneous transmission of the same visual information to consumers having such varying display devices as Internet, SDTV or HDTV systems. In literature this method of media content delivery is sometimes referred to as a *quality of service* (QoS) framework. A current framework used for simultaneous transmission of monoscopic moving-image sequences can be seen from Fig. 1.1(a). As observed from this



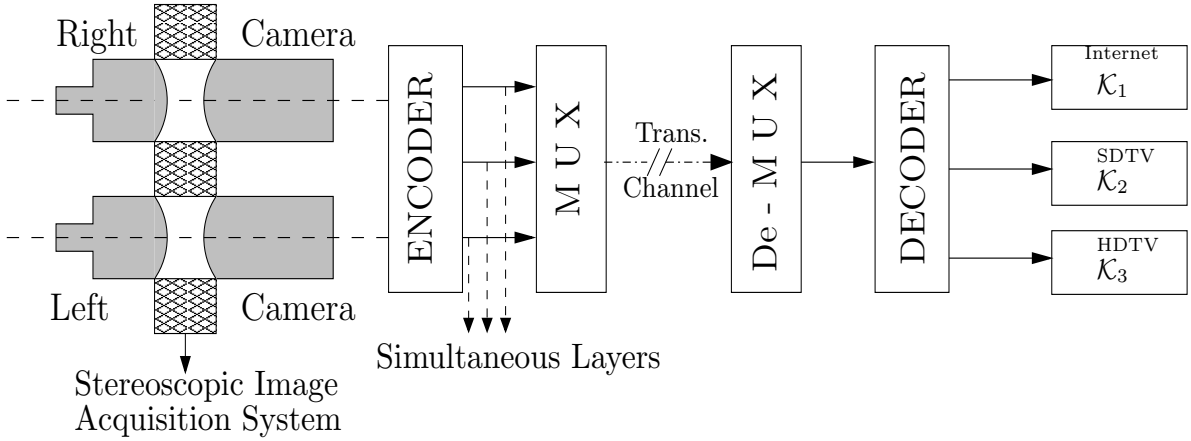
(a) Independent Mode



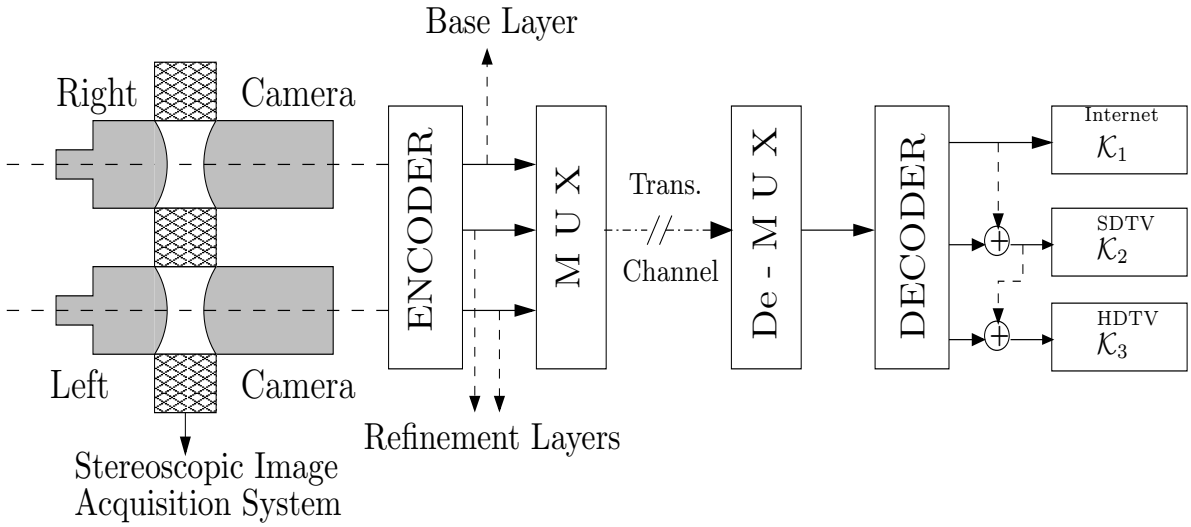
(b) Embedded Mode

Fig. 1.1: QoS frameworks for transmission of monoscopic video content in (a) independent and (b) embedded simulcast modes. $\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{K}_3$ represent bit-rates.

figure, three versions of the same data, at different spatial resolutions, are independently generated and transmitted. This is commonly referred to as an *independent* simulcast



(a) Independent Mode



(b) Embedded Mode

Fig. 1.2: QoS frameworks for transmission of stereoscopic video content in (a) independent and (b) embedded simulcast modes. $\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{K}_3$ represent bit-rates.

(i.e., simultaneous-telecast) QoS-framework. This is not an optimal framework for data transmission. Separate data streams must be generated for each level of service leading to a highly redundant operation. Content meant for Internet and SDTV systems are *downsampled* versions of HDTV system content. Hence, transmission of such redundant visual content can place enormous constraints on available bandwidth.

An alternative framework for simultaneous transmission of monoscopic data is shown in Fig. 1.1(b). This is referred to as an *embedded* simulcast QoS-framework in this thesis. Unlike the independent simulcast framework, a *base layer* of information is generated. This layer, though specifically intended for webcasting, can be used in SDTV and HDTV displays as well. A set of *refinement layers* are also generated. Hence, a SDTV display would have a base layer along with the addition of a single refinement layer. On the other hand, a HDTV display would have the base layer with both refinement layers added to it.

These frameworks can be extended to encode stereoscopic imagery as well (Fig. 1.2). Specifics of these frameworks are discussed in later chapters of this thesis. It is also shown that an embedded QoS framework is more suitable for transmission of stereoscopic imagery than its independent counterpart (both in terms of rate-distortion (R-D) and perceptual quality).

The overall performance of a stereoscopic moving-image transmission system depends individually on the performance of the image acquisition system, encoder, multiplexer, transmission channel, de-multiplexer, and display system. This thesis addresses the problem of design specifications for encoding and decoding stereoscopic moving-image content. It is assumed that stereoscopic images, such as these used in the thesis, have been acquired from "reasonably good" imaging systems. Furthermore, it is also assumed that these images can be viewed on "visually pleasing" display systems. Design of a multiplexer/de-multiplexer system is (generally) a hardware related problem and hence

not discussed in this thesis. An error-free transmission channel is assumed when evaluating the performance of the decoder.

Given this background information, the specific problem addressed in this thesis is described in the following section.

1.2 Summary of proposed research work

This section is organized into two parts. The first part highlights salient features and sections of the proposed algorithm. The second part provides a justification for performing the research work described in this thesis.

1.2.1 Problem definition

As mentioned in the previous section, the scope of this thesis is limited to the design of an efficient coding system for stereoscopic imagery. Initially, a novel stereoscopic still-image codec is presented. Underlying principles from this codec are subsequently applied in designing a codec for encoding stereoscopic moving-images.

From a compression and coding point of view, various features sought in designing this codec are outlined as follows:

- *Embedded coding*: A bit-stream is said to be embedded if subsets from it contain complete to near-complete information about an image. This is a highly desired feature in image coding as it enables users to specify any arbitrary bit-rate during decoding. Higher bit-rates imply adding a series of "refinement" layers to an original "base-layer" of information. This feature marks out the current JPEG2000 image coding standard [1] over conventional standards.

- *Spatial-scalability*: A bit-stream is said to be spatially scalable if it contains the same visual information at different spatial resolutions. As seen from Fig.1.2(b) (and discussed in the later chapter) an embedded framework is a qualitatively

and quantitatively superior technique for obtaining spatial-scalability. A dyadic subsampling structure is generally used to reduce computational complexities of *full-search* (FS) motion- or disparity-estimation techniques [2]. In the proposed algorithm, the subsampling structure of a discrete wavelet transform (DWT) is exploited to obtain finite levels of spatial-scalability.

- *SNR-scalability*: As previously defined, subsets of an embedded bit-stream contains nearly all relevant information about an image. Assuming that the spatial resolution is kept constant, these subsets, when decoded and viewed at the given spatial resolution, will have a particular SNR. In order to improve this SNR, additional bits need to be decoded. This decoded information can be added to previously decoded information, making the bit-stream SNR-scalable. Evidently, this is a desirable feature in any embedded stereoscopic image codec. Wavelet-based transforms have inherent capabilities of embedded image coding with high levels of SNR-scalability (i.e., progressive coding). Hence this feature is also present in the proposed algorithm.

- *Asymmetrical coding*: From psycho-visual experiments, it has been deduced [3] that both views of a stereoscopic image pair need not be displayed at full perceptual quality. This led to the concept of asymmetrical coding wherein one image view is displayed at a higher SNR than the other view. Within certain limits, when viewing both images in a stereoscopic mode, the overall quality is entirely dependent on the quality of the image having a higher SNR. This is a useful feature from a compression point of view. As a result, this concept is incorporated in the design of this codec.

- *Miscellaneous features*: Current standards of moving-image encoding facilitate *object-scalability*. This involves selectively decoding various regions of an image at dif-

ferent time instants and at varying perceptual qualities. In order to achieve this, the proposed codec incorporates a feature for limited object-scalability. This is (generally) a first stage when implementing similar techniques discussed in literature [4]. Finally, a feature in any moving-image encoding technique is a desire to achieve *temporal-scalability*. This involves viewing an image sequence at a low *frame-rate* and progressively improving the quality by increasing this rate. This is not discussed during the course of this thesis. However, as shown in Chapter 4, this feature can be implicitly derived when implementing the proposed codec structure.

Having underlined various features in the proposed codec, reasons and justification are presented as to the need for undertaking this research project.

1.2.2 Justifying the proposed research work

Various algorithms have been proposed for encoding stereoscopic still- and moving-images. It is well established that independent coding of both image views is not an optimal solution. On the other hand, algorithms that exploit *inter-view* redundancies between both images have been shown to produce better results. Previous work, [5], [6], [7], [8], relied on DCT-based coding techniques. However it has been shown, [9], [10], that wavelet based encoding techniques provide superior results than their DCT-based counterparts.

Use of wavelet-based techniques in stereoscopic still- and moving-images have been reported in literature. These algorithms have varying degrees of success. However they leave scope for further improvement. Notable amongst these are work by Bolugouris and Strintzis [11] and Frajka and Zeger [12]. The codec proposed in this thesis relies on work discussed in these papers. However both these algorithms have some drawbacks. The codec structure, presented in [11] relies on an *embedded zerotree wavelet* (EZW) coding technique. This has been superseded by other algorithms, [10], [13]. The algorithm in

[12] utilizes a *multigrid embedding* (MGE) of wavelet coefficients in encoding. Results have been presented that prove the superiority of this algorithm when compared with the one presented in [11].

The proposed research work attempts at a hybrid solution. In [11], a closed-loop formulation has been proposed for optimal stereoscopic still-image encoding. Use of a MGE algorithm in [12] stems from previously published results [14]. Further improvements can be made in this algorithm. EZW [9] and SPIHT [10] rely on identification of *zerotrees* in subbands. In [15] this is termed *inter-scale* correlation. The MGE algorithm abandons this correlation in favor of *intra-scale* correlation amongst subbands. During this research, it was conjectured that an algorithm utilizing both inter- and intra-scale correlation amongst subbands would provide superior results. Hence a novel stereoscopic coding technique is presented that utilizes an *adaptively scanned wavelet-difference-reduction* (ASWDR) technique [16].

As mentioned previously, object-scalability is a desired feature in current moving-image encoding techniques. The algorithms described in previous paragraphs of this sub-section do not have such features. Preliminary work in this context have been reported by Shukla and Radha [17]. The algorithm proposed in this thesis uses a *variable block-based* disparity-estimation scheme, similar to the one proposed in [17]. Such a scheme forms a first-stage, when implementing other sophisticated techniques used for object-scalability [4]. *Partition-artifacts* are a problem with any disparity related coding scheme. In algorithms cited in previous paragraphs, these are referred to as *blocking artifacts*, as *fixed block-based* disparity-compensation is used in all of them. Optimal solutions have been reported in literature that overcome such artifacts. Notable amongst them would be an *overlapped block disparity compensation* (OBDC) scheme, proposed by Woo and Ortega [18].

Unfortunately this scheme cannot be extended to the algorithm proposed in this the-

sis. Instead, *loop-filtering* constitutes a viable alternative. At the time of writing this document, no specific references have been found that addressed this issue in conjunction with variable block-based compensation techniques. As an original contribution, a novel scheme is presented that alleviates the problem of arbitrarily shaped partitions in disparity-compensation. An *edge preserving noise reduction* (EPNR) filter, originally proposed to clean images corrupted with Gaussian noise [19], is adapted as a loop filter. Hence the proposed algorithm combines a closed-loop formulation, ASWDR embedded encoding scheme and an EPNR loop-filter to effectively encode stereoscopic still-images. Results shown in Chapter 3 indicate the superiority of this algorithm when compared with similar methods indicated in [12] and [17].

In literature, very few references have been found that address the problem of wavelet-based stereoscopic moving-image coding. The closest work that has been identified is by Chang and Wu [20]. This is an improvement over current industry standards for stereoscopic moving-image coding [21]. High levels of SNR-scalability cannot be obtained from the latter, while the former technique does not provide scope for embedded moving-image coding. In addition, both these formulations have no scope for object-scalability. This stems from the picture hierarchy used in encoding such moving-images. As an original contribution, a novel picture hierarchy is proposed. This is combined with the algorithm used in encoding stereoscopic still-images. This involves motion-estimation between successive pictures of both streams. Due to similarities in motion- and disparity-estimation[1] the algorithm can be seamlessly used to encode moving-images as well. This formulation ensures high levels of *drift*-free SNR-scalability during encoding and decoding such moving-images.

As mentioned in the previous sub-section, spatial-scalability is a desired feature in the proposed codec. During the literature survey, no references have been found that

[1]Explained in Chapter 2

provide scope for spatial-scalability in conjunction with encoding stereoscopic imagery. In this thesis, a dyadic subsampling structure of a discrete wavelet transform (DWT) is exploited to obtained finite-levels of spatial scalability. This is different from related work presented for monoscopic moving-image encoding [22]. In this, images need to be *explicitly* downsampled before encoding them at different spatial resolutions. However in the algorithm described in this thesis, the downsampling operation is *intrinsic* in nature. A detailed discussion is presented in later chapters.

Finally, for the sake of completeness, preliminary subjective results have been presented when encoding stereoscopic still- and moving-images in an asymmetrical (or mixed-SNR-resolution) framework. Psycho-visual experiments conducted by Tam *et al.* [23] have revealed that *visual fatigue* may arise in the HVS when continuously viewing asymmetrically coded stereoscopic image data. As such the authors in this work proposed a novel *temporal interleaving* of such asymmetrically data. As a result, they conjectured that visual fatigue can be reduced. In a coding framework this is tantamount to degrading the quality of images from one stream with respect to the other.

To achieve this goal, the authors have proposed a Gaussian blurring of one image stream. This blurred image stream is subsequently encoded using a state-of-the-art monoscopic moving-image coding technique. The other image stream is also independently coded, but at a higher bit-rate than the Gaussian blurred image stream. However, as previously described, independent coding of stereoscopic images is not an optimal solution [3]. Furthermore, disparity-estimation between two images at varying perceptual qualities may lead to biased results. This can affect overall coding performances. Hence, this scheme cannot be incorporated in the algorithm presented in this thesis.

However, removal of visual fatigue is still a desirable feature. To achieve this, a novel solution is proposed. This involves temporal-interleaving of stereoscopic moving-images at arbitrary time instances. This is different from the scheme presented in [23], where

such an interleaving is implemented at *scene-cuts* only. Degradation in image quality is achieved exclusively by using the progressive encoding and decoding feature of an ASWDR algorithm. Unlike [23], *a priori* blurring of images is not implemented. Limited subjective results indicate that the HVS is not able to readily differentiate between regular and the proposed temporal-interleaving of asymmetrically coded stereoscopic moving-image data.

Having defined the problem, the concluding section of this chapter describes the general organization of this thesis.

1.3 Thesis organization

As indicated in Sec. 1.2.1, various features have been included in designing the algorithm proposed in this thesis. A structure for coding and decoding stereoscopic still-images is presented. This structure is then incorporated in a new structure used for encoding time-varying stereoscopic imagery. To facilitate a better understanding of concepts, each chapter is preceded by an abstract highlighting its contents. A detailed literature review, pertinent to aspects covered in a chapter is then presented. The chapter then concludes by providing a detailed discussion on relevant concepts.

Thus, the remaining chapters in this thesis are organized as follows:

- **Chapter 2 :** In this chapter, the reader is introduced to some concepts on stereoscopic imaging. A brief discussion of wavelets is also provided. The concept of lifting in wavelet analysis is also discussed. Appendix A completes this discussion. Next, a summary of relevant motion- and disparity-estimation techniques is provided. The drawbacks of current disparity- and motion-compensation techniques is also presented, followed by a discussion on *hierarchical-search* strategies in motion- and disparity-estimation. Similarities and subtle differences in estimating motion- and disparity-vectors using this algorithm are presented.

- **Chapter 3 :** This chapter begins with a discussion on the concepts of progressive image coding. Next, a brief review of current state-of-the-art embedded coding techniques is presented. Limitations of such schemes, in the context of stereoscopic imaging is discussed and how an ASWDR algorithm can overcome such limitations. This is followed by a review of an ASWDR algorithm. For the sake of completeness, some comparative results are provided in Appendix B.

- **Chapter 4 :** A survey of current stereoscopic still-image encoders is presented. This is followed by a discussion on optimal conditions for stereoscopic still-image encoding. This chapter is concluded by a discussion on two current algorithms [11, 12] used in stereoscopic still-image coding. Useful features and drawbacks (in the context of this research work) of both algorithms are presented.

- **Chapter 5 :** This chapter introduces the reader to current motion and disparity compensation techniques. Limitations of these techniques are presented, followed by a discussion on a novel EPNR filtering scheme. Justification is also provided in using this as a loop-filter to smooth disparity- and motion-compensated images. This is followed by a discussion on the proposed algorithm, when encoding stereoscopic still-image pairs. Comparative objective results between algorithms presented in [12], [17] and the proposed algorithm are provided. In addition, limited subjective results are also presented when encoding stereoscopic color-images. A conclusion and scope for further research work rounds up this chapter.

- **Chapter 6 :** This chapter presents a discussion on various picture hierarchies employed, when encoding monoscopic and stereoscopic moving-image sequences. A survey of existing stereoscopic moving-image encoding systems (using these picture hierarchies) is presented. This is followed by a discussion on perceived limitations of these systems. As previously indicated in this chapter, no suitable references have

been found that addressed the problem of spatial-scalability in conjunction with stereoscopic moving-image coding. Hence an algorithm [22], that addresses the problem of spatial-scalability in the context of monoscopic moving-image encoding is discussed briefly.

- **Chapter 7 :** Drawbacks of picture hierarchies, introduced in Chapter 6, are presented here. To alleviate these limitations, a novel picture hierarchy is proposed. It is shown that this picture hierarchy can faithfully be used to obtain high-levels of SNR-scalability when viewing a moving-image sequence, either in monoscopic or stereoscopic modes.

  Objective results are presented when encoding various pictures of a test stereoscopic moving-image sequence. In addition, a limited subjective discussion is presented that compares the performance of encoded versions of this sequence, with and without temporal interleaving. This sequence[2] does not have any scene cuts and, hence, is a key factor in differentiating the proposed algorithm when compared with the scheme presented in [23].

- **Chapter 8 :** This chapter summarizes the salient features of the proposed algorithm when encoding stereoscopic still and moving-images. In doing so, it highlights various contributions made during the course of this research work. Finally, a discussion is provided that highlights future research topics that have emerged during the course of this research work.

[2]Please refer to the enclosed CD-ROM.

Chapter 2

Preliminaries on Stereoscopic Imaging and Wavelets

Overview

A discussion is presented on some aspects of stereoscopic imaging techniques. Concepts of disparity and motion in stereoscopic sequences are presented. This is followed by a discussion on relevant concepts of wavelets and lifting-based implementations of a DWT. A justification and discussion is also provided on the efficacy of hierarchical-search strategies for disparity- or motion-vector estimation.

2.1 Concepts of stereoscopic imaging

VISUAL information, as perceived by the HVS, is basically three-dimensional in nature. Traditional monoscopic imaging systems offer extensive detail about real world scenes. However they are unable to provide a viewer with the sense of "*depth*" in perceiving a scene. This is possible with binocular imaging and hence led to the development of stereoscopic imaging systems. In such systems, visual information about a scene is recorded by two different cameras (Fig. 2.1) as opposed to a single camera. This is analogous to a binocular HVS. In general both image views contain nearly the same visual content. However, there are some areas in one view that are absent from the other; these are generally referred to as *occluded regions*. To better appreciate the flow of discussion, the following paragraphs describe notations used in conjunction with

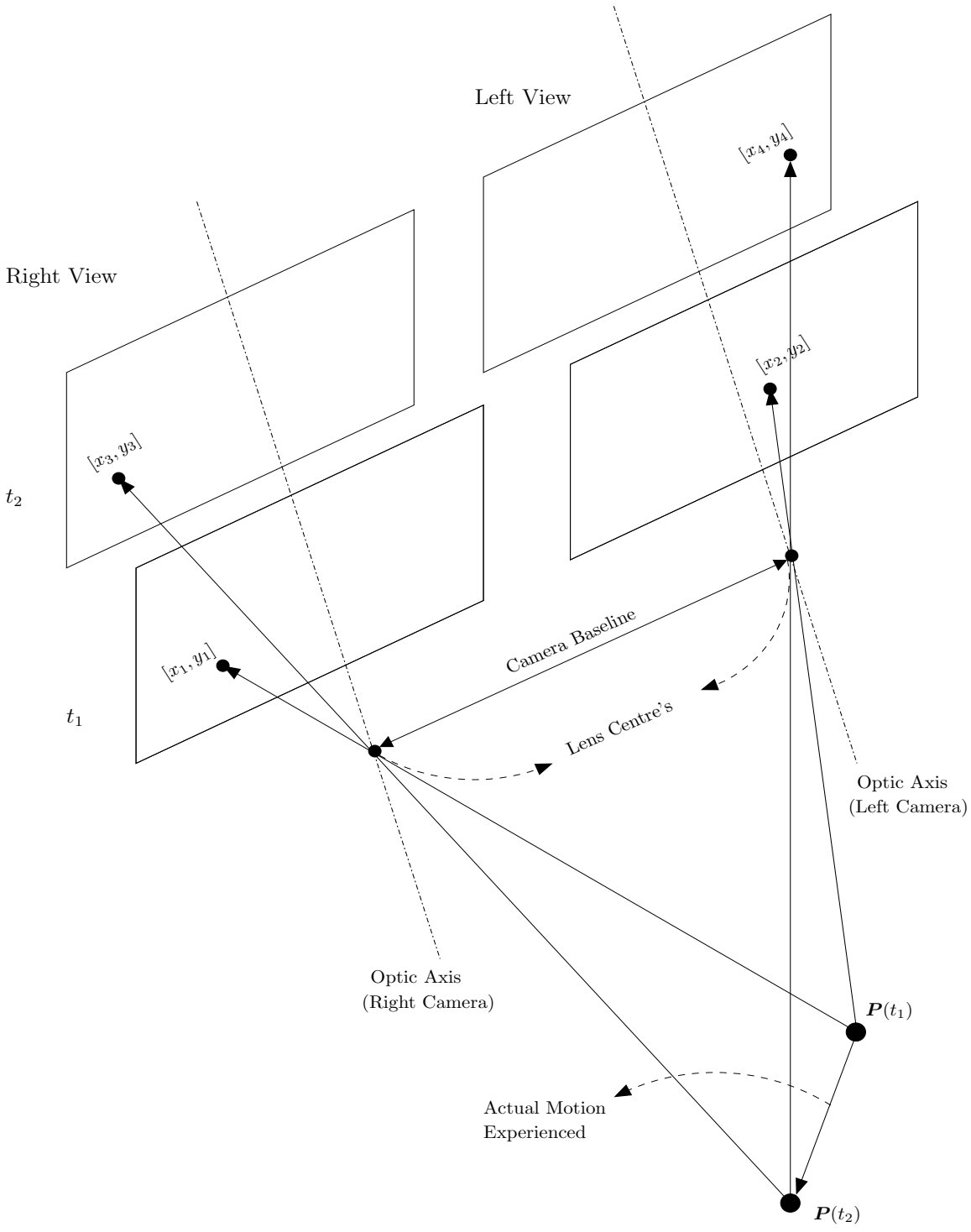


Fig. 2.1: Schematic of a Binocular Stereoscopic Imaging System. Proper optical arrangements are incorporated so as to prevent inversion of images. Both camera's are assumed to be stationary.

stereoscopic imaging in this thesis.

From Fig. 2.1, consider point $\boldsymbol{P}$ in a 3-D space. Let this point be projected into a 2-D continuous space, and be indicated as $\boldsymbol{C}$. Due to resolution-dependent imaging systems, this 2-D image must be sampled at discrete points. In addition, systems imaging these discrete points should have capabilities to acquire three separate channels of information. This corresponds to tri-stimulus color values of a HVS and are usually acquired and displayed in the RGB domain. Let this sampled color-image be represented as

$$\underline{\boldsymbol{I}} = \{\boldsymbol{I_R}, \boldsymbol{I_G}, \boldsymbol{I_B}\}$$

where $\boldsymbol{I_R}, \boldsymbol{I_G}$, and $\boldsymbol{I_B}$ are 2-D discrete matrices. Due to the redundant nature of information contained in the individual matrices, it is convenient to transform them into a luminance/chrominance space. This is is known as the YCbCr domain. At a particular discrete point $[x, y]$ in the color image, YCbCr values can be obtained from the corresponding RGB values as per the transformation shown below[1]:

$$\begin{bmatrix} I_Y \\ I_{Cb} \\ I_{Cr} \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.5 \\ 0.5 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} I_R \\ I_G \\ I_B \end{bmatrix} \tag{2.1}$$

Here I_R, I_G, and I_B represent intensity values of matrices $\boldsymbol{I_R}, \boldsymbol{I_G}$, and $\boldsymbol{I_B}$ at co-ordinates $[x, y]$. Generally, these values lie between [0,1]. Let the transformed matrices be represented as $\boldsymbol{I_Y}, \boldsymbol{I_{Cb}}$, and $\boldsymbol{I_{Cb}}$. The discussion provided in this thesis generally corresponds to the *luminance* component $\boldsymbol{I_Y}$ of an image. Unless otherwise indicated, the symbol $\boldsymbol{I}$ generally refers to $\boldsymbol{I_Y}$.

Let $\boldsymbol{I_r}$ represent an intensity-only, right-view image and let $\boldsymbol{I_l}$ represent such an intensity-only image obtained from the left camera. Let the point $\boldsymbol{P}$ be imaged by both these cameras. Intensity values from both images are indicated as $I_r(x_1, y_1)$ and $I_l(x_2, y_2)$, where $[x_1, y_1]$ and $[x_2, y_2]$ indicate spatial co-ordinates of the projected point in both cameras. In an ideal scenario, both intensity values should be equal. However, this

[1] Assuming that RGB values are gamma-corrected

may not be exactly true due to illumination conditions and problems associated with optical or electronic components in the camera. Hence $I_r(x_1, y_1) \approx I_l(x_2, y_2)$.

Assume that the right image-view $\boldsymbol{I_r}$ constitutes a *reference* view while the left-view $\boldsymbol{I_l}$ constitutes a *target* view. If approximate equality between intensity values is satisfied, then the relative difference in spatial co-ordinates indicates the *disparity* of the point $\boldsymbol{P}$ in the target-view with respect to the reference-view. Hence

$$\boldsymbol{SDV}(x_2, y_2) = [x_1 - x_2, y_1 - y_2] \tag{2.2}$$

Thus, every spatial co-ordinate $\boldsymbol{x}$ in the target image will have a disparity vector $\boldsymbol{SDV}(\boldsymbol{x})$ associated with its corresponding location in the reference image, while satisfying the approximate equality condition. This can be mathematically represented as

$$I_l(\boldsymbol{x}) \approx I_r(\boldsymbol{x} + \boldsymbol{SDV}(\boldsymbol{x}))$$

This discussion is valid if $\boldsymbol{P}$ is stationary. Assume that this point now experiences a displacement over time. An additional time variable t must be introduced in the aforementioned expression. Thus, at time instant t,

$$I_l(\boldsymbol{x}, t) \approx I_r(\boldsymbol{x} + \boldsymbol{SDV}(\boldsymbol{x}, t), t)$$

where $\boldsymbol{SDV}(\boldsymbol{x}, t)$ represents the disparity-vector field between the reference and target images at time instant t.

Let $\boldsymbol{P}(t_1)$ represent the position of the point at time instant t_1. Assume that at time instant t_2 the point has been displaced and is indicated as $\boldsymbol{P}(t_2)$. As seen from the figure, this point is captured by both cameras and the image of this point has moved to co-ordinates $[x_3, y_3]$ and $[x_4, y_4]$ in the right- and left-views, respectively. If the approximate equality in intensity is extended then

$$\begin{aligned} I_r(x_1, y_1, t_1) &\approx I_r(x_3, y_3, t_2) \\ I_l(x_2, y_2, t_1) &\approx I_l(x_4, y_4, t_2) \end{aligned}$$

where an additional index of t_i reflects the displacement experienced by the point. This apparent displacement in two consecutive images indicates the *motion* experienced by the point. Two such motion-vectors, corresponding to both views, can be identified as:

$$\begin{aligned} \boldsymbol{MV}(x_3, y_3) &= [x_3 - x_1, y_3 - y_1] \\ \boldsymbol{MV}(x_4, y_4) &= [x_4 - x_2, y_4 - y_2] \end{aligned} \tag{2.3}$$

As with a disparity-vector field, there exists two such motion-vector fields for both images at time instant t_2, with respect to t_1. The approximate equality condition can be similarly expressed as

$$I_k(\boldsymbol{x}, t_2) \approx I_k(\boldsymbol{x} + \boldsymbol{MV}(\boldsymbol{x}, t_2), t_1)$$

where $I_k(.)$ indicates either a right or left image stream. The motion vector field at time instant t_j is given by $\boldsymbol{MV}(\boldsymbol{x}, t_j)$.

Eq. 2.2 defines disparity between two views at the *same time instant*. This is termed as a *standard disparity-vector* ($\boldsymbol{SDV}$) in this thesis. In addition, a vector is defined identifying the relative displacement of $\boldsymbol{P}$ between images at *different views* and at *different time instants*. This is termed as a *displaced disparity-vector* ($\boldsymbol{DDV}$). If $I_r(\boldsymbol{x}, t_1)$ is assumed to be the reference then, from the principle of approximate equality of intensities

$$I_l(\boldsymbol{x}, t_2) \approx I_r(\boldsymbol{x} + \boldsymbol{DDV}(\boldsymbol{x}, t_2), t_1)$$

where $\boldsymbol{DDV}(\boldsymbol{x}, t)$ indicates the displaced disparity-vector field for the target image at t_2 with respect to the reference image at t_1. For the point $\boldsymbol{P}$ shown in the figure

$$\boldsymbol{DDV}(x_4, y_4, t_2) = [x_4 - x_1, y_4 - y_1]$$

Usefulness of these vectors in stereoscopic moving-image compression will be established in a future chapter.

As previously mentioned, real world imaging systems have discrete spatial-sampling

structures. In this thesis, non-interlaced (sometimes called progressive[2]) imaging systems having rectangular sampling structures are considered. However, some commercial imaging systems have non-rectangular sampling structures that give rise to *interlaced* images which introduce *interlacing artifacts* [2]. Current video systems rely on interlaced transmission, due to the widespread use of interlaced display systems. It is expected that in future, progressive display systems will predominate over their interlaced counterparts. This forms the justification of using progressive images in discussing the performance of the proposed algorithm in this thesis research.

2.2 Wavelets and multiresolution analysis

Most state-of-the-art image compression algorithms use some form of transform-based analysis. A widely used standard, designed in the 1990's, is the *Joint Photographic Experts Group* (JPEG) compression algorithm. This is based on the *discrete cosine transform* (DCT) [24]. This algorithm yields reasonably good results for moderate compression ratios. However at higher compression ratios, the underlying block structure used in the DCT begins to manifest itself in the compressed image. These distortions are referred to in literature as *blocking artifacts*, and the HVS is extremely perceptive to them.

In the late 1990's, work was undertaken on a new compression standard that utilized the discrete wavelet transform (DWT) as an analysis tool. Wavelet methods involve *overlapping* transforms with a set of variable-length basis functions. Due to the overlapped nature of wavelet transforms, perceptually discomforting blocking artifacts are completely eliminated. In addition, the *multiresolution* character of such transforms leads to superior energy compaction and visually pleasing compressed images. These factors were responsible for incorporating the DWT as a transform tool in the new

[2]Not to be confused with the term progressive coding used in Chapter 3.

JPEG-2000 image coding standard [1]. Design specifications for wavelets are beyond the scope of this thesis. The concerned reader is directed to books by Daubechies [25], Mallat [26] and Chui [27] for a rigorous mathematical description of wavelets.

As stated previously, wavelets help in transformation of signals at multiple scales. Let $\boldsymbol{c_m}$ represent a one-dimensional discrete signal at scale-m. It can be transformed into its detail $\boldsymbol{d_{m-1}}$ and approximate $\boldsymbol{c_{m-1}}$ signals as

$$\begin{aligned} c_{m-1}[n] &= \frac{1}{\sqrt{2}} \sum_{k=2n}^{2n+K_1-1} c_m[k]\ \tilde{h}[k-2n], \\ d_{m-1}[n] &= \frac{1}{\sqrt{2}} \sum_{k=2n}^{2n+K_2-1} c_m[k]\ \tilde{g}[k-2n]. \end{aligned} \tag{2.4}$$

where $\tilde{\boldsymbol{h}}$ and $\tilde{\boldsymbol{g}}$ represent K_1- and K_2-coefficient filters (sometimes loosely referred to as wavelet-filters). Eq. 2.4 is otherwise referred to as *multiresolution analysis* equations.

Evidently, it is necessary to recover $\boldsymbol{c_m}$ from its detail and approximate components. This involves a combination of signals at coarse resolutions, sometimes referred to as *multiresolution synthesis*. This is obtained as

$$c_m[n] = \frac{1}{\sqrt{2}} \sum_{k=\lceil \frac{(n-K_2+1)}{2} \rceil}^{\lfloor \frac{n}{2} \rfloor} c_{m-1}[k]\ h[n-2k] + \frac{1}{\sqrt{2}} \sum_{k=\lceil \frac{(n-K_1+1)}{2} \rceil}^{\lfloor \frac{n}{2} \rfloor} d_{m-1}[k]\ g[n-2k] \tag{2.5}$$

It can be observed that filter lengths have been reversed when implementing a multiresolution synthesis. This is explained shortly. Fig. 2.2 depicts a multiresolution analysis and synthesis operation. There are several methods for designing these wavelet filters,

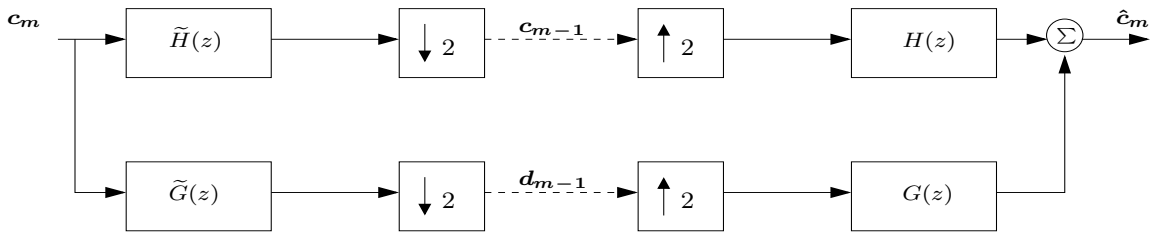


Fig. 2.2: 2-channel filter-bank.

e.g., those based on spectral factorization [28, 29], lattice structure [30], time-domain optimization [31] and quadratic-constrained least-squares [32]. Filters shown in Fig. 2.2 are *perfect-reconstruction* (PR) filters if they satisfy the following properties [26]:

$$\begin{aligned} \tilde{H}(z)H(z) - \tilde{H}(-z)H(-z) &= 2z^{-(2L+1)} \\ \tilde{G}(z) &= H(-z) \\ G(z) &= -\tilde{H}(-z) \end{aligned} \tag{2.6}$$

The filters used in these equations are related to each other as:

$$\begin{aligned} g[k] &= (-1)^{k-1}\tilde{h}[k] \\ h[k] &= (-1)^{k-1}\tilde{g}[k] \end{aligned} \tag{2.7}$$

In the literature, this criterion leads to the design of *biorthogonal filters*. This explains the difference in lengths when implementing an analysis or synthesis operation. A special scenario occurs when

$$\tilde{g}[k] = (-1)^k\, \tilde{h}[2K+1-k], \quad k = 1, 2, \ldots, 2K. \tag{2.8}$$

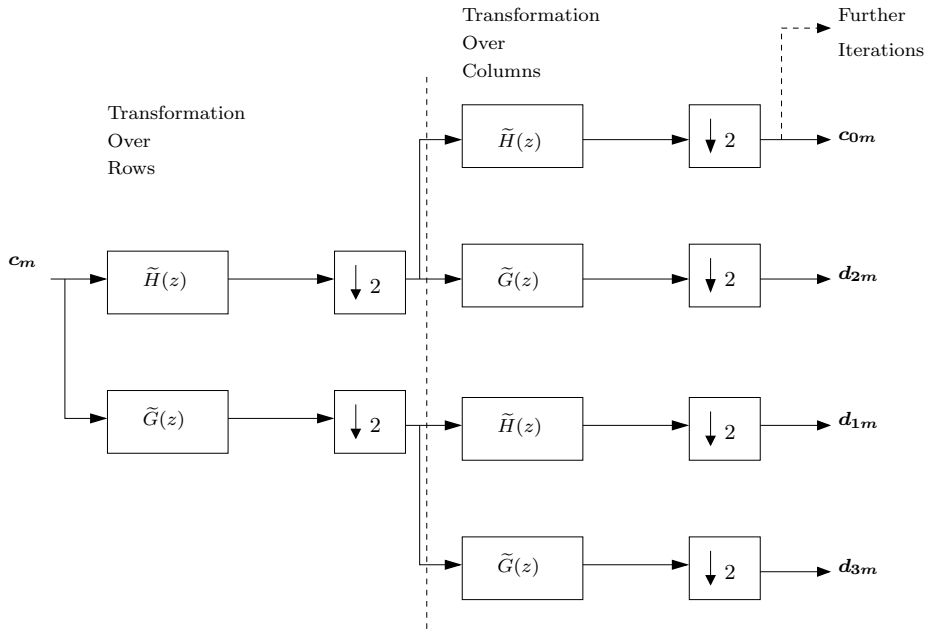
Filters satisfying this criterion are termed *orthogonal filters*. If this is true, all four filters can be computed from a single *mother wavelet*. Orthogonal filters would thus form a natural candidate in wavelet-based image compression. These filters should also be short in length, so as to speed up computation. Linear-phase properties are preserved when these filters are cascaded together (e.g., in a pyramidal decomposition). Aside from the *Haar* wavelet filter [26], non-trivial symmetric filters with *real* coefficients, satisfying Eqs. 2.7 and 2.8, do not exist. Symmetric filters are an asset in image compression as they maintain the correct spatial and time positions of coefficients [16].

As a result, biorthogonal filters are invariably used in state-of-the-art still- and moving-image coding algorithms. The most popular amongst these is the "CDF-9/7"

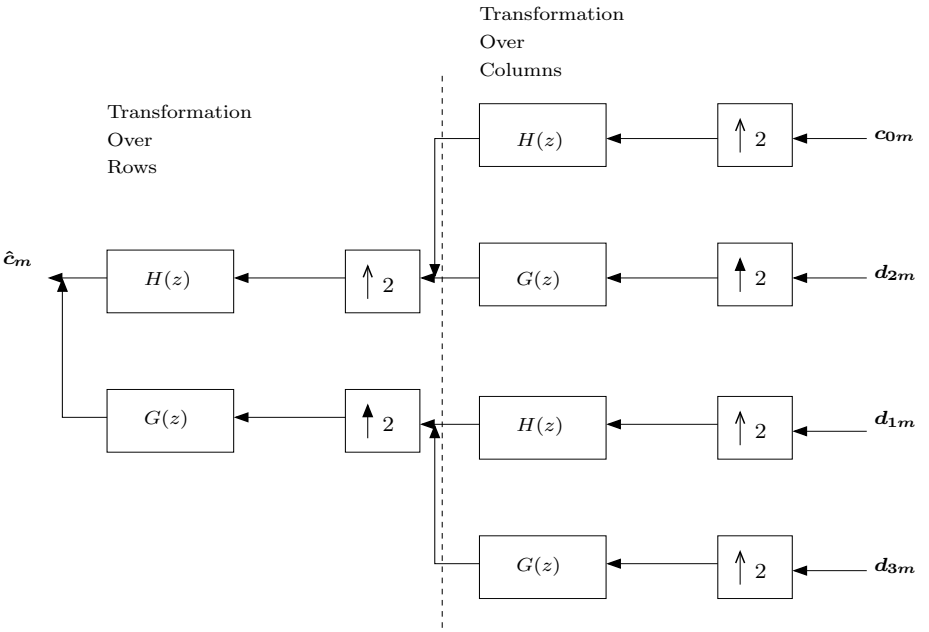
filter [25, p 279]. This is a symmetric filter having 9 and 7 filter taps in $\tilde{h}[n]$ and $\tilde{g}[n]$. This is also a part of the JPEG2000 image coding standard [1]. Consequently, most results presented in this thesis use a "CDF-9/7" filter. However, additional filters have been proposed that result in improved compression performance in a R-D context. In this thesis two such filters are used. These are the "Odegard-9/7" and "Cooklet-17/11" filters. Analysis-stage coefficients of these three filters can be found in Tables A.1, A.3 and A.5, listed in Appendix A.

The classic algorithm by Mallat [33] extended the 1-D DWT analysis scheme for transforming images. This involves an iterative and separate transformation of rows and columns of an image. This is shown in Fig. 2.3. Initially the 2-D discrete signal, $\boldsymbol{c_m}$, is transformed. This involves separately applying Eq. 2.4 on each row. The next stage involves applying the same set of equations on the columns generated from the first pass. The resulting coefficients are arranged in a "*Mallat-order*". These operations are continued until the iteration levels have been exhausted. Fig. 2.4 depicts a three-scale wavelet decomposition using Mallat's algorithm. In the literature, the subband having coefficients $\boldsymbol{c_{0i}}$ is generally referred to as a *LL*-subband; $\boldsymbol{d_{1i}}$ as a *HL*-subband; $\boldsymbol{d_{2i}}$ as a *LH*-subband and $\boldsymbol{d_{3i}}$ as a *HH*-subband. As rule-of-thumb, it is assumed that the top-left corner of an image is the reference point. This co-ordinate system is also followed when dealing with wavelet-transformed images (as seen from Fig. 2.4). Throughout this thesis, it is assumed that image dimensions are dyadic in nature. This limits the actual number of scales of decomposition. Non-separable versions of this algorithm have been proposed [34]. As two-dimensional convolution is involved, these transforms do not find much use in practical applications.

Mathematically speaking, the transform previously discussed is implemented on data sets of infinite length. However in real-world applications, implementing this algorithm on finite datasets introduces *edge artifacts*. Symmetric extension is commonly used



(a) Analysis-stage



(b) Synthesis-stage

Fig. 2.3: 1-level, 2-D separable, forward and inverse wavelet transform using Mallat's algorithm. $\boldsymbol{c_m}$ is a 2-D discrete signal.

