Université d'Ottawa • University of Ottawa

PERMISSION DE REPRODUIRE ET DE DISTRIBUER LA THÈSE

PERMISSION TO REPRODUCE AND DISTRIBUTE THE THESIS

NOM DE L'AUTEUR / NAME OF AUTHOR: Shabnam Sodagari

ADRESSE POSTALE / MAILING ADDRESS:
4th Floor
65, Golsan Nakhjavan, Mojdah Street
IRAN

GRADE / DEGREE: M.A.Sc.(Electrical Engineering)

ANNÉE D'OBTENTION / YEAR GRANTED 2003

TITRE DE LA THÈSE / TITLE OF THESIS:
Variable Block-size Disparity Estimation in Stereo Imagery

L'auteur permet, par la présente, la consultation et le prêt de cette thèse en conformité avec les règlements établis par le bibliothécaire en chef de l'Université d'Ottawa. L'auteur autorise aussi l'Université d'Ottawa, ses successeurs et cessionnaires, à reproduire cet exemplaire par photograhie ou photocopie pour fins de prêt ou de vente au prix coûtant aux bibliothèques ou aux chercheurs qui en feront la demande.

Les droits de publication par tout autre moyen et pour vente au public demeureront la propriété de l'auteur de la thèse sous réserve des règlements de l'Université d'Ottawa en matière de publication de thèses.

N.B. LE MASCULIN COMPREND ÉGALEMENT LE FÉMININ

The author hereby permits the consultation and the lending of this thesis pursuant to the regulations established by the Chief Librarian of the University of Ottawa. The author also authorizes the University of Ottawa, its successors and assignees, to make reproductions of this copy by photographic means or by photocopying and to lend or sell such reproductions at cost to libraries and to scholars requesting them.

The right to publish the thesis by other means and to sell it to the public is reserved to the author, subject to the regulations of the University of Ottawa governing the publication of theses.

04.11.03
DATE

Sodagari
(AUTEUR) SIGNATURE (AUTHOR)

Université d'Ottawa · University of Ottawa

Université d'Ottawa · University of Ottawa

FACULTÉ DES ÉTUDES SUPÉRIEURES ET
POSTDOCTORALES

FACULTY OF GRADUATE AND
POSTDOCTORAL STUDIES

SODAGARI, Shabnam
AUTEUR DE LA THÈSE - AUTHOR OF THESIS

M.A.Sc. (Electrical Engineering)
GRADE - DEGREE

School of Information Technology and Engineering
FACULTÉ, ÉCOLE, DÉPARTEMENT - FACULTY, SCHOOL, DEPARTMENT

TITRE DE LA THÈSE - TITLE OF THE THESIS

Variable Block-Size Disparity Estimation in Stereo Imagery

Éric Dubois
DIRECTEUR DE LA THÈSE - THESIS SUPERVISOR

EXAMINATEURS DE LA THÈSE - THESIS EXAMINERS

R. Laganière

A. Cuhadar

J.-M. De Koninck, Ph.D.
LE DOYEN DE LA FACULTÉ DES ÉTUDES
SUPÉRIEURES ET POSTDOCTORALES

SIGNATURE

DEAN OF THE FACULTY OF GRADUATE
AND POSTDOCTORAL STUDIES

VARIABLE BLOCK-SIZE DISPARITY ESTIMATION IN STEREO IMAGERY

Shabnam Sodagari

A Thesis submitted to the Faculty of Graduate and Postdoctoral
Studies in partial fulfillment of the requirements for the degree of
Master of Applied Science, Electrical Engineering

April 2003

Ottawa-Carleton Institute for Electrical and Computer Engineering
School of Information Technology and Engineering
University of Ottawa
Ottawa, Ontario, Canada

0-612-79373-7

To My Mother and My Father

Contents

List of Figures

Abstract

This thesis addresses the problem of developing and implementing in software a variable block size(quadtree splitting) disparity estimation algorithm that is optimized for use in compression of stereo image pairs and studying its performance over a range of rate/distortion values for a variety of images.

First the constrained optimization problem is converted to an unconstrained one using the Lagrange multiplier approach. Then by solving the optimization problem using dynamic programming, the optimal variables representing the optimal quadtree structure and the quantizer for each node are determined. The experimental results show the improvements of this method over simple intraframe JPEG coding and over fixed block-size disparity estimation.

Acknowledgements

First and foremost I would sincerely like to thank Dr Eric Dubois for supervising and funding this thesis and introducing me to the field of Image processing. It was a pleasure for me to be a student of such a knowledgeable, gentle and patient person. His devotion and ideas was a great support for me. I am indebted to the help and guidance of such an insightful professor.

I would like to thank members of VIVA (Vision, Image, Video and Audio) lab for their friendly manner.

I also wish to thank the School of Information Technology and Engineering at the University of Ottawa for providing me with an excellent environment for studying and research during my degree.

Shabnam Sodagari

Ottawa, Canada, February 2003

Chapter 1

Introduction

Stereoscopic imagery, which involves using binocular disparity to recognize depth information and hence giving a feeling of three-dimensional scene perception, is important in several aspects, since the real world that is being presented to us by our brain processing mechanisms is three-dimensional.

When two pictures are taken of a scene with a camera shift (as exactly horizontal as possible) and the right and left pictures are represented to the right and left eyes respectively, a sense of stereo vision will be given to the viewer. The two images