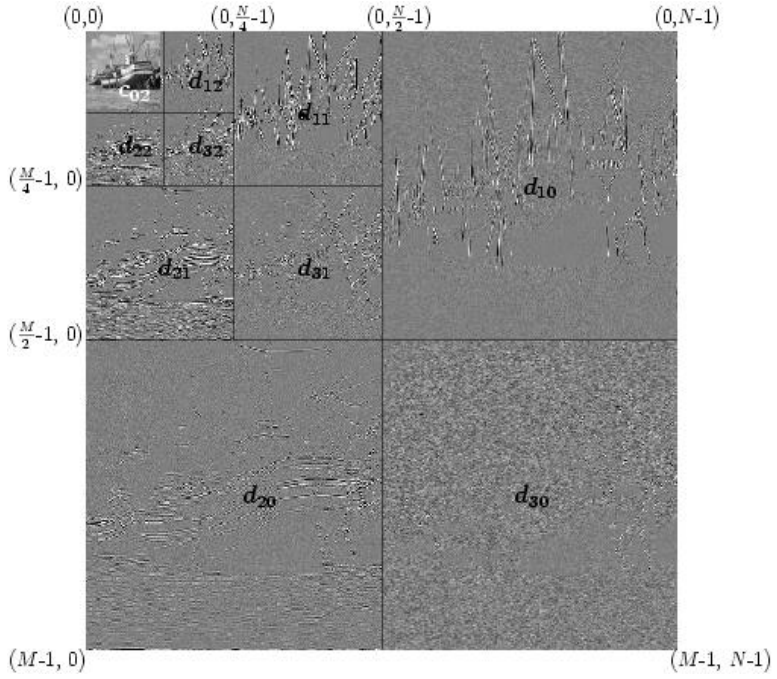
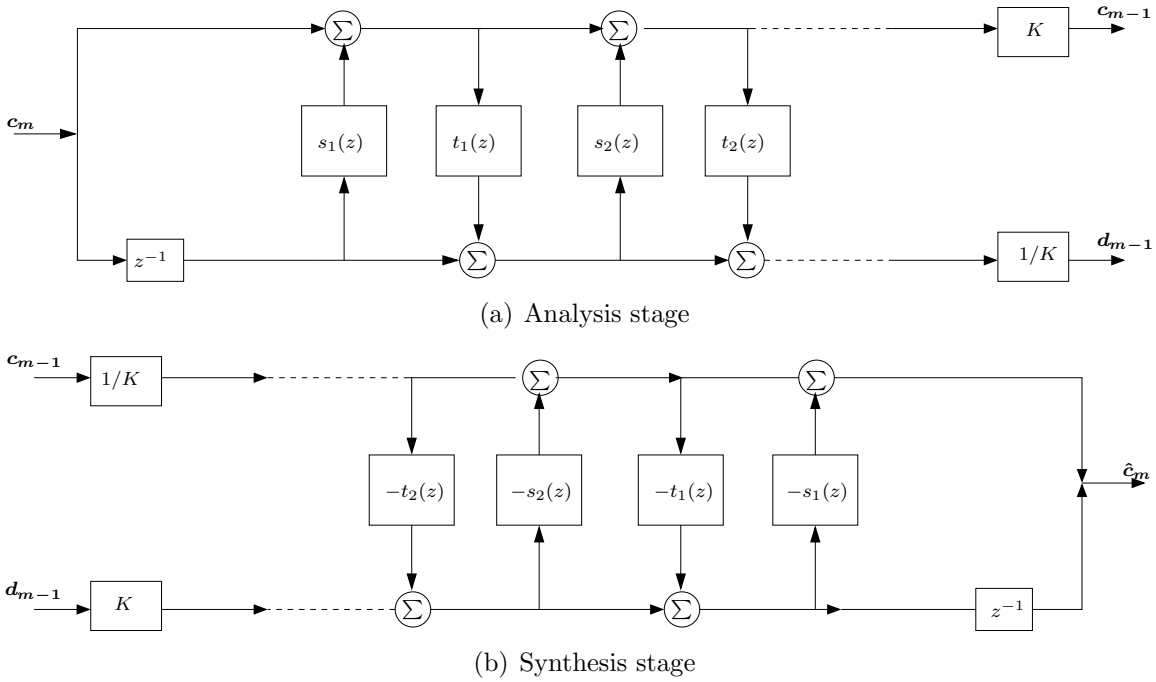


Fig. 2.4: 3-scale wavelet decomposition of a 2-D image (having dyadic dimensions), using Mallat's algorithm. Image dimensions are (M, N).

[35, 36, 37, 38] to minimize this problem in reconstructed images. As reported in [39], using symmetric extension introduces artificial discontinuities at edges. This tends to introduce edge artifacts in image subbands. As part of this thesis research, an extrapolated DWT was proposed [40]. In this a one-dimensional Burg extrapolation was implemented on the rows and columns, prior to a wavelet analysis or synthesis. This was an improvement on the polynomial-extrapolation technique presented in [39]. However, it is computationally more intensive than a symmetric extension technique.

A new approach to the DWT was proposed by Sweldens [41] to further simplify the transformation process. This approach is known as *Lifting*. It attempts to predict the approximate data from its detail counterpart and updates it in the first step. Subsequently the detail data is predicted from its approximate counterpart in the next step. A diagram for this implementation can be seen in Fig. 2.5. The non-unique nature of polynomial division, associated with the lifting step generation, can lead to many dif-



(a) Analysis stage

(b) Synthesis stage

Fig. 2.5: Lifting-based implementation of a 1-scale DWT, previously shown in Fig. 2.2

ferent implementations of the wavelet transform [42]. When compared with a standard implementation, there is a significant reduction in computational complexity when implementing a lifting-based DWT [42].

Let $\tilde{\boldsymbol{P}}(z)$ represent polyphase components of the analysis filters[41]

$$\tilde{\boldsymbol{P}}(z) = \begin{bmatrix} \tilde{H}_e(z) & \tilde{H}_o(z) \\ \tilde{G}_e(z) & \tilde{G}_o(z) \end{bmatrix} \tag{2.9}$$

Then, the lifting steps shown in Fig. 2.5(a) are related to $\tilde{\boldsymbol{P}}(z)$ by

$$\tilde{\boldsymbol{P}}(z) = \left\{ \prod_{i=1}^{n} \begin{bmatrix} 1 & s_i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t_i(z) & 1 \end{bmatrix} \right\} \begin{bmatrix} K & 0 \\ 0 & 1/K \end{bmatrix} \tag{2.10}$$

In a similar manner, lifting steps from the synthesis stage can be represented as

$$\boldsymbol{P}(z) = \begin{bmatrix} 1/K & 0 \\ 0 & K \end{bmatrix} \left\{ \prod_{i=1}^{n} \begin{bmatrix} 1 & -t_i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -s_i(z) & 1 \end{bmatrix} \right\} \tag{2.11}$$

A similar strategy can be employed in performing a 2-D separable transform. Due to these advantages, a lifting-based strategy is employed when implementing a DWT in

this thesis. For the sake of completeness, lifting-steps associated with the filters listed in Tables A.1-A.5, can be found in Appendix A.

2.3 Summary of disparity- and motion-estimation algorithms

2.3.1 Justification for disparity and motion estimation in stereoscopic moving-image coding

As mentioned in a previous section, stereoscopic image pairs acquired from two different cameras have nearly the same visual content. In any given view, there is also a strong correlation amongst images acquired at different time instants. Removal of such *inter-view* (former) and *intra-view* (latter) redundancies are an essential part of any state-of-the-art moving-image encoding standard.

A straightforward solution would be to independently encode all images acquired from a camera. However this has been shown to be inefficient when encoding monoscopic moving-images [2]. This led to the formulation of current industry standard moving-image encoding techniques. Such standards fall under the scope of H.264 and *Moving Picture Experts Group*-4 (MPEG-4) specifications. More information about current MPEG-4 standards can be found in [43, 44].

The standards specify that various contiguous images can be efficiently encoded by applying *prediction-based* techniques. For example, in Fig. 2.1 it is observed that $\boldsymbol{P}$ is present in both pictures (as projections) of the right image-view. Rather than encoding both images separately, MPEG standards specify estimating disparity- and motion-vector fields. These vector fields are used to generate disparity or motion compensated images. Residual images, generated by subtracting these compensated images from their originals, are instead encoded. The following sub-section summarizes relevant motion and disparity estimation techniques.

2.3.2 Summary of relevant algorithms for disparity- and motion-estimation

In an ideal scenario, it would be pertinent to estimate motion- or disparity-vectors for all possible pixel locations. However, this is a computationally expensive operation. Various solutions have been proposed to overcome this drawback. Due to the nature[3] of disparity- and motion-vectors in motion- or disparity-vector estimation, a generic review of some algorithms is presented.

A comprehensive review of motion-estimation algorithms, in the context of video coding, can be found in the paper by Stiller and Konrad [45]. When encoding video data, motion information constitutes *overhead*. Hence the goal of any motion-estimation algorithm is minimization of some objective criterion. Furthermore, as a sub-optimal solution, *region-based* estimation techniques [4] have been proposed to overcome the computational complexity of estimation over all pixel locations. This presupposes the fact that all pixels in the region being estimated have a constant motion- or disparity-vector. When performing region-based estimation, it should be remembered that information about a region *must* be made available at the decoder. Hence in addition to motion- or disparity-vector information, information about regions used in estimating these vectors should be transmitted. This tends to further increase the overhead information.

Region-based techniques can generally be classified into two distinct categories: *block-based* and *arbitrary-shape-based* estimation techniques. In the former, the image to be estimated is predicted from non-overlapping rectangular blocks from the reference image. If these blocks are of equal size and partition the image, then it is referred to as *fixed-block-based* estimation. On the other hand, if these blocks assume a finite number of rectangular shapes (e.g., 2×2 - 32×32) it falls under the category of *variable-block-based* estimation. A quadtree-partitioning is effected on the reference image in order to obtain these variable-shaped blocks. Generally, this is often used as a first stage when devel-

[3]Subtle differences between them are explained in a later part of this chapter.

oping other arbitrary-shaped (otherwise referred to as object-based) estimation schemes [4].

Due its relative simplicity, and ease of encoding, fixed block-based estimation algorithms are used in commercial video encoders [2] as well as disparity-estimation schemes [18, 11, 12]. Object-scalability is a desired feature of the new MPEG-4 video coding standard [43, 44]. In such scenarios, fixed block-based estimation would be deemed unsuitable when compared with their variable-block-based counterparts. It has also been reported that variable-block-based techniques outperform their fixed-block-based counterparts in a R-D sense.

Compared with their fixed-block-based counterparts, variable-block-based estimation schemes *preferentially* segment images. In other words, a few large blocks may be sufficient to represent large non-textured areas. On the other hand, smaller sized blocks would be utilized to represent textured regions. This results in an uneven distribution of bits when encoding disparity- or motion-vectors. Compared to this, both textured and non-textured regions are predicted from blocks of similar sizes in fixed-block-based estimation. Consequently, this uneven block-size distribution guarantees generation of "better" compensated images. This in turn insures residual images with sufficiently less energy content. The reader is directed to [46] and [17] for numerical results, justifying this fact.

To better explain block-matching and related notations, the reader's attention is directed towards Fig. 2.6. Let $\boldsymbol{B_i}$ represent a block from the target image. Assume that a best-matched block is found from the reference image. Let the displacement (motion or disparity) vector of $\boldsymbol{B_i}$, with respect to this best-matched block, be indicated as $\boldsymbol{v_i}$. On application of an approximate equality of intensity between both blocks

$$I_l(\boldsymbol{x}) \approx I_r(\boldsymbol{x} + \boldsymbol{v_i}), \quad \boldsymbol{x} \in \boldsymbol{B_i}, \quad i = 1, 2, \ldots, N_b$$

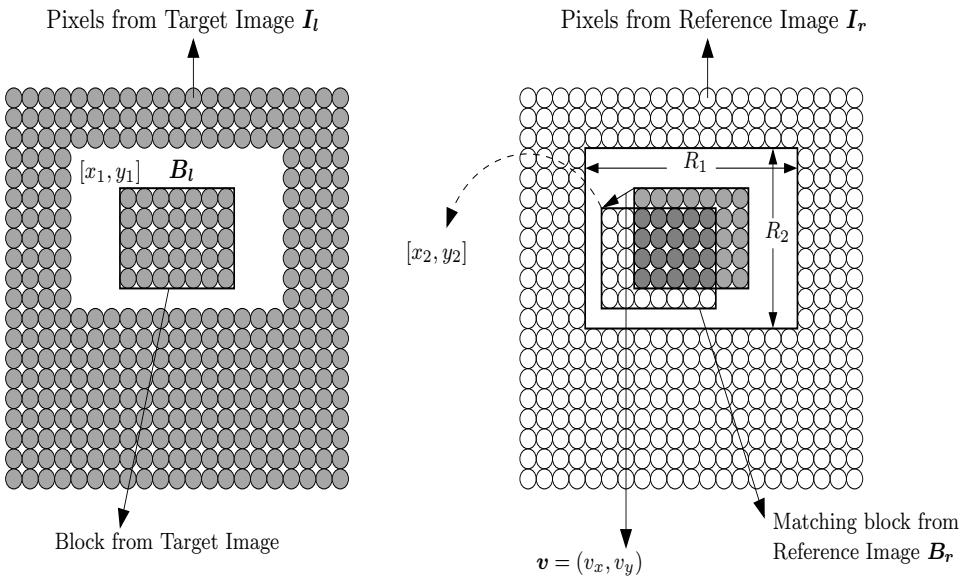


Fig. 2.6: Block disparity- or motion-vector estimation

where

$$\boldsymbol{B_i} = \{\boldsymbol{x} | \boldsymbol{x} \text{ within the } i^{th} \text{ block boundary}\}$$

and N_b is the number of blocks possible from the target image. In other words, the intensity of a block (positioned at $\boldsymbol{x}$) in the target-image is approximately equal to the intensity of a corresponding block from the reference image, but displaced by a vector $\boldsymbol{v}$. Due to this approximate equality, additional residual information needs to be transmitted. To achieve this, a *displaced frame difference* (DFD) is generated as

$$\boldsymbol{d}(\boldsymbol{x}, \boldsymbol{v}_i) = I_l(\boldsymbol{x}) - I_r(\boldsymbol{x} + \boldsymbol{v}_i) \tag{2.12}$$

where $\boldsymbol{v}_i$ is the actual displacement vector[4]. An objective function, *sum-of-absolute-difference* (SAD), is defined as

$$\text{SAD}(i, \boldsymbol{v}_i) = \sum_{x \in \boldsymbol{B}_i} |\boldsymbol{d}(\boldsymbol{x}, \boldsymbol{v}_i)| \tag{2.13}$$

[4]The subscript t, omitted in these expressions, is implicitly assumed.

An estimate $\widehat{\boldsymbol{v}}_i$ can be obtained by minimizing this objective function as

$$\widehat{\boldsymbol{v}}_i = \arg \min_{|v_x| \leq R_1, |v_y| \leq R_1} \text{SAD}(i, \boldsymbol{v}_i) \tag{2.14}$$

Here (v_x, v_y) are the horizontal and vertical components of the displacement vector $\boldsymbol{v}$, while (R_1, R_2) indicate the limits of the search area. Having estimated displacement vectors for all blocks in the target image, the DFD is encoded using an embedded image coding scheme.

Estimating these displacement-vectors involves positioning the reference image block at each pixel location in a region, and comparing it with the block under consideration from the target image. This exhaustive search strategy is referred to as a *full-search* (FS). This provides an optimal solution when estimating $\boldsymbol{v}$. However this benefit is outweighed by its computationally prohibitive costs in real-time applications. As a result, many sub-optimal strategies have been proposed to replace this. Some of these include a 2D-logarithmic search (2DLS) [47], three-step search (3SS) [48], four-step search (4SS) and a fast full-search [49] algorithm. A *hierarchical-search* strategy is an efficient solution that optimizes computational speed with respect to coding efficiency. This uses a combination of fewer search locations in addition to fewer pixels in determining $\boldsymbol{v}$. This technique has been successfully ported for real-time stereoscopic moving-image coding [50]. The following section presents a discussion of this technique, in conjunction with the previously described wavelet transforms.

2.4 Hierarchical-search strategy

Hierarchical estimation involves a two-step process. Initially, both reference and target images are repeatedly downsampled by a factor of two in each dimension [51, 2]. Next, a coarse disparity- or motion-estimate is made at a coarse scale using a FS strategy. Vectors obtained from this stage are scaled by a factor of two and refined in successive scales. A limited number of picture co-ordinates need to be scanned during this refinement process.

As seen from Fig. 2.3, the intrinsic downsampling structure of a 2D-separable DWT can be employed in hierarchical-search framework. An inverse wavelet transform using Mallat's 2-D separable transform is performed on locally quantized[5] reference and target images. As seen from Fig. 2.7, this estimation is performed on the all low-pass subband (LL). Assume that a displacement-vector (disparity or motion) $\widehat{\boldsymbol{v}}_i^k$ is obtained at a coarse scale k, subject to the criterion presented in Eq. 2.13. This vector, at the next fine-scale $k-1$, is estimated as

$$\widehat{\boldsymbol{v}}_i^{k-1} = \underbrace{2\,\widehat{\boldsymbol{v}}_i^k}_{\text{Coarse Estimate}} + \underbrace{\widehat{\Delta}_i^{k-1}}_{\text{Refinement}} \tag{2.15}$$

This is tantamount to first positioning the reference block at spatial co-ordinates indicated by $2\widehat{\boldsymbol{v}}_i^k$. A FS is implemented on a reduced area. This generally is ± 1 and ± 2 pixels[6] in either dimension when estimating motion- and disparity-vectors, respectively. It would suffice to transmit $\widehat{\boldsymbol{\Delta}}_i^{k-1}$ in order to decode $\widehat{\boldsymbol{v}}_i^{k-1}$, provided $\widehat{\boldsymbol{v}}_i^k$ is known. This process is continued until all remaining scales have been exhausted. This search strategy is indicated as RS in Fig. 2.7. The concerned reader is directed to [2, p 128-134] to better appreciate the computational savings obtained, when using this search strategy. Thus the output of a hierarchical-search strategy consists of vectors from a coarse-scale and refinement-vectors, with reduced magnitude, for all successive scales.

By reducing the spatial-resolution of images, motion or disparity between images are reduced by a factor of 2^m. Previously discussed algorithms (e.g., 2DLS, 3SS, etc.) have a high probability of getting trapped in local minima during disparity- or motion-vector estimation. Low-pass filtering used in generating a wavelet pyramid significantly reduces this probability [2].

A major drawback of hierarchical search strategies lies in its ability to reduce the disparity or motion between contiguous pictures. As the number of levels of decomposition

[5]Explained in a later chapter
[6]Empirical values

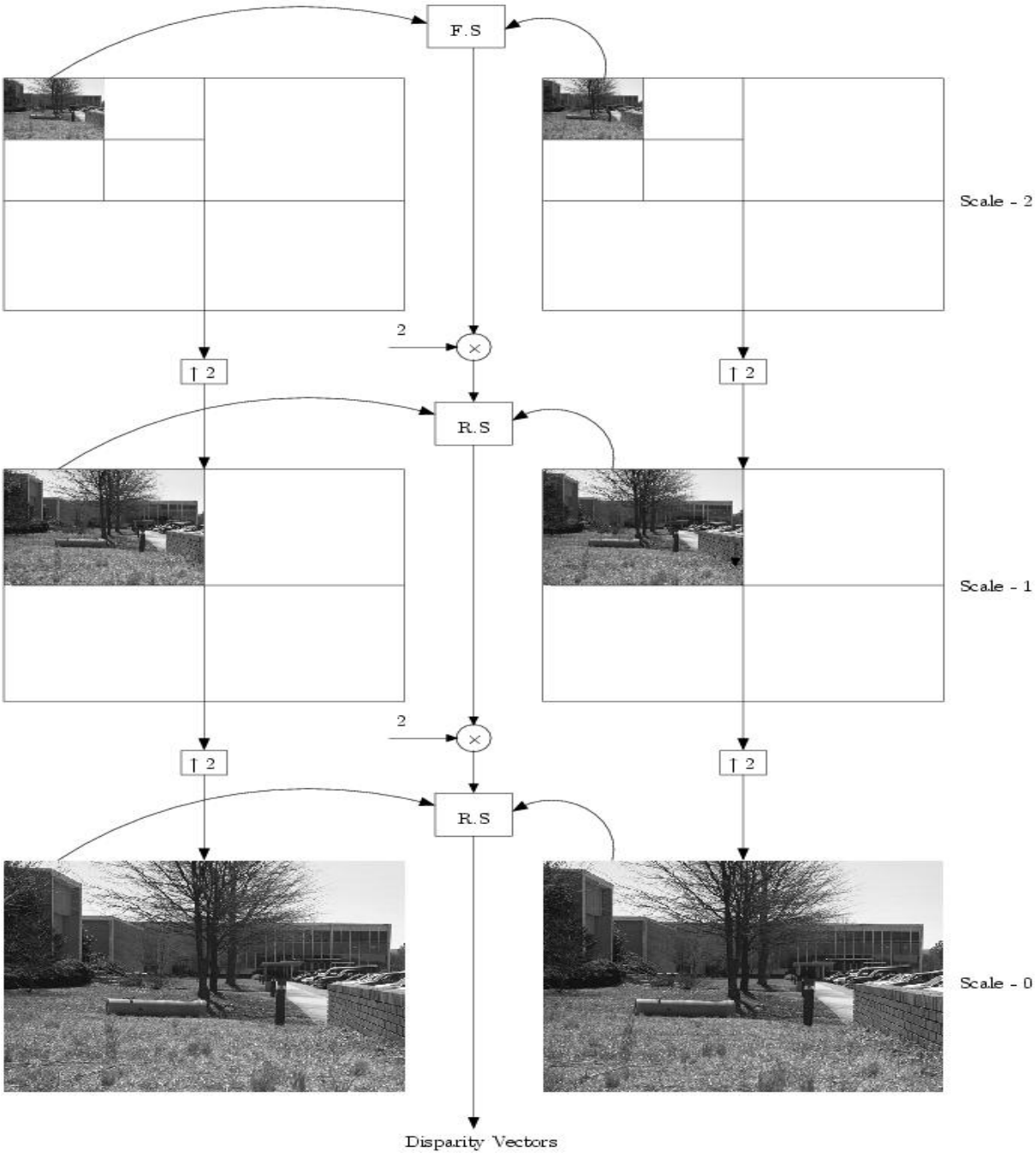


Fig. 2.7: Hierarchical disparity-vector estimation, in a DWT framework. ↑2 indicates a dyadic upsampling performed in a 2-D separable DWT. Disparity-estimation is performed between the all low-pass subband at each scale. Disparity-vectors from coarse-scales are scaled by a factor of two, when moving to a finer scale. A similar strategy can be used when estimating motion-vectors.

increases, more features are removed from the image. This removes any semblance of disparity or motion between these pictures. In this thesis, disparity- or motion-estimation is limited to three scales. Fig. 2.7 indicates a 3-scale hierarchical variable-block-based disparity estimation. Another major drawback is its inherent inability to track regions containing small objects.

To conclude this chapter, subtle differences between disparity- and motion-vector estimation is discussed. Disparity between images is primarily dependent on camera-geometry. Due to this fact, further simplification can be made when estimating disparities. From Fig. 2.6 the vertical co-ordinates, R_2, can be limited to at most a few pixels (e.g., ± 2) when performing a FS estimation at a coarse scale. This, eventually reduces the computational complexity of the proposed algorithm.

The following chapter presents a review of pertinent still-image, wavelet-based, coding algorithms with special emphasis on an ASWDR algorithm.

Chapter 3

Adaptively-Scanned Wavelet-Difference-Reduction Algorithm

Overview

This chapter introduces the concept of progressive image coding and decoding. A review of zerotree and non-zerotree based algorithms is also presented. Limitations of such algorithms, in the context of stereoscopic image coding are presented. This leads to the justification in using an adaptively-scanned wavelet-difference-reduction (ASWDR) algorithm in encoding stereoscopic imagery. This chapter is concluded by a detailed discussion of steps involved in implementing an ASWDR algorithm. Results from Appendix B complement this discussion.

3.1 Progressive coding of still images

PSYCHO-VISUAL experiments have revealed that the HVS is less sensitive to perturbations in high-frequency components of an image than in the low-frequency components. In a compression strategy this implies that high-frequency components in an image can be more coarsely represented than their lower-frequency counterparts. This reasoning applies to gray-scale image coding. Redundancies due to color are discussed in a later chapter.

Mean-squared error (MSE) is a widely used objective metric to test the performance

of any image coding scheme. The efficiency of any coding algorithm can be gauged by its ability to reduce the MSE between the original and reconstructed image for a given bit-rate. Let the original intensity image be represented as $\boldsymbol{I}$ and let the reconstructed image be represented as $\hat{\boldsymbol{I}}$. Thus MSE is defined as

$$\text{MSE} = \frac{1}{MN} \sum_{x_i, y_i} \Big(I(x_i, y_i) - \hat{I}(x_i, y_i) \Big)^2$$

Here (M, N) represent the dimensions of the image being encoded.

Generally, MSE is expressed in a different form. This is known as peak signal-to-noise (PSNR) ratio and defined as

$$\begin{aligned} \text{PSNR} &= 10\, log_{10}\Big(\frac{k^2}{\text{MSE}}\Big) \\ k &= \begin{cases} 1, & \boldsymbol{I} \in [0,1] \\ 255, & \boldsymbol{I} \in [0,255] \end{cases} \end{aligned}$$

Progressive image coding may be defined as a sequential transmission of wavelet-transformed coefficients, wherein coefficients with higher magnitudes[1] are reconstructed prior to lower-magnitude components. Consider Fig. 3.1. This depicts a section of a histogram of wavelet-transformed coefficients. It should be indicated that positions of these coefficients cannot be deduced from this histogram. Coefficients with higher magnitudes are significantly less frequent than those with lower magnitudes. Consider two representative threshold values T_1 and T_2 with $T_1 > T_2$. Progressive image coding implies that all coefficients with magnitude greater than T_1 are encoded prior to coefficients whose magnitude is greater than T_2 but less than or equal to T_1. The energy-compaction properties of wavelet transforms [26] insures that (nearly) all regions of an image can be efficiently reconstructed with a limited set of coefficients.

As indicated in the previous paragraph, positions of these wavelet coefficients cannot be inferred from Fig. 3.1. Benefits of wavelet-based transforms in image coding becomes evident if positions of these "significant coefficients" can be encoded efficiently. This

[1] From any subband

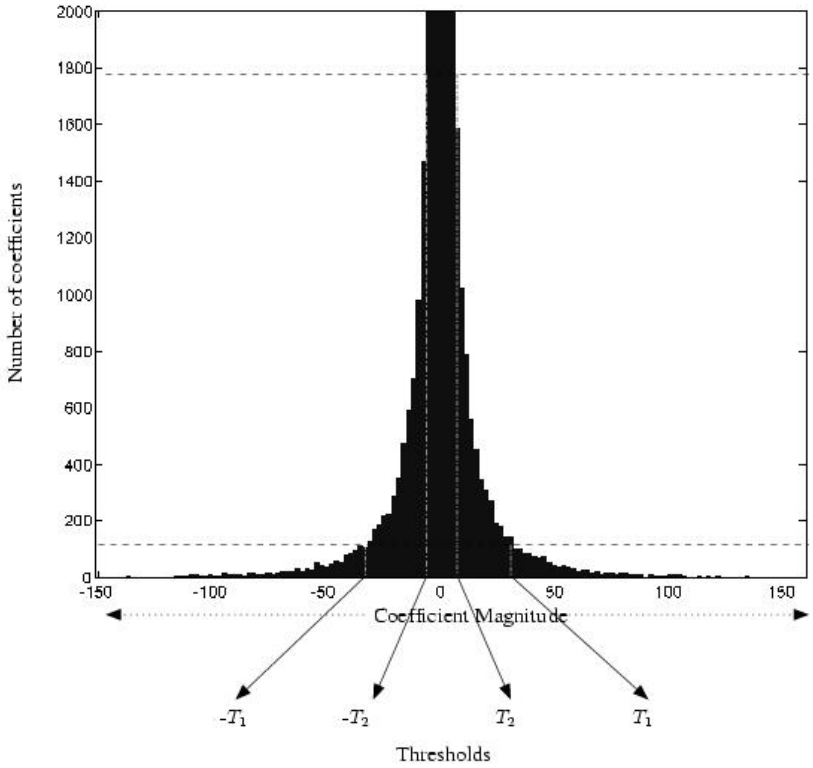


Fig. 3.1: Progressive image coding of wavelet-transformed coefficients. The x-axis indicates magnitudes of coefficients. $|T_1| > |T_2|$. The y-axis indicates the number of coefficients satisfying a "*greater than threshold*" criterion.

forms the basis of any modern embedded image coding algorithm. A trivial solution would be to independently encode each subband. However more sophisticated encoding schemes have since been proposed. As the reference view in a stereo-image pair is encoded using a state-of-the-art coding algorithm, the following section summarizes some widely used coding algorithms. In addition, a discussion is provided that highlights one such algorithm used in this thesis.

3.2 Summary of wavelet-based image coding schemes

In his classic paper [9], Shapiro identified a *parent-child* dependency among wavelet coefficients. For example, in a 3-scale wavelet-transformed image (Fig. 2.4), a parent

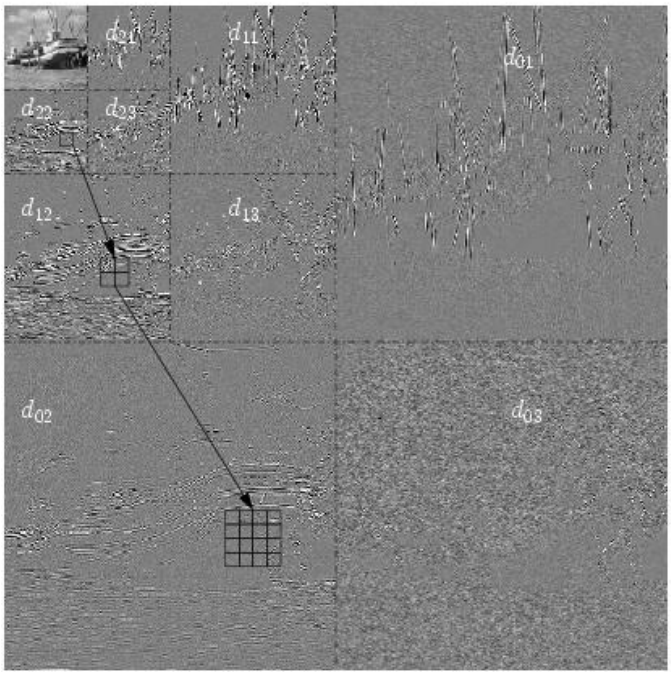


Fig. 3.2: Parent-child or *inter-scale* relationship between wavelet coefficients at different subbands. Coefficients have been scaled for display purpose.

coefficient in $\boldsymbol{d_{22}}$ induces four child coefficients in $\boldsymbol{d_{12}}$. Each child coefficient further induce four additional child coefficients in $\boldsymbol{d_{02}}$. This parent-child relationship exists at all subbands except for the all low-pass subband $\boldsymbol{c_{02}}$. This is visually depicted in Fig. 3.2. Shapiro conjectured that in such pyramidal image representations, if a coefficient at a given scale (e.g., $\boldsymbol{d_{12}}$) is deemed *insignificant* with respect to a certain threshold value, then descendants of this coefficient (i.e., 4 children, 16 grandchildren etc.) lying in finer scales and at same spatial orientations would most *likely* be insignificant.

Shapiro termed this as a *decaying spectrum hypothesis*. Consequently, he coined the term *zerotree* in order to identify such coefficients and its descendents. The ease with which zerotrees can be identified determines the overall performance of an embedded image coding algorithm. As a result, positions of these descendants need not be encoded. This crucial fact, in identifying locations of zerotrees, led to the development of Shapiro's classic *Embedded Zerotree Wavelet* (EZW) coding algorithm [9].

A simplified representation of zerotrees in an embedded bit-stream was made possible by the pioneering work of Said and Pearlman [10] referred to as *Set Partitioning in Hierarchical Trees* (SPIHT). This is very similar to Shapiro's original work. However a new feature, called *state-transitions*, was incorporated when classifying zerotrees in subbands [16, Sec. 6.3.3]. This results in a bit-stream containing three symbols in the dominant pass (+, - and S_t) where S_t indicates a state-transition. This significantly improves the coding performance, when compared with Shapiro's EZW technique having 4 symbols in the dominant pass (+, -, I, ZT). Here I refers to an isolated significant coefficient, while ZT indicates a zerotree root. The concerned reader is directed to [16] for an explanation of the aforementioned terms as well as a comparative study of both these algorithms.

As seen from Fig. 3.2, zerotree-based algorithms rely on *inter-scale* correlation among wavelet-transformed coefficients. This assumption stems from the decaying-spectrum hypothesis. In reality this is partially true. As noted by Shapiro [9, Pg. 3451]

> "... Experiments run on about 30 images of all different types, show that the correlation coefficient between the square of a child and the square of its parent tends to be between 0.2 and 0.6 with ..."

This indicates that inter-scale correlation amongst coefficients *need not be an optimal criterion for encoding images*. An alternate criterion, in which bit-plane correlation rather than inter-scale correlation is used to encode coefficients. This involves significance determination with respect to coefficients in all subbands at a given scale. Hence, this criterion is termed as *intra-scale* correlation and illustrated in Fig. 3.3.

This was identified by Lan and Tewfik [15]. The algorithm proposed by these authors was termed *Multigrid Embedding* (MGE) of wavelets. In this, positions of coefficients at a given level are encoded by quadtree-partitioning. A similar scheme, based on encoding positions of significant coefficients via quadtrees, was independently proposed by

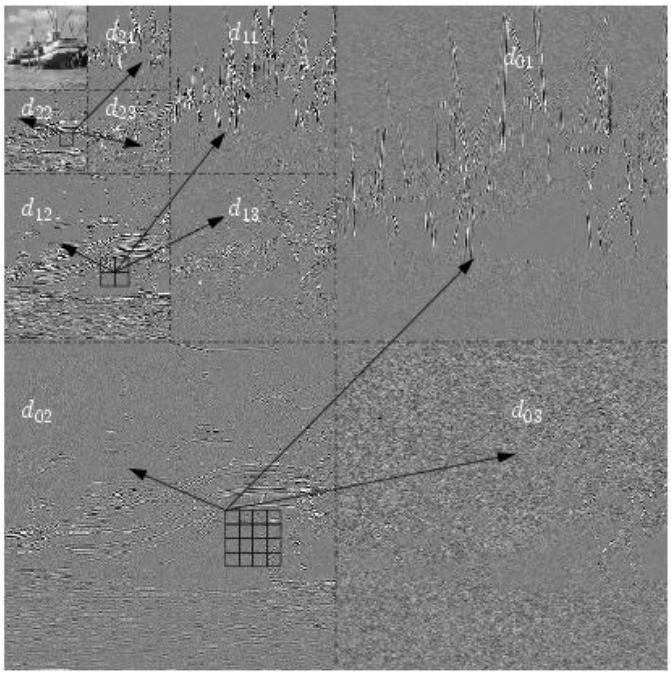


Fig. 3.3: *Intra-scale* relationship between wavelet coefficients in any given subband. Coefficients have been scaled for display purpose. The direction of arrows indicate that all coefficients at a particular scale are examined before a coefficient can be deemed significant.

Munteanu *et al.* [52]. It has been reported that an MGE algorithm is useful in encoding natural images [15] and disparity compensated residual images [12] (without entropy coding) when compared with traditional SPIHT algorithms. This fact forms the basis for the discussion presented in the next section.

3.3 Justification of using an adaptively-scanned wavelet-difference-reduction algorithm

Independent of the authors in [15] and [52], a technique that utilizes intra-scale correlation in encoding images was proposed by Tian and Wells [53]. This was termed a *wavelet-difference-reduction* (WDR) algorithm. Unlike the MGE algorithm, this technique does not rely on quadtrees to locate the positions of significant coefficients, making it a computationally simpler process. Instead, the coefficients are accumulated in a pre-

determined scan order. Positions of significant coefficients in the scan order are encoded using a *binary-difference-reduction* technique, explained later in this chapter. As with MGE, WDR also has been shown to provide improved perceptual and quantitative results when encoding images with high-frequency content [53] (without entropy coding) when compared with SPIHT.

From Figs. 3.2 and 3.3 it is observed that the aforementioned techniques sacrifice either intra- or inter-scale correlation when encoding coefficients. Non-exploitation of intra-scale correlation affects the performance of SPIHT when encoding images with high frequency content. Similarly, non-exploitation of inter-scale correlation amongst subbands affects the performance of both MGE and WDR when encoding natural images.

To alleviate this problem, Walker and Nguyen proposed an improved version of the WDR algorithm [16]. This is referred to as an *adaptively-scanned wavelet-difference-reduction* (ASWDR) algorithm. Initially, coefficients are scanned in a pre-determined order. Significant coefficients are identified in a manner similar to that of a WDR algorithm. These significant coefficients are then used to adjust the positions of remaining insignificant coefficients. Insignificant child coefficients induced from significant parent coefficients are scanned prior to other coefficients at a particular scale. This new arrangement of coefficients is used to modify positions of coefficients at finer scales. This can be observed from Fig. 3.4. Details of this algorithm is presented in the following section.

3.4 Steps implemented in an ASWDR algorithm

To better understand the discussion, some notations are first presented.

- ***ICS*** is a 1-D vector, containing all coefficients from the transformed image and scanned in a certain order,

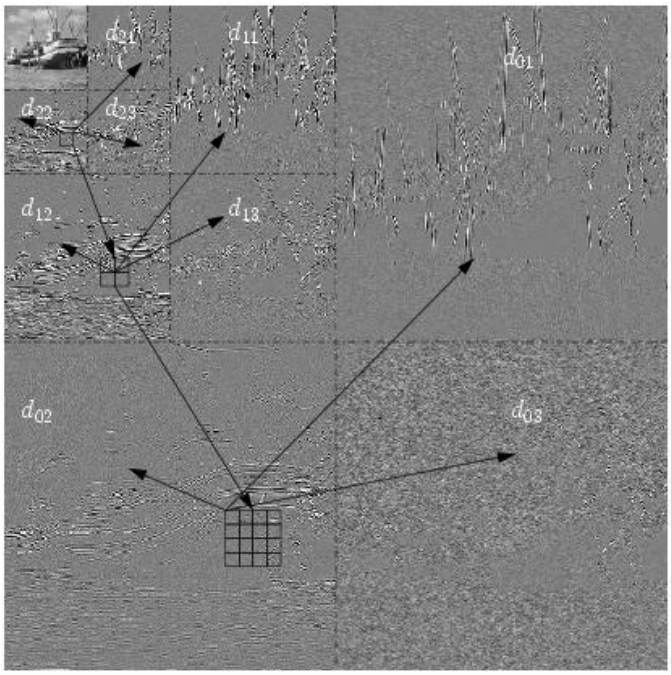


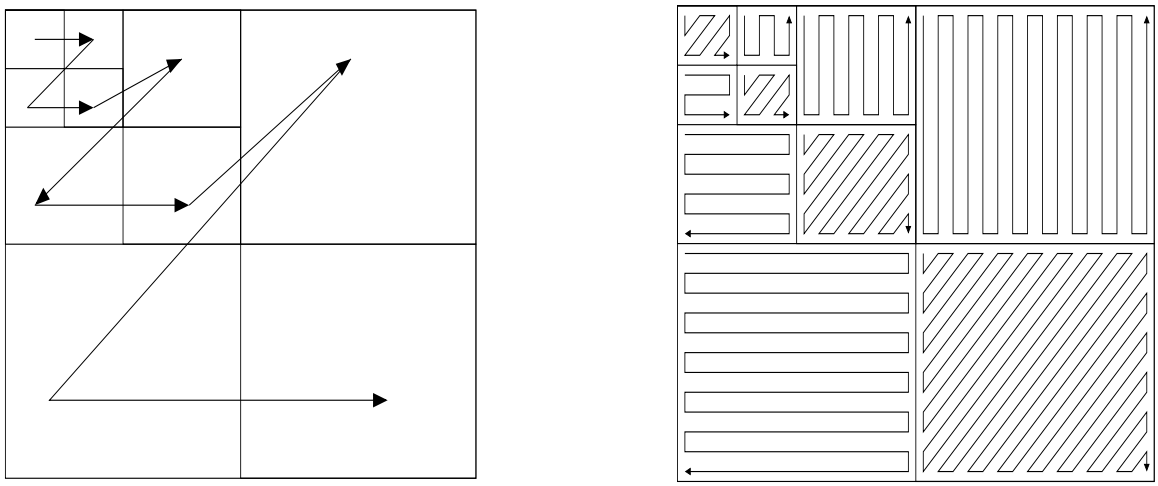
Fig. 3.4: Methodology used in an ASWDR algorithm. Significance of coefficients is determined via *intra-scale* correlation. *Inter-scale* correlation is used to "bring forward" descendants of previously identified significant coefficients. This reduces overall bits required to encode positions of significant coefficients.

- ***SCS*** is a 1-D vector, containing coefficients that have been deemed significant during dominant passes,
- ***TPS*** is a temporary 1-D vector, containing scan-updated coefficients,
- w indicates a wavelet-transformed coefficient, and
- γ is the wavelet coefficient, from the transformed image, with the largest magnitude.

The ASWDR encoding algorithm for a grey-scale image can be explained in a 7-step procedure. To better understand these steps, Shapiro's example of an 8×8 matrix, consisting of wavelet transformed coefficients [9, Fig.8], shown in Fig. 3.5, is considered. The steps are as follows:

63	-34	49	10	7	13	-12	7
-31	23	14	-13	3	4	6	-1
15	14	3	-12	5	-7	3	9
-9	-7	-14	8	4	-2	3	2
-5	9	-1	47	4	6	-2	2
3	0	-3	2	3	-2	0	4
2	-3	-6	-4	3	6	3	6
5	11	5	6	0	3	-4	4

Fig. 3.5: Shapiro's 8×8 image having three levels of wavelet transform



(a) Global (b) Individual subbands

Fig. 3.6: Scan order employed in accumulating coefficients.

- **Step 1**: A 2-D separable lifting wavelet transform (Fig. 2.3), using Mallat's algorithm [33] and lifting procedures outlined by Daubechies and Sweldens [54] is performed on the image.

- **Step 2**: The wavelet coefficients are globally scanned as per the order shown in Fig. 3.6(a). The coefficients in individual subbands are scanned in a:

  - JPEG-style zigzag order [24] in the all low pass (LL) subband,

  - Column-wise order in HL subbands,

– Row-wise order in LH subbands, and

– JPEG-style zigzag order in HH subbands.

This is shown in Fig.3.6(b). These coefficients, having been scanned in the aforementioned order, are placed in the $\boldsymbol{ICS}$ list. Additionally, an empty $\boldsymbol{SCS}$ list is also initialized. For this example,

$$\begin{aligned}\boldsymbol{ICS} \;=\; [&63, -34, -31, -23,\\ &49, 14, -13, 10, 15, 14, -7, -9, 3, -14, -12, 8,\\ &7, 3, 5, 4, -2, -7, 4, 13, -12, 6, 3, 3, 2, 9, -1, 7,\\ &-5, 9, -1, 47, 2, -3, 0, 3, 2, -3, -6, -4, 6, 5, 11, 5,\\ &4, 3, 6, -2, -2, 3, 0, 6, 0, 2, 4, 3, 3, -4, 6, 4]\end{aligned}$$

- **Step 3**: γ is calculated as

$$\gamma = \arg \max_{i \in [1, MN]} |ICS[i]| \tag{3.1}$$

where MN is the length of $\boldsymbol{ICS}$. (M, N) are the dimensions of the image. An initial threshold

$$T = 2^{\lfloor \log_2 \gamma \rfloor} \tag{3.2}$$

is chosen. Finally, a counter $j = 0$ is initialized. In this example

$$\gamma = ICS[1] = 63$$

and

$$T = 32$$

- **Step 4**: A *dominant scan* is performed on $\boldsymbol{ICS}$. This involves scanning all coefficients in an ascending index $(ICS[1], ICS[2], \ldots, ICS[MN])$ order. During this process if the magnitude of a coefficient $ICS[k]$, at position k, is greater than T it

is removed from $\boldsymbol{ICS}$. It is modified and placed in the $\boldsymbol{SCS}$ list, at position $l+1$. Here l is the current length of the $\boldsymbol{SCS}$ list. In other words

$$SCS[l+1] = \begin{cases} |ICS[k] - T|, & |ICS[k]| > T \\ \text{no change}, & |ICS[k]| \leq T \end{cases}$$

The original position of this significant coefficient, k, is encoded along with the sign of the coefficient. This is achieved by a *binary reduction* process. This involves a two stage process. In the first stage a binary representation of the (unsigned) integer is obtained with a minimum number of bits. For example

$$\begin{aligned} 5 &= 101 \\ 24 &= 11000 \end{aligned}$$

If observed, the most significant bit (MSB) of these binary representations is always equal to one. This can be omitted and instead, the sign of the coefficient can be used as a separator between two successive position values. During decoding, the MSB removed during the encoding process can be appended back in the binary representation. This correctly identifies the position of the coefficient in the list. In addition, the sign of the coefficient can also be decoded from this reduced binary representation. It has been shown [55], that a binary-reduced value is the shortest length by which a signed integer can be represented.

Positions of all coefficients, after this currently identified significant coefficient is decremented by one as

$$ICS[k+n] \leftarrow ICS[k+n+1], \quad n = 1, 2, \ldots, \text{last index position of the list}$$

and *shrinks* the list. In other words, the position index of the next significant coefficient is calculated *relative* to the position of the just extracted significant coefficient. The steps involved in extracting significant coefficients, in the first

dominant pass, for the given example are as follows:

$$\begin{aligned}
\texttt{pos.} &= +1, \quad \texttt{sig.coeff.} = 63 \\
\boldsymbol{ICS} &= [-34, -31, -23, 49, \ldots, 7, 3, \ldots, -5, 9, \ldots, 47, 2, \ldots, 4, 3, \ldots, 6, 4] \\
\texttt{SCS} &= [31] \\
\texttt{pos.} &= -1, \quad \texttt{sig.coeff.} = 31 \\
\boldsymbol{ICS} &= [-31, -23, 49, \ldots, 7, 3, \ldots, -5, 9, \ldots, 47, 2, \ldots, 4, 3, \ldots, 6, 4] \\
\boldsymbol{SCS} &= [31, 2] \\
\texttt{pos.} &= +3, \quad \texttt{sig.coeff.} = 49 \\
\boldsymbol{ICS} &= [-31, -23, 14, \ldots, 7, 3, \ldots, -5, 9, \ldots, 47, 2, \ldots, 4, 3, \ldots, 6, 4] \\
\boldsymbol{SCS} &= [31, 2, 17] \\
\texttt{pos.} &= +31, \quad \texttt{sig.coeff.} = 47 \\
\boldsymbol{ICS} &= [-31, -23, 14, \ldots, 7, 3, \ldots, -5, 9, \ldots, 2, -3, \ldots, 4, 3, \ldots, 6, 4] \\
\boldsymbol{SCS} &= [31, 2, 17, 15]
\end{aligned}$$

The *relative* nature of index positions can be guaged by observing the original $\boldsymbol{ICS}$ list. For example, 47 has an *absolute* index position of 5 while the actual value encoded is +3 (where the + indicates the sign of the coefficient). Hence, at the end of the first dominant scan, the following position indices, [+1, -1, +3, +31], are output. A reduced binary representation of these signed integers would be as follows:

$$+ \; - \; 1 + \; 1111+$$

An *end-of-scan* (EOS) indicator is also needed to enable the decoder to differentiate between two successive dominant or refinement scans. This is achieved by outputting a binary-reduced value

$$C = MN - C_{sec} + 1$$

followed by a + symbol, at the end of a dominant scan. This value is chosen, as it is guaranteed to be out-of-range when compared with the maximum index position in the modified list. Here MN is the length of the original $\boldsymbol{ICS}$ list, prior to encoding. C_{sec} indicates the length of the $\boldsymbol{SCS}$ list at the end of the current dominant scan. At the end of the first dominant scan, four coefficients are deemed significant. The maximum index position of the modified $\boldsymbol{ICS}$ list is 60. Thus, $C = 61$ making it just out-of-range with respect to 60. Hence, the complete bit-stream at the end of the first dominant scan would be as follows:

$$\texttt{bit}-\texttt{stream} = +\ -\ \texttt{1} + \ \texttt{1111} + \ \texttt{11101}+$$

- **Step 5** : This is a *refinement scan* and is implemented only if $j > 0$. This is similar to the refinement scans implemented in current embedded image coding algorithms like EZW, SPIHT, MGE, etc.

  All elements of $\boldsymbol{SCS}$ obtained from previous dominant scans, except for the one just concluded, are examined. For a particular coefficient $SCS[l]$ a refinement bit R is generated, subject to the following conditions:

$$R = \begin{cases} 1, & SCS[l] > T \\ 0, & SCS[l] \leq T \end{cases}$$

  Consequently the values of these coefficients are also modified as:

$$SCS[l] \leftarrow \begin{cases} SCS[l] - T, & R = 1 \\ \text{unchanged}, & R = 0 \end{cases}$$

  In this example, the first refinement scan occurs after the second dominant scan, where $T = 16$. Hence, the bit-stream and the modified $\boldsymbol{SCS}$ list would be as follows:

$$\begin{aligned} \texttt{bit}-\texttt{stream} &= \texttt{1010} \\ \boldsymbol{SCS} &= [\texttt{15}, \texttt{2}, \texttt{2}, \texttt{15}] \end{aligned}$$

- **Step 6**: Here, the scan-order of coefficients remaining in $\boldsymbol{ICS}$ is modified. As seen from Step 4, the minimum number of bits required for a binary representation of an integer increases with increasing magnitude. In this step, descendants and siblings of coefficients previously deemed significant are "*brought forward*". This is tantamount to reducing magnitudes of position indices of coefficients that may be deemed significant in future dominant scans. Justification for using these steps can be found in [16, 56].

  Initially, the remaining insignificant coefficients from the all low-pass subband are scanned and placed in $\boldsymbol{TPS}$. The following steps are implemented to update the scan order:

  - Significant parents at scale k, generated from previous scans, are noted. Here, significance is derived with respect to the *current* threshold value. The first part of the scan, at the next fine scale $k-1$, contains the insignificant values lying amongst the children induced by these parents. The remaining insignificant values (parents) at level k are then scanned and joined to $\boldsymbol{TPS}$.
  - The second part of the scan order at scale $k-1$ consists of insignificant child values, *having at least one significant sibling.*
  - The third part of the scan order at scale $k-1$ consists of insignificant child values, *having no significant sibling.*

  In order to reflect the hierarchical importance of coefficients at this bit-plane, three different chains are created. At the completion of the current scale (bit-plane), these chains are sequentially joined to the $\boldsymbol{TPS}$. Subsequently, this modified ordering is used to update the scan order of coefficients in the next fine scale. This process is repeated until all levels have been exhausted. $\boldsymbol{ICS}$ is subsequently replaced with $\boldsymbol{TPS}$.

- **Step 7**: The counter j is incremented by 1, while the present threshold T is scaled by a factor of $\frac{1}{2}$. Subsequently Steps 4-6 are repeated until a specified bit-budget is exhausted.

There are 3 possible symbols that are output during the course of the aforementioned steps:

- Dominant Scan : 0, 1, and 'S'. Here 0, 1 are used to represent values from the binary-reduced sequence. 'S' indicates the occurrence of a separator between two consecutive positions. The sign of the coefficient is determined by the bit following 'S'. A positive value is inferred if this bit is 0 while a 1 indicates the occurrence of a negative number. This extra bit is omitted when outputting the binary reduced values for an EOS.

- Refinement Scan : 0, 1. These are the only two possible values that emanate during this scan.

These symbols are losslessly encoded using a context-based arithmetic coder (CAC) [57]. Similar to other state-of-the-art algorithms (e.g., SPIHT, EZW, WDR, etc.) the encoded bit-stream is preceded by a header information. This generally consists of the image dimensions and initial threshold T.

During decoding, the steps described above are recapitulated and a quantized output is produced. An inverse wavelet transform is performed on these quantized values to obtain a final decompressed image.

3.5 An example

To conclude this chapter, a complete (unencoded) bit-stream[2] is provided, when encoding Shapiro's image in Fig. 3.5. The main objective of this example is to show the "bring-forward" effect in an ASWDR algorithm when compared with a WDR counterpart. For

[2]It should be emphasized that position indices are eventually represented as binary-reduced values.

additional results and qualitative discussion between ASWDR, SPIHT and JPEG2000, the concerned reader is directed to Appendix B.

Closely observing both tables, it can be noticed that "position indices" of significant coefficients are represented as smaller numbers in Table 3.2 when compared with Table 3.1. As an example consider the last value at the end of each dominant scan. In most instances, this value is less in Table 3.2 than in Table 3.1. This indicates that using a scan-update procedure (Step 6 of Sec. 3.4) insures that, in general, more number of coefficients are encoded. This fact may not be easily perceived from this trivial example. In order to explain this the reader is directed to results presented in Appendix B.

Table 3.1: Output data stream for the matrix shown in Fig. 3.5 using a WDR encoding process. The column on the left indicates the current pass (D_i = dominant pass, S_i = refinement pass). The column on the right indicates the threshold for the current pass. Boldfaced numbers indicates an occurrence of EOS .

D_1	1+1-3+31+**61+**	32.0
D_2	1-1+**59+**	16.0
S_1	1010	16.0
D_3	1+1-1+1+1+2-2-1-9+1-5+4+12+ **46+**	8.0
S_2	100100	8.0
D_4	1-2+1+2+3-2+5+1-8+2+1+1+3+5+7+ **31+**	4.0
S_3	1001011001100010000	4.0
D_5	1+1+1+2+1+1+5-2+2-1-1+1+3+4+1+1+1-1+ **13+**	2.0
S_4	1001010001000000001011000100000000	2.0
D_6	1-1+3+2+1-1-3+ **6+**	1.0
S_5	0000001010100000000010001001100111001100000010010001	1.0
D_7	1-1- **4+**	0.5
S_6	1011011011100010111011101101111111111111010011111101011100 1	0.5

Table 3.2: Output data stream for the matrix shown in Fig. 3.5 using an ASWDR encoding process. The column on the left indicates the current pass (D_i = dominant pass, S_i = refinement pass). The column on the right indicates the threshold for the current pass. Boldfaced numbers indicates an occurrence of EOS .

D_1	1+1-3+31+**61+**	32.0
D_2	1-1+**59+**	16.0
S_1	1010	16.0
D_3	1+1-1+1+1+2-2-1-5+11+1-5+9+ **46+**	8.0
S_2	100100	8.0
D_4	1-2+1+3+3-4+2+1-8+3+6+2+1+3+4+ **31+**	4.0
S_3	1001011001100010000	4.0
D_5	1+1+1+1+2+1+4+3-1-2-1+2+4+1+1-1+1+1+ **13+**	2.0
S_4	1001010001000000001011000100000000	2.0
D_6	1-1+4+1+1-2-2+ **6+**	1.0
S_5	0000001010100000000010001000111011001100000000100110	1.0
D_7	1-2- **4+**	0.5
S_6	1011011011100011011011101101111111111111000111101110111001	0.5

Chapter 4

Stereoscopic Still-Image Coding - A summary

Overview

A review of current trends in stereoscopic still-image coding is presented. This is followed by a mathematical derivation of optimal solutions when encoding such imagery. The concept of asymmetrical coding of stereoscopic imagery is also discussed. Two specific algorithms [11, 12], pertinent in the proposed research work, are also discussed here. Limitations of these algorithms are presented. This provides a justification in designing the algorithm proposed in this thesis.

4.1 Introduction

IMAGES in a stereoscopic pair essentially depict the same scene, but imaged from two slightly different points of view (e.g., Fig. 2.1). Hence, independent storage or transmission of these images is an extremely redundant operation [3]. Instead, methods that exploit the disparity between them have been shown to produce better results in a rate-distortion (R-D) framework. The concepts of disparity and disparity-vector estimation in stereoscopic imagery were previously discussed in Chapter 2.

From an encoding point of view it has been shown [3] that, when viewed in a stereoscopic mode, both images need not be displayed at full perceptual quality. This fact is validated from extensive subjective results presented in [3, 58, 59]. This also led to the

development of asymmetrical coding in stereoscopic imagery, a fact discussed later on in this chapter.

In [60] a simple wavelet-transform-based method was presented for encoding stereoscopic still-images, without disparity-estimation. This system uses Lloyd-Max quantizers while exploiting properties of the HVS. Lack of disparity-estimation makes it an inefficient system. In addition, Lloyd-Max quantizers have been reported to provide sub-optimal results when encoding wavelet coefficients [61].

Superiority of disparity compensated coding techniques have been previously demonstrated. In [5, 6, 7], DCT-based methods have been proposed. Due to inherent drawbacks of DCT-based systems, as discussed in Chapter 2, DWT-based methods were developed. In [8] a sophisticated wavelet-based coding scheme has been proposed. This uses disparity-compensation in a wavelet domain and subspace projection techniques for encoding wavelet coefficients. This method is computationally expensive as different basis functions need to be constructed for representing each block of wavelet coefficients. It addition, this technique does not support progressive transmission of wavelet coefficients.

To overcome these drawbacks, Boulgouris and Strintzis [11] proposed a novel solution. This scheme guarantees high levels of SNR-scalability, subject to some conditions. In addition, a very simple disparity-estimation is performed making it computationally efficient. Fixed-block-based disparity-estimation techniques were used by the authors of that paper. A similar concept has been proposed in [17], but variable-block-based disparity-estimation is used. The authors of [17] have shown the superiority of using such a technique when compared with fixed-block-based counterparts.

Frajka and Zeger [12] proposed an alternative theory with respect to encoding stereoscopic imagery. They argued that global encoding of residual images may not produce superior results in a R-D framework. Their hypothesis was based on work, previously reported by Moellenhoff and Maier [14], that identified the distinctive nature of disparity

compensated residual images. The results, reported by the authors [12], were obtained using a conditional coder (CONCOD) [3]. This has been shown to be slightly inferior when compared with a closed-loop structure [11]. Details of this technique are also discussed in this chapter.

4.2 Solutions for disparity estimation and compensation

To better understand the algorithms discussed in this chapter, as well as the algorithm proposed in this thesis, some background information must be discussed. These pertain to disparity estimation and disparity compensation.

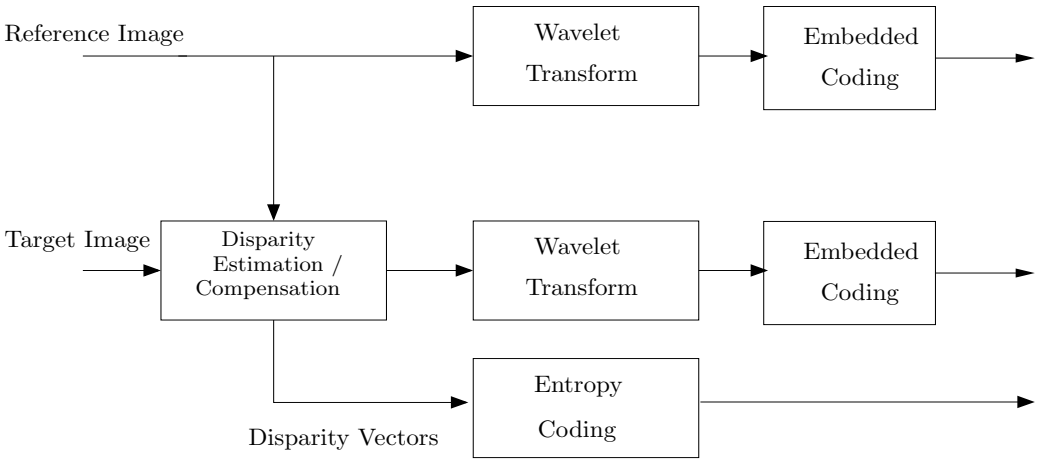
4.2.1 Disparity estimation

In the state-of-the-art of the algorithms [12, 11] discussed in this chapter, disparity estimation is performed between two images having *full perceptual quality* and at *full spatial resolution*. In Chapter 2 the efficacy of using a multiresolution-based hierarchical disparity estimation has been shown. The following paragraphs argue against the use of full-perceptual-quality images in disparity estimation.

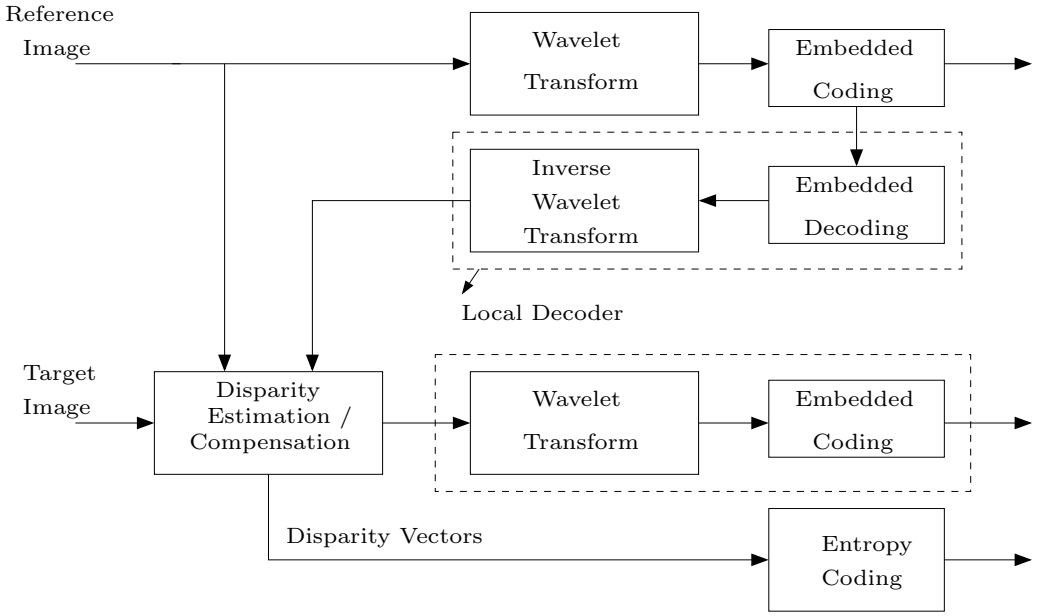
Figs. 4.1(a) and 4.1(b) illustrate typical disparity estimation that can be performed between images, having full perceptual quality (i.e., highest SNR-resolution). These are discussed shortly. These are useful if only stereoscopic still-image coding is desired. If on the other hand the proposed encoding structure is extended to stereoscopic moving-image coding, some problems may arise. As explained in Chapter 6, certain images from the reference stream *will not* be available at full perceptual quality. This implies that disparity estimation *should not* be performed at full perceptual resolution.

Assume that both images at full perceptual quality are represented at 8.0 bits-per-pixel (bpp). This can be restated as disparity estimation being performed when images are *represented at the same bit-rate*. If this criterion is used, the problem described in

the previous paragraph can be overcome by *using locally decoded versions of both images*



(a) Open-loop (CONCOD) disparity encoder structure



(b) Closed-loop disparity encoder(CLDC) structure

Fig. 4.1: Two distinct stereoscopic still-image coding hierarchies

for disparity estimation. While this scheme may introduce some additional quantization noise during decoding, it also insures that *unbiased* disparity estimation is performed between both images.

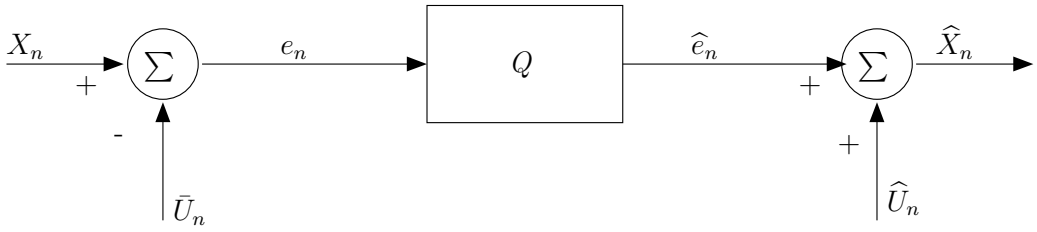


Fig. 4.2: Residual error quantization

4.2.2 Disparity compensation

In order to explain the process of efficient disparity compensation, consider Fig. 4.2 ([62, p 204, Fig. 7.1]). X_n indicates the signal to be quantized while $\bar{U}_n$ represents a signal that is subtracted from X_n. The unquantized error signal e_n is represented as

$$e_n = X_n - \bar{U}_n,$$

A quantized version of this signal, $\widehat{e}_n$ is obtained by passing it through a quantizer Q as

$$\widehat{e}_n = Q(e_n),$$

Let $\widehat{U}_n$ represent a signal that is added to $\widehat{e}_n$ in order to obtain a quantized version of X_n as

$$\widehat{X}_n = \widehat{e}_n + \widehat{U}_n,$$

The distortion $\mathcal{E}$ incurred as a result of this quantization process is expressed as

$$\begin{aligned}
\mathcal{E} &= E[(X_n - \widehat{X}_n)^2] \\
&= E[((e_n + \bar{U}_n) - (\widehat{e}_n + \widehat{U}_n))^2] \\
&= E[((e_n - \widehat{e}_n) + (\bar{U}_n - \widehat{U}_n))^2] \\
&= E[(e_n - \widehat{e}_n)^2] + E[(\bar{U}_n - \widehat{U}_n)^2] + \boxed{2E[(e_n - \widehat{e}_n)(\bar{U}_n - \widehat{U}_n)]}
\end{aligned} \tag{4.1}$$

The dashed-box term in Eq. 4.1 can be assumed to be negligible as it represents the expected value of the product of two uncorrelated random variables. As a consequence,

the distortion $\mathcal{E}$ in Eq. 4.1 can be approximated as

$$\mathcal{E} \approx E[(e_n - \widehat{e}_n)^2] + \boxed{E[(\bar{U}_n - \widehat{U}_n)^2]} \tag{4.2}$$

The distortion is minimized if $\bar{U}_n$ equals $\widehat{U}_n$ when generating $\widehat{X}_n$.

In the context of disparity compensation, this implies that *distortion in a reconstructed target image is minimized if the reference image, used in disparity compensation at the encoder, is the same as the reference image used in disparity compensation at the decoder*. This concept is further explained when describing the algorithms proposed in [11, 12].

4.3 Summary of algorithms

A widely used structure for encoding stereoscopic still-images is shown in Fig. 4.1(a). In the literature this is sometimes referred to as a conditional coder (CONCOD) [3]. Disparity estimation is performed between images having *full perceptual quality*. These disparity vectors are used by the reference image to generate a disparity compensated image. A residual image is generated by subtracting a generated disparity compensated image from the target image at *full perceptual quality*. Both reference and residual images are encoded using any state-of-the-art embedded image coding algorithm. It is observed that the dashed-box term in Eq. 4.2 can *never* be minimized from this structure.

It is also observed from Fig. 4.1(a), that a residual image is generated from a full perceptual quality target image. During decoding, a full perceptual quality reference image is not available. As a result, the disparity compensated image obtained during decoding is not the same as that generated during encoding. When information from the residual bit-stream is added to this *mismatched* disparity compensated image, the reconstructed target image is said to suffer from drift.

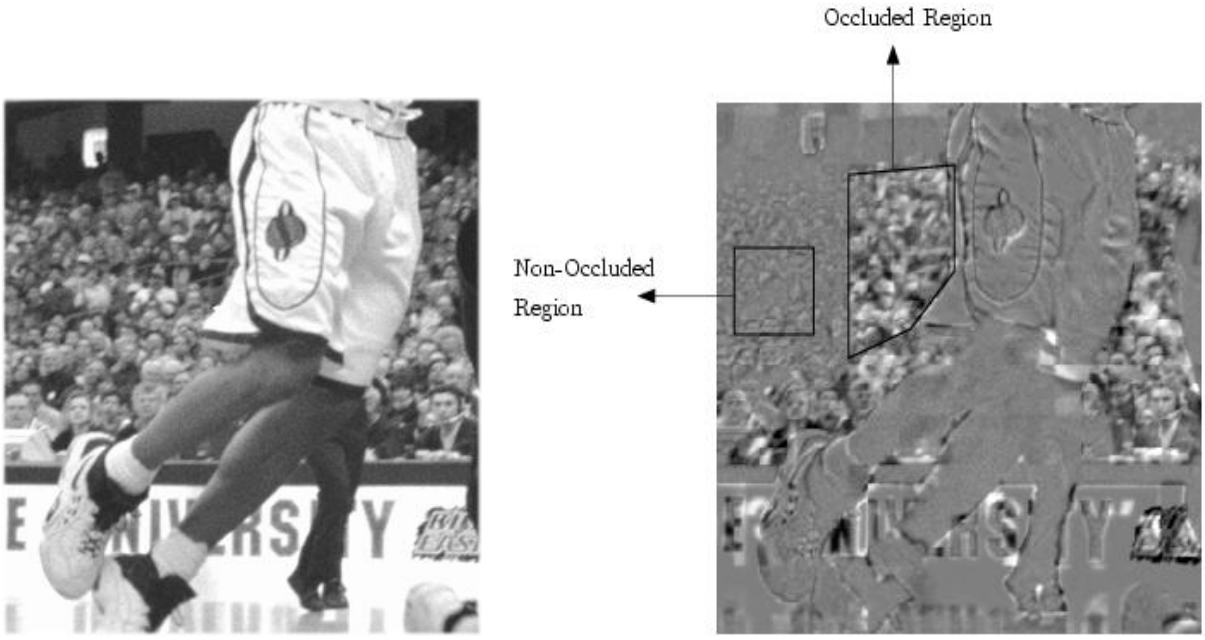
In order to overcome this problem, Boulgouris and Strintzis proposed a *closed-loop disparity codec* (CLDC) structure to encode these image pairs [11]. A block diagram of

this structure can be found in Fig. 4.1(b). The additional feature in this structure is a local decoding of the reference image. A disparity compensated image is generated using this locally decoded reference image. As such this image is also available when decoding bits from the reference image. Thus, the expression in the dashed-box of Eq. 4.2 is minimized as $\bar{U}_n = \widehat{U}_n$, which in this case would be the same disparity compensated images. This makes a CLDC a better structure for encoding stereoscopic still-images. It should be emphasized here that problems due to drift can arise in this structure. This occurs if disparity-compensation is performed with a decoded reference image at a bit-rate less than that version used to generate a compensated image. Hence, this structure is limited by its *a priori* dependence on bit-rates of reference image-views.

The authors in [11] proposed an EZW codec for encoding residual as well as reference images. As previously mentioned, fixed-block-based disparity estimation is performed between the images. In addition, they also proposed a novel yet *ad-hoc* approach to obtain asymmetrical coding. This involves a biased scaling of reference image wavelet coefficients when comparison with those from the residual image. This insures that during decoding, perceptual quality of reference images improves more rapidly than target images.

Frajka and Zeger [12] proposed an alternate approach to efficiently encode these disparity compensated residual images. As an example consider Fig. 4.3(a). Two representative regions have been indicated. *Occluded* regions consists of areas present in one image view and absent in another. It is evident that occluded regions have very similar characteristics to that of natural images. However. non-occluded regions consist of near zero-intensity coefficients. The authors in [12] conjectured that embedded coding of residual images does not guarantee high levels of SNR-scalability in stereo-image coding.

They propose that occluded and non-occluded regions in an residual image warrant



(a) Raw image (b) Residual image

Fig. 4.3: Raw and residual versions of an extracted portion from the "*basketball*" target (left) image. Images have been scaled for display purposes.

separate analysis. Such a scheme is termed *mixed-transform analysis*. Results presented by the authors in [12] have been obtained using an overlapped-block disparity compensation (OBDC) technique [18]. As reported in [12, Sec. 3.3], residual blocks that contain significant high-frequency information (i.e., non-occluded blocks) are transformed using a DCT. On the other hand, occluded blocks are transformed using a three-scale Haar wavelet decomposition.

The authors in [12] based their algorithm on results previously published by Moellenhoff and Maier [14]. These results indicate that pixels in residual images are less correlated than those in natural images. Frajka and Zeger extended this observation by analyzing local correlation of pixels (restricted to 1-pixel) across block-boundaries in these residual images. They observed that this local-correlation drops significantly at block boundaries. Hence this provides a justification of using different transform-based analysis on separate blocks of residual images.

Furthermore, they proposed using an MGE algorithm [15] to encode these residual images. As discussed in Chapter 3, this algorithm relies on intra-level correlation amongst subbands in generating embedded bit-streams. Such a scheme produces perceptually better results when the image being encoded consists of large areas of high-frequency content. To *classify* occluded blocks the authors proposed evaluating the energy (i.e., estimation-error) of all blocks in the residual image. Any block having an energy above a certain threshold is classified as an occluded block.

Unlike [11], no ad-hoc scaling is implemented on reference and target images. Notwithstanding this, the authors have been able to obtain superior results, in a R-D framework, when compared with the results presented in [11] [12, see Fig. 8]. It should be emphasized that the authors do not use any computationally intensive scheme for disparity estimation. As previously indicated, they use a sub-optimal CONCOD structure, shown in Fig. 4.1(a), to obtain their results It can only be concluded that use of an MGE algorithm (that exploits intra-scale correlation amongst wavelet coefficients) justifies the improved results of Frajka and Zeger.

4.4 Asymmetrical Coding

To conclude this chapter, the reader is introduced to the concept of asymmetrical coding of stereoscopic imagery. In his paper, Perkins argued [3] that in a stereoscopic viewing mode, the HVS is relatively insensitive to perturbations in one image (i.e., target-view) when viewed simultaneously with a higher perceptual quality reference image. From a R-D point of view, this involves encoding the reference-view at a higher bit-rate than the target-view.

Consider Fig. 4.4. In this, a Gaussian-blur has been applied on the target-view prior to encoding while the reference-view is maintained at full perceptual quality. This coding formulation of stereoscopic imagery has been presented in [23]. As high-frequency

(a) *A priori* Gaussian-blurred target image (b) Reference image at full perceptual quality

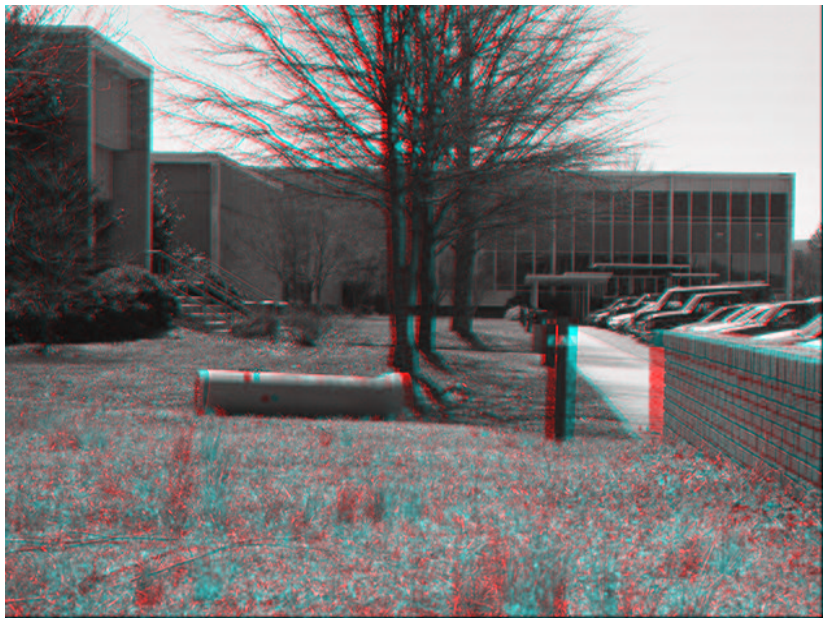
Fig. 4.4: An example of Gaussian-blurred target image, with a higher perceptual quality reference image.

information is removed from the target image, it can be more efficiently encoded at low bit-rates. This insures an asymmetrical coding framework whereby, the reference-view can be encoded at a higher bit-rate than the target view. Fig. 4.5 depicts some representative examples of anaglyphs of these images. The left (target) image in all cases is blurred with a 5×5, zero-mean, Gaussian filter.

This framework is however unsuitable for the algorithm proposed in this thesis. From a qualitative point of view, disparity-estimation between two image views at different perceptual qualities may lead to biased estimation results. This was explained previously in Sec. 4.2.1. For example in Fig. 4.4 the railings near the steps in the target-view cannot be distinguished. Hence correct disparity-estimation cannot be made when comparing with its corresponding view from the reference image, where the railings and steps are clearly distinguishable.

As indicated in the previous section, Boulgouris and Strintzis proposed an ad-hoc scaling of residual wavelet coefficients prior to encoding them. No mathematical justification is provided for selecting these scaling values [11, Table 1]. In addition, these

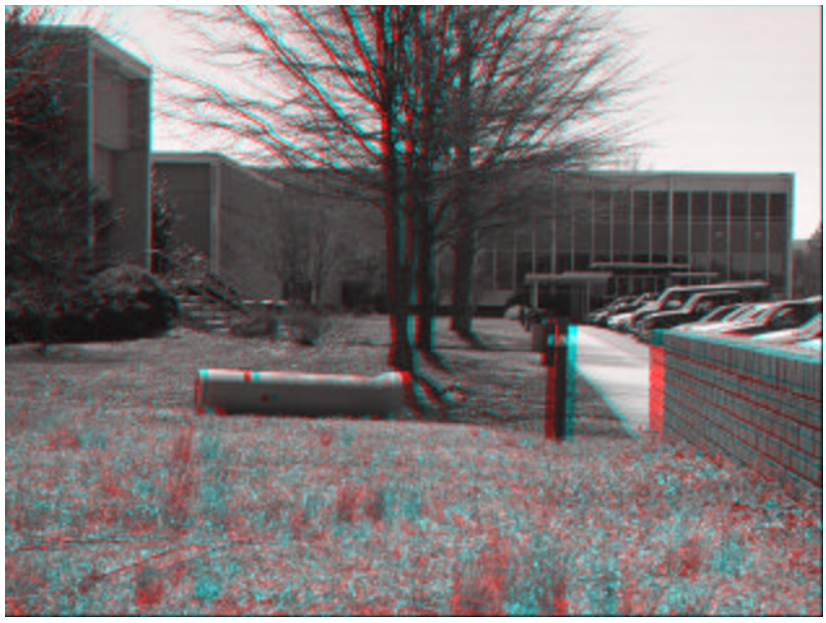
(a) Original target image



(b) Anaglyph with full-resolution reference image

Fig. 4.5: Raw target (left view) image and an anaglyph with a full-resolution reference image from the "*outdoors*" stereo-image pair (dimensions = 640×480).

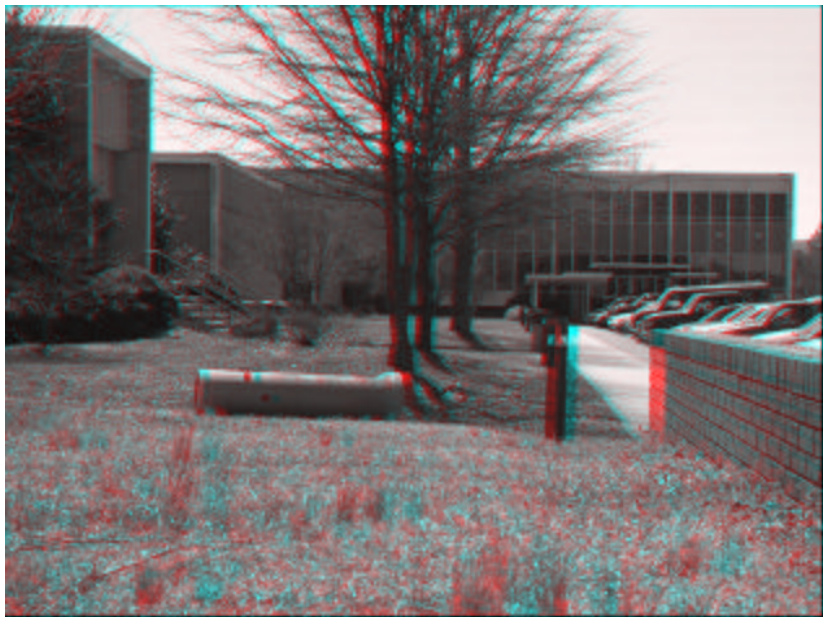
(c) Blurred target image; $r = 1.5$



(d) Left-image blurred; $r = 1.5$

Fig. 4.5: Medium level of blur applied on the target image, and a corresponding anaglyph with a full-resolution reference image. Target (left) image has been blurred using a 2-D Gaussian filter $G(x, y) = \frac{1}{2\pi r^2}\, e^{-\frac{x^2+y^2}{2r^2}}$.

(e) Blurred target image; $r = 5.0$



(f) Anaglyph with full resolution reference image

Fig. 4.5: High level of blur applied on the target image, and a corresponding anaglyph with a full-resolution reference image. Target (left) image has been blurred using a 2-D Gaussian filter $G(x, y) = \frac{1}{2\pi r^2}\, e^{-\frac{x^2+y^2}{2r^2}}$.

values are scale-dependent. Hence, this technique is also avoided in the design of the proposed algorithm. Frajka and Zeger have reported results in an asymmetrical coding framework that relies entirely on bits reconstructed from the target image stream, without any ad-hoc scaling. Consequently, this approach is adapted in this thesis.

Chapter 5

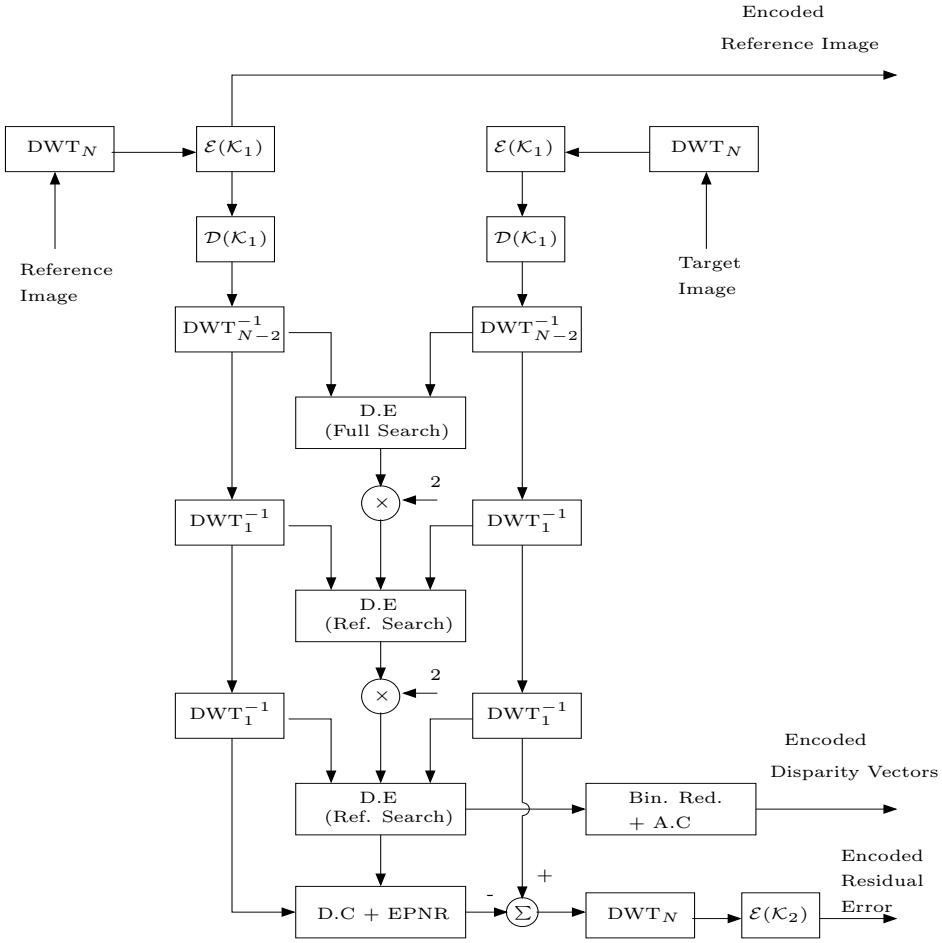
Proposed Wavelet-Based Scalable Stereoscopic Still-Image Codec

Overview

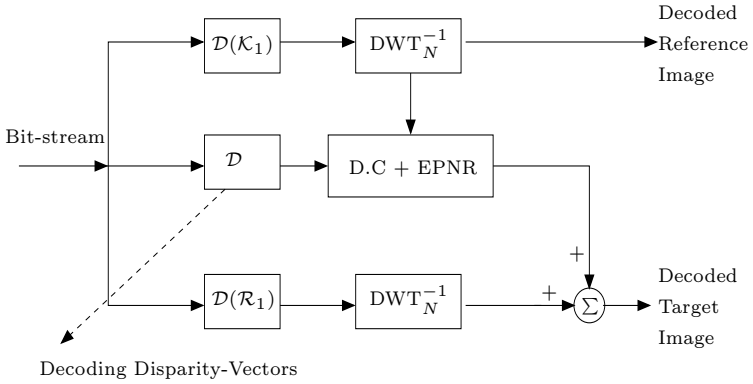
In this chapter a novel stereoscopic still-image codec structure is presented. Differences of this structure, when compared to those presented in Chapter 4, are also highlighted. This is followed by a discussion of a novel loop-filtering scheme, using an edge-preserving noise-reduction (EPNR) filter. It is shown that partition-artifacts are suppressed to a large extent in the generated residual images. This in turn improves PSNR values of reconstructed target images. Comparative results indicate that the proposed algorithm outperforms Frajka and Zeger's [12] and Shukla and Radha's [17] algorithms by at least 0.3-1.2 dB. Finally, a method is presented to obtain discrete levels of spatial-scalability with the proposed algorithm.

5.1 Proposed codec

BASED on observations made in Chapters 2-4, a novel stereoscopic still-image codec is described below. This is organized into two parts. The first part deals with means to achieve SNR-scalability from the codec. The second part outlines a method to obtain discrete levels of spatial scalability.



(a) Encoder



(b) Decoder

Fig. 5.1: Block diagram of proposed codec, with SNR-scalability, at a specified spatial-resolution.

5.1.1 SNR-scalability

A block diagram of the new structure used for compression with SNR-scalability can be seen in Fig. 5.1. It is assumed that images will be reconstructed and displayed at *full* spatial-resolution (i.e., scale-0). It is also assumed at this point that gray-scale images are being processed. The steps implemented in this structure are outlined as follows:

- **Step 1**: Both the reference and the target images are decomposed using an N-level, lifting-based, 2-D separable wavelet transform. This is indicated as DWT_N in the block diagram.

- **Step 2**: In Chapter 4 it was argued that disparity-estimation should be performed between images having similar features. In [11, 12], raw images are used for disparity-estimation. In other words, images represented at 8.0 bpp[1] are used for disparity-estimation. Here, it is proposed that these images be encoded at a high bit-rate using the remaining steps of an ASWDR encoding scheme, as previously discussed in Chapter 3. In the proposed codec, this is represented as $\mathcal{E}(\mathcal{K}_1)$ ($\mathcal{K}_1 < 8.0$). The value $\mathcal{K}_1$ within the parentheses indicates the bit-rate used in representing both images. Bits generated from the reference image are transferred to the overall bit-stream. However, no such operation is performed with bits generated from the target image.

- **Step 3**: Bits generated from both images are locally decoded in this step. This is obtained by using the inverse ASWDR steps $\mathcal{D}$ (described in Chapter 3). It should be emphasized that this decoding is performed at the same bit-rate $\mathcal{K}_1$. (i.e., the entire bit-stream is decoded)

- **Step 4**: An inverse wavelet transform (DWT_{N-2}^{-1}) is performed on these locally quantized coefficients such that images are reconstructed until a scale $N-2$ is

[1] Assuming raw gray-scale images are represented at 8.0 bpp

reached. As an example, the reader is referred to Fig. 2.7.

- **Step 5**: A hierarchical-block-based disparity estimation is performed in this step. Details of this algorithm can be found in Chapter 2. If a variable-block-based disparity estimation[2] is used, then the following steps are also implemented:
  - A quadtree partition is implemented on the all low-pass subband of the target image, at scale-2. Details of this partitioning scheme are outlined later in this chapter.
  - The quadtree-map generated from this partitioning scheme is encoded, using a binary-reduction technique previously described in Chapter 3. Bits generated from this process are transferred to the main output data-stream.

  No such quadtree-map encoding is necessary if fixed-block-based disparity estimation is used.

- **Step 6**: Disparity-vectors at scale-0 are encoded using a binary-reduction technique, followed by a context-based arithmetic coding scheme. Bits generated in this process are transferred to the main output data-stream.

- **Step 7**: A disparity compensated image is generated from the locally decoded reference image and disparity vectors at scale-0. An edge-preserving noise reduction (EPNR) filter is applied on this compensated image in order to reduce perceptually visible partition artifacts. Justification and details of this filter are described shortly.

- **Step 8**: A disparity compensated residual image is generated by subtracting this filtered compensated image from the locally decoded target image. This residual image is first transformed using an N-level DWT and subsequently encoded at a

[2]This is not explicitly shown in Fig. 5.1.

bit-rate of $\mathcal{K}_2$ using the remaining steps of an ASWDR algorithm (indicated as $\mathcal{E}(\mathcal{K}_2)$ in Fig. 5.1(a)). Bits generated in this process are transferred to the main output data-stream.

Thus, a final data-stream would consist of bits from the reference image, quadtree-map (if variable-block-based disparity estimation is performed), disparity vectors at scale-0 and residual image.

As images are displayed at full spatial-resolution, the decoding process is quite straightforward. As shown in Fig. 5.1(b), bits from the reference image are decoded. Disparity compensation is performed using a reference image decoded at a bit-rate of $\mathcal{K}_1$. Reasons for this have been explained in Chapter 4. Loop-filtering (with the same set of parameters used during encoding) is performed here as well. This is used to smooth the generated disparity compensated image. After this step, bits from the residual image are decoded at any arbitrary bit-rate $\mathcal{R}_1$. This is obtained by inverting the ASWDR steps, indicated as $\mathcal{D}(\mathcal{R}_1)$. An N-scale inverse DWT is performed and the generated residual image is added to the loop-filtered disparity compensated image in order to obtain a decoded target image.

As seen from Fig. 5.1(a), the progressive decoding of bits from the residual image determines the SNR-scalability obtained from the proposed codec. SNR-scalability can also be obtained by progressively decoding bits from the reference image stream. However this is limited by the fact that a bit-rate of $\mathcal{K}_1$ bpp must be satisfied prior to generating any bits from the residual stream. As previously mentioned $\mathcal{K}_1$ is a "high" bit-rate. Hence differences in decoded reference images are visually imperceptible when compared with their raw versions. Thus, any additional bits added to this bit-stream would be redundant in nature. Thus, *SNR-scalability obtained while decoding a stereo-image pair is limited to bits from the residual image stream only.*

5.1.2 Spatial-scalability

A hierarchical disparity estimation scheme is generally employed to reduce the complexity of a full search (FS) scheme. This was the justification provided in Sec. 2.4. However, an unexpected benefit can be derived by using such an estimation. Fig. 5.1(a) depicts the generation and subsequent encoding of a disparity compensated residual image at scale-0. Such an encoding can also be performed at scales 2 and 1.

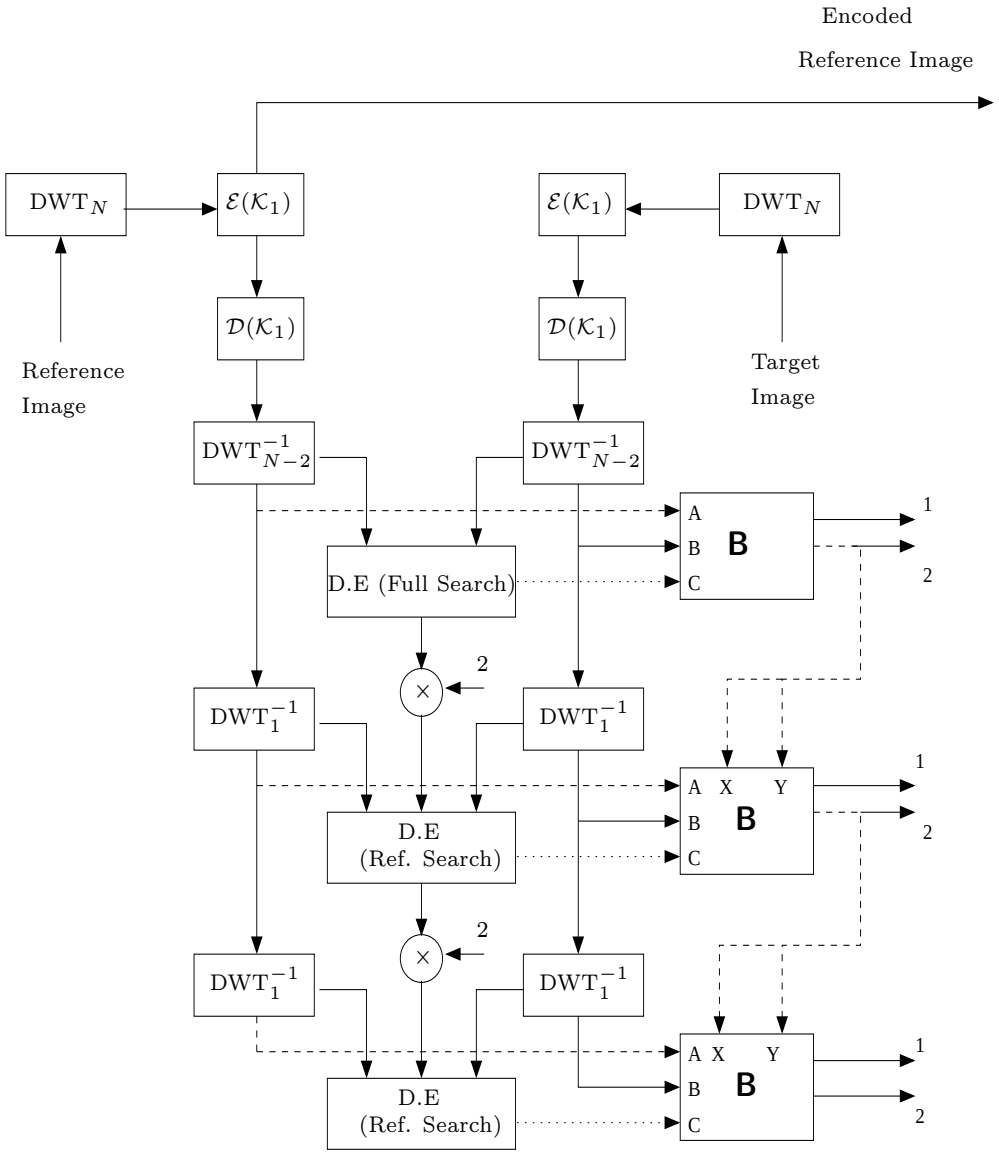
Assume that a residual image is generated, from the all low-pass subband only, at scale-2. Assume that this is encoded and subsequently decoded at a bit-rate of $\mathcal{K}_{22}$. This locally decoded residual image can be used for encoding the residual image at scale-1. This would entail the following steps:

- **Step 1**: Generate a residual image at scale-1.
- **Step 2**: Perform a single stage DWT on the residual image obtained from Step 1.
- **Step 3**: Subtract the residual image, obtained at scale-2, from the all low-pass subband of the wavelet transformed image in Step 2. This is visually depicted in the dotted box of Fig. 5.2(b).
- **Step 4**: Encode (and locally decode) this reduced energy, transformed residual image at a bit-rate of $\mathcal{K}_{21}$.

The above steps can be repeated for all succeeding levels. The nature of a DWT facilitates discrete levels of spatial-scalability on the reference image. Due to problems associated with drift, this operation cannot be implemented in a straightforward manner on the target image. The steps outlined above alleviate this problem. These steps are highlighted in Fig. 5.2(a).

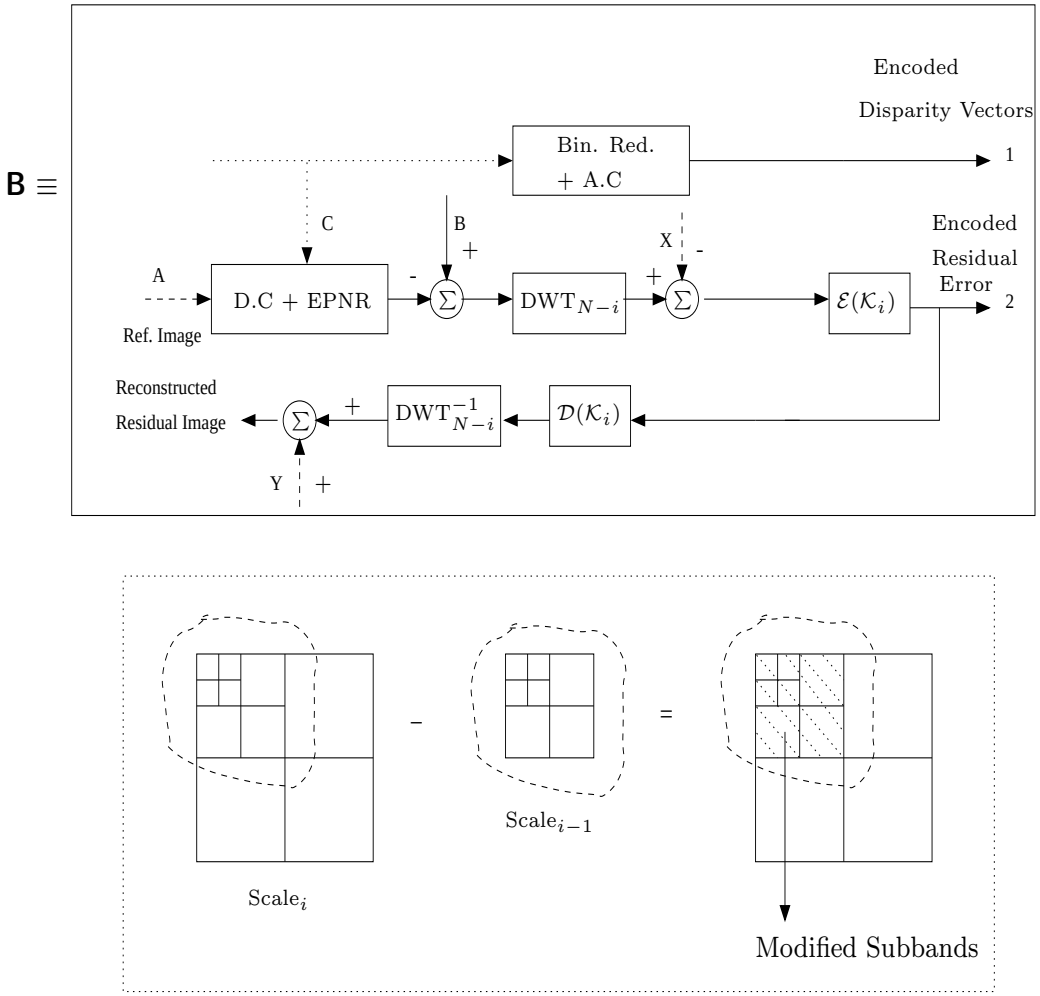
As a result, the embedded nature of this algorithm also becomes evident. Bits allocated for scale-2 target images *must* be decoded prior to those bits earmarked for scale-1

images and so on. During decoding, scale-2 target images are also available at scales 1 and 0. As this involves upsampling scale-2 images, SNR values of these target images would be considerably less than images at scale-2. This explains the subtle difference between SNR- and spatial-resolution in embedded image coding. A drawback of this system is that it is restricted to a dyadic framework only.



(a) Fig. 5.1(a) when simultaneously encoding stereo image pairs at different spatial resolutions. The block **B** is expanded in the next part of this figure.

Fig. 5.2: Global structure for spatial scalability



(b) Expansion of B and reduced energy residual generation

Fig. 5.2: contd. The dotted box depicts the procedure in which energy of a residual image at a finer scale is minimized, using a locally decoded version of a residual image from a coarse scale. The modified residual image is encoded and subsequently decoded using an ASWDR encoding/decoding scheme, at a bit-rate of $\mathcal{K}_i$. These are indicated as $\mathcal{E}(\mathcal{K}_i)$ and $\mathcal{D}(\mathcal{K}_i)$, respectively. A previously encoded residual image, at a coarse scale, is subtracted (X) and subsequently added (Y) to regenerate the residual at the current scale. Other notations are similar to that shown in Fig. 5.1(a).

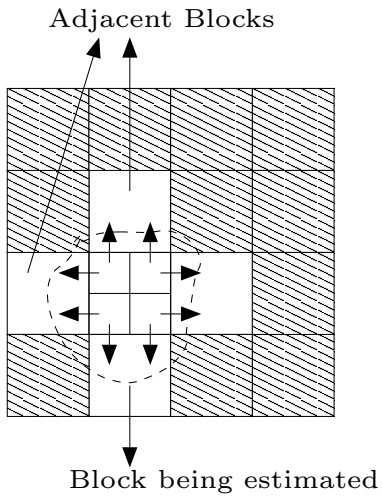


Fig. 5.3: Overlapped-block disparity compensation (OBDC). All blocks *must* have same dimensions. Different regions of the block (enclosed within the dashed line) are estimated from different neighbours.

5.2 Justification for a new loop-filtering scheme

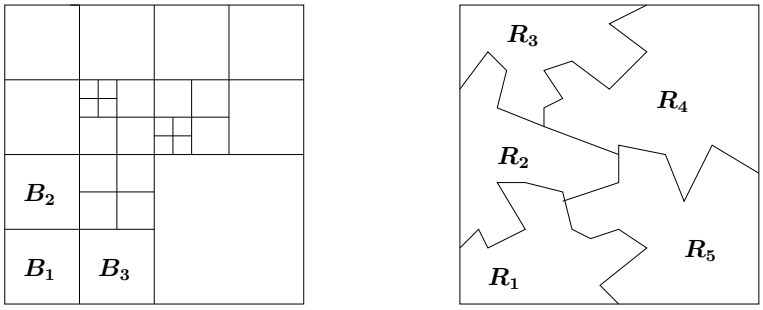
A widely used motion-compensation technique is *overlapped block motion compensation* (OBMC) proposed by Auyeung *et al.,* [63]. This was extended to disparity-compensation by Woo and Ortega [18] and referred to as *overlapped block disparity compensation* (OBDC). Consider Fig. 5.3. Assume that a block $\boldsymbol{B_1}$, having dimensions 8×8, is being estimated. Thus, from an OBDC scheme, pixels of $\boldsymbol{B_1}$ are estimated as [2, p 248]:

$$\widehat{B}_1(x,y) = \frac{H_1(x,y)\, q(x,y) + H_2(x,y)\, r(x,y) + H_3(x,y)\, s(x,y) + 4}{8} \tag{5.1}$$

where $\boldsymbol{H_1}, \boldsymbol{H_2}$ and $\boldsymbol{H_3}$ are scaling matrices having dimensions equal to that of the block being compensated. $q(.)$, $r(.)$ and $s(.)$ are pixels from the reference picture obtained from three motion vectors. These are the motion vectors of the current block and two of its four neighboring blocks. This can be seen from Fig. 5.3. As observed, this scheme is only valid if fixed-block-based disparity estimation is performed between images.

In the proposed algorithm, arbitrarily-shaped region-based disparity-estimation will be performed. A few representative examples of such estimation schemes can be seen in

Fig. 5.4. Consequently, Eq. 5.1 cannot be used as surrounding blocks are not guaranteed



(a) Variable-block (b) Arbitrarily-shaped

Fig. 5.4: Examples of region-based disparity-compensation

to be of similar dimensions (Fig. 5.4(a)) or they may be arbitrarily-shaped regions (Fig. 5.4(b)). ODBC is used to visually suppress "*partition artifacts*". *Blocking artifacts* are a special case in which the partitions are blocks of fixed size.

Such partition artifacts occur in disparity-compensation schemes shown in Fig. 5.4 as well. In the proposed algorithm of this thesis, a variable-block-based disparity compensation scheme (Fig. 5.4(a)) is used. The generic term, partition artifacts, is used throughout the following discussion. This is meant to differentiate with blocking artifacts, a term widely referred to in the literature in conjunction with fixed-block-based disparity- (or motion-) compensation.

In order to remove partition-artifacts from non fixed-block-based disparity compensation schemes, *loop-filtering* is used. This involves smoothing the disparity- (or motion-) compensated image prior to generating a residual image. Loop-filtering forms the basis of various video coding standards such as MPEG-1, etc. [2]. In [64] a sophisticated loop-filter has been reported. Unfortunately, this is adapted for fixed-block-based compensation schemes. Additionally, these filters are dependent on the block size and have been described for DCT-based coding. During the literature survey, no suitable references have been found addressing the problem of loop-filtering in DWT-images.

It is envisaged that any loop-filtering scheme designed for a specific compensation

scheme (e.g., variable-block-based) should be compatible with other schemes as well. In other words, this implies that any loop-filtering scheme should be region-independent. In addition, any loop-filtering scheme should suppress partition-artifacts while preserving as much as possible, useful features from compensated images. In this regard, moving-average (MA) and median-filters [65, p. 191] can form possible candidates.

However, neither of these filters should be used in conjunction with partition-artifact suppression. MA filters introduce blurring in the image by degrading relevant edge information. On the other hand median-filtering is effective only when the noise in images consists of strong spike-like components. This forces points with distinct intensities to be more like their neighbors. A useful feature of median-filtering is that natural edges in images are well preserved.

5.3 Edge-preserving noise-reduction filter

In order to reduce partition-artifacts from disparity compensated images, a filtering scheme that incorporates selective blurring of regions, coupled with natural edge-preservation and unconstrained by region size dependency needs to be developed. This was achieved by adapting an *edge-preserving noise-reduction* (EPNR) filter, proposed by Kröener and Ramponi [19]. This filter was proposed to clean images, corrupted with Gaussian noise. Here, it is assumed that pixel intensities lie between [0, 1].

Let $\boldsymbol{f}$ and $\boldsymbol{g}$ represent 2-D input and output signals, respectively. This filter is essentially a cubic-polynomial operator whereby $g[n_1, n_2]$, the filtered output pixel, is expressed as

$$g[n_1, n_2] = \alpha\, f[n_1, n_2] + \frac{1}{6}(1-\alpha)\,\beta \tag{5.2}$$

where

$$\begin{aligned} \alpha &= \frac{1}{2}\left(\boxed{(f[n_1-1, n_2] - f[n_1+1, n_2])^2 + (f[n_1, n_2-1] - f[n_1, n_2+1])^2}\right)\lambda + (1-\lambda) \\ \beta &= f[n_1, n_2-1] + f[n_1-1, n_2] + f[n_1, n_2+1] + f[n_1+1, n_2] + 2\, f[n_1, n_2] \end{aligned} \tag{5.3}$$

This filtering operation is a simplified version of the original scheme proposed by Ramponi in [66]. The selective smoothing property of this filter is made possible due to the boxed term in Eq. 5.3. As indicated in [66], a sum of squared difference of sample values at the extremes of the filter mask is evaluated. A sufficiently large value for this term implies that the mask is positioned at a partition-artifacts boundary, making the frequency response of the filter operator less selective (i.e., more blurring at boundaries).

Here λ represents the "*smoothing parameter*". This is used to control the amount of blurring in the disparity compensated images. Based on extensive experimental results, λ is set equal to 1.35. This is a representative value. A range of [0.8,1.35] is sufficient for most images analyzed during the course of this thesis.

Partition-artifacts in disparity compensated images are low-contrast high frequency regions. Smoothing these regions would imply "spreading" the energy contained in them and distributing it in surrounding regions. As noted by the authors in [19], the above filter acts as a simple linear filter when the surrounding regions do not contain any low-contrast regions. If however, such a low-contrast region is encountered in the image, a difference in the gray-scale values between the outer and inner edges of the filter-mask is created. Hence, from Eq. 5.2, $\alpha \rightarrow 1$. This results in the filtered value being equal to the central value $f[n_1, n_2]$. Thus, applying this filter on disparity compensated images results in suppression of visible partition-artifacts. Additionally, some useful low-contrast high-frequency content is also removed. This has a beneficial effect when the corresponding residual image is encoded.

This filter can be applied iteratively on an image. With each successive iteration partition-artifacts are gradually suppressed. This can be correlated with reasons provided in the previous paragraph. This implies that a *reasonable* number of iterations would suffice in smoothing disparity compensated images. In this thesis, two filter iterations are used for smoothing disparity compensated images. It should be emphasized that this

is an experimentally determined value.

Results shown in [14] indicate that residual images contain large areas of zero-intensity values, smaller areas of vertical edge information and some occluded regions. Due to this uneven distribution of regions, Frajka and Zeger [12] conjectured that the popular "CDF-9/7" wavelet filter set would be unsuitable for transforming this image. The use of an EPNR filter in this algorithm alters the distribution of the aforementioned regions in the residual image. There are significantly larger areas of *smooth* edge information while partition-artifacts are suppressed to a very large extent. As indicated in Chapter 3, the ASWDR algorithm can reconstruct such edge information more effectively than currently used embedded image coding algorithms.

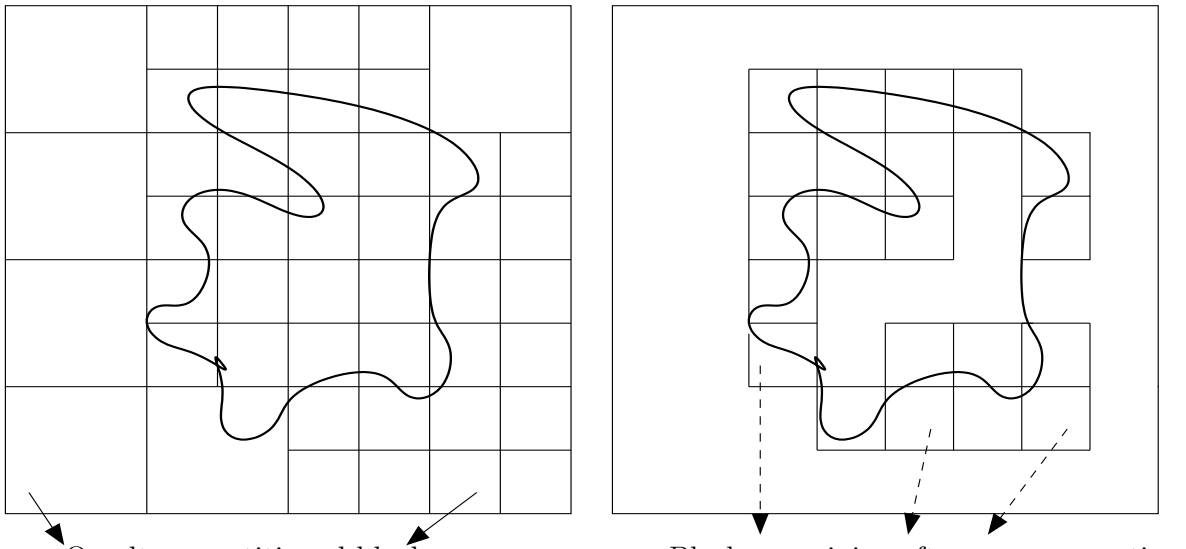
5.4 Variable-block-based partitioning schemes

Justification for using a variable-block-based disparity estimation scheme has been provided in Chapter 2. In addition, the previous section describes a novel loop-filter that can be used to suppress partition-artifacts arising in variable-block-based estimation schemes. This section describes some commonly used quadtree-partitioning schemes to obtain blocks of variable sizes. A few representative examples of quadtree-partitioning schemes can be seen from Fig. 5.5.

5.4.1 Rate-distortion constrained quadtree-partitioning schemes

In [46], the authors presented a multiresolution variable-block-based coding scheme. In it, the *mean absolute difference* (MAD) of a disparity compensated block with respect to the original target image block was calculated. From Eq. 2.13, MAD for the i^{th} block is defined as

$$\mathrm{MAD}(i, \boldsymbol{v}_i) = \frac{1}{B_x B_y}\mathrm{SAD}(i, \boldsymbol{v}_i) \tag{5.4}$$



(a) Split quadtree partitioning (b) Split and merge quadtree partitioning

Fig. 5.5: Representative examples of quadtree-partitioning schemes

where (B_x, B_y) are the dimensions of the block being estimated. Hence, the authors in [46] specified that if

$$\text{MAD}(i, \boldsymbol{v}_i) > M_t$$

then the block can be further subdivided. Here M_t is an ad-hoc threshold value. The subdivision of blocks stops if this criterion is not met or the minimum possible block size has been reached. The basic theory of this partitioning scheme lies in representing a block with four disparity vectors in place of a single vector, increasing the bit-rate incurred in representing these vectors. This scheme is primarily undertaken so as to minimize the energy of residual images.

The authors in [67, 68] and [17] address this problem in a R-D framework by solving a Lagrangian cost function. More specifically, this can be written as

$$L_i(\lambda^*) = D_i^{res} + R_i\lambda^* \tag{5.5}$$

where λ^* is a Lagrange parameter, D_i^{res} the coding distortion of the residual block (shown previously in Eq. 2.12) being considered and R_i is the bits expended in encoding the

block co-ordinates. The authors in [17] observed that:

- the pruning structure used in [67, 68] does not account for neighboring blocks having similar Lagrangian cost functions, and
- overall coding performance can be improved if these blocks, having similar Lagrangian cost functions, are combined.

The pruning criterion, used in the algorithms mentioned above, was originally proposed by Ramchandran and Vetterli [69]. An example of both partitioning schemes can be found in Fig. 5.5. These schemes, however, have some limitations and are discussed as follows.

In [17, Sec. 2] the authors observed that the total distortion of all split and merged blocks, $D^*(\lambda^*)$, corresponds to the minimum distortion possible, for the given rate

$$R^*(\lambda^*) = \sum_i R_i$$

However, distortions caused by partition-artifacts are not taken into account by this scheme. As discussed in earlier chapters, a relatively higher bit-rate is needed to encode residual images with partition-artifacts . It has also been shown, earlier in this chapter, that an EPNR filter is used to smooth out partition-artifacts arising due to imperfect disparity compensation. Hence, the above stated criterion for R-D optimization cannot be used in conjunction with the proposed algorithm. To further emphasize this fact, it is shown in the next section that fixed-block-based estimation can outperform this variable-block-based scheme (in a R-D framework) subject to certain conditions.

Current MPEG-4 standards provide for object (or content)-scalability in encoded and decoded moving-image sequences. A quadtree-partitioned image is unsuitable, on its own, to provide good object-scalability. This is because the resulting boundaries are jerky and do not align well with object discontinuities [4]. However, a quadtree-partition

can provide a coarse approximation of an object. Additional partitioning can be applied on the remaining blocks in order to obtain a close approximation for the object. One such scheme is *wireframe-partitioning* [4]. Object segmentation is beyond the scope of this thesis. The concerned reader is directed to an excellent review paper by Strintzis and Malassiotis [4] for more details on object segmentation, in the context of stereoscopic image coding.

Due to limitations associated with current R-D constrained quadtree-partitioning scheme, a more general scheme is explored. It should be emphasized here that the quadtree-partitioning scheme implemented in this research work:

- is entirely dependent on image content,
- does not rely on R-D constraints associated with current partitioning schemes. In other words, there *may* exist fixed-block-based solutions that may provide similar R-D performance, and
- is performed, while acknowledging the fact that more efficient partitioning schemes can be implemented.

5.4.2 Image content based quadtree-partitioning

A split-only quadtree-partitioning scheme, based entirely on image content, can be implemented using the technique suggested by Vaisey and Gersho [70]. In this, the variance of pixels is used to classify a block either as homogeneous, textured, or edge-related. In order to simplify this process, the following discussion pertains in classifying a block as homogeneous or non-homogeneous. Let σ_b^2 be the variance of pixels of the block under consideration. Vaisey and Gersho state that a block is homogeneous if

$$\sigma_b^2 < V_t$$

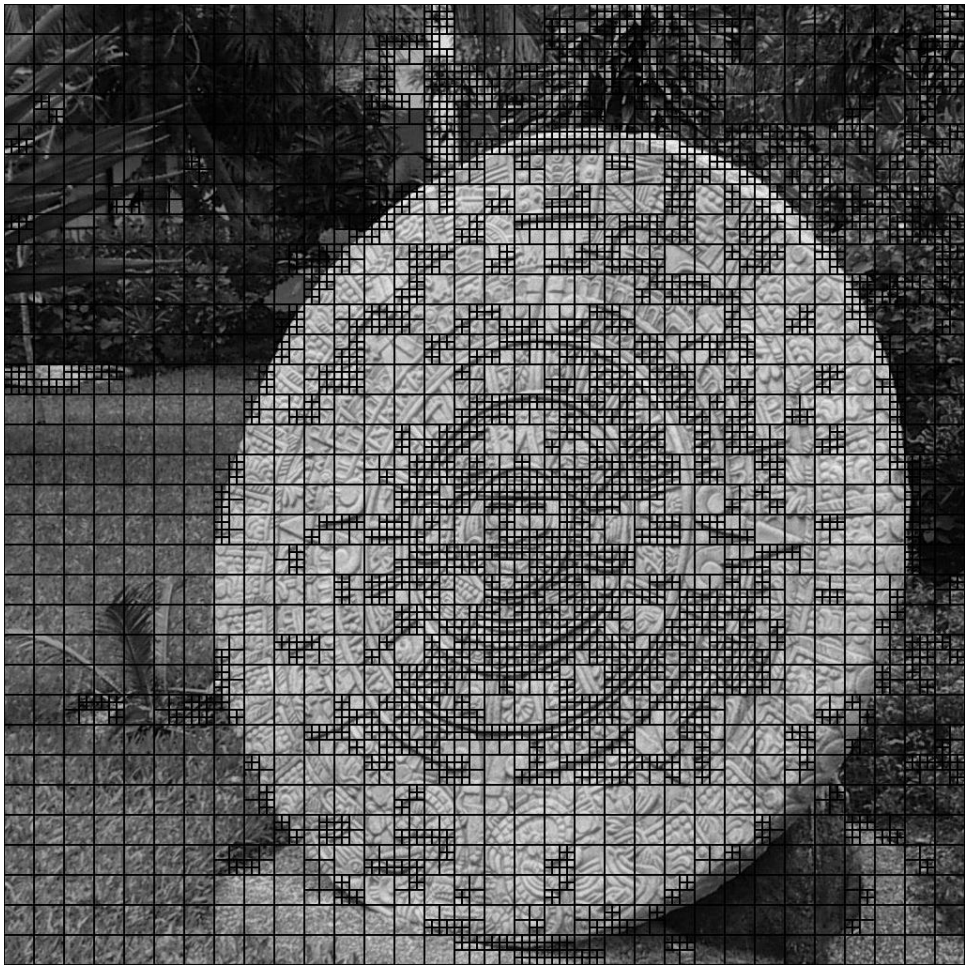


Fig. 5.6: Quadtree-partitioning of Y-component of a textured image, with quadtree-map generated at scale-0 (i.e., at original spatial resolution). $V_t = 30$. Block dimensions range from 8×8 - 32×32. Image dimensions are 1024×1024.

where V_t is an empirically determined threshold value. In the context of quadtree-partitioning, if a block is deemed homogeneous then further partitioning in the block is stopped. By restricting classification to only homogeneous blocks, there is a chance that image regions having sufficiently high texture are partitioned. This may be useful if such regions occur at object boundaries. However, additional partitioning may not be required if surrounding regions of a block are also textured. As an example, consider Fig. 5.6.

Higher levels of partitioning are required in identifying the boundary of the medallion.

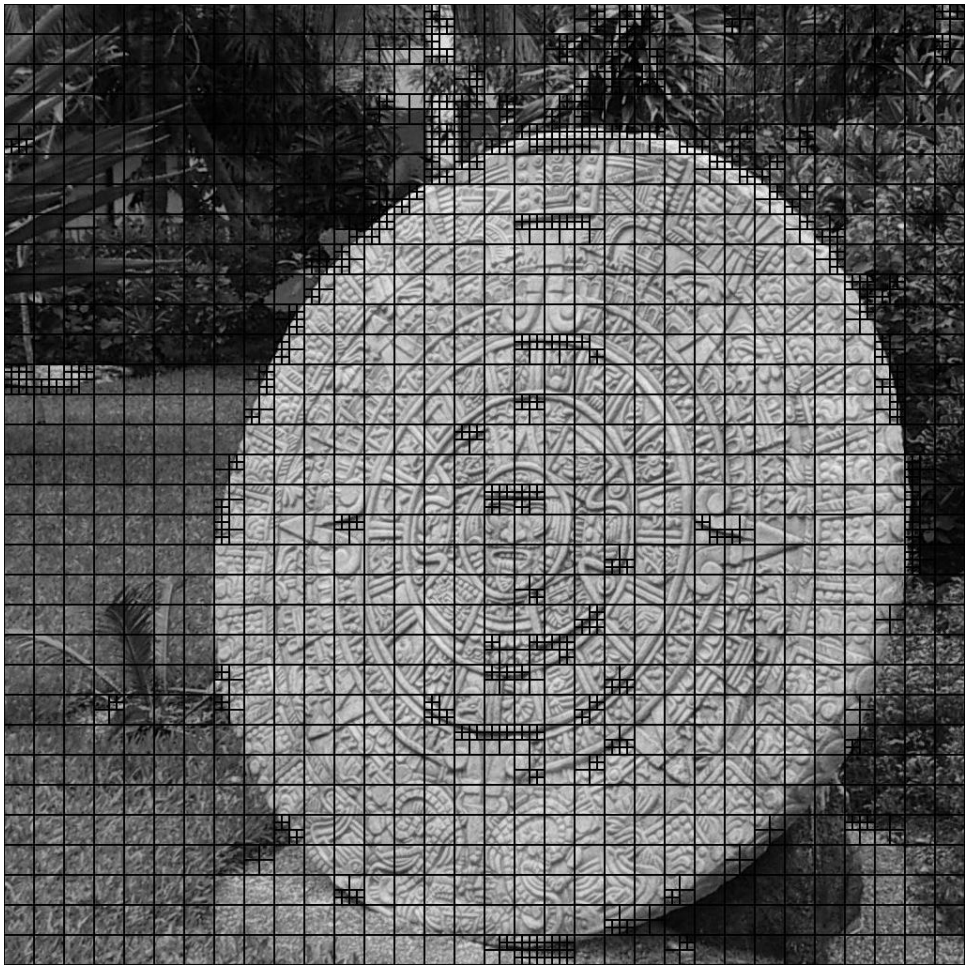


Fig. 5.7: Image from Fig. 5.6, partitioned using a quadtree-map generated at scale-2 (e.g., as in Fig. 2.7) with a threshold $V_t = 120$. Block dimensions range from 8×8 - 32×32. Image dimensions are 1024×1024.

However, fine partitioning is not required when identifying regions on the medallion. One way to avoid this redundant partitioning is to classify textured blocks, as per the algorithm shown in [70]. A simplification to this process can be achieved by exploiting the hierarchical nature of a 2-D separable DWT. On observing Fig. 2.7, it can be inferred that the all low-pass image at scale-2 contains significantly less detail than its scale-0 counterpart. An assumption is made whereby the variance threshold at scale-2 is increased by a factor of 4, when compared with its scale-0 counterpart. In Fig. 5.6, $V_t = 30$. Hence, $V_t = 120$ when partitioning the all low-pass subband at scale-2.

The image is Fig. 5.6 was initially divided in "*superblocks*" having dimensions equal to 32×32. Regular quadtree-partitioning was effected on these superblocks. In order to achieve a similar size of blocks, the all low-pass subband at scale-0 is divided into superblocks, having dimensions equal to 8×8. The quadtree map generated at this stage is subsequently scaled by a factor of four when scale-0 is reached. This can be seen from Fig. 5.7. It is observed that object boundaries (e.g., medallion and the background) are clearly defined in both images. However, partitions in textured regions (e.g., features on the medallion) are significantly reduced in Fig. 5.7 compared to Fig. 5.6.

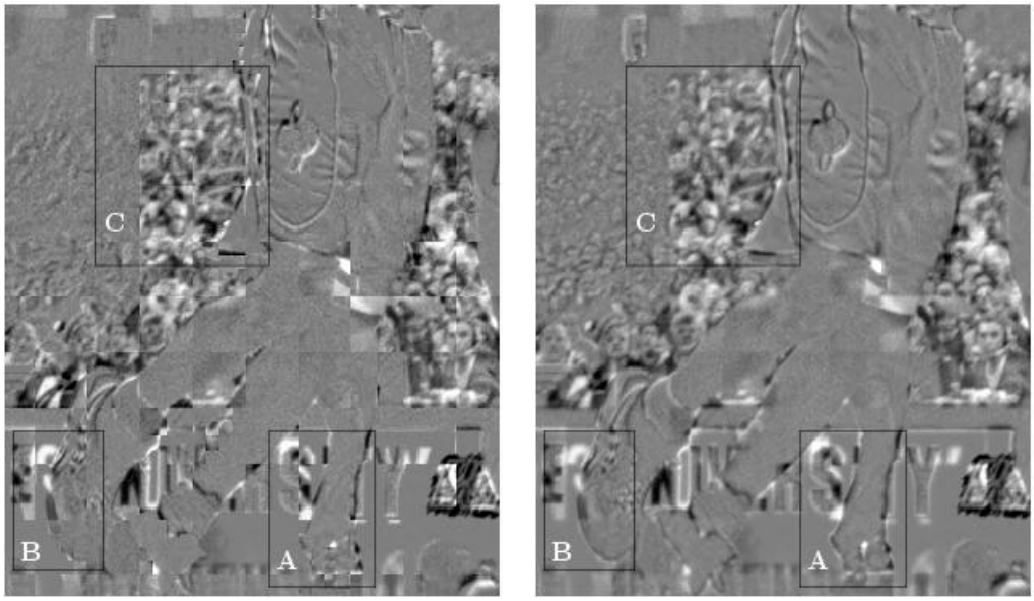
As previously indicated, this quadtree partition is not optimal in a R-D sense. There exist fixed-block-based counterparts that may produce similar or better performance as this partitioning scheme. These are enumerated in the following section. Furthermore, it is acknowledged that a *merge* operation of these partitioned blocks would lead to improved performance of reconstructed images. This is beyond the scope of this thesis, and is left as future research work.

5.5 Results and analysis

Qualitative and quantitative results when encoding gray-scale images, are presented in this section. Subsequently, limited subjective results are presented when encoding stereoscopic color-images.

5.5.1 Performance evaluation when using a loop filter

Fig. 5.8 qualitatively explains the difference when an EPNR filter is used in the algorithm. The picture represents a section of the "*basketball*" stereo-image pair. It can be seen from Fig. 5.8(a) that partition-artifacts are clearly perceptible near the referee's leg and the university advertising board (indicated as **A** in the figure). Another region of interest (indicated as **B** in the figure) would be the player's right shoe and the advertising



(a) Without loop-filtering (b) With loop-filtering, $\lambda = 1.35$

Fig. 5.8: Sections of disparity compensated residual images when encoding the "*basketball*" stereo-image pair. The images have been scaled for display purposes. A raw version of this image section can be seen from Fig. 4.3(a).

board in the background. It should also be noted that these artifacts are very perceptible at transitions between occluded and non-occluded regions (e.g., region **C**).

On examining Fig. 5.8(b), it is observed that partition-artifacts in these regions and others have been suppressed to a large extent. It can also be observed that transitions between perfectly compensated and occluded areas have also been smoothed. Partition-artifacts require a significant amount of bits when encoding such residual images. If disparity compensated images are smoothed, such artifacts are suppressed. Hence coefficients from other regions of residual images can be encoded. This tends to increase PSNR values of reconstructed target images. This is validated by observing the results presented in the latter part of this section.

5.5.2 Qualitative results when using fixed-block and variable-block disparity estimation

Qualitative results are presented to complement the discussion about using variable-block-based disparity estimation rather than a fixed-block-based scheme. The textured stereoscopic image, shown in Fig. 5.7, is predicted by disparity estimation and compensation. A 3-scale hierarchical search strategy, with fixed- and variable-block-based disparity estimation is used. Compensated images are smoothed using an EPNR filter, with $\lambda = 1.35$ and two filter iterations. Resulting residual images can be seen in Figs. 5.9 and 5.10. Two representative regions, **A** and **B**, are shown in both images. It is evident, perceptually, that a variable-block-based disparity estimation scheme generates reduced energy residual images. This factor helps in increasing PSNR values of reconstructed target images, as shown in the next sub-section.

5.5.3 Experimental results with monochrome images

SNR-scalability

A quantitative evaluation of the proposed algorithm is undertaken in this section. It is compared with Frajka and Zeger's results on the "*outdoors*" (Figs. 5.11(a) and 5.11(b)) and "*fruits*" (Figs. 5.11(c) and 5.11(d)) stereo-image pairs [71]. These results have been provided by Dr. Tamás Frajka. A summary of this algorithm can be found in Chapter 4, while extensive details can be found in [12]. Comparative results are also provided with Shukla and Radha's R-D constrained, variable-block-based algorithm on the "*arch*" stereo-image pair (Figs. 5.11(e) and 5.11(f)) [72]. These results have been provided by Rahul Shukla at the EPFL, Switzerland. It should be indicated that the results presented here compare the performance of the proposed algorithm at the same bit-rates used in [12, 17]. The following notations are used when explaining the results obtained:

- FZ indicates results obtained using Frajka and Zeger's algorithm [12],

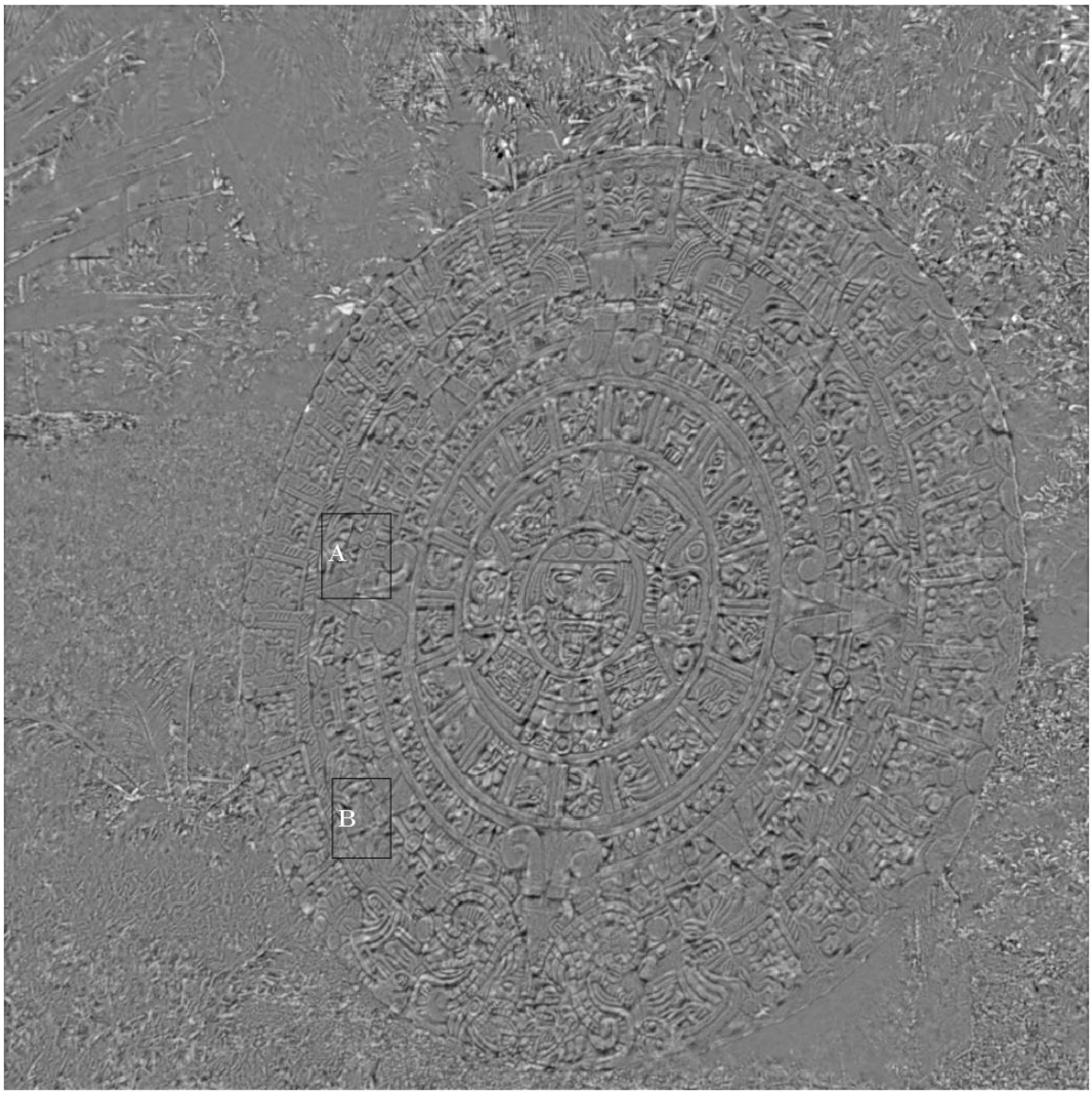


Fig. 5.9: Residual image obtained when predicting image shown in Fig. 5.7. A 3-scale hierarchical fixed-block-based disparity estimation scheme is used, with scale-0 block size of 16×16. Image has been scaled for display purposes.

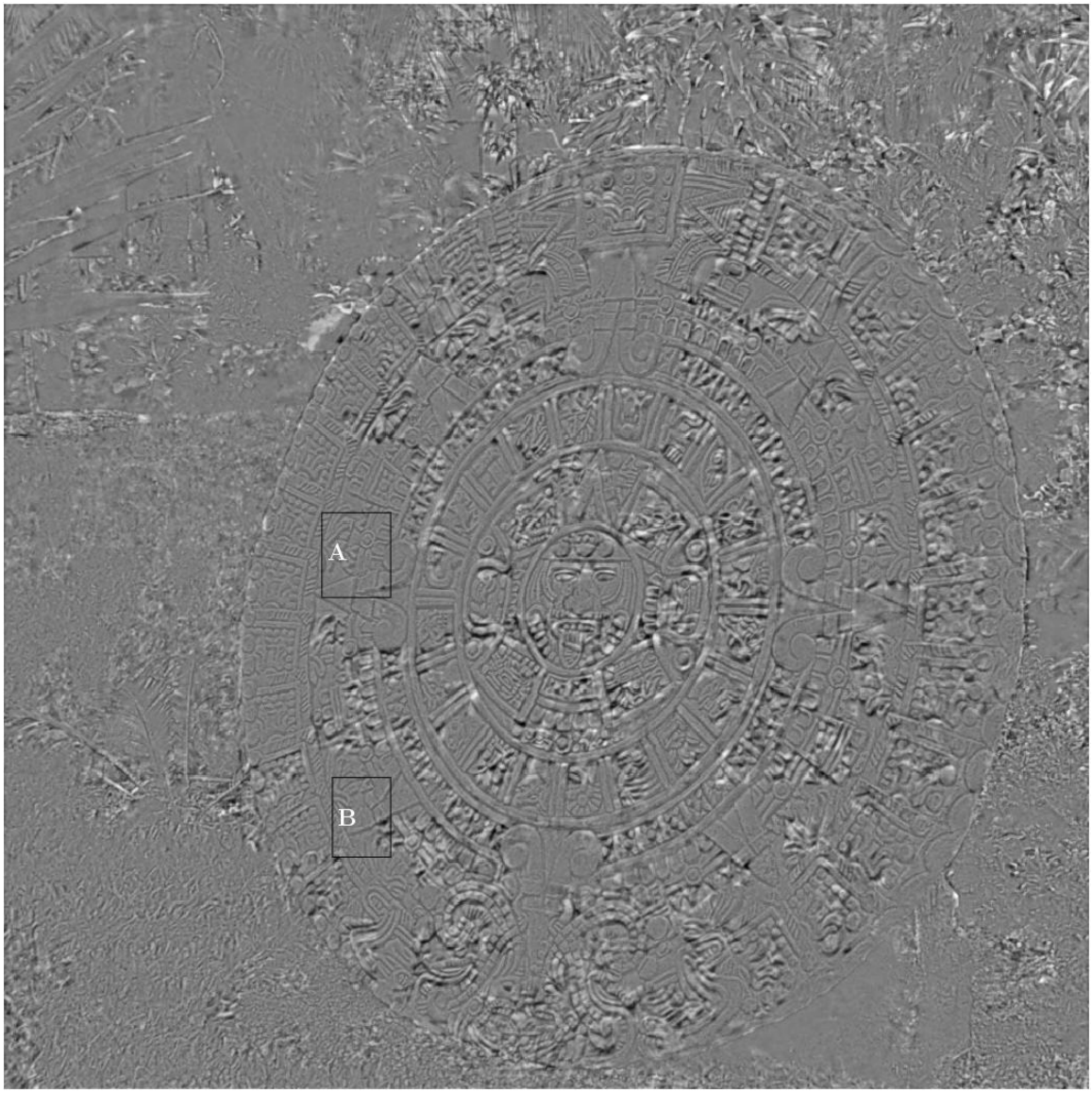
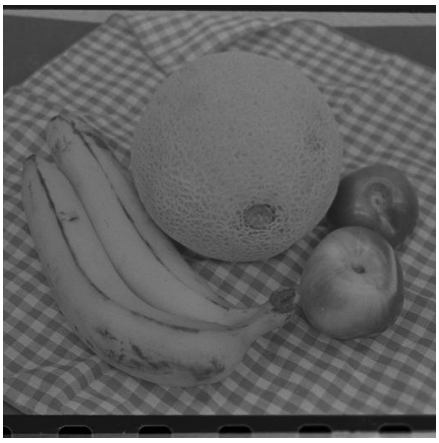


Fig. 5.10: Residual image obtained when predicting image shown in Fig. 5.7. A 3-scale hierarchical variable-block-based disparity estimation scheme is used, with a scale-0 block sizes ranging from 8×8 - 32×32. Image has been scaled for display purposes.

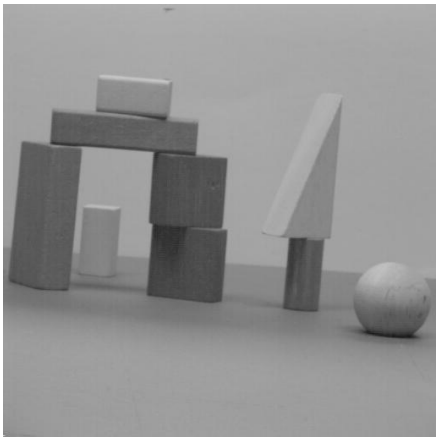
(a) Left-view (tar.), 640×480

(b) Right-view (ref.), 640×480

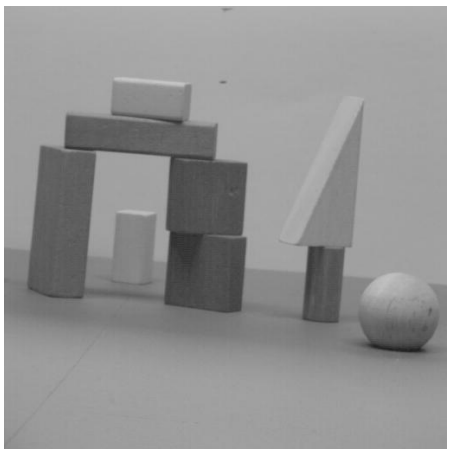


(c) Left-view (tar.), 512×512

(d) Right-view (ref.), 512×512



(e) Left-view (ref.), 512×512



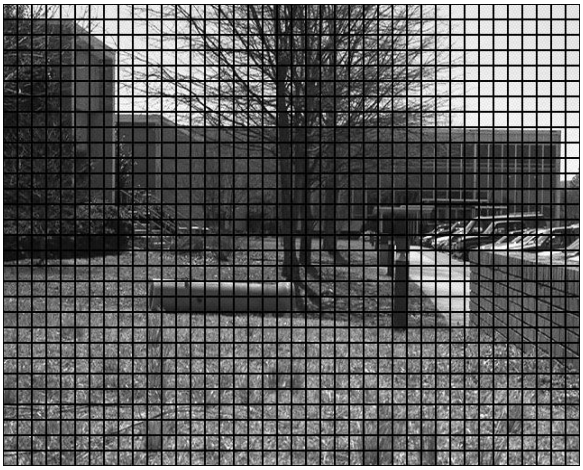
(f) Right-view (tar.), 512×512

Fig. 5.11: "*outdoors*", "*fruits*" and "*arch*" stereo-image pairs.

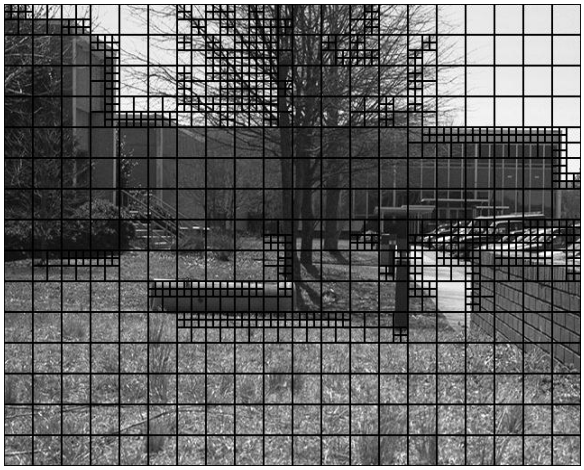
- RS indicates results obtained using Shukla and Radha's algorithm [17],
- ND_{nf} indicates results when using the proposed algorithm in conjunction with "CDF-9/7" filters. Loop-filtering is not used.
- ND_1 indicates results when using the proposed algorithm in conjunction with "CDF-9/7" filters. An EPNR filter is used, with $\lambda = 1.35$, and two filter iterations,
- ND_2 indicates results when using the proposed algorithm in conjunction with "Odegard-9/7" filters. An EPNR filter is used, with $\lambda = 1.35$, and two filter iterations,
- ND_3 indicates results when using the proposed algorithm in conjunction with "Cooklet-17/11" filters. An EPNR filter is used, with $\lambda = 1.35$, and two filter iterations,

The "*outdoors*" stereo-image pair is decomposed using a 4-scale DWT, while a 5-scale DWT is used in decomposing the "*fruits*" and "*arch*" stereo-image pairs, in order to match results published in [12, 17]. The bit-rate for a target image represents a sum total of bits required for decoding a (scale-2) quadtree-map, (scale-0) disparity-vectors and (scale-0) residual image data. Results of encoding these stereo image pairs are presented in Tables 5.1-5.13. For each test condition, the highest PSNR values is shown in bold. Details of these tables are described as follows:

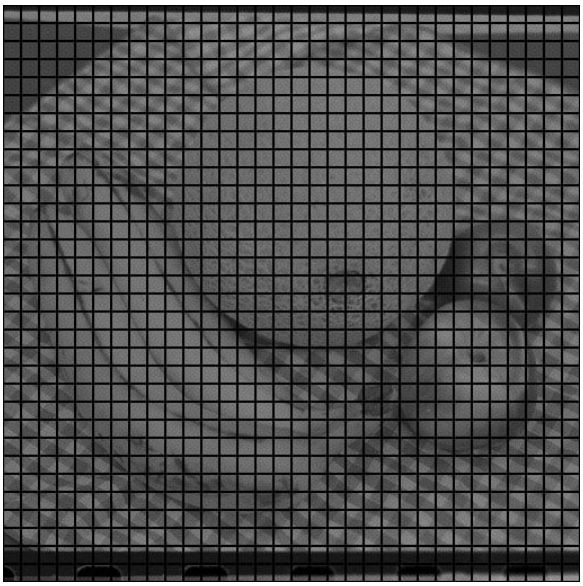
- Tables 5.1, 5.2, 5.6, 5.7 and 5.11 compare the performance of the proposed algorithm with those presented in [12] and [17]. A variable-block-based disparity estimation scheme is used, with quadtree maps generated at scale-2. "CDF-9/7" wavelet filters are used in transforming images. Results are presented with and without using an EPNR filter ($\lambda = 1.35$ and two filter iterations). The results



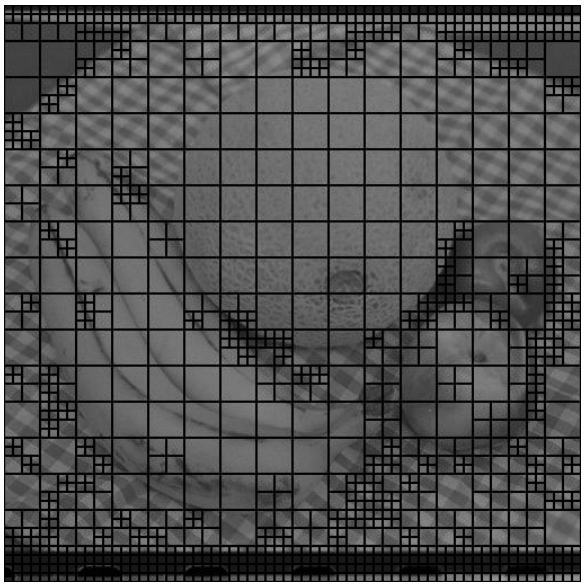
(a) Fixed-block-based, 16×16



(b) Variable-block-based, 8×8 - 32×32, V_t at scale-2 = 120



(c) Fixed-block-based, 16×16



(d) Variable-block-based, 8×8 - 32×32, V_t at scale-2 = 48

Fig. 5.12: Block structure of "*outdoors*" and "*fruits*" target image-views, when using fixed- and variable-block-based disparity estimation.

from these tables clearly indicate that the proposed algorithm outperforms Frajka and Zeger's and Shukla and Radha's algorithms by 0.3-1.5 dB., if loop filtering is used. Even without loop-filtering, the proposed algorithm outperform's Shukla and Radha's algorithm. However, it under-performs against Frajka and Zeger's algorithm as they used an overlapped-block disparity compensation (OBDC) technique to smooth disparity compensated images.

- Tables 5.3, 5.4, 5.8, 5.9 and 5.12 present results when comparing the proposed algorithm with the different wavelet filters used in this research work. Variable-block-based disparity estimation is used in each instance. An EPNR filter, with $\lambda = 1.35$ and two filter iterations, is used to smooth disparity compensated images for each result. Comparable results are obtained when using longer and smooth wavelet-filters. These results are superior to the Haar-DCT combination reported in [12]. This is further discussed, shortly.

- Tables 5.5, 5.10 and 5.13 compare the efficacy of using variable-block or fixed-block disparity-estimation schemes. A "CDF-9/7" wavelet filter is used to transform images. An EPNR filter, with $\lambda = 1.35$ and two filter iterations, is used to smooth disparity compensated images for each result. In most cases, a variable-block disparity estimation scheme provides superior PSNR results when compared with its fixed-block-based counterparts.

A detailed discussion is presented that expands on the summary presented above. In order to begin this discussion a statement, with regards to the "CDF-9/7" wavelet filter, from [12, p. 4] is quoted as follows:

> "..It is preferred for its regularity and smoothing properties. With image pixels less correlated in residual images shorter filters can better capture the local changes.."

Table 5.1: Ref. (right) view of "*outdoors*" stereo-image pair encoded at 2.00 bpp.

bpp (tar.)	FZ (dB)	ND_{nf} (dB)	ND_1 (dB)
0.20	21.87	20.25	**22.15**
0.25	22.38	20.73	**22.69**
0.30	22.78	21.14	**23.14**
0.35	23.18	21.64	**23.52**
0.40	23.57	22.04	**23.90**
0.45	23.95	22.31	**24.21**
0.50	24.34	22.59	**24.62**
0.60	25.09	23.17	**25.45**
0.75	26.18	24.08	**26.60**

Table 5.2: Ref. (right) view of "*outdoors*" stereo-image pair encoded at 5.33 bpp.

bpp (tar.)	FZ (dB)	ND_{nf} (dB)	ND_1 (dB)
0.20	21.82	20.28	**22.21**
0.25	22.32	20.76	**22.80**
0.30	22.72	21.18	**23.20**
0.35	23.12	21.69	**23.59**
0.40	23.49	22.08	**23.97**
0.45	23.89	22.37	**24.30**
0.50	24.27	22.65	**24.76**
0.60	25.02	23.24	**25.62**
0.75	26.14	24.23	**26.78**

Table 5.3: Encoding "*outdoors*" stereo-image pair with different wavelet filters. Ref. image at 2.00 bpp.

bpp (tar.)	ND_1 (dB)	ND_2 (dB)	ND_3 (dB)
0.20	22.15	**22.26**	22.10
0.25	22.69	**22.78**	22.68
0.30	23.14	**23.19**	23.15
0.35	23.52	**23.61**	23.49
0.40	23.90	**23.97**	23.88
0.45	24.21	**24.33**	24.20
0.50	24.62	**24.77**	24.55
0.60	25.45	**25.46**	25.42
0.75	26.60	26.58	**26.63**

Table 5.4: Encoding "*outdoors*" stereo-image pair with different wavelet filters. Ref. image at 5.33 bpp.

bpp (tar.)	ND_1 (dB)	ND_2 (dB)	ND_3 (dB)
0.20	22.21	**22.28**	22.13
0.25	22.80	**22.81**	22.72
0.30	23.20	**23.24**	23.18
0.35	23.59	**23.67**	23.54
0.40	23.97	**24.05**	23.92
0.45	24.30	**24.43**	24.26
0.50	24.76	**24.88**	24.65
0.60	**25.62**	25.60	25.54
0.75	26.78	26.77	**26.79**

Table 5.5: Encoding "*outdoors*" stereo-image pair with fixed-block (**F.B**) and variable-block-based (**V.B**) disparity estimation using "CDF-9/7" filters. Ref. image at 2.00 bpp.

bpp (tar.)	ND_1(**V.B**) (dB)	ND_1(**F.B** - 16×16) (dB)	ND_1(**F.B** - 8×8) (dB)
0.20	**22.15**	22.05	20.22
0.25	**22.69**	22.59	21.11
0.30	**23.14**	23.08	21.75
0.35	**23.52**	23.44	22.31
0.40	**23.90**	23.82	22.86
0.45	**24.21**	24.15	23.23
0.50	**24.62**	24.54	23.62
0.60	**25.45**	25.38	24.27
0.75	**26.60**	26.55	25.54

Table 5.6: Ref. (right) view of "*fruits*" stereo-image pair encoded at 2.00 bpp.

bpp (tar.)	FZ (dB)	ND_{nf} (dB)	ND_1 (dB)
0.20	34.61	33.78	**35.52**
0.25	35.58	34.56	**36.46**
0.30	36.33	35.28	**37.11**
0.35	36.88	35.84	**37.71**
0.40	37.39	36.22	**38.22**
0.45	37.84	36.63	**38.74**
0.50	38.25	37.02	**39.01**
0.60	38.87	37.62	**39.86**
0.75	39.74	38.72	**40.67**

Table 5.7: Ref. (right) view of "*fruits*" stereo-image pair encoded at 5.33 bpp.

bpp (tar.)	FZ (dB)	ND_{nf} (dB)	ND_1 (dB)
0.20	34.55	33.75	**35.53**
0.25	35.53	34.52	**36.50**
0.30	36.26	35.25	**37.17**
0.35	36.80	35.82	**37.79**
0.40	37.30	36.21	**38.32**
0.45	37.74	36.64	**38.86**
0.50	38.15	37.04	**39.22**
0.60	38.74	37.71	**40.01**
0.75	39.60	38.81	**40.90**

Table 5.8: Encoding "*fruits*" stereo-image pair with different wavelet filters. Ref. image at 2.00 bpp.

bpp (tar.)	ND_1 (dB)	ND_2 (dB)	ND_3 (dB)
0.20	35.52	**35.67**	35.18
0.25	36.46	**36.59**	36.09
0.30	37.11	**37.24**	37.00
0.35	37.71	**37.80**	37.55
0.40	38.22	**38.32**	38.08
0.45	38.74	**38.71**	38.56
0.50	39.07	**39.20**	39.10
0.60	39.86	**39.90**	39.87
0.75	40.67	**40.73**	40.70

Table 5.9: Encoding "*fruits*" stereo-image pair with different wavelet filters. Ref. image at 5.33 bpp.

bpp (tar.)	ND_1 (dB)	ND_2 (dB)	ND_3 (dB)
0.20	35.53	**35.71**	35.18
0.25	36.50	**36.63**	36.10
0.30	37.17	**37.31**	37.00
0.35	37.79	**37.89**	37.59
0.40	38.32	**38.45**	38.14
0.45	38.86	**38.83**	38.63
0.50	39.22	**39.33**	39.21
0.60	40.01	**40.07**	39.98
0.75	40.90	**40.99**	40.89

Table 5.10: Encoding "*fruits*" stereo-image pair with fixed-block (**F.B**) and variable-block-based (**V.B**) disparity estimation using "CDF-9/7" filters. Ref. image at 2.00 bpp.

bpp (tar.)	ND_1(**V.B**) (dB)	ND_1(**F.B** - 16×16) (dB)	ND_1(**F.B** - 8×8) (dB)
0.20	35.52	**35.78**	29.06
0.25	36.46	**36.75**	34.20
0.30	37.11	**37.38**	35.41
0.35	37.71	**37.95**	36.34
0.40	38.22	**38.43**	37.08
0.45	38.74	**38.91**	37.66
0.50	39.07	**39.32**	38.17
0.60	39.86	**39.99**	39.07
0.75	40.67	**40.79**	40.11

Table 5.11: Ref. (left) view of "*arch* stereo-image pair encoded at 0.25 bpp.

bpp (tar.)	RS (dB)	ND_{nf} (dB)	ND_1 (dB)
0.0885	40.28	40.95	**41.61**
0.1132	41.32	41.60	**42.32**
0.1379	41.99	42.04	**42.89**
0.1626	42.38	42.43	**43.30**
0.1873	42.80	42.81	**43.52**
0.2120	43.08	43.01	**43.78**

Table 5.12: Encoding "*arch*" stereo-image pair with different wavelet filters. Ref. image at 0.25 bpp.

bpp (tar.)	ND_1 (dB)	ND_2 (dB)	ND_3 (dB)
0.0885	**41.61**	41.57	41.07
0.1132	42.32	**42.34**	42.08
0.1379	**42.89**	42.79	42.68
0.1626	**43.30**	43.20	43.19
0.1873	**43.52**	43.44	43.45
0.2120	**43.78**	43.65	43.66

Table 5.13: Encoding "*arch*" stereo-image pair with fixed-block (**F.B**) and variable-block-based (**V.B**) disparity estimation using "CDF-9/7" filters. Ref. image at 0.25 bpp.

bpp (tar.)	ND_1(**V.B**) (dB)	ND_1(**F.B** - 16×16) (dB)	ND_1(**F.B** - 8×8) (dB)
0.0885	**41.61**	40.29	37.66*
0.1132	**42.32**	41.94	37.66*
0.1379	**42.89**	42.65	37.66*
0.1626	**43.30**	43.11	37.66*
0.1873	**43.52**	43.45	37.66*
0.2120	**43.78**	43.66	39.63

Results in the tables shown previously clearly contradict this statement. "CDF-9/7" filters clearly outperform a Haar-DCT combination reported in [12]. In addition to this, "Odegard-9/7" and "Cooklet-17/11" filters also show results that are comparable to those obtained when using a "CDF-9/7" filter. This improvement in PSNR values can be attributed to the following reasons:

- Variable-block-based disparity estimation, performed at multiple scales, greatly reduces bits required for encoding disparity-vectors. In contrast, Frajka and Zeger have reported using fixed-block-based disparity estimation with dimensions equal to 16×16. The proposed algorithm outperforms Frajka and Zeger's algorithm, when used with 16×16 fixed-block and variable-block-based (8×8 - 32×32) disparity estimation.

- Using an EPNR filter reduces partition-artifacts. This in-turn improves the correlation amongst pixels at partition (block) boundaries. In other words, transitions between partitions become smooth. As a result, smooth bi-orthogonal filters like "CDF-9/7", "Odegard-9/7" and "Cooklet-17/11" can be used to transform such images.

- As discussed in Chapter 3, and experimentally proved in Appendix B, an ASWDR algorithm is able to encode more high-frequency wavelet coefficients than a non-adaptive WDR algorithm. An MGE algorithm (used by the authors in [12]) exploits only intra-correlation amongst coefficients for encoding wavelet coefficients. This is similar to a WDR algorithm.

It is also observed that higher bit-rates for reference images do not guarantee higher PSNR values for reconstructed target images, when using the FZ algorithm. This reflects the inherent problem of drift associated with an open-loop structure (Fig. 4.1(a)). These results support mathematical proof provided in Chapter 4 that explains the sub-optimal

nature of this codec. On the other hand, the proposed algorithm uses a closed loop structure. This, coupled with an ASWDR algorithm guarantees that more high-frequency components are reconstructed and hence explains improved PSNR values in all instances.

As indicated from Tables 5.11 - 5.13, the proposed algorithm outperforms the state-of-the-art variable-block-based algorithm, proposed by Shukla and Radha [17]. Reasons for this are highlighted as follows:

- In [17], blocks generated from a quadtree-partitioning scheme are *independently* encoded using a disparity-compensated JPEG2000 algorithm. An overall distortion metric for the complete image is obtained by summing these individual distortions. This is a formulation that has been previously reported in [67, 68] and [69]. A major disadvantage of using this formulation is that correlation amongst nearest neighbor blocks is not exploited. This affects the encoding process. In addition, this formulation does not take into account partition (blocking) artifacts occuring due to imperfect disparity compensation. From Table 5.11 it can be observed that the performance of the algorithm in [17] is nearly comparable to the proposed algorithm, when loop-filtering is not used. From the same table it can be observed that exploiting regional correlation amongst blocks i.e., encoding the image globally and not block-wise, greatly improves PSNR values.

- In [73], qualitative and quantitative results are provided explaining the superior performance of an ASWDR algorithm in comparison with JPEG2000, when encoding natural images. Results shown in Table 5.11 validate these findings when encoding disparity compensated residual images. As discussed in Appendix B, JPEG2000 uses an EBCOT image coding algorithm that relies exclusively on intra-scale correlation when encoding wavelet coefficients. Furthermore, the blocky nature of EBCOT introduces "tiling-artifacts" in reconstructed images. These have similar characteristics to partition-artifacts .

- The algorithm in [17] relies on scale-0 quadtree partitioning. In the proposed algorithm, a scale-2 partitioning is effected. Qualitative results, shown previously, clearly demonstrate that a reduced number of blocks are required when using a hierarchical approach for quadtree-map generation. This in turn improves PSNR values.

In spite of these advantages, the proposed technique has an inherent *drawback*. An *a priori* knowledge of the threshold variance, V_t, is necessary in order to perform adequate levels of quadtree partitioning. In simulations described previously the following variance values have been used when partitioning the all low-pass subband at scale-2:

- $V_t = 120$, when encoding the "*outdoors*" stereo image pair and
- $V_t = 48$, when encoding the "*fruits*" and "*arch*" stereo image pairs.

These are empirical values and have been found suitable for images considered in these simulations. For example, if V_t is assigned a small value then the all low-pass subband may be *over-partitioned* leading to an increase in bits required to transmit disparity vectors. On the other hand, if V_t is assigned a large value then the all low-pass subband may be *under-partitioned*. This may degrade the quality of residual images affecting the number of bits required to encode them.

From Table 5.10 it can be seen that the variable-block scheme is outperformed by an equivalent 16×16 fixed block-based scheme for the "*fruits*" stereo-image pair. On the flip side, Table 5.13 reveals that using an 8×8 fixed-block-based estimation scheme does not produce any change in PSNR values! This is due to the fact that *bits earmarked for residual image encoding are wasted in encoding disparity vectors only*. As previously stated, further research work is needed to ascertain variance threshold values for a large class of stereoscopic image data (textured, smooth and combination of both features). Results cited by Vaisey and Gersho [70] can be used as an initial reference point.

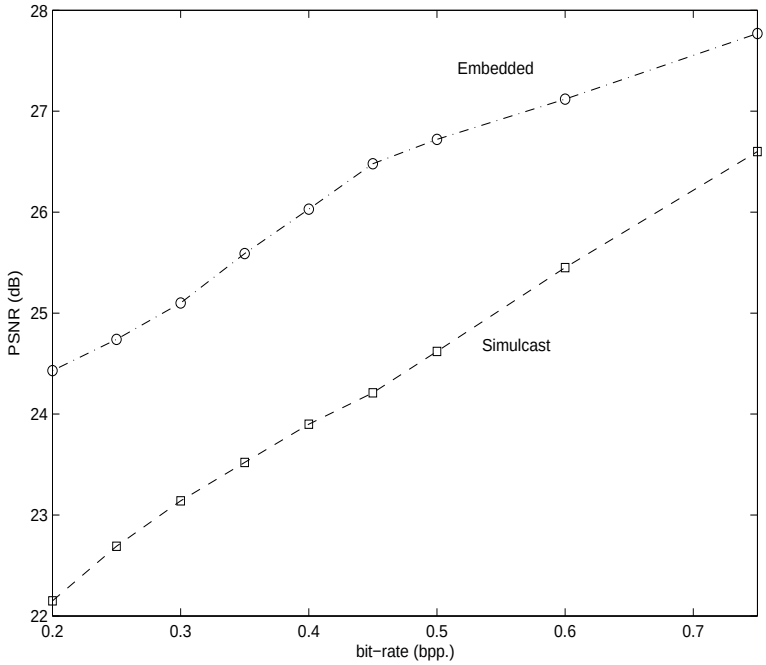
Notwithstanding this limitation, the proposed image content-based partitioning scheme is useful in a R-D sense. In addition, suitable choice of variance values can efficiently segment objects in a region (e.g., the drainage pipe, outline of the tree in Fig. 5.12(d)). This is not possible with a fixed-block-based disparity estimation scheme. As previously indicated, improvements in this encoding scheme can be obtained by merging blocks of an area having similar texture. This is left as a topic for future research.

Spatial-scalability

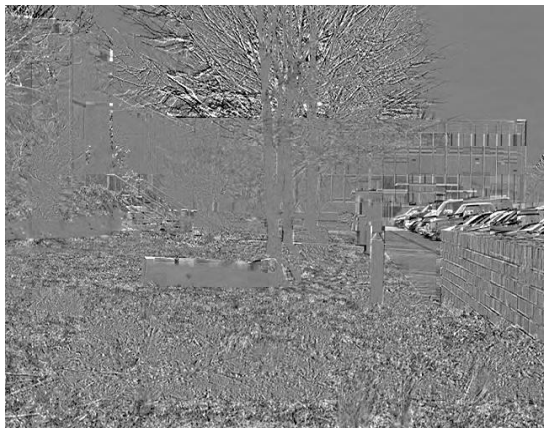
In order to explain the performance of the proposed algorithm in a spatially-scalable framework, the following example is considered.

Example : From Fig. 5.1(a), let $\mathcal{K}_{22}$, $\mathcal{K}_{21}$ and $\mathcal{K}_{20}$ be overall bit-rates allocated for target images from the"*outdoors*" stereo-image pair, obtained from scale-2, -1 and -0, respectively. Consequently, results from Table 5.1 may be interpreted as PSNR values of reconstructed target images at scale-0, in presence of images from scale-2 and scale-1. The algorithm listed in Sec. 5.1.2 presents an alternative approach, whereby three *layers* of images are transmitted simultaneously in an embedded framework. PSNR values of reconstructed target images at scale-0, obtained in both embedded and independent simulcast modes, can be seen from Fig. 5.13(a).

A quality constraint is imposed on images obtained at scales 2 and 1. This is to provide an *unbiased* distribution of bits, when transmitting images at different spatial resolutions. Thus, images at scales 2 and 1 should be encoded at "sufficiently high bit-rates". This being a subjective criterion, no numerical values are presented here. However, for experimental purposes K_2 and K_1 are assumed to be equal to 2.0 bpp. Other parameters used in obtaining results shown in Table 5.1 are kept constant. Qualitatively speaking, the improved performance of an embedded mode when compared with an independent simulcast mode can be attributed to the type of disparity compensated residual image being encoded. It can clearly be seen that Fig. 5.13(c) contains signifi-



(a) PSNR values of reconstructed target images from the "*outdoors*" stereo-image pair, in independent (simulcast) (□) and embedded (○) modes. "CDF-9/7" filters are used for transforming images.



(b) Residual Image - Independent Mode, $\lambda = 1.35$

(c) Residual Image - Embedded Mode, $\lambda = 1.35$

Fig. 5.13: PSNR plots and residual images at scale-0 when encoding the "*outdoors*" stereo-image pair. Variable-block-based disparity estimation, 4-scale DWT and EPNR filter (with $\lambda = 1.35$ and two filter iterations) have been used. Images have been scaled for display purposes.

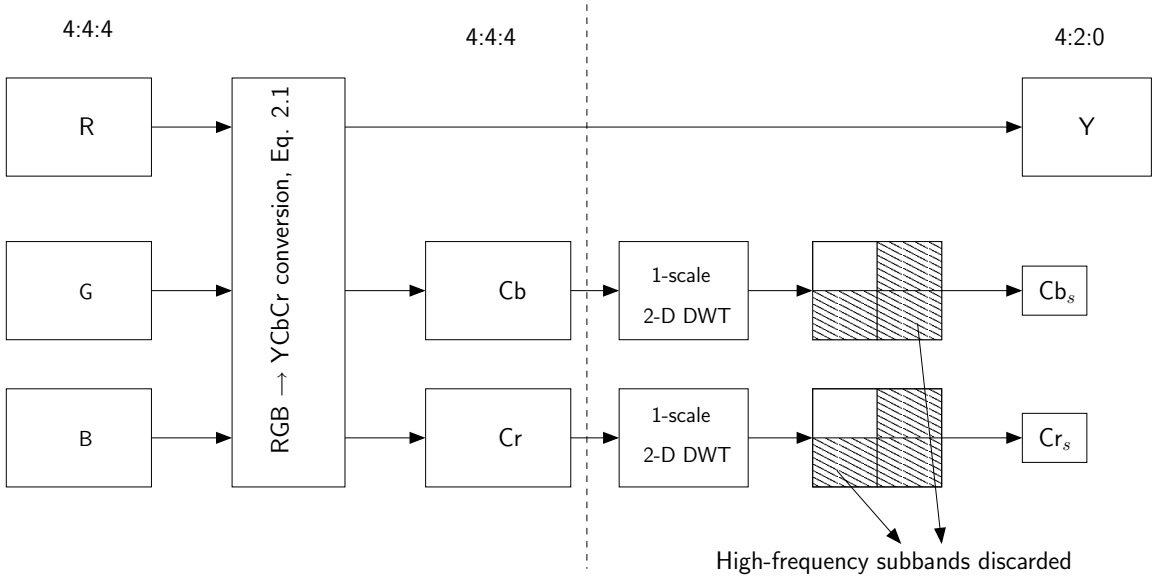


Fig. 5.14: 4:4:4 RGB to 4:2:0 YCbCr conversion. Cb_s and Cr_s represent downsampled versions of Cb- and Cr-components.

cantly less information than Fig. 5.13(b). This in turn explains the PSNR values shown in Fig. 5.13(a).

5.5.4 Results for encoding stereoscopic color images

The results discussed in previous sub-sections correspond to stereoscopic gray-scale images. To conclude this section, preliminary subjective results are presented when encoding stereoscopic color images. As indicated in Chapter 2, stereoscopic color images are acquired in an RGB space. This is transformed into a YCbCr space, as per Eq. 2.1 [2]. The HVS is more sensitive to perturbations in the Y component than in the Cb and Cr components. From a compression point of view, this implies that Y components *must* be represented at *significantly* higher bit-rates when compared with Cb and Cr components. This fact is exploited when encoding monoscopic images.

YCbCr images acquired at a sampling ratio of 4:4:4, can be transformed, and downsampled as per a 4:2:0 ratio and shown in Fig. 5.14. These ratios (i.e., 4:4:4 and 4:2:0) are used in video-coding [2] and *indicate* whether sub-sampling has been performed on

different components of an image. Thus a ratio 4:4:4 indicate that no sub-sampling has been performed, while a ratio 4:2:0 indicate that Cb and Cr components are sub-sampled by a factor of two in both horizontal and vertical dimensions.

This involves performing a 1-scale 2-D separable DWT on the Cb and Cr components. The all low-pass subband of these components are retained while the high-frequency subbands are discarded (Fig. 5.14). This in fact reduces actual number of coefficients from Cb and Cr components that need to be encoded. Hence if a bit-rate k bpp is allocated for encoding these downsampled image components, it actually implies that a bit-rate of $k/4$ bpp is used to represent it at full spatial resolution.

Distortions perceived in color stereo-image pairs can be partially classified as:

- Coding artifacts introduced by the successive-approximation scheme of an ASWDR algorithm,
- Blurring artifacts introduced by an EPNR filter,
- Color bleeding due to coarse quantization of chrominance components, and
- Visual fatigue due to improper depth perception.

When presented with a stereo-image pair, the HVS tries to fuse objects from both views. As a result, distortions from the target image are "masked" out by the HVS when this image is viewed simultaneously (in context) with a higher quality reference image. However this is true only up to a certain threshold [23]. In these experiments, an attempt is made to subjectively determine this threshold value. This is based entirely on bit-rates needed to encode each component of an image. The scenarios of perceptual ringing artifacts, blurring, and color bleeding in the decoded stereo image pair are considered as parameters in this subjective evaluation. Results from two representative stereo-image pairs have been presented. The "*medallion*" image pair contains large textured areas,

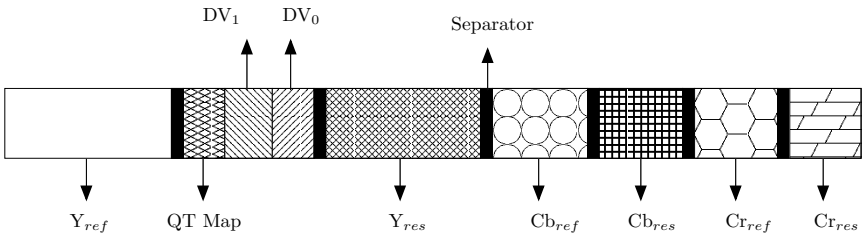


Fig. 5.15: File structure of an encoded color stereo-image pair (independent simulcast mode).

while the "*bull*" image pair generally has large smooth regions. Due to space and display constraints actual images cannot be shown here. Instead, the reader is directed to the enclosed CD ROM for individual images as well as *anaglyphs* obtained by fusing both image views.

Due to varying sensitivities of the HVS to individual components, the bit-rate ratio $\mathcal{K}_Y$:$\mathcal{K}_{Cb}$:$\mathcal{K}_{Cr}$ is varied from 32:1:1 (for highly textured images) to 80:1:1 (for non-textured images). Disparity estimation and compensation is performed on the Y component as per the algorithm shown in Fig. 5.1(a). As a hierarchical variable-block based disparity estimation (HBDE) scheme is employed, disparity vectors obtained at scale-1 are used to compensate both the chrominance components. A single file stream is generated for storing all resulting bits. Fig. 5.15 visually depicts various components in an embedded bit-stream when encoding a color stereo-image pair. DV indicates the disparity vectors from scale-1 and scale-0. QTMap indicates bits required for encoding the quadtree map. It should also be emphasized that these images are encoded in an independent simulcast mode. "CDF-9/7" wavelet filters have been used in transforming images.

Other conditions used in encoding monochrome stereo-image pairs (e.g., filter smoothing parameter λ, etc.) are repeated here. The decoded images were informally analyzed by several test subjects experienced in viewing stereoscopic images. Tables 5.14 and 5.15 indicate results from the "*medallion*" and "*bull*" stereo-image pairs. To limit the scope of this experiment, the subjects were only asked if *coding artifacts* (CA) or *color-bleeding*

Table 5.14: Subjective results when viewing decoded images from the "*medallion*" stereo-image pair in a stereoscopic mode.

$\mathcal{K}_Y$ (Ref.)	*ratio* (Ref.)	$\mathcal{K}_Y$ (Tar.)	*ratio* (Tar.)	C.A	C.B
2.0	32:1:1	0.5	32:1:1	N	N
		0.25		N	N
		0.20		Y	N
1.6	32:1:1	0.5	32:1:1	N	N
			128:1:1	N	N
			256:1:1	N	N
			512:1:1	N	N
		0.4	512:1:1	N	N
		0.25	32:1:1	Y	N
1.0	32:1:1	0.5	32:1:1	N	N
		0.4		Y	N

Table 5.15: Subjective results when viewing decoded images from the "*bull*" stereo-image pair in a stereoscopic mode.

$\mathcal{K}_Y$ (Ref.)	*ratio* (Ref.)	$\mathcal{K}_Y$ (Tar.)	*ratio* (Tar.)	C.A	C.B
2.0	80:1:1	0.5	512:1:1	N	N
		0.25		N	N
		0.20		N	N
		0.125		N	N
		0.1		Y	N
1.0		0.25		N	N
		0.125		Y	N

(CB) was perceptible in the decoded pairs when compared with their uncompressed versions. Images were displayed on a CRT computer screen. These were viewed with both red-blue (anaglyph) and time-sequential (stereoscopic) glasses. User responses are indicated as Y/N (yes/no). The following notations have been used in these tables:

- $\mathcal{K}_Y$ in bits-per-pixel (bpp),
- *ratio* indicates $\mathcal{K}_Y$:$\mathcal{K}_{Cb}$:$\mathcal{K}_{Cr}$ at *full spatial resolution*.

Thus, a 'Y' indicates that subjects were able to detect blurring and/or other coding artifacts in reconstructed images, when compared with their corresponding uncompressed versions. An important point was ascertained from these limited results. Subjects were not able to detect color-bleeding when viewing both images simultaneously, at extremely small bit-rates. However, they were clearly able to identify perceptible bleeding when individually viewing reconstructed target images (e.g., the reconstructed target image in Table 5.15 when $\mathcal{K}_Y$ (Ref.) $= 0.8$ bpp.). This fact can form the basis for effective compression of stereoscopic moving color-image sequences. However, no specific details can be inferred with regards to optimum bit-rates for encoding reference images. A future research work, involving a larger number of images and observers, is envisaged to determine this fact.

All images referred to in Tables 5.14 and 5.15 can be viewed on the accompanying CD ROM to verify the observations made in the tables. Results for another stereo image pair ("*burial-ground*") are also presented in the CD ROM. A pair of anaglyph glasses, adapted to the anaglyph images, is also provided with this thesis. Anaglyphs shown in these tables have been created using the algorithm presented in [74].

To better appreciate these results, a few representative examples of individual and fused anaglyph images of the "*bull*" stereo-image pair are presented. A bit-rate of 8.0 bpp is assumed for each component of raw-versions of these images. Anaglyphs of these stereo pairs consists of encoded reference and (disparity-compensated) target image views, having the following bit-rates:

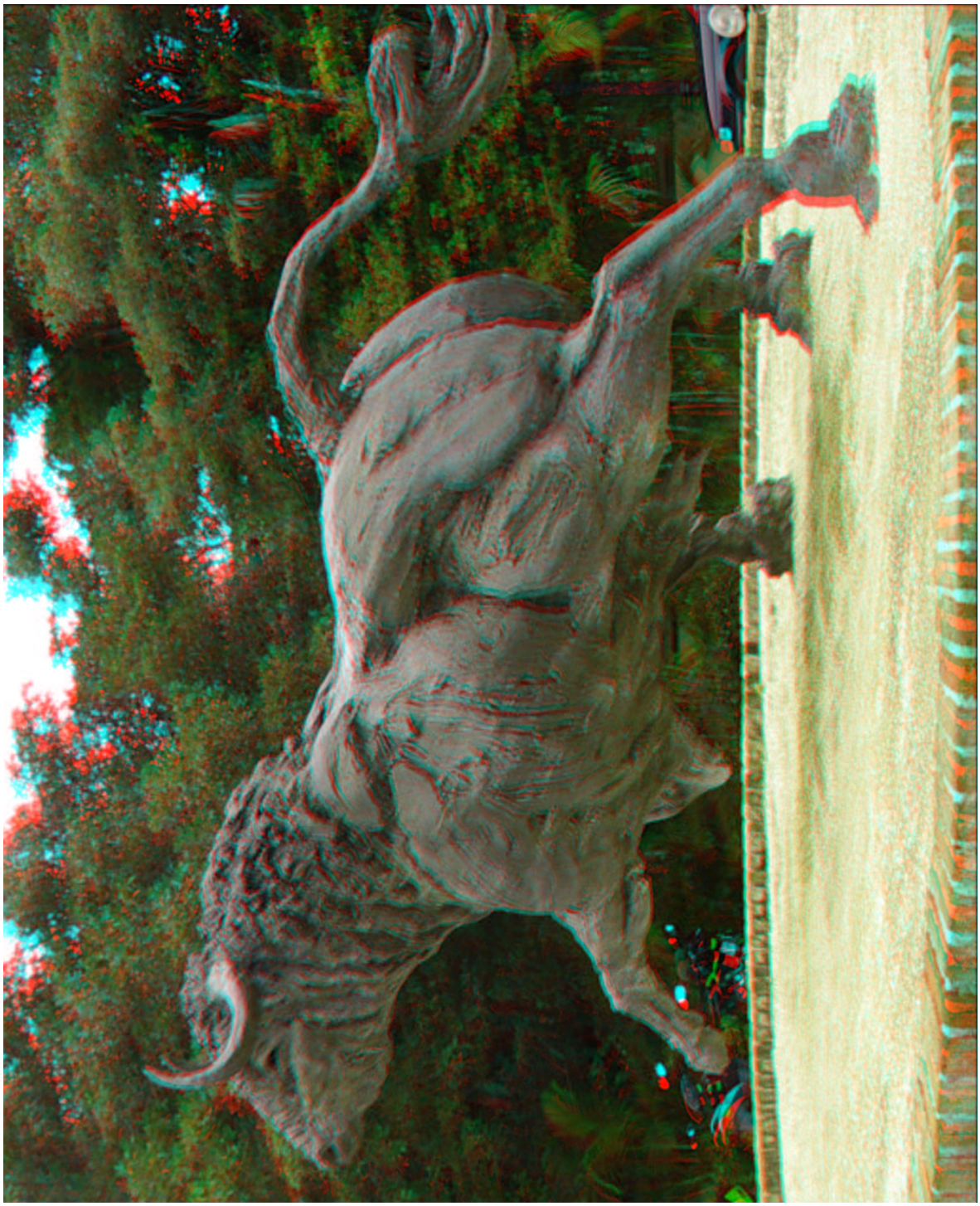
- "*Bull*"
  - Reference image : $\mathcal{K}_Y = 1.0$ bpp, $\mathcal{K}_Y : \mathcal{K}_{Cb} : \mathcal{K}_{Cr} = 80 : 1 : 1$,
  - Target image : $\mathcal{K}_Y = 0.125$ bpp, $\mathcal{K}_Y : \mathcal{K}_{Cb} : \mathcal{K}_{Cr} = 512 : 1 : 1$,

(a) Raw target image from the "*bull*" stereo-image pair. Image has been scaled for display purposes

Fig. 5.16: Representative examples of individual target images (raw and encoded) and anglyphs (raw and encoded) from the "*bull*" stereo-image pair.

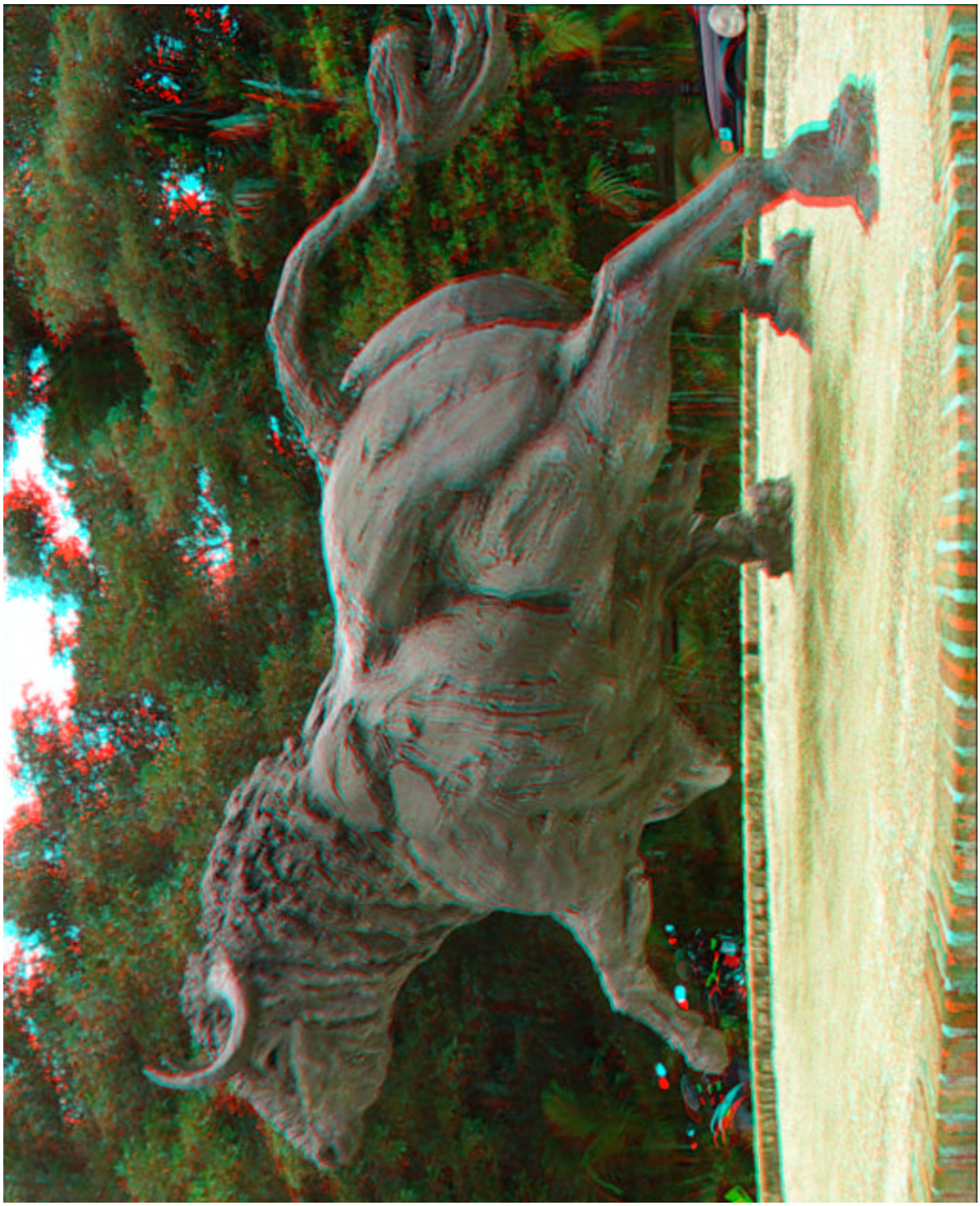
(b) Encoded target image. Image has been scaled for display purposes

Fig. 5.16: contd.



(c) Anaglyph - raw. Image dimensions are 1260×1024. An offset-correction of 20 pixels has been effected on the images for comfortable viewing.

Fig. 5.16: contd.



(d) Anaglyph - encoded. Image dimensions are 1260×1024. An offset-correction of 20 pixels has been effected on the images for comfortable viewing.

Fig. 5.16: contd.

On comparing Fig. 5.16(b) with Fig. 5.16(a) it can be observed that significant details in the former image have been blurred out (e.g., the body of the bull, background foliage, etc.). Color bleeding is also observed in the image (e.g., purple patches predominate the body of the bull in Fig. 5.16(b) when compared with uniform brown color in Fig. 5.16(a)). These artifacts occur entirely due to embedded coding of images and differs from previously discussed *a priori* Gaussian blurring effected on target images (Chapter 4).

Figs. 5.16(d) and 5.16(c) indicate encoded and raw versions of anaglyphs. It can be observed that distortions perceived in the encoded target image (Fig. 5.16(b)), in a monoscopic mode, are absent when viewing the same image in a stereoscopic mode (Fig. 5.16(d)). This justifies the efficacy of asymmetrical coding of stereoscopic image pairs.

Chapter 6

Summary of Stereoscopic Moving-Image Encoding and Decoding Algorithms

Overview

In this chapter a review of relevant stereoscopic moving-image coding algorithms is undertaken. Some aspects of MPEG picture hierarchies are also presented. This is followed by a discussion on the relative disadvantages of such hierarchies. SNR- and spatial-scalability are desired features in a stereoscopic moving-image coding scheme. Limited references of these features are present in literature, with respect to stereoscopic moving-image coding. As a consequence, monoscopic moving-image coding schemes are considered to discuss these features. This chapter is concluded by a discussion on temporal-interleaving, when encoding stereoscopic moving-image sequences.

6.1 Introduction

REMOVAL of temporal redundancies is a critical component in any moving-image coding algorithm. As discussed in Chapter 2, disparity- and motion-vector estimation techniques are very similar to each other. In stereoscopic moving-image sequences, both components need to be estimated. However, disparity estimation can be avoided if encoding both streams independently.

Highlights of various motion estimation techniques can be found in [45]. As indicated

in Chapter 2, a hierarchical-search strategy is employed in estimating motion-vectors. An early implementation of a joint disparity-/motion-vector estimation can be found in [51]. In addition, a hardware implementation of real-time stereoscopic moving image encoder (DISTIMA project [50]) has justified the use this search strategy.

An early implementation of a stereoscopic moving-image encoder can be found in [75]. Refinements to this technique have been proposed by Chang and Wu [20] and Thanapirom *et al.* [76]. All these algorithms are wavelet-based. In contrast, Puri *et al.* [77] reviews some aspects of DCT-based stereoscopic moving-image encoding. As mentioned previously, DCT-based techniques have been replaced by their DWT-based counterparts when encoding still-images. Moving-image encoding involves transmission of intra-pictures and residual images. Current MPEG-4 standards envisage using DWT-based coding for intra-pictures. Hence, unless otherwise mentioned, this chapter focuses on DWT-based moving-image coding systems. Other terminologies (e.g., intra-pictures, etc.) are discussed shortly. In addition, drawbacks of these wavelet-based systems are also discussed later in this chapter.

SNR- and spatial-scalability are desired features when encoding moving-images. However, no suitable references have been found addressing these issues, in the context of stereoscopic moving-image coding. However, such features have been addressed in the context of monoscopic moving-image coding. In Arnold et al. [78], a DCT-system is presented that achieves drift-free SNR-scalability. A corresponding structure can be found in Domanski *et al.* [22], addressing the problem of spatial-scalability. The scope of this chapter is limited to stereoscopic moving-image encoding. Hence these techniques are not discussed in detail. The concerned reader is directed to the references presented above for further details.

In [3] it has been stated that, when viewing a stereo-image pair, both images need not be displayed at full perceptual quality (i.e., full SNR-resolution). Psycho-visual

experiments [79], [59], [23] have validated this fact. It was conjectured in [23] that prolonged exposure to these asymmetrically coded stereo-image pairs *might* lead to *visual fatigue*. As a result, *temporal interleaving* was proposed. This involves interchanging the perceptual qualities of both views, preferentially at the occurrence of scene-cuts in a moving-image sequence. Drawbacks of this technique are discussed in a later part of this chapter.

6.2 Current picture hierarchies and their drawbacks

A summary of notations involved in monoscopic moving-image encoding is provided below. These notations are appropriately extended to the context of stereoscopic moving-image encoding. MPEG coding standards provide a classification of contiguous pictures in a monoscopic moving-image stream. These are highlighted as follows [2]:

- *I-Picture* : These are encoded using intra-picture coding techniques (e.g., any state-of-the-art embedded image coding scheme). They provide for fast random access but offer only moderate compression rates. Due to their being reference points for future pictures (both in the reference and target streams), they have to be encoded at "sufficiently high" bit-rates. This limits the SNR-scalability that can be obtained for such pictures. In the proposed algorithm, the perceptually efficient ASWDR algorithm is used for encoding such I-pictures.

- *P-Picture* : These are estimated from previously encoded I-pictures using motion-compensated prediction (in the reference stream). Additionally, these P-pictures from the reference stream are used to predict current and future pictures from the target stream. Such predicted frames are also classified as P-pictures. As a relatively large number of pictures are predicted from P-pictures when compared with I-pictures, limited SNR scalability can be obtained when encoding such pictures.

- *B-Picture* : These are estimated from:
  - I- and P-pictures using motion-compensated (in the reference stream) or disparity-compensated (in the target stream) prediction and
  - Two P-pictures using motion-compensated (in the reference stream) or disparity-compensated (in the target stream) prediction

  Generally, these pictures are not used to predict any other pictures. Hence, SNR-scalability for any prediction-based compressed moving-image stream, is determined by the coding performance of B-pictures.

An MPEG-2 compliant, contiguous group of pictures (GOP) can be seen in Fig. 6.1. Such a GOP is sometimes referred to as *closed GOP* [2, p 193]. In [77] a set of picture

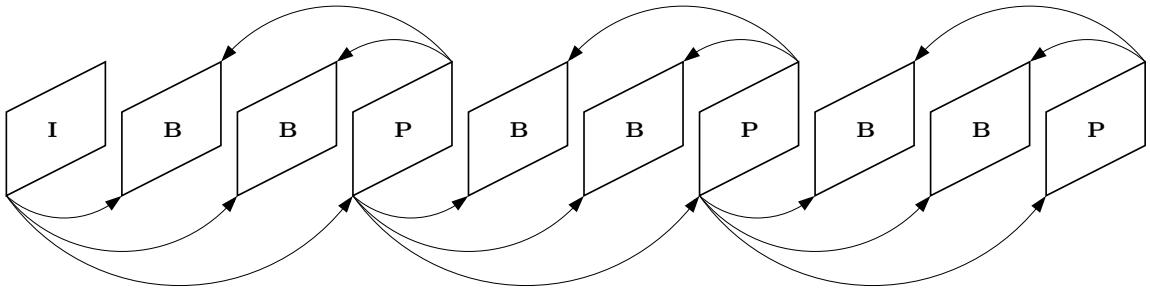


Fig. 6.1: Encoding and display hierarchy of contiguous pictures in a, MPEG-2 compliant, monoscopic moving-image sequence (GOP = 10).

hierarchies for encoding stereoscopic moving-pictures have been presented. Fig. 6.2 depicts one such structure. This has been incorporated in the MPEG-2 multiview profile (MVP) [21, 2] standards. A modified version of this hierarchy forms the framework for encoding stereoscopic moving-image pairs in [76, Fig. 2]. This scheme of contiguous pictures suffers from a major drawback. All pictures are used for predicting current, past or future pictures. In order to prevent artifacts due to drift, SNR-scalability has to be sacrificed completely. Hence this framework is not used in the current proposed

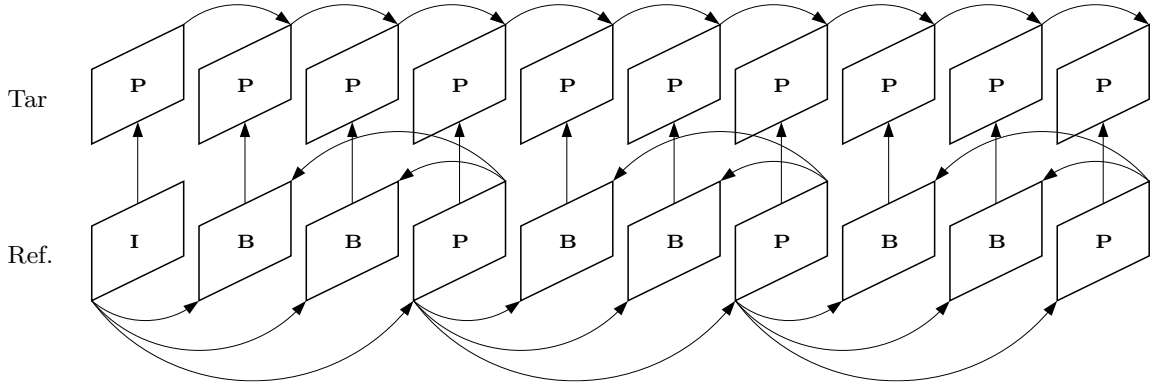


Fig. 6.2: MPEG-2 complaint multiview picture hierarchy for encoding stereoscopic imagery.

algorithm. The authors in [77] term the above structure as a *disparity- and motion-compensated* framework. On the other hand the authors in [75] and [20] use a simplified version of Fig. 6.2. This is referred to as a *disparity-compensated* structure and shown in Fig. 6.3. Evidently, SNR-scalability can be obtained when *decoding target pictures only*,

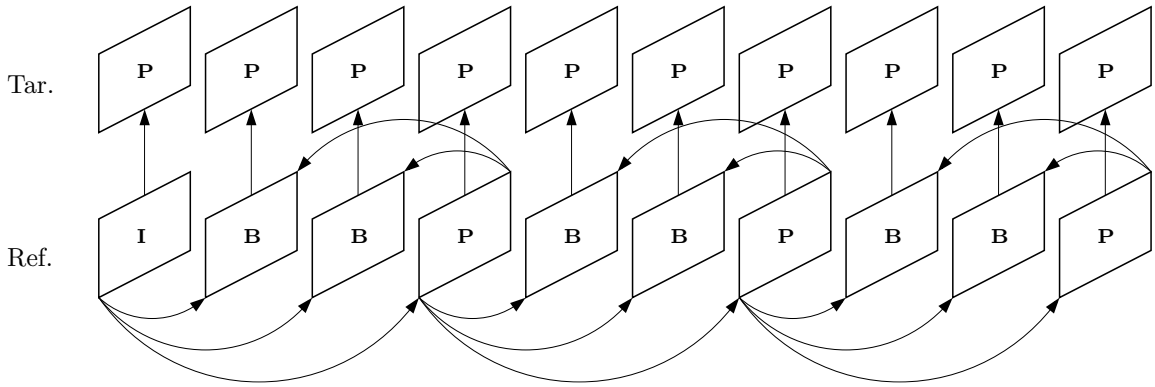


Fig. 6.3: Disparity-compensated multiview picture hierarchy for encoding stereoscopic imagery.

when using this framework. This is due to the fact that pictures from the target stream are not used for prediction. However, this structure also suffers from a subtle drawback.

In an asymmetrical coding framework the HVS can comfortably perceive depth while masking out artifacts from the target picture stream. This presupposes the fact that the reference picture stream *has been* encoded at a higher perceptual quality than the target

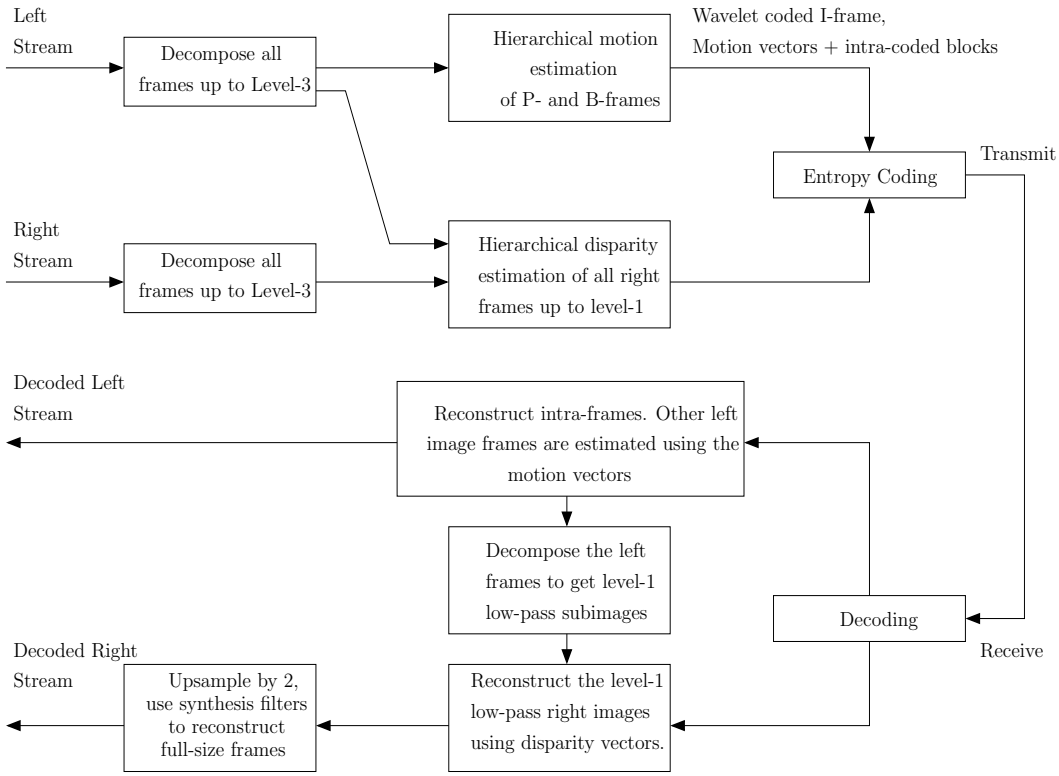


Fig. 6.4: Stereoscopic moving-image codec structure proposed by Sethuraman, Siegel and Jordan

stream. It has also been stated that while encoding a monoscopic picture sequence, SNR-scalability is obtained from B-pictures. Evidently, it is not possible to obtain both these features when using the structure shown in Fig. 6.3. B-pictures from the reference stream are used for predicting pictures from the target stream. Hence they *must* be encoded at a fixed SNR-resolution in order to prevent drift in decoded target pictures.

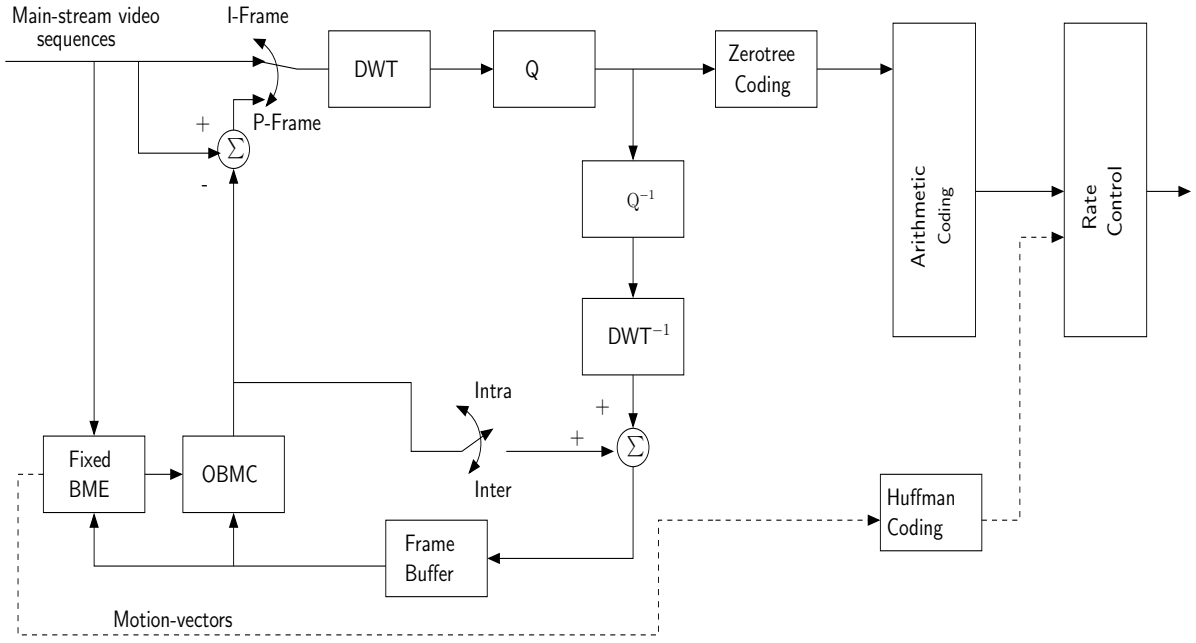
6.3 Selected stereoscopic moving-image encoding algorithms

As indicated at the beginning of this chapter, no references have been found that address issues of spatial-, SNR-, content- and temporal-scalability in a united manner, when encoding stereoscopic moving-images.

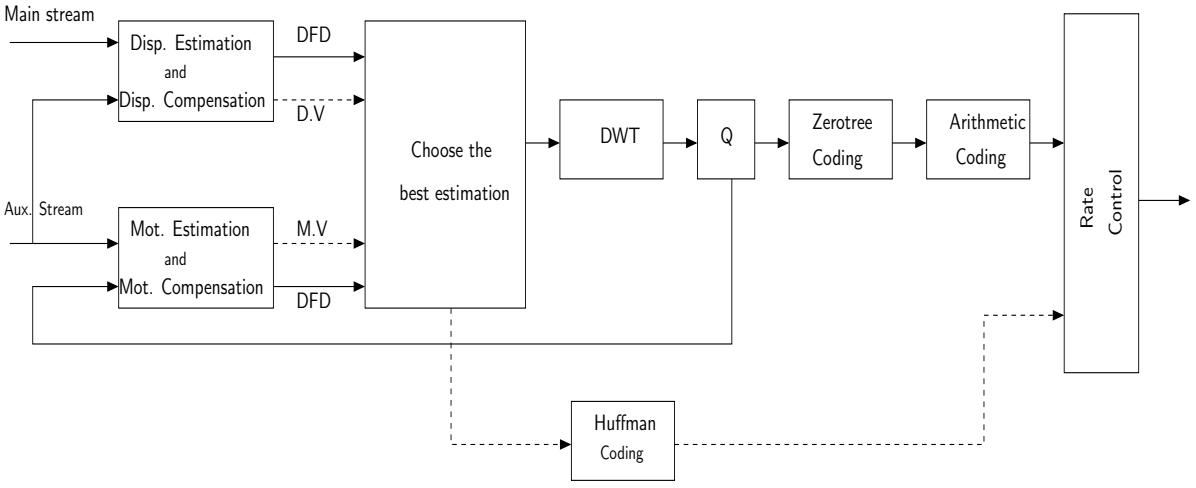
Fig. 6.4 illustrates a generic framework used by Sethuraman et.al [75] and Chang and Wu [20]. Unfortunately, a non progressive coding structure is used in encoding both dis-

parity and motion compensated pictures/frames [20, Fig. 12 & 14]. In addition, OLDC structures have been used in encoding disparity-compensated residual images. As proved in Chapter 4, this is a sub-optimal framework when encoding such residual images.

Recently, a zerotree-based stereoscopic moving-image coding structure has been proposed by Thanapirom et al. [76]. As the name indicates, a zerotree-based structure is



(a) Reference-stream encoding



(b) Target-stream encoding

Fig. 6.5: Stereoscopic moving-image encoding structure proposed by Thanapirom, Fernando and Edirisinghe

used to encode disparity compensated residual images. Fixed-block-based disparity and motion estimation is performed between corresponding and successive pictures of both streams. An OBMC technique is used for compensation, followed by Shapiro's EZW algorithm for encoding generated residual images. The picture hierarchy shown in Fig. 6.2 is used in this structure. As indicated in [76, Fig. 3], both disparity and motion estimation is performed on target stream pictures. Residual images in both instances are generated, with the residual image having a lower-energy content eventually encode. This makes it a redundant operation. As with previously discussed algorithms, this technique also relies on a sub-optimal OLDC structure for generating and subsequently encoding disparity compensated residual images. A schematic of this algorithm can be seen in Fig. 6.5.

Limited SNR- and spatial-scalability can be obtained from the above algorithms. The non-progressive nature of Chang and Wu's algorithm [20] and problems due to drift, arising from all algorithms in [75, 77, 20, 76], justifies the development of new stereoscopic moving-image coding algorithms. In [78] a scheme is presented that achieves two levels of SNR-scalability when encoding monoscopic moving-image sequences. In this, input images are quantized at two separate SNR-resolutions, DCT-transformed and subsequently encoded. Due to the progressive nature of any wavelet-based scheme, this operation of separate quantization of images becomes redundant. For the sake of completeness, Fig. 6.6 depicts this encoding scheme. This can also be found in [78, Fig. 13].

As previously indicated, no suitable references have been found that deals with the issue of spatial-scalability in the context of stereoscopic moving-image coding. Some references have been found in the context of monoscopic moving-image encoding. One such work has been reported by Domanski et al. [22]. Images are downsampled using 2-tap linear-phase filters. The all low-pass subbands, generated from this process, are DCT-transformed and subsequently encoded using motion-compensation techniques.

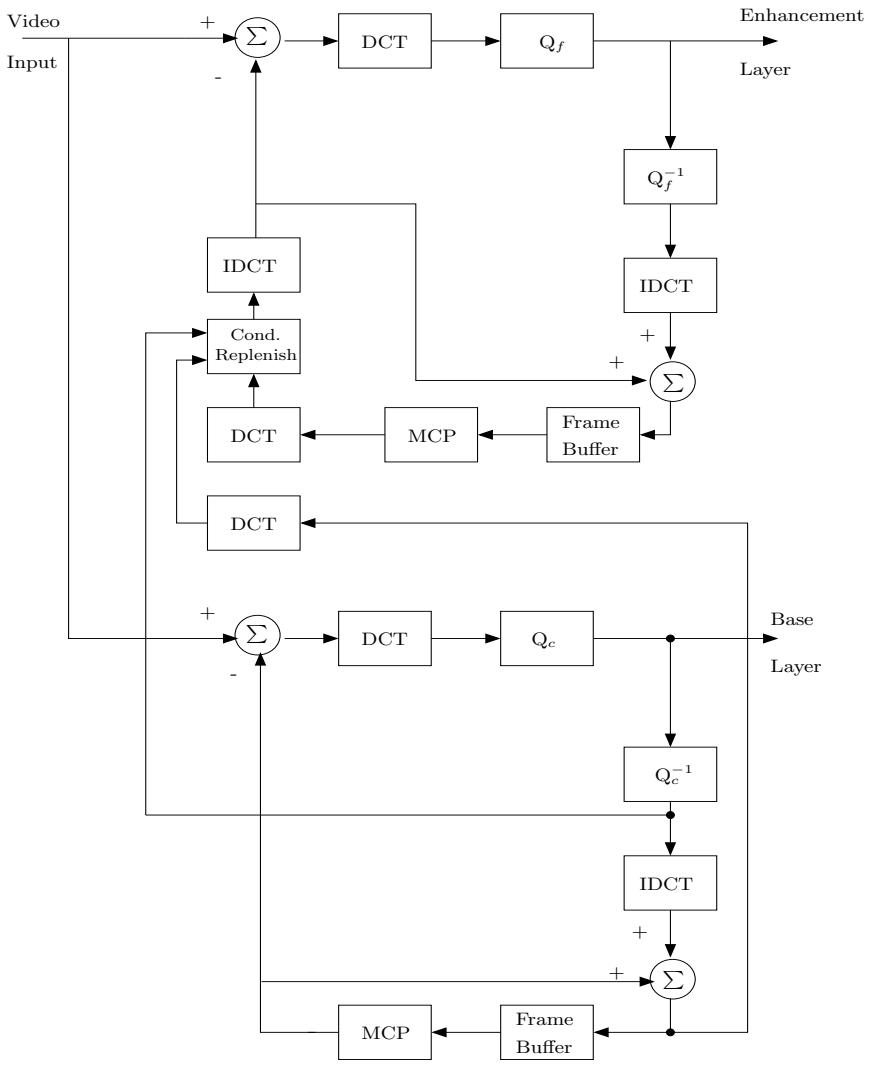


Fig. 6.6: Two-loop, DCT-based SNR-scalable encoder, proposed by Arnold, Frater and Wang.

As shown in the next chapter, this becomes a redundant operation, if a dyadic spatial framework is required. Hence, the methodology reported in this paper is not used in this thesis.

The discussion presented in the previous paragraphs pertains to prediction-based moving-image encoding. Due to the separable nature of wavelet-transforms, alternative techniques have been proposed. These fall under the category of *3-dimensional moving-image* coding techniques. Drift is a serious problem when encoding monoscopic

moving-images. 3-D structures eliminate the need for prediction-based encoding as the temporal domain is considered as a third dimension, in addition to the horizontal and vertical dimensions of images. A 3-D separable wavelet-transform is implemented on a GOP, followed by embedded image encoding techniques in three dimensions. This can achieve high levels of SNR-scalability, similar to what can be achieved when encoding a simple 2-D image. A major advantage of this technique is the absence of artifacts due to drift during decoding. However, these techniques have some inherent drawbacks.

Compared with their prediction-based counterparts, 3-D techniques require large buffers to process incoming pictures. Current pictures have to be stored in a buffer while encoding them using 3-D embedded coding techniques. This would typically be the size of the GOP being considered. In addition, future pictures have to be stored in buffers while encoding of current pictures takes place. This problem is accentuated when stereoscopic moving-images are considered. A trivial solution would be separately encode both streams using these 3-D coding techniques. As explained in previous chapters, this process does not account for inter-view redundancies in both image streams. For more information on this evolving framework of moving-image coding, the concerned reader is directed to [80, 81, 82].

6.4 Temporal interleaving in stereoscopic moving-image encoding

To conclude this chapter, the concept and justification of *temporal interleaving* in stereoscopic moving-image encoding is introduced.

In previous chapters, it was identified that when viewing stereoscopic moving-image pairs, both views need not be displayed at full SNR-resolution (assuming that spatial-resolution of both views is kept constant). This forms the premise for any state-of-the-art stereoscopic moving-image encoding structure [76, 20, 75]. This implies that pictures

from the reference view *should* be encoded and displayed at higher SNR-resolution than pictures from the target view. In [23], Stelmach and Tam hypothesized that prolonged exposure to such asymmetrically coded stereoscopic moving-image data may lead to *visual fatigue* in the HVS. To alleviate this problem, the authors proposed a *cross-switch* of these asymmetrically coded images at some time intervals. From experimental results, they concluded that this cross-switching was not perceptible by the HVS if implemented at scene cuts. This intermittent cross-switching of asymmetrically coded stereo image-pairs is referred to as temporal-interleaving. However, this technique has some inherent drawbacks and these are highlighted in the following paragraph.

The authors in [23] have not reported the use of disparity-compensation in encoding both views of the stereo-image sequence. Instead, they apply an *a priori* Gaussian blur on the target image stream (e.g., Fig. 4.4). They justify this by stating that this blurred image stream can be independently encoded at sufficiently low bit-rates (e.g., MPEG-2 coding techniques). In doing so, the authors fail to exploit inter-view redundancies between both image streams. It has been indicated in Chapter 4, that disparity estimation between images at different SNR-resolutions may lead to biased results qualitatively as well as quantitatively.

In addition, this technique assumes that a moving-image sequence is guaranteed to have scene-cuts. However, there are instances like remote robotic applications, telemedicine, etc., that do not have scene-cuts. This necessitates the formulation of alternate strategies to achieve temporal-interleaving in stereoscopic moving-image encoding. This is discussed in the next chapter.

Chapter 7

Proposed Wavelet-Based Scalable Stereoscopic Moving-Image Codec

Overview

In this chapter the algorithm proposed in Chapter 5 is extended to encode stereoscopic moving images. The first part of this chapter consists of a detailed description of the proposed codec structure. Next, a comparative study is made of its relative advantages compared with present techniques, as discussed in Chapter 6. This chapter is concluded with simulation results obtained by implementing the proposed algorithm on the "*redcar*" stereoscopic moving-image sequence.

7.1 New picture hierarchy

IN this section, a new picture hierarchy is proposed. This overcomes drawbacks of current picture hierarchies that were discussed in Chapter 6. Current MPEG-2 MVP coding standards use the hierarchy shown in Fig. 6.2. As discussed previously, Thanapirom et al., [76] use this picture hierarchy to encode a contiguous set of stereoscopic pictures. This involves individually estimating motion- and disparity-compensated residual images. The residual image with minimum energy is encoded (Fig. 6.5). This determines whether disparity- or motion-vectors are encoded.

In Chapter 2, the concept of displaced-disparity-vector (DDV) was introduced. The proposed picture hierarchy utilizes only SDV's and DDV's when estimating pictures from

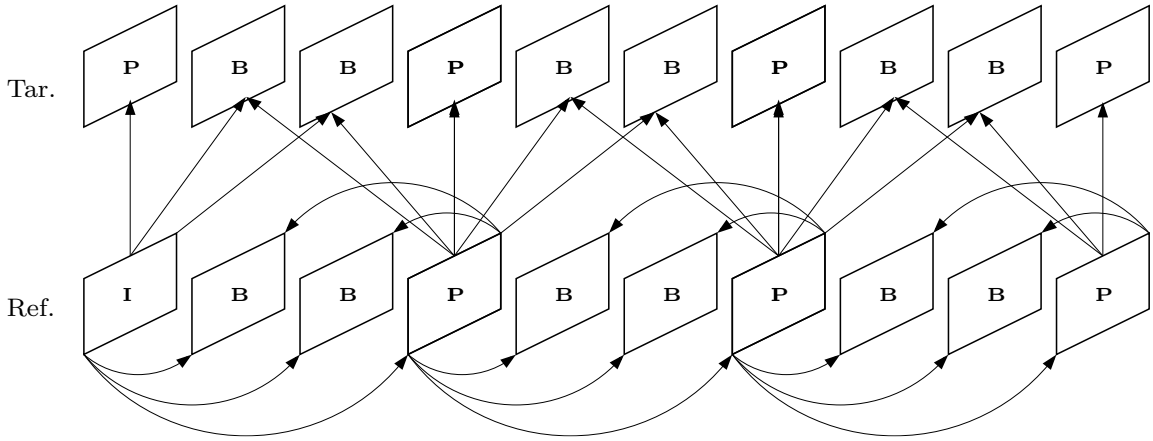


Fig. 7.1: Proposed contiguous picture hierarchy, used in stereoscopic moving-image encoding (GOP = 10)

the target stream and MV's in estimating pictures from the reference stream. Justification for this is provided shortly. The proposed picture-hierarchy can be seen in Fig. 7.1.

As previously mentioned, in an asymmetrical coding framework the HVS can comfortably perceive depth while masking out some artifacts from the target picture stream. This presupposes that the reference picture stream *has been* encoded at a higher perceptual quality than the target stream. It has also been stated that while encoding a monoscopic picture sequence, SNR-scalability is obtained from B-pictures. Evidently it is not possible to obtain both these features when using the structure shown in Fig. 6.3. B-pictures from the reference stream are used for predicting pictures from the target stream. Hence, they *must* be encoded at a fixed SNR-resolution in order to prevent drift during decoding.

In addition, an implicit benefit can be derived from this hierarchy. In [2, p 214], the authors have used the terms *base-layer* and *enhancement-layer* in conjunction with stereoscopic moving-image coding. This suggests that a user *should* have the flexibility of viewing a moving-image sequence, either in monoscopic or stereoscopic modes. The contiguous picture hierarchy shown in Fig. 7.1 satisfies this criterion. Unlike previously shown hierarchies (Figs. 6.2 and 6.3), Fig. 7.1 can have both SNR- and spatial-scalability

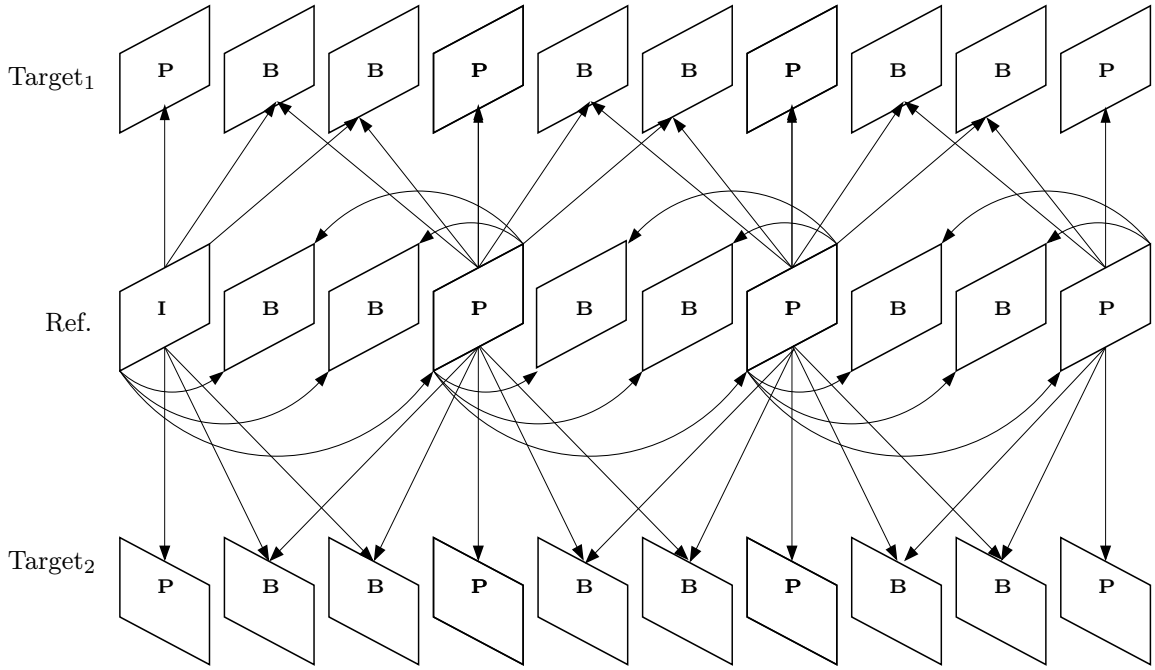


Fig. 7.2: Proposed contiguous picture hierarchy, when used in multi-view (i.e., more than 2 views) moving-image encoding (GOP = 10)

when viewed in monoscopic or stereoscopic modes.

From Fig. 2.1 and Fig. 2.6 it can be seen that DDV-estimation requires the largest search area. The complexity encountered at this stage (i.e., scale-2) is however offset by the hierarchical nature of estimation at scale-1 and scale-0. These vectors *may not* lead to optimal residual images. However, this small limitation is overshadowed by the various advantages accrued by using this picture hierarchy. Consider Fig. 16 in [77] showing a multi-view imaging system. The proposed picture hierarchy can be similarly extended to encode pictures from other target streams in a manner shown in Fig. 7.2.

Hence, in summary it can be stated that using the above picture hierarchy:

- Asymmetrical coding of stereoscopic or multi-view imagery is possible,
- SNR-scalability can be obtained from both reference and target picture streams, and

- Problems due to drift, associated with other similar picture hierarchies, are avoided during decoding.

7.2 Design characteristics of the proposed codec

Having established a contiguous picture hierarchy, the following discussion explains design characteristics of the proposed codec structure. The reader is also directed to Chapter 5 for terms, notations and encoding parameters described in this section.

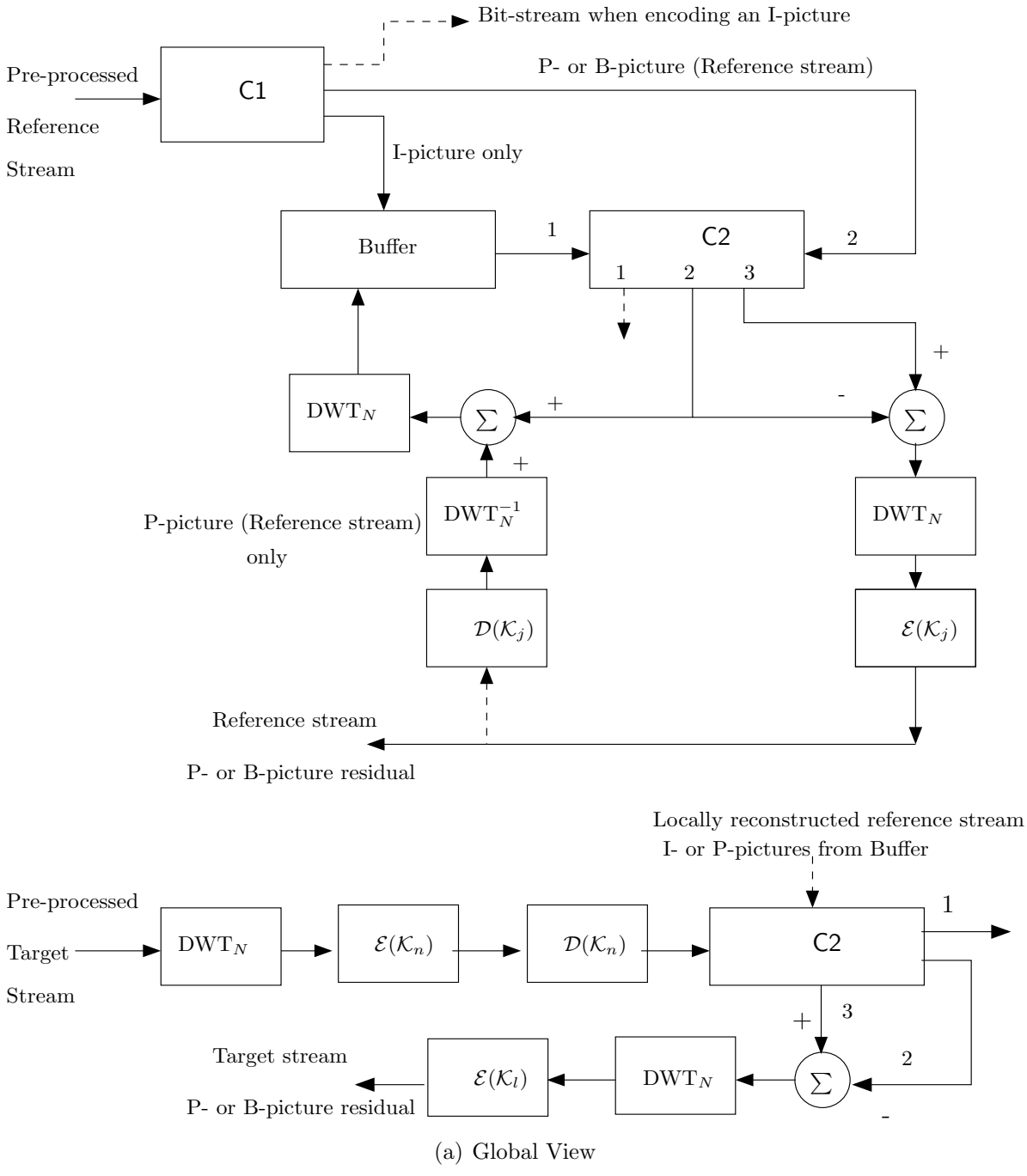
7.2.1 SNR-scalability

From Fig. 7.1, it can be observed that locally quantized I- and P-pictures from the reference stream, are used for predicting future, past or current pictures from both reference and target streams. Hence these I- and P-pictures need to be stored in *buffers*. A P-picture from the reference stream is used for predicting (at least) ten pictures while a corresponding I-picture is needed to predict six pictures. At least one buffer is required for each level of spatial scalability. In the proposed algorithm, as three levels of spatial scalability are computed, a minimum of three buffers are required for proper synchronization of pictures during encoding as well as decoding.

A generalized structure for encoding different pictures of a stereoscopic moving-image sequence, subject to the hierarchy of Fig. 7.1, can be seen in Fig. 7.3(a). The following paragraphs explain the performance of this structure when encoding moving-image sequences in monoscopic as well as stereoscopic modes. It should be emphasized that the images are encoded at the highest spatial resolution (i.e., absence of spatial-scalability).

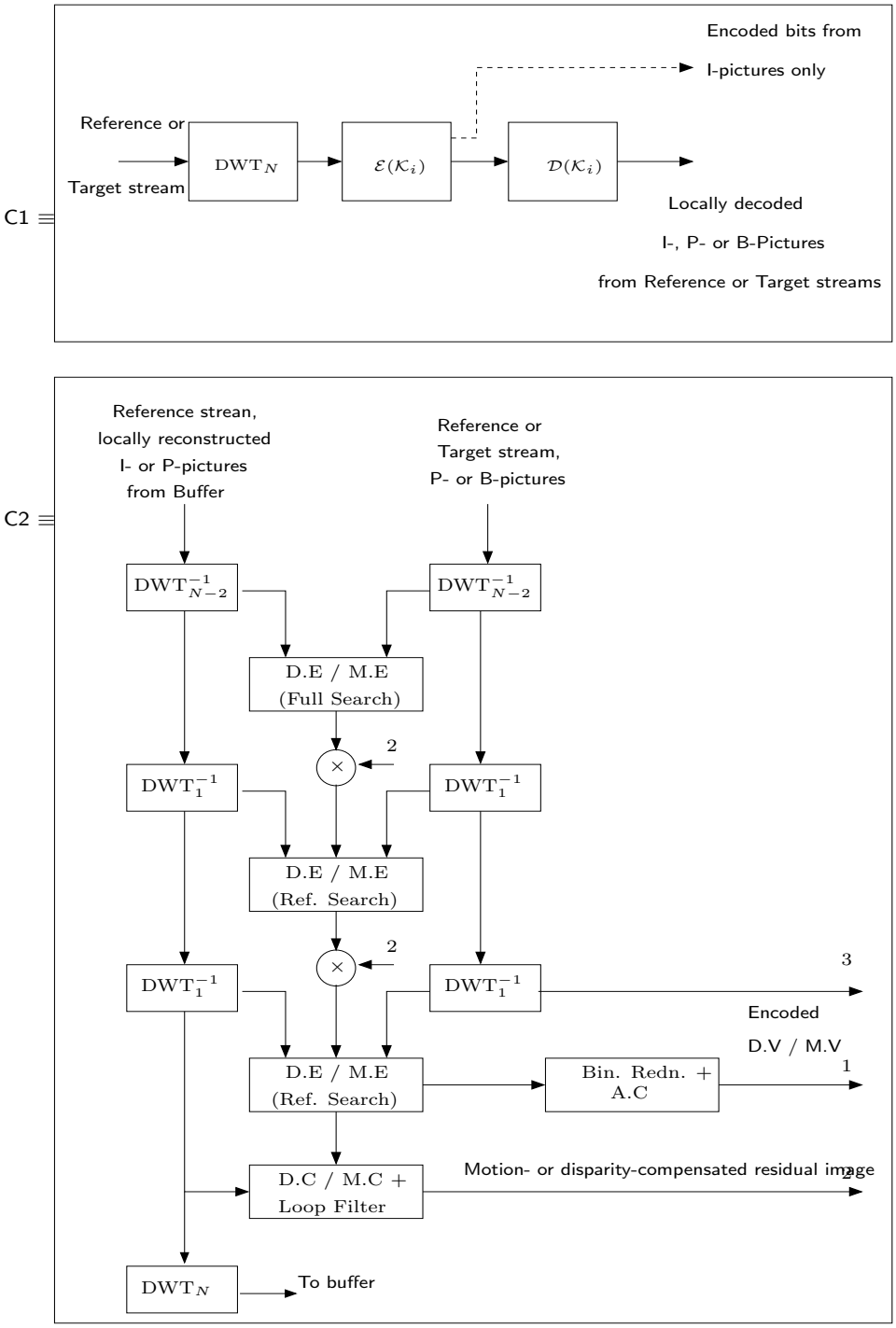
Monoscopic mode

As per the picture hierarchy shown in Fig. 7.1, an I-picture is encoded using an adaptively scanned wavelet difference reduction (ASWDR) technique. This is indicated as $\mathcal{E}(\mathcal{K}_i)$ in Fig. 7.3(b), where $\mathcal{K}_i$ indicates the bit-rate expended in generating the bit-stream.



(a) Global View

Fig. 7.3: Fundamental Structure employed when encoding different pictures of a stereoscopic moving-image sequence at the highest spatial resolution.



(b) Expanded views of **C1** and **C2** from Fig. 7.3(a) with additional components.

Fig. 7.3: cont.

Next, a locally decoded version of this image at the same bit-rate $\mathcal{K}_i$ is generated and stored in the buffer. This is indicated as $\mathcal{D}(\mathcal{K}_i)$ in Fig. 7.3(b).

The process described above is repeated when encoding the P-picture. However, bits generated by intra-coding of this image are not transmitted to the bit-stream. Instead, a 3-scale hierarchical variable-block based motion estimation is performed between this image and previously decoded I- or P-pictures. Scale-0 motion vectors (indicated as 1 in block labeled as C2) are transmitted to the output bit-stream. In addition, bits generated from encoding a motion-compensated residual image are also transmitted. An ASWDR algorithm is used for this purpose. This is indicated as $\mathcal{E}(\mathcal{K}_j)$ where $\mathcal{K}_j$ indicate the bit-rate expended in generating these bits.

Next, bits allocated for motion-compensated residual images are reconstructed at the same bit-rate $\mathcal{K}_j$ (indicated as $\mathcal{D}(\mathcal{K}_j)$) and added to the motion-compensated image (indicated as 2 in Fig. 7.3(a)). An N-level DWT is performed on this image and transferred to the buffer. It should also be pointed out that an N-level DWT is performed on the I-picture (shown in Fig. 7.3(b)) and stored in the buffer so that it can be used for predicting future pictures.

The process used for predicting P-pictures is repeated for B-pictures as well. As shown in Fig. 7.1, these pictures can be estimated from *previously encoded I-pictures or P-pictures or from both*. At this point, it is duly acknowledged that using both I- or P-pictures can produced improved motion compensated B-pictures [2]. SNR-scalability of the moving-image sequence in a monoscopic mode is thus determined by the coding performance of these B-pictures. As no pictures are estimated from these pictures, a bit-rate $\mathcal{K}_m$ can be chosen to encode them. Typically, this rate is much lower than either $\mathcal{K}_i$ or $\mathcal{K}_j$.

Stereoscopic mode

In proposing the novel picture hierarchy in Fig. 7.1, it was stated that a user should have a capability to view a moving image sequence either monoscopically or stereoscopically. This statement implies that *stereoscopic mode of viewing is always preceded by its monoscopic counterpart*. Eventually, this boils down to designing an efficient multiplexing and de-multiplexing system (as indicated in Chapter 1).

The process used in estimating reference stream P-pictures is repeated. This is observed from Fig. 7.3(a) and is very similar to the structure used in encoding stereoscopic still-images (described in Chapter 5). The only constraint imposed in the proposed codec structure is that $\mathcal{K}_n$ *should be approximately equal* to the overall bit-rate of a corresponding picture from the reference stream (e.g., when an I-picture is used for prediction $\mathcal{K}_n = \mathcal{K}_i$). As these estimated target stream pictures are not used for predicting future pictures, any suitable bit-rate $\mathcal{K}_l$ can be chosen for encoding generated disparity compensated residual images.

Hence, the final bit-stream consists of data alternating between pictures from both streams. This is explained in the next part of this sub-section.

7.2.2 Encoding color stereoscopic moving-images

When processing color images, disparity- or motion-vectors generated for the Y-component can be used to estimate Cb- and Cr-components as well. This has been explained in Chapter 5 wherein, disparity (or motion) vectors at scale-1 should be available to the decoder. This is required as the Y-, Cb-, and Cr-components follow the 4:2:0 sampling structure (explained in Chapter 5). A single file stream is generated for all components. This is shown in Fig. 7.4. As shown in the file stream, information from reference stream pictures are followed by residual bit-streams from target pictures. This enables the user to seamlessly *alternate* between monoscopic and stereoscopic modes during decoding. As

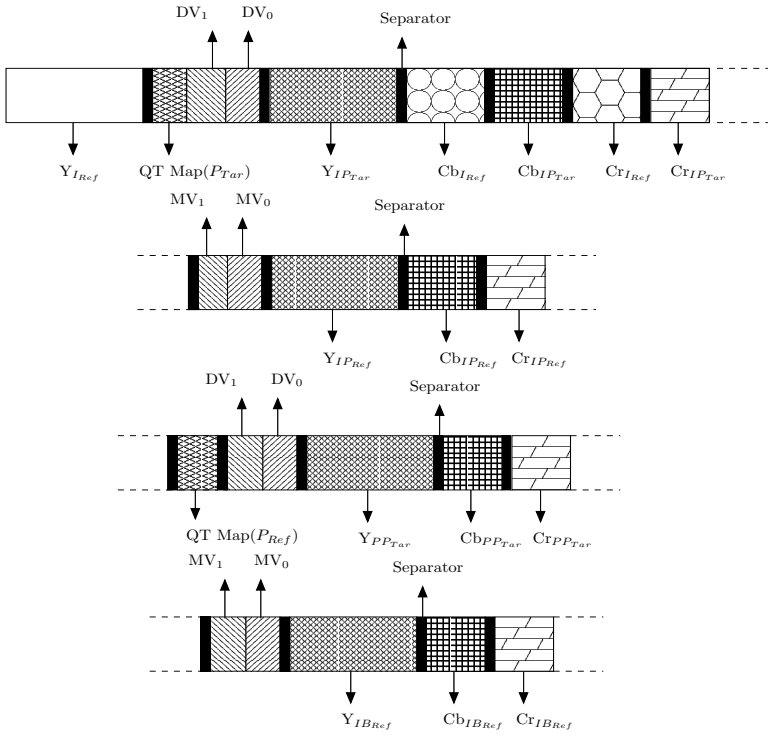


Fig. 7.4: Bit-streams of various components when encoding a stereoscopic moving-image sequence at a specific spatial resolution. e.g., $Y_{IP_{Tar}}$ indicates the Y-component of the disparity compensated residual image between a reference I-picture and a target P-picture. QT Map(.) indicates the quadtree map for the picture that is being estimated. DV_X, MV_X indicates disparity- and motion-vectors as per notations previously introduced in Chapter 2.

per previously indicated constraints (Chapter 5), all bits allocated for reference stream P-pictures *must* be decoded, before any further bits are decoded. This insures removal of artifacts arising due to drift during decoding.

The file stream shown in Fig. 7.4 represents decoded pictures at a single spatial-resolution (in this case at the highest spatial resolution). Using a similar strategy to that discussed in Chapter 5, discrete levels of spatial-scalability can be obtained when encoding these pictures. This is discussed in the following sub-section.

7.2.3 Spatial-scalability

From [2, Table 6.8, p 211], a description of various profiles employed in scalable MPEG-2 video coding can be found. Assume high and low spatial resolutions of 1440×1152 and 352×288, with an intermediate resolution of 704×576. In a modern perspective these can correspond to spatial resolutions of video content distributed over the internet (low), SDTV (intermediate) and HDTV systems (high), respectively. As indicated in Chapter 6, previous techniques (e.g., [22]) have relied on explicit downsampling or quantization of pictures before generating separate bit-streams for each resolution. This becomes redundant if these spatial resolutions have a dyadic relationship between them.

In Chapter 5, a method has been proposed to exploit the dyadic subsampling structure of a 2-D separable DWT, in order to obtain discrete levels of spatial-scalability when encoding stereoscopic still-images. A hierarchical search strategy is employed to reduce the computational complexity of a FS estimation algorithm. It can be seen from Fig. 7.3(b) that disparity- or motion-compensation is performed at the highest spatial resolution (i.e., scale-0) before a residual image is generated and subsequently encoded. Such an embedded encoding can be performed at scale-2 and scale-1 as well. This is similar to the concept shown in Fig. 5.2(b). The only additional feature would be generation of reference-stream P-pictures at every scale.

Assume that a residual image has been generated at scale-2, encoded and subsequently decoded at a bit-rate of $\mathcal{R}_1$. This locally decoded residual image can be used for encoding a residual image at scale-1. This would entail the following steps:

- **Step 1**: Generating a residual image at scale-1.
- **Step 2**: Performing a 1-level DWT on the residual image obtained from Step 1.
- **Step 3**: Subtracting the residual image, obtained at scale-2, from the all low-pass subband of the wavelet transformed image from Step-2.

- **Step 4**: Further transforming this reduced energy all low-pass subband.
- **Step 4**: Encoding (and locally decoding) this, reduced energy, residual image from scale-1 at a bit-rate of $\mathcal{R}_2$.

The above steps can be repeated when encoding residual images at scale-0. The implicit sub-sampling nature of a DWT facilitates obtaining discrete levels of spatial-scalability on the reference image. Due to problems associated with drift, this operation cannot be implemented in a straightforward manner on target images (which in this case might be P- or B-pictures from reference or target streams). The steps outlined above alleviate this problem. The file stream shown in Fig. 7.4 can be appropriately modified to incorporate additional information from various scales. Bits earmarked for scale-2 and scale-1 images *must* be decoded before any bits allocated for scale-0 images are decoded. In addition, residual image bit-streams follow any disparity or motion vectors generated at a particular scale.

7.3 Results and analysis

A quadtree partition, discussed previously in Chapter 5, is effected on the all low-pass subband of target images at scale-2. The following results illustrate performance of the proposed algorithm, when encoding monochromatic as well as color images. Lack of suitable reference algorithms makes it impossible to provide any comparative results. However, wherever applicable, qualitative discussion is provided that justifies the superiority of the proposed algorithm when compared with existing counterparts.

7.3.1 Experimental results with monochrome images

Experimental results presented in Sec. 5.5.3 justify the superior performance of the proposed algorithm when encoding I- and target stream P-pictures. PSNR results are presented when encoding

- Reference stream P-pictures from I-pictures and
- Target stream B-pictures from I-pictures. In this instance only forward prediction is considered.

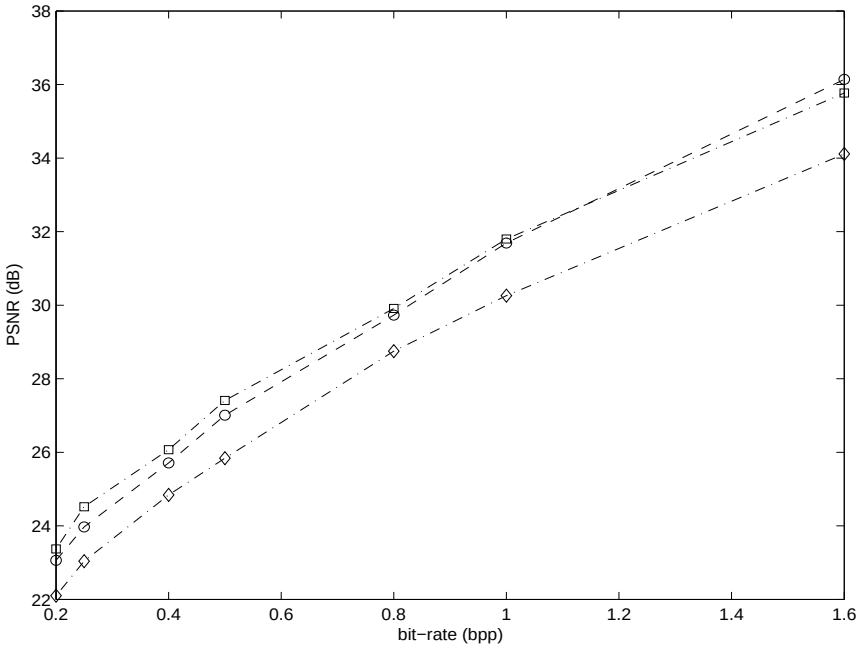
Four consecutive pictures from the "*redcar*" stereoscopic moving-image sequence are considered. Image dimensions are 704 × 576. Other factors used in this simulation are similar to that used in encoding stereoscopic still-images. The reader is directed to Chapter 5 for exhaustive details. A summary of various parameters are presented as follows:

- "CDF-9/7" wavelet-filters are used for transforming pictures,
- 5-levels of wavelet decomposition,
- Smoothing parameter $\lambda = 1.35$, with two filter iterations,
- Quadtree-partitioning at scale-2, with threshold $V_t = 120$, followed a 3-scale hierarchical variable-block-based disparity and motion estimation,
- I-pictures encoded (and locally decoded) at a "high" bit-rate of 2.5 bpp.

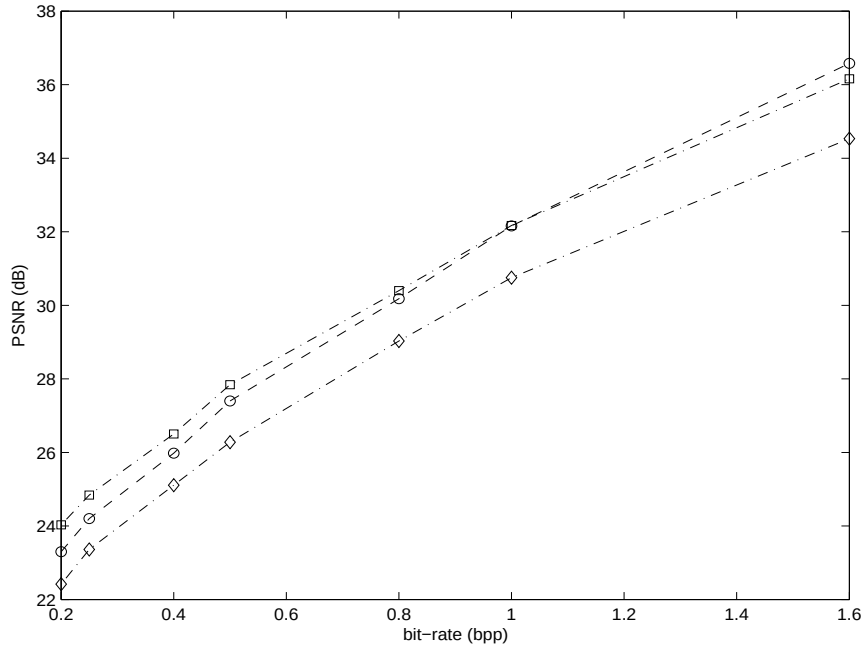
Fig. 7.5(a) indicates PSNR values when encoding reference image P-pictures, predicted from I-pictures only. Fig. 7.5(b) depicts values when encoding target stream B-pictures, predicted only from I-pictures. These results pertain to Y-components of images. Evidently, the proposed algorithm when used with a loop-filter outperforms independent coding of residual images. However at high bit-rates, independent coding[1] of these residual images is advantageous. It should be stressed that these PSNR values are completely dependent on image content.

PSNR values when encoding Y-components of these pictures in an embedded mode

[1]Otherwise known as intra-picture (analogous to intra-frame coding in MPEG-2) coding

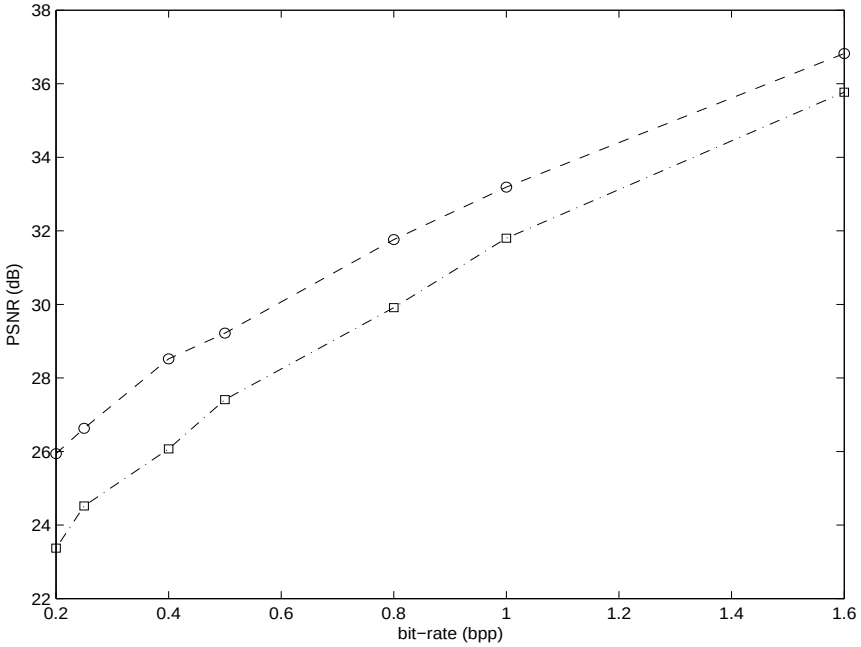


(a) Reference stream P-picture predicted from an I-picture

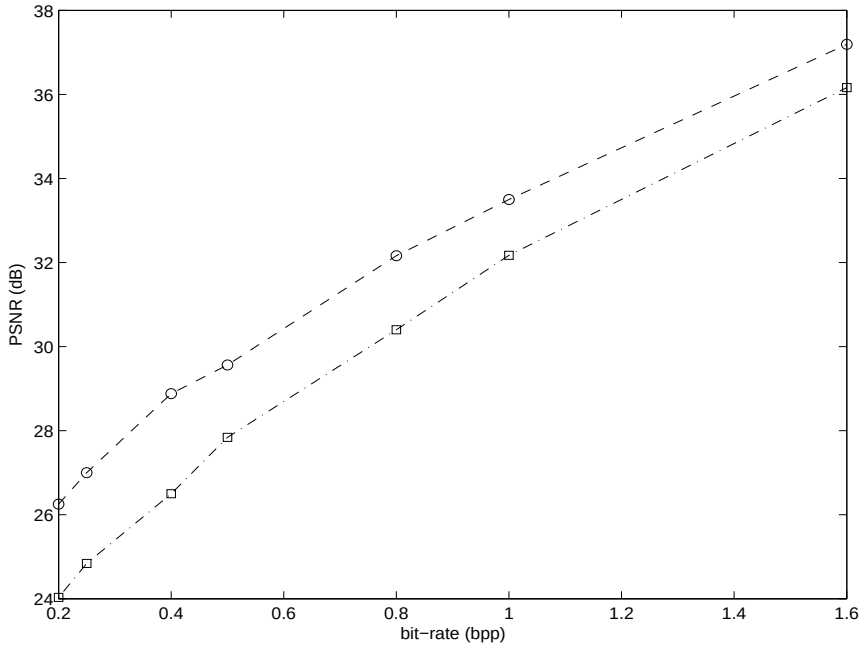


(b) Target stream B-picture predicted from an I-picture

Fig. 7.5: PSNR plots when encoding motion- and disparity compensated residual images with loop-filtering (□), independent ASWDR coding (○) and without loop-filtering (◇) in an independent simulcast mode. Image dimensions are 704×576.



(a) Reference stream P-picture predicted from an I-picture



(b) Target stream B-picture predicted from an I-picture

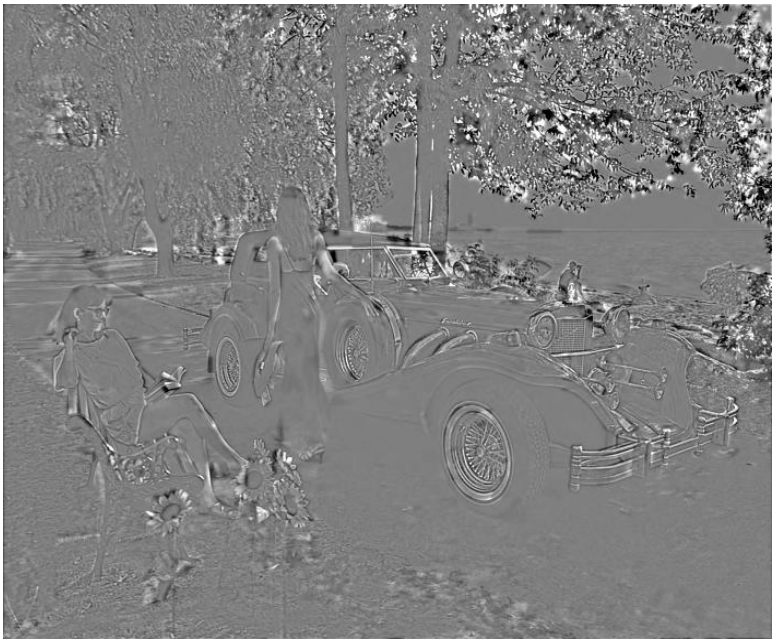
Fig. 7.6: Comparative PSNR plots when encoding motion- and disparity compensated residual images in independent (□) and embedded simulcast modes (○). Image dimensions are 704×576.

are shown in Fig. 7.6. In order to achieve this, residual images at scale-2 and scale-1 are encoded (and locally decoded) at *high* bit-rates. This is to provide an *unbiased* framework for distribution of bits, and is exactly similar to the strategy proposed in Sec. 5.5.3. An empirical value of 2.0 bpp has been chosen for these simulations. This is purely speculative and depends entirely on the content of pictures being analyzed. The superiority of an embedded when compared with an independent simulcast mode can be appreciated by observing Figs. 7.7 and 7.8.

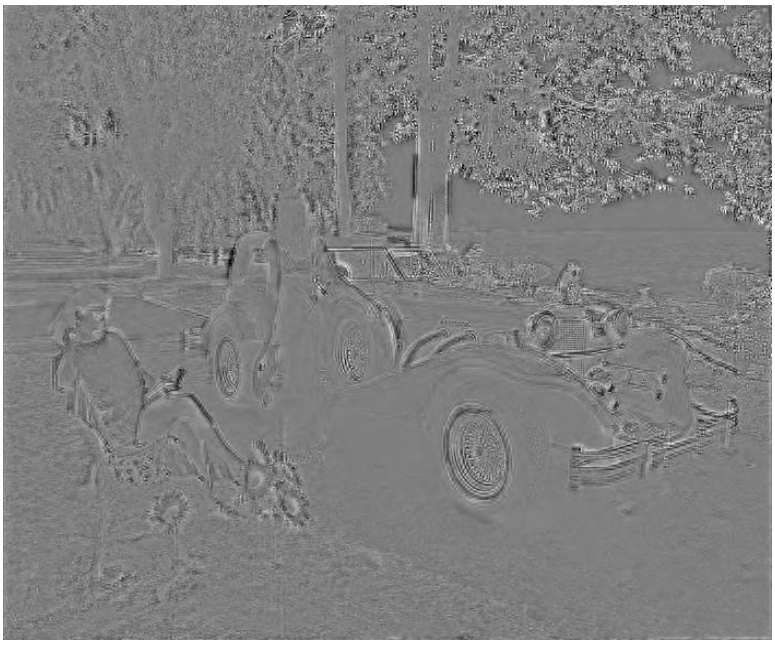
Evidently, Figs. 7.7(b) and 7.8(b) contain less energy content than their independent simulcast counterparts Figs. 7.7(a) and 7.8(a). This also justifies the PSNR values shown in Fig. 7.6. Similar to their disparity compensated counterparts, motion-compensated residual images contain partition-artifacts. In an embedded coding framework, these regions require a large number bits to be encoded. As such this degrades the quality of image reconstruction and is indicated by the results shown in Fig. 7.5. The EPNR filter proposed in Sec. 5.2, effectively reduces these artifacts thus improving the quality of decompressed images. It should be emphasized that these images have been scaled for display purposes whereby modified pixel values lie between [0,1].

7.3.2 Informal results when encoding color stereoscopic moving-image sequence

Contiguous stereo image pairs from the "*redcar*" sequence, having dimensions 704 × 576, are used in this experiment. To simplify the encoding process, forward-prediction is only used as per the structures shown in Fig. 7.9. Psycho-visual analysis has revealed that the HVS is very sensitive to changes in the Y-component of an image [65]. It is less sensitive to large perturbations in the Cb- and Cr-components. Hence in encoding color images, a ratio between various components of an image need to be specified. This is very similar to the formulation for stereoscopic still-images, previously explained in

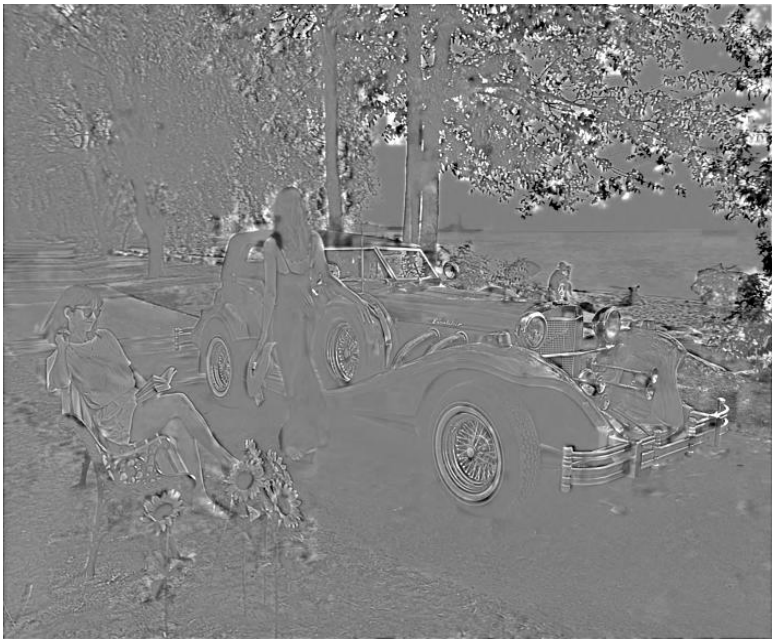


(a) I- and P- pictures, Independent Simulcast Mode

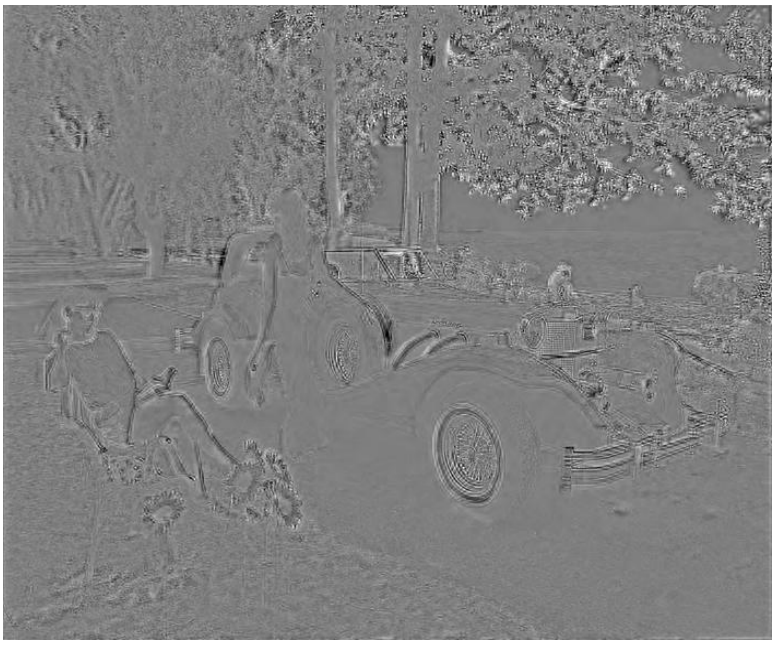


(b) I- and P- pictures, Embedded Mode

Fig. 7.7: Residual images when encoding P-pictures from I-pictures. Image dimensions are 704×576. Images have been scaled for display purposes.

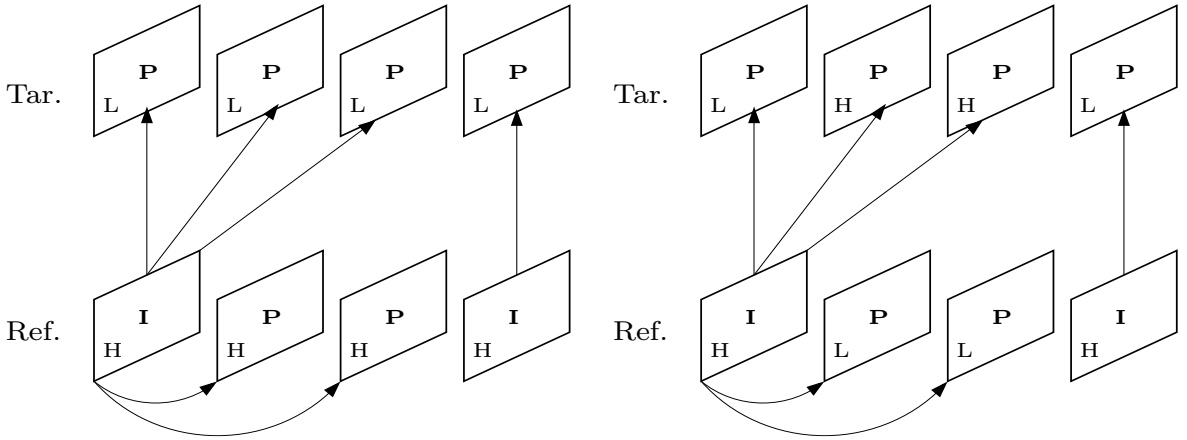


(a) I- and B- pictures, Embedded Simulcast Mode



(b) I- and B- pictures, Embedded Mode

Fig. 7.8: Residual images when encoding B-pictures from I-Pictures. Image dimensions equals 704×576. Images have been scaled for display purposes.



(a) Regular asymm. coding (b) Interleaved asymm. coding

Fig. 7.9: Modified asymmetrical coding frameworks for stereoscopic moving-images. "H" and "L" indicates overall high and low bit-rates when encoding reference and target pictures.

Chapter 5. As defined in that chapter, let $\mathcal{K}_Y$ indicate the (overall) bit-rate allocated for a Y component of an image. Other parameters are defined as follows:

- A 4:2:0 sampling structure for the Y, Cb and Cr components,
- $\mathcal{K}_Y$:$\mathcal{K}_{Cb}$:$\mathcal{K}_{Cr}$ = 80:1:1, with $\mathcal{K}_Y$ = 2.0 bpp when encoding I-pictures,
- $\mathcal{K}_Y$:$\mathcal{K}_{Cb}$:$\mathcal{K}_{Cr}$ = 128:1:1, with $\mathcal{K}_Y$ = 0.8 bpp when encoding pictures marked as "L",
- $\mathcal{K}_Y$:$\mathcal{K}_{Cb}$:$\mathcal{K}_{Cr}$ = 128:1:1, with $\mathcal{K}_Y$ = 1.6 bpp when encoding pictures marked as "H", except for I-pictures,
- 5- and 4-levels of wavelet decomposition for luminance and chrominance components, respectively,
- $\lambda = 0.85$, with two filter iterations,
- Quadtree-partitioning threshold, $V_t = 120$ at scale-2 and
- Independent simulcast mode picture coding.

Three sequences (having 28 contiguous stereo image pairs) and encoded at 25 fps were generated in this simulation. Specifications for these are as follows:

- "*Sequence-A*" : This was generated as per an interleaved asymmetrical coding , shown in Fig. 7.9(b),
- "*Sequence-B*" : This was generated as per regular asymmetrical coding, shown in Fig. 7.9(a) and
- "*Sequence-Raw*": This was a raw image sequence consisting of uncompressed pictures

These sequences were then viewed, informally, by four test subjects experienced in viewing stereoscopic images. These sequences were viewed on a CRT display. A general consensus amongst these viewers was that they were not able to differentiate between sequences "*Sequence-A*" and "*Sequence-B*". However with respect to "*Sequence-raw*" they were, generally, able to identify slight changes in color (primarily on the chassis of the car). This is expected as bit-rates for chrominance components in these sequences are very small. Notwithstanding this, the subjects were not able to determine perceptible color-bleeding in either sequence. It should be worth mentioning that the question of visual comfort was not posed during this subjective testing. In order to analyze these results, two separate discussions are necessary.

7.3.3 Sequences when viewed in monoscopic mode

In Chap. 5, it was stated that the HVS is more sensitive to perturbations in low-frequencies of an image than in higher frequencies. However, this reasoning cannot be applied in isolation to color-images. From psycho-visual experiments it has been proved that the HVS is more sensitive to changes in the Y-component of an image [65] when compared with Cb- or Cr-components. As such, PSNR cannot be used as an objective

function to evaluate performance of color-images [83].

At very low bit-rates for chrominance components, experienced viewers were not able to detect any significant changes in encoded images. Recent research has emphasized the design of *perceptual metrics* in context of moving-image coding (e.g., JNDmetrix® by Sarnoff corporation [84] and standards evolved by the *video quality experts group* (VQEG) [85]). Due to the limited scope of this topic in the context of the overall research work of this thesis, extensive subjective results are not presented with respect to color moving-image encoding. The values used in the above simulation are purely speculative in nature. With a high bit-rate allocated for the Y-component it can (generally) be inferred that chrominance components eventually determine the compression ratio of any encoding technique.

7.3.4 Sequences when viewed in a stereoscopic mode

Preliminary results with respect to encoding color stereoscopic still-images have been presented in Sec. 5.5.4. From these, it can be inferred that target images can be compressed at lower perceptual qualities (i.e., lower SNR-resolution) than reference images. A straightforward extension to this technique can be applied to moving-image coding. The reference stream *must* be encoded at a "high" bit-rate with respect to the target stream.

In [23] it was argued that when exposed to asymmetrically coded stereo image pairs (e.g., Fig. 7.9(a)) the HVS *may* experience visual fatigue. In order to rectify this problem, the authors proposed a *temporal interleaving* process. It involved switching the reference and target image views at scene cuts. In other words, at a particular scene cut, the relative perceptual qualities of both streams were interchanged. The authors in [23] imposed some constraints in applying such a technique. Independent coding of both streams was proposed with an *a priori* Gaussian blurring of the target image stream.

They justified this by arguing that such low-pass filtered images can be encoded at "high" compression ratios.

The limitations of this framework were presented in Chapter 4. It was stated that such a framework may lead to biased results when estimating disparity between the reference and target streams. Secondly, there are occurrences of stereo-image sequences without any scene cuts. Notable examples would be robot vision in surveillance and mining, medical images in remote surgery applications and some forms of teleconferencing. If the hypothesis of visual fatigue is applied then the principle of scene-cuts, applied by the authors in [23], to interchange the relative qualities of both streams cannot be sustained in these examples.

On the other hand, the proposed temporal interleaving scheme is not limited by these constraints. "*Sequence-A*" was found to be indistinguishable from "*Sequence-B*". The authors in [23] have stated that when viewing stereo image sequences having a temporal interleaving structure shown in Fig. 7.9(b), observers were able to perceive "jerky" motion between images. "Jerky" motion in this context may be defined as the ability of the HVS to perceive distortions in contiguous pictures of a stereoscopic moving-image sequence. Limited results, derived from this informal testing, do not support this observation. The HVS, in effect, tries to mask out imperfections from both image streams. If the time instants between successive pictures in a sequence is large then it is possible for the HVS to perceive "jerky" motion. No inference about threshold limits, however, can be derived from these simulations. Extensive subjective testing is necessary from a large subject pool, involving a variety of stereoscopic image sequences.

However, the observation by the authors in [23] that temporal interleaving at points other than scene-cuts would result in "jerky" motion have been challenged by results obtained from this informal subjective testing. Pictures from the target stream have not been uniformly blurred. Instead, coding artifacts and low-pass filtering effects of the pro-

posed loop-filter forms a major part in target image quality degradation. In other words, local rather than global degradation is achieved. Due to the nature of the ASWDR encoding process, this degradation is predominant in high-frequency regions of an image rather than low-frequency regions. As previously mentioned, the HVS is less sensitive to perturbations in high-frequency regions. Hence, distortions in these regions are masked out by the HVS when viewed simultaneously with perceptually higher-quality regions from the reference view. Another advantage of arbitrary temporal interleaving[2] process is that both streams can be independently viewed in a monoscopic mode, with high levels of SNR scalability. This has been explained in an earlier part of this chapter.

[2]Fig. 7.9(b) is a representative example of such temporal interleaving

Chapter 8

Conclusion and Future Work

8.1 Summary of proposed algorithm

IN this thesis, a novel algorithm has been proposed for encoding stereoscopic still-images. With suitable modifications, this algorithm has been extended to encode stereoscopic moving-images as well. The following sub-sections summarizes the algorithm, when encoding both stereoscopic still and moving images.

8.1.1 Stereoscopic still-image coding

In Chapter 4, two state-of-the-art algorithms are discussed. Boulgouris and Strintzis [11] propose a closed-loop structure to encode disparity compensated residual images. This has been shown to be better than its open-loop counterpart, as per the discussion presented in their paper, as well as in Chapter 4. On the other hand, Frajka and Zeger conjectured that mere generation of reduced energy, residual images does not necessarily guarantee improved stereoscopic image coding results. Instead, they exploit the unique nature of disparity compensated residual images, a fact previously presented by Mollenhoff and Maier [14]. Frajka and Zeger suggest using an algorithm that is able to encode high-frequency and edge information prevalent in residual images. A multi-grid embedding of wavelet coefficients (MGE), used by them [12] provided superior results when compared with a zerotree algorithm, used by Boulgouris and Strinzis's algorithm [11].

As discussed in Chapter 3, the MGE algorithm relies entirely on intra-scale correlation to encode images. An adaptively-scanned wavelet-difference-reduction (ASWDR) algorithm is instead proposed to encode wavelet transform images. This exploits both intra- as well as inter-scale correlation when encoding wavelet-transformed images. This algorithm is used as an embedded image encoder, instead of the MGE algorithm proposed by Frajka and Zeger. In addition, a closed loop structure for disparity compensated residual image generation is used.

Shukla and Radha have shown [17] that variable-block-based disparity estimation schemes outperform their fixed-block-based counterparts in a rate-distortion (R-D) framework. This motivates the use of a variable-block-based disparity estimation scheme in the proposed algorithm. This is in contrast to the fixed-block-based schemes reported in [11] and [12]. As demonstrated in Chapter 5, partition-artifacts must be eliminated in disparity compensated images, prior to generation of residual images. Due to uneven block sizes, loop-filtering is used instead of the currently used overlapped-block disparity compensation (OBDC) scheme [18].

Finally, the hierarchical search technique used in disparity-estimation facilitates spatial-scalability when encoding such stereo image pairs. Advantages of embedded techniques, when compared with their independent simulcast counterparts, are also shown in Chapter 5. Hence the algorithm, proposed in Chapter 5, can be summarized as having the following features:

- ASWDR algorithm to encode and decode reference and residual images,
- Variable-block-based disparity estimation,
- Hierarchical search strategy to estimate disparity vectors and provide scope for spatial-scalability and

- An edge-preserving noise-reduction (EPNR) filter to minimize partition artifacts in disparity compensated images

8.1.2 Stereoscopic moving-image coding

The above algorithm has been extended to encode time-varying stereoscopic imagery. Chapter 7 details this implementation. This is made possible by the novel picture hierarchy shown in Fig. 7.1. This removes the inherent limitations of current picture hierarchies that are discussed in Chapter 6. Use of a closed-loop structure insures drift-free target picture reconstruction, during decoding. This is a departure from traditional open-loop structures proposed in [76, 20] and [75]. In addition, the picture hierarchy shown in Fig. 7.1 eliminates the redundant nature of the algorithm shown in Fig. 6.5. Motion vector estimation is not performed for target pictures. Instead, these pictures are obtained by displaced-disparity-vector (DDV) estimation. Principles of DDV estimation have been discussed in Chapter 2.

As with its still-image counterpart, spatial-scalability can be obtained when encoding stereoscopic moving-images. This is possible due to the hierarchical search strategy used in disparity- and motion-estimation. In addition, the use of the novel loop-filter insures that partition artifacts are minimized to a large extent, thus improving peak signal-to-noise ratio (PSNR) values of reconstructed images. Finally, temporal interleaving at arbitrary time-instants have been proposed when encoding stereoscopic moving-images in an asymmetrical coding framework.

8.2 Summary of original contributions made in the thesis

The original contributions that have been made during the course of this thesis are highlighted as follows:

- In [12], a MGE algorithm was proposed to encode disparity compensated residual images. In Chapter 3, the limitations of using this algorithm has been shown. A perceptually superior ASWDR algorithm was instead proposed to encode, natural as well as residual images. This algorithm is effective in exploiting intra- as well as inter-scale correlation amongst wavelet coefficients. However, there exists room for further improvement in this algorithm. This is discussed in the next section.

- An EPNR filter, originally used to clean images [19] corrupted with Gaussian noise, has been adapted in this algorithm as a loop filter. The current standard in generating disparity- or motion-compensated images is an overlapped-block-compensated technique [18, 63]. As shown in Chapter 5, this algorithm is restricted to fixed block-based disparity- or motion-estimation techniques only. Due to its lack of region-size dependency, an EPNR filter can be used to smooth compensated images obtained from arbitrary region-based estimation schemes. Hence, this can be efficiently incorporated in current MPEG-4 coding standards where object scalability is a desired feature.

- In [11, 12] and other literature surveyed in this thesis, it has been reported that disparity estimation is performed between reference and target images at full perceptual quality. This is generally true when encoding still-images. However, this scenario is not valid in some instances; e.g., estimating a target P-picture from a previously encoded reference P-picture. Hence it was proposed to locally quantize reference and target images (Fig. 5.1) before estimating disparity or motion vectors. Qualitatively speaking, this produces unbiased results during estimation.

- A scheme has also been presented to obtain discrete levels of spatial scalability. This is possible by exploiting the inherent nature of hierarchical search strategies in motion- and disparity-vector estimation. As discussed in Chapter 7, differ-

ent spatial resolutions in MPEG-2 video coding profiles (generally) have a dyadic relationship between them. Unlike current state-of-the-art schemes, no explicit downsampling is necessary [22] to encode images at different spatial resolutions.

- Current standards for encoding stereoscopic moving-images [21] have scope for limited SNR-scalability. As shown in Fig. 6.2, every picture in a sequence is used for predicting current, future or past pictures. This limits SNR-scalability when decoding such pictures. Consequently, a new picture hierarchy has been proposed in this thesis. This can be seen from Fig. 7.1. Target pictures in this hierarchy are predicted *only* from their reference view counterparts. When using this structure for decoding users have the flexibility of viewing the sequence in both monoscopic and stereoscopic modes, without sacrificing SNR-scalability. As shown in Fig. 7.2, this picture hierarchy can be extended to encode pictures from multi-view (i.e., more than two views) imaging systems while preserving useful features of the aforementioned algorithm.

- Modifications in existing temporal-interleaving schemes for viewing asymmetrically coded stereoscopic moving-image data [23] have been proposed. The present scheme is only valid for sequences having scene-cuts. In this algorithm, the limitation of scene cuts has been removed. This insures that temporal-interleaving can be achieved in sequences without scene-cuts (e.g., telemedicine applications).

- A limited discussion, based on informal subjective testing, has also been provided for encoding color stereoscopic images (still and moving). This is possible by exploiting psychovisual characteristics of the HVS. Results from Tables 5.14 and 5.15 indicate that chrominance components of target images can be quantized very coarsely. Notwithstanding this, test subjects with experience in viewing stereoscopic images were unable to perceive color-bleeding and coding-artifacts when

viewing these images in a stereoscopic mode (along with perceptually higher quality reference images). The same subjects were also not able to detect "jerky" motion (i.e., change in perceptual quality) between successive images of a stereoscopic moving-image sequence.

8.3 Scope for future research work

The research work presented in this thesis advances the concept of stereoscopic image coding (still and time-varying). This has led to the identification of additional topics that can form part of future research work. These are highlighted as follows:

- In Chapter 1, an ideal scenario of an error-free transmission channel is assumed. For all practical purposes this *cannot* be assumed when transmitting data over noisy channels (e.g., wireless networks). As a result, source codes generated from the encoding process described in Chapters 5 and 7 must be "protected" adequately before transmission can take place. In this regard, current work is focused on protecting codes generated from zerotree-based techniques. More information of these techniques can be found in an excellent review paper by Wang et. al. [86]. Being a relatively new technique, research work needs to be undertaken to generate *error-resilient* codes when using an ASWDR algorithm. One approach would be to output "*markers*" in both dominant and refinement scans. These can be output when scanning between different subbands (e.g., after the last coefficient of $\boldsymbol{c_{20}}$ and the first coefficient of $\boldsymbol{d_{21}}$). The arithmetic coding framework, used in the present scheme, can be improved to included localized context-based modeling of coefficients. Details of some of these proposals can be found in [73]. It is hoped that improved arithmetic coding coupled with error-resilient code generation will make an ASWDR image coding scheme competitive with current techniques.

- Quadtree-partitioning is not an efficient approach to segment textured and non-textured regions of an image. There exists other partitioning techniques [87] that can be effected on images that are more efficient. This is primarily an *a-posteriori* approach in segmenting an image. The DWT is the preferred method of transforming images. An *a-priori* approach to this method can be obtained from the work by Cinkler and Mertins [88]. In this method, edge information is obtained from images before implementing a DWT. This information can be used to segment images into "*non-overlapping regions*". The authors in [88] have shown that this approach leads to improved image reconstruction in a R-D framework. In the context of stereoscopic still- or moving-image coding, these non-overlapping regions can be used for disparity or motion estimation instead of rectangular blocks generated from a quadtree-partitioning scheme. A recent algorithm utilizing this concept for monoscopic moving-image coding can be found in [89].

- In this research work, preliminary results have been presented when encoding stereoscopic color images (still and moving). In Chapter 5 it was observed that chrominance components of a target image can be represented at an extremely low bit-rate. It was also observed that the luminance component of reference images should be generally represented at higher bit-rates. A similar finding has been reported in Chapter 7, when encoding stereoscopic color moving-images. This can form the basis of a subsequent research topic. Extensive subjective analysis is needed in order to develop mathematical models, to ascertain threshold values of bit-rates for each component of both reference and target images. This information can also help in determining threshold values when implementing a temporal interleaving process. This implies that users should just be unable to perceive distinct changes in SNR-resolution of contiguous stereo-image pairs.

Appendix A

"CDF-9/7", "Odegard-9/7", "Cooklet-17/11" - Lifting Steps

Coefficients for "CDF-9/7" can be found from [25, p. 279] while those of "Odegard-9/7" can be found from the "*Wavelet Image Compression Construction Kit*" [90]. Theoretical work pertaining to the design of this filter can be found in [91]. Coefficients for "Cooklet-17/11" can be found in [92].

Lifting steps of filters used in this thesis are outlined in the following tables. The polynomial division algorithm, proposed by Daubechies and Sweldens [42] is used to generate these steps. The following pages also outlines these lifting steps. The following notations are used:

- $c_m[n]$: 1-D input signal being analyzed,
- $d_{m-1}[n]$: 1-D signal containing the detail coefficients and
- $c_{m-1}[n]$: 1-D signal containing the approximate coefficients.

Table A.1: "CDF-9/7" Analysis filter coefficients

n	$\tilde{h}_n\ z^n$	$\tilde{g}_n\ z^n$
0	0.60294901823636	0.55754352622850
±1	0.26686411844288	-0.29563588155712
±2	-0.07822326652899	-0.02877176311425
±3	-0.01686411844288	0.04563588155713
±4	0.02674875741081	

Table A.2: Lifting coefficients - "CDF-9/7"

i	$S_i(z),\ T_i(z) = a_1 z^1 + a_2 + a_3 z^{-1}$		
	a_1	a_2	a_3
$S_1(z)$	-1.586134342	-1.586134342	0
$T_1(z)$	0	-0.05298011854	-0.05298011854
$S_2(z)$	0.8829110762	0.8829110762	0
$T_2(z)$	0	0.4435068522	0.4435068522
K	1.149604398		

Lifting steps for this filter pair are as follows:

$$
\begin{aligned}
c_{m-1}[n] &\leftarrow c_m[2n] \\
d_{m-1}[n] &\leftarrow c_m[2n+1] \\
d_{m-1}[n] &\leftarrow d_{m-1}[n] + s_1[1]\ (c_{m-1}[n] + c_{m-1}[n+1]) \\
c_{m-1}[n] &\leftarrow c_{m-1}[n] + t_1[1]\ (d_{m-1}[n-1] + d_{m-1}[n]) \\
d_{m-1}[n] &\leftarrow d_{m-1}[n] + s_2[1]\ (c_{m-1}[n] + c_{m-1}[n+1]) \\
c_{m-1}[n] &\leftarrow c_{m-1}[n] + t_2[1]\ (d_{m-1}[n-1] + d_{m-1}[n]) \\
c_{m-1}[n] &\leftarrow K\, c_{m-1}[n] \\
d_{m-1}[n] &\leftarrow \frac{1}{K}\, d_{m-1}[n]
\end{aligned}
\tag{A.1}
$$

Table A.3: "Odegard-9/7" Analysis filter coefficients

n	$\tilde{h}_n\ z^n$	$\tilde{g}_n\ z^n$
0	0.78751377152779	0.81678063499211
±1	0.38697186387262	-0.44030170672499
±2	-0.09306926370358	-0.05483692690278
±3	-0.03341847327935	0.08674831613171
±4	0.05286576853296	

Table A.4: Lifting coefficients - "Odegard-9/7"

i	$S_i(z),\ T_i(z) = a_1 z^1 + a_2 + a_3 z^{-1}$		
	a_1	a_2	a_3
$s_1(z)$	-1.581932486	-1.581932486	0
$t_1(z)$	0	-0.071678341	-0.071678341
$s_2(z)$	0.825773750	0.825773750	0
$t_2(z)$	0	0.523072245	0.523072245
K	1.079383836		

Lifting steps for this filter pair are as follows:

$$
\begin{aligned}
c_{m-1}[n] &\leftarrow c_m[2n] \\
d_{m-1}[n] &\leftarrow c_m[2n+1] \\
d_{m-1}[n] &\leftarrow d_{m-1}[n] + s_1[1]\ (c_{m-1}[n] + c_{m-1}[n+1]) \\
c_{m-1}[n] &\leftarrow c_{m-1}[n] + t_1[1]\ (d_{m-1}[n-1] + d_{m-1}[n]) \\
d_{m-1}[n] &\leftarrow d_{m-1}[n] + s_2[1]\ (c_{m-1}[n] + c_{m-1}[n+1]) \\
c_{m-1}[n] &\leftarrow c_{m-1}[n] + t_2[1]\ (d_{m-1}[n-1] + d_{m-1}[n]) \\
c_{m-1}[n] &\leftarrow K\, c_{m-1}[n] \\
d_{m-1}[n] &\leftarrow \frac{1}{K}\, d_{m-1}[n]
\end{aligned}
\tag{A.2}
$$

Table A.5: "Cooklet-17/11" Analysis filter coefficients

n	$\tilde{h}_n\ z^n$	$\tilde{g}_n\ z^n$
0	0.8402696692	0.7568252267
±1	0.4090630083	-0.4226067872
±2	-0.1073757602	-0.033145604
±3	0.0533641923	0.0814830079
±4	0.0073357876	0.0082864076
±5	-0.0135767155	-0.0124296114
±6	-0.0006712263	
±7	0.0010068394	

Table A.6: Lifting coefficients - "Cooklet-17/11"

i	$S_i(z),\ T_i(z) = a_1 z^1 + a_2 + a_3 z^{-1} + a_4 z^{-2}$			
	a_1	a_2	a_3	a_4
$S_1(z)$	0	-1.5	0	0
$T_1(z)$	0.187500000	0.187500000	0	0
$S_2(z)$	0	-0.266667113	-0.266667113	0
$T_2(z)$	-0.219726335	-0.219726335	0	0
$S_3(z)$	0	0.898246006	0.898246006	0
$T_3(z)$	-0.057113653	0.395736546	0.395736546	-0.057113653
K	1.190916752			

Lifting steps for this filter pair are as follows:

$$
\begin{aligned}
c_{m-1}[n] &\leftarrow c_m[2n] \\
d_{m-1}[n] &\leftarrow c_m[2n+1] \\
d_{m-1}[n] &\leftarrow d_{m-1}[n] + s_1[1]\ (c_{m-1}[n] + c_{m-1}[n+1]) \\
c_{m-1}[n] &\leftarrow c_{m-1}[n] + t_1[1]\ (d_{m-1}[n-1] + d_{m-1}[n]) \\
d_{m-1}[n] &\leftarrow d_{m-1}[n] + s_2[1]\ (c_{m-1}[n] + c_{m-1}[n+1]) \\
c_{m-1}[n] &\leftarrow c_{m-1}[n] + t_2[1]\ (d_{m-1}[n-1] + d_{m-1}[n]) \\
d_{m-1}[n] &\leftarrow d_{m-1}[n] + s_3[1]\ (c_{m-1}[n] + c_{m-1}[n+1]) \\
c_{m-1}[n] &\leftarrow c_{m-1}[n] + t_3[1]\ d_{m-1}[n+1] + t_3[2]\ d_{m-1}[n] + t_3[3]\ d_{m-1}[n-1] + t_3[4]\ d_{m-1}[n-2] \\
c_{m-1}[n] &\leftarrow K\, c_{m-1}[n] \\
d_{m-1}[n] &\leftarrow \frac{1}{K}\, d_{m-1}[n]
\end{aligned}
\tag{A.3}
$$

Appendix B

ASWDR algorithm - Some Results

Overview

Numerical results are presented that explains the "brings-forward" principle of an ASWDR algorithm when compared with a WDR counterpart. Some results are also presented that highlight the differences between ASWDR encoded images with that of SPIHT and JPEG2000 algorithms. In all instances, "CDF-9/7" wavelet filters have been used.

B.1 Comparison between WDR and ASWDR algorithms

Frajka and Zeger justified using an MGE embedded image encoder[1] for their algorithm [12]. As discussed in Chapter 3, WDR and MGE algorithms use similar principles, but different methodologies in encoding positions of wavelet coefficients. To illustrate the performance of WDR and ASWDR algorithms, the right-view image from a biomedical stereo-image pair ("*angioMR*") is considered.

These images contain large areas of high-frequency content interspersed with low-frequency regions. For example, in Fig. B.1 features in the heart valve are clearly visible along with veins and arteries that constitute low-contrast regions. Fig. B.2 represents a disparity compensated residual image obtained from the algorithm described in Chapter 3. Heart-valve features are no longer conspicuous. Occluded regions consist of arteries

[1]Explained in Chapter 4

and veins that are present in the left-view. As noted, these are low-contrast areas and generally constitute high-frequency regions.

A four-scale wavelet analysis, using a "CDF-9/7" wavelet-filter pair, is performed on both images. These transformed images are subsequently encoded (using WDR and ASWDR) at a bit-rate of 1.0 bpp with context-based arithmetic coding. Tables B.1 and B.2 indicate number of significant coefficients when implementing a WDR and ASWDR decoding.

It is clearly evident that an ASWDR algorithm is able to decoded more significant coefficients, than a WDR algorithm, for the same bit-rate. A similar inference can be drawn when comparing the performance of an ASWDR algorithm with a MGE algorithm. This in-turn explains results shown in Tables 5.1, 5.2, 5.6 and 5.7.

B.2 Comparison with JPEG2000 and SPIHT

To conclude this appendix, a qualitative discussion is presented by comparing the performance of an ASWDR algorithm with SPIHT and JPEG2000. For an exhaustive discussion on this topic, the concerned reader is directed to [73].

The use of an ASWDR algorithm is justified by its ability to effectively reconstruct low-contrast high-frequency regions. To compare the performance of this algorithm two popular images, "*Barbara*" and "*mandrill*", are selected. Images have been compressed with SPIHT[2] and JPEG2000[3].

On observing Fig. B.3(b) it can be seen that stripes on the right leg of Barbara's pants are visually more perceptible than from Figs. B.4(a) and B.4(b). A similar inference can be drawn by observing her scarf. In fact in Fig. B.4(b), large regions of her pant, scarf and left eye are blurred. This confirms results shown in Table 5.11 in which disparity compensated residual images were encoded using JPEG2000 [17].

[2] `http://www.cipr.rpi.edu/research/SPIHT/`
[3] *IrfanView* with a JP2 plugin from `http://www.luratech.com`

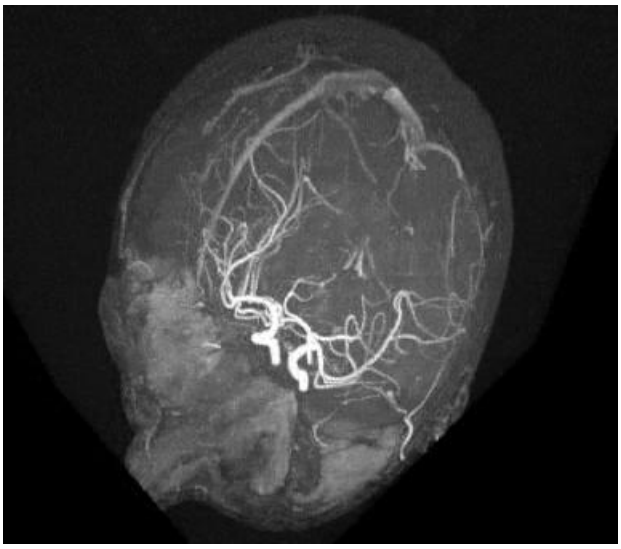


Fig. B.1: Right image-view from the "*angioMR*" stereo-image pair. Image dimensions equals 384 × 352.

Table B.1: Significant coefficients obtained when decoding an encoded version of the image shown in Fig. B.1

Bit-rate (bpp)	Sig. coeff. (WDR)	Sig. coeff. (ASWDR)	% change w.r.t WDR
0.125	2682	2852	+6.33
0.20	4224	4734	+12.07
0.25	5476	5959	+8.82
0.30	6507	6750	+3.73
0.35	7202	8013	+11.26
0.40	8727	9819	+12.51
0.45	10175	11149	+9.57
0.50	11360	12618	+11.07
0.55	12541	13714	+9.35
0.60	13889	14251	+2.60
0.80	18334	20058	+9.40
1.00	24159	26078	+7.94

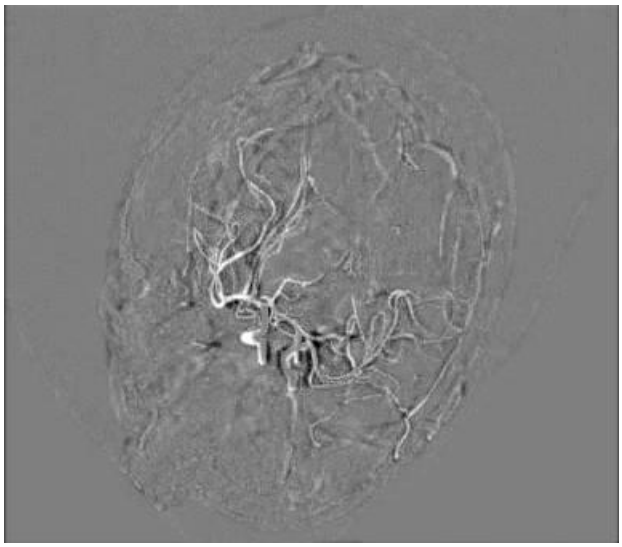


Fig. B.2: Disparity-compensated residual image from the "*angioMR*" stereo-image pair. Image dimensions equals 384 × 352. Image has been scaled for display purposes.

Table B.2: Significant coefficients obtained when decoding an encoded version of the image shown in Fig. B.2

Bit-rate (bpp)	Sig. coeff. (WDR)	Sig. coeff. (ASWDR)	% change w.r.t WDR
0.125	2645	2714	+2.60
0.20	4622	4816	+4.19
0.25	5863	6034	+2.91
0.30	6946	7252	+4.40
0.35	7695	8038	+4.45
0.40	9270	9856	+6.32
0.45	10771	11379	+5.64
0.50	12023	12919	+7.45
0.55	13358	14326	+7.24
0.60	14728	15463	+4.99
0.80	19081	20297	+6.37
1.00	25038	26829	+7.15

(a) Original image at 8.0 bpp

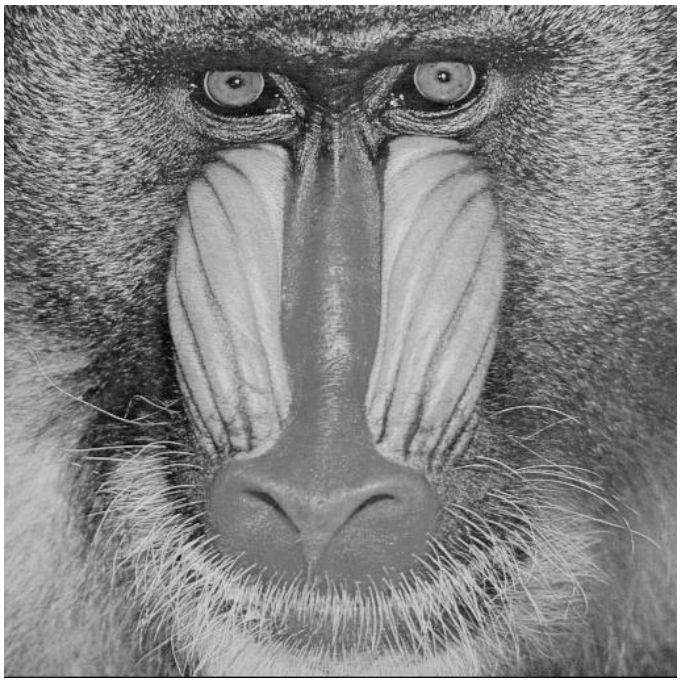
(b) ASWDR - encoded at 0.125 bpp

Fig. B.3: Original and ASWDR encoded "*Barbara*" image. Image dimensions are 512×512

(a) SPIHT - encoded at 0.125 bpp

(b) JPEG2000 - encoded at 0.125 bpp

Fig. B.4: "*Barbara*" image encoded with SPIHT and JPEG2000. Image dimensions are 512×512



(a) Original image at 8.0 bpp

(b) ASWDR - encoded at 0.125 bpp

Fig. B.5: Original and ASWDR encoded "*mandrill*" image. Image dimensions are 512×512

(a) SPIHT - encoded at 0.125 bpp

(b) JPEG2000 - encoded at 0.125 bpp

Fig. B.6: "*Mandrill*" image encoded with SPIHT and JPEG2000. Image dimensions are 512×512

Similar conclusions can be inferred on comparing Fig. B.5(b) with Figs. B.6(a) and B.6(b). For example, large regions of the mandrill's mane have been blurred in Fig. B.6(a). In addition, the region joining its nostrils are blurred in the same figure. However these regions are fairly easy to differentiate in Fig. B.5(b). Another example would be features on its facial skin. These are perceptually more visible in Fig. B.5(b) than in Fig. B.6(a). Once again, Fig. B.6(b) has large areas of the above regions blurred.

It should be mentioned that these images have been encoded with entropy coding. It is well established [16], [73] and [15] that SPIHT outperforms all other algorithms in a R-D framework primarily due to its arithmetic coding scheme. Notwithstanding this, results from Figs. B.3(b) and B.5(b) indicate a perceptually superior performance of an ASWDR algorithm. As mentioned in Sec. 3.3, SPIHT relies on zerotrees (i.e., inter-scale correlation) amongst wavelet coefficients for efficient encoding. On the other hand, ASWDR relies on *vision principles* (i.e., intra- and inter-scale correlation) amongst subbands to encode coefficients. From this it can be inferred that significant values at edges at a particular scale imply significant values at a next higher resolution [73].

JPEG2000 relies on an *embedded block coding with optimized truncation* (EBCOT) algorithm [1]. This is a block-based transform and produces "*tiling-artifacts*" in decompressed images. Post-processing smoothing algorithms are required to remove these artifacts [1]. This affects overall image quality. In addition JPEG2000 has a *feature-rich bit-stream* capability [13]. This requires storage of overhead information. Hence less bits are allocated for actual image information. This also affects overall quality of decompressed images.

Appendix C

Software, Hardware and CD ROM Details

THIS supplementary chapter presents details of software used in generating and hardware used in viewing images and sequences, shown in Chapters 5 and 7. In addition, details of contents in the enclosed CD ROM are also presented.

C.1 Software

The codec structure, reported in this thesis, has been developed using JAVA®. Anaglyph's, reported in Chapter 5, have been generated using an algorithm described in [74]. Operating system in all instances was Windows-2000®. In addition, these images were viewed on applications designed with OpenGL® technology.

C.2 Hardware

. In case of moving-image sequences a frame-rate of 25 fps was used. Individual image-pairs as well as complete image sequences have been displayed on hardware with the following specifications:

- *Display system*: CRT screen with a screen resolution of 2560×1024 and a refresh rate of 120 Hz,
- *Video card*: F 980 NVidia Quadro4 980XGl®,

- *Processor*: Pentium-4® processor at 2.0 GHz,
- *Memory*: 1.0 Gigabyte DDR,
- *Shuttered glasses*: Crystal Eyes® from Stereographics Corp.[1]

C.3 CD ROM

The CD ROM enclosed with this thesis contains:

- Individual images as well as anaglyph's, shown in Tables 5.14 and 5.15 in Chapter 5, and
- An additional table containing individual images and anaglyph's from the "*burial-ground*" stereo-image pair.

The concerned reader is requested to open the "index.html" file present in the CD-ROM in order to obtain and view these images. Fused stereo-image sequences are not presented due to software constraints. The display device for these image sequences is left to the discretion of the reader.

The "*medallion*", "*bull*" and "*burial-ground*" stereo-image pairs are copyright of Eric Dubois at the University of Ottawa, Canada. Kodak® corporation[2] are copyright holders of the "*basketball*" stereo-image pair, used in Chapters 4 and 5. IMAX® corporation[3] are exclusive copyright holders of the "*redcar*" stereo-image sequence used in Chapter 7.

[1] `http://www.stereographics.com`
[2] `http://www.kodak.com`
[3] `http://www.imax.com`

References

[1] D. S. Taubman and M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, 1st ed. Boston, USA: Kluwer Academic Press, 2001.

[2] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards: Algorithms and Architectures*, 2nd ed. Norwell, MA: Kluwer Academic Publishers, 1997.

[3] M. G. Perkins, "Data compression of stereopairs," *IEEE Trans. Commun.*, vol. 40, no. 4, pp. 684–696, April 1992.

[4] M. Strintzis and S.Malassiotis, "Object-based coding of stereoscopic and 3D image sequences," *IEEE Signal Process. Mag.*, pp. 14–28, May 1999.

[5] W. H. Kim, J. Y. Ahn, and S. W. Ra, "An efficient disparity estimation algorithm for stereoscopic image compression," *IEEE Trans. Consum. Electron.*, vol. 43, no. 2, pp. 165–172, 1997.

[6] D. Tzovaras and M. Strintzis, "Disparity estimation using rate-distortion theory for stereo image sequence coding," in *Proc. Int. Conf. on DSP*, vol. 1, July 1997, pp. 413–416.

[7] C. W. Lin, E. Y. Fei, and Y. C. Chen, "Hierarchical disparity estimation using spatial correlation," *IEEE Trans. Consum. Electron.*, vol. 44, no. 3, pp. 630–637, 1998.

[8] Q. Jiang, J. Lee, and M. H. Hayes, "A wavelet based stereo image coding algorithm," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, Phoenix, AZ, March 15-19 1999, pp. 3137–3160.

[9] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3462, 1993.

[10] A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchial trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, 1996.

[11] N. V. Boulgouris and M. G. Strintzis, "A family of wavelet-based stereo image coders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 898–903, Oct. 2002.

[12] T. Frajke and K. Zeger, "Residual image coding for stereo image compression," *Optical Engineering*, vol. 42, no. 1, pp. 1–8, Jan. 2003.

[13] D. Taubman, "High performance scalable image compression with ebcot," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, July 2000.

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

[15] T. Lan and A. Tewfik, "Multigrid embedding (mge) image coding," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Kobe, Japan, Oct. 1999, pp. 369–373.

[16] J. Walker and T. Nguyen, "Wavelet-based image compression," in *The Transform and Data Compression Handbook*, K. R. Rao and P. Yip, Eds. Boca Raton, FL: CRC Press, 2000, ch. 6, pp. 267–312.

[17] R. Shukla and H. Radha, "Disparity dependent segmentation based stereo image coding," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, Sept. 2003.

[18] W. Woo and A. Ortega, "Overlapped block disparity compensation with adaptive windows for stereo image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 2, pp. 194–200, March 2000.

[19] S. Kröener and G. Ramponi, "Edge preserving noise smoothing with an optimized cubic filter," in *Proc. COST-254 Workshop*, Ljubljana, Slovenia, Nov. 1998.

[20] P. Chang and M. Wu, "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, 2000.

[21] ISO/IEC International Standard 13818-2/ITU-T Rec. H.262, "Generic coding of moving pictures and associated audio information:video, Amendment 3:for the Multiview profile," Geneva, Sept. 1996.

[22] M. Domański, A. Luczak, and S. Maćkowiak, "Spatio-temporal scalability for MPEG video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 7, pp. 1088–1093, October 2000.

[23] W.J.Tam, L. Stelmach, and S.Subramaniam, "Stereoscopic video: asymmetrical coding with temporal interleaving," in *Electronic Imaging*, San Jose, CA, Jan. 2001.

[24] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*, 1st ed. N.Y, USA: Van-Nostrand Reihnhold, 1993.

[25] I. Daubechies, *Ten Lectures on Wavelets*, 1st ed. Society For Industrial and Applied Mathematics (SIAM), May 1992.

[26] S. Mallat, *A Wavelet Tour Of Signal Processing*, 1st ed. San Diego, CA: Academic Press, 1998.

[27] C. K. Chui, *An introduction to Wavelets*, 1st ed. San Diego, CA: Academic Press, 1992.

[28] F. Mintzer, "Filters for distortion-free two-band multirate filter banks," *IEEE Trans. Acoust. Speech Signal Process.*, pp. 626–630, 1985.

[29] M. J. T. Smith and T. P. Barnwell III, "Exact reconstruction techniques for tree-structure subband coders," *IEEE Trans. Acoust. Speech Signal Process.*, pp. 434–441, 1986.

[30] P. P. Vaidyanathan and P. Q. Hoang, "Lattice structures for optimal design and robust implementation of two-channel perfect-reconstruction qmf baks," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, pp. 81–94, 1988.

[31] ——, "Time-domain filter bank analysis: A new design theory,," *IEEE Trans. Signal Process.*, vol. 40, pp. 1412–1429, June 1992.

[32] T. Q. Nguyen, "A quadratic constrained least-squares approach to the design of digital filter banks," *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 1344–1347, 19992.

[33] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.

[34] J. Kovačević and M. Vetterli, "Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for $\mathcal{R}^n$," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 533–555, March 1992.

[35] G. Karlsson and M. Vetterli, "Extension of finite length signals for subband coding," *Signal Processing*, vol. 17, no. 2, pp. 161–166, June 1989.

[36] M. Smith and S. Eddins, "Analysis/synthesis techniques for subband image coding," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 28, pp. 1446–1456, Aug. 1990.

[37] H. Kiya, K. Kiyoshi, and M. Iwahashi, "A development of symmetric extension method for subband image coding," *IEEE Trans. Image Process.*, vol. 3, no. 1, pp. 78–81, Jan. 1994.

[38] V. Silva and L. Sa, "General method for perfect reconstruction subband processing of finite length signals using linear extensions," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2572–2575, Sept. 1999.

[39] J. Williams and K. Amaratunga, "A discrete wavelet transform without edge effects using wavelet extrapolation," *Journal of Fourier Analysis and Applications*, vol. 3, no. 4, pp. 435–449, 1997.

[40] S. Nath and E. Dubois, "Minimization of edge effects in images using an extrapolated discrete wavelet transform," in *Proc. SPIE Video Technologies for Multimedia Applications*, vol. 1, Denver, CO, Aug. 2001, pp. 1–12.

[41] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM J. Math. Anal.*, vol. 29, no. 2, pp. 511–546, 1997.

[42] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *Journal of Fourier Analysis and Applications*, vol. 4, no. 3, pp. 245–267, 1998.

[43] International organisation for standardisation, "ISO/IEC JTC1/SC29/WG11 N4668, Coding Of Moving Pictures And Audio," March 2002.

[44] I. E. G. Richardson, *H.264 and MPEG-4 Video compression: Video coding for next-generation multimedia*, 1st ed. West Sussex, UK: Wiley, 2003.

[45] C. Stiller and J. Konrad, "Estimating motion in image sequences: A tutorial on modeling and computation of 2D motion," *IEEE Signal Process. Mag.*, vol. 16, pp. 70–91, July 1999.

[46] S. Sethuraman, M. Siegel, and A. Jordan, "A multiresolutional region based segmentation scheme for stereoscopic image compression," in *Digital Video and Compression: Algorithms and Technologies, IS&T/SPIE Electronic Imaging*, San Jose, CA, February 1995.

[47] J. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. Commun.*, vol. COM-29, pp. 1799–1808, Dec. 1981.

[48] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion-compensated interframe coding for video conferencing," in *Proc. NTC81*, New Orleans, LA, Dec. 1981, pp. C9.6.1–9.6.5.

[49] M. Brünig and W. Niehsen, "Fast full search block matching," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 2, pp. 241–247, Feb. 2001.

[50] M. Ziegler, L. Falkenhagen, R. ter Horst, and D. Kalivas, "Evolution of stereoscopic and three-dimensional video," *Signal Processing: Image Communication*, vol. 14, pp. 173–194, 1998.

[51] D.Tzovaras, M.G.Strintzis, and H.Sahinoglou, "Evaluation of multiresolution techniques for motion and disparity estimation," *Signal Processing: Image Communication*, vol. 6, no. 1, June 1994.

[52] A. Munteanu, J. Cornelis, G. V. Auwera, and P. Cristea, "Wavelet image compression - The quadtree coding approach," *IEEE Trans. Inf. Tech. Biomed.*, vol. 3, no. 3, pp. 176–185, Sept. 1999.

[53] J. Tian and R.O. Wells Jr., "Embedded image coding using wavelet difference reduction," in *Wavelet Image and Video Compression*, P.Topiwala, Ed. Norwell, MA: Kluwer Academic Publications, 1998, pp. 289–301.

[54] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *Journal of Fourier Analysis and Applications*, vol. 4, no. 3, pp. 245–267, 1998.

[55] P. Elias, "Universal codeword sets and representations of the integers," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 194–203, March 1975.

[56] J. Walker and T. Nguyen, "Adaptive scanning methods for wavelet difference reduction in lossy image compression," in *Proc. IEEE Int. Conf. Image Processing*, Vancouver, Canada, Sept. 2000.

[57] I. Witten, R. Neal, and J. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, vol. 30, pp. 520–540, June 1987.

[58] L. B. Stelmach and W. J. Tam, "Stereoscopic image coding: Effect of disparate image-quality in left- and right-eye views," *Signal Processing: Image Communication*, vol. 14, pp. 111–117, 1998.

[59] T. Mitsuhashi, "Subjective image position in stereoscopic TV systems - Considerations on comfortable stereoscopic images," in *Human Vision, Vis. Proc., and Digital Disp. V*, vol. 2179, 1994, pp. 259–266.

[60] W.D. Reynolds Jr. and R. Kenyon, "The wavelet transform and the supression theory of binocular vision for stereo image compression," in *Proc. IEEE Int. Conf. Image Processing*, Lausanne,Switzerland, Sept. 1996.

[61] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Mag.*, vol. 15, pp. 23–50, Nov. 1998.

[62] A. Gersho and R. M. Gray, *Vector quantization and signal compression*, 1st ed. Boston, MA: Kluwer Academic press, 1991.

[63] C. Auyeung, J. Kosmach, M.Orchard, and T.Kalafatis, "Overlapped block motion compensation," in *Proc. SPIE, Visual Comm. Image Processing*, vol. 1818, Boston, MA, Nov. 1992, pp. 561–571.

[64] B. Tao, M.Orchard, and B. Dickinson, "Joint application of overlapped block motion compensation and loop filtering for low bit-rate video coding," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Washington, D.C, Oct. 1997, pp. 626–629.

[65] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 1st ed. Addison Wesley Longman, 1993.

[66] G. Ramponi, "The rational filter for image smoothing," *IEEE Trans. Signal Process. Let.*, vol. 3, no. 3, pp. 63–65, March 1996.

[67] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Object-based coding of stereo image sequences using joint 3-D motion/disparity compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 312–327, Apr. 1997.

[68] D. Tzovaras, S. Vachtsevanos, and M. G. Strintzis, "Optimization of quadtree segmentation and hybrid two-dimensional and three-dimensional motion estimation in

a rate-distortion framework," *IEEE J. Sel. Areas Commun.*, vol. 15, no. 9, pp. 1726–1738, Dec. 1997.

[69] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate distortion sense," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 160–175, Apr. 1993.

[70] J. Vaisey and A. Gersho, "Image compression with variable block size segmentation," *IEEE Trans. Signal Process.*, vol. 40, no. 8, pp. 2040–2060, Aug. 1992.

[71] http://code.ucsd.edu/~frajka/images/stereo/stereo_images.html.

[72] http://vasc.ri.cmu.edu/idb/html/stereo.

[73] J. Walker, "Wavelet based image-processing," in *Proc. Int. Soc. Anal. App. Com. (ISAAC)*, Brazil, October 2004.

[74] E. Dubois, "A projection method to generate anaglyph stereo images," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, vol. 3, Salt Lake City, UT, May 2001, pp. 1661–1664.

[75] S. Sethuraman, A. Jordan, and M. Siegel, "Multiresolution based hierarchical disparity estimation for stereo image pair compression," in *Proc. of the Symposium on Application of subbands and wavelets*, March 1994.

[76] S. Thanapirom, W. Fernando, and E. Edirisinghe, "A zerotree stereo video encoder," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 2, May 2003, pp. 608–611.

[77] A. Puri, R. Kollarits, and B. Haskell, "Basics of stereoscopic video, new compression results with MPEG-2 and a proposal for MPEG-4," *Signal Processing: Image Communication*, vol. 10, pp. 201–234, 1997.

[78] J. Arnold, M. Frater, and Y. Wang, "Efficient drift-free signal-to-noise ratio scalability," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 1, pp. 70–82, February 2000.

[79] S.Pastoor, "Human factors of 3D displays in advanced image communications," *Displays*, vol. 14, pp. 150–157, 1993.

[80] B. Kim, Z. Xiong, and W. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 8, pp. 1374–1387, Dec. 2000.

[81] J. Xu, Z. Xiong, S. Li, and Y. Zhang, "Three-dimensional embedded subband coding with optimized truncation (3-D ESCOT)," *App. Comp. Har. Anal.*, vol. 10, pp. 290–315, 2001.

[82] S. Hsiang and J. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank," *Signal Processing: Image Communication*, vol. 16, pp. 705–724, 2001.

[83] B. Girod, "What's wrong with mean-squared error," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 207–220.

[84] http://www.sarnoff.com.

[85] http://www.vqeg.org.

[86] Y. Wang, S. Wenger, J. Wen, and A. Katsaggelos, "Error resilient video coding techniques: Real-time video communications over unreliable networks," *IEEE Signal Process. Mag.*, pp. 61–82, July 2000.

[87] N. Lu, *Fractal Imaging*, 1st ed. San Diego, CA: Academic Press, 1997.

[88] K. Cinler and A. Mertins, "Edge sensitive subband coding of images," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, Atlanta,GA, May 1996, pp. 2365–2368.

[89] S. Li and W. Li, "Shape-adaptive discrete wavelet transforms for arbitrary shaped visual object coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 5, pp. 725–743, Aug. 2000.

[90] http://www.geoffdavis.net/dartmouth/wavelet/wavelet.html.

[91] J. Odegard and C. Burrus, "Smooth biorthogonal wavelets for applications in image compression," in *Proc. of DSP Workshop*, Loen, Norway, Sept. 1996. [Online]. Available: http://citeseer.nj.nec.com/odegard96smooth.html

[92] L. Winger and A. N. Venetsanopolos, "Biorithogonal nearly coiflet wavelets for image compression," *Signal Processing: Image Communication*, vol. 16, pp. 859–869, 2001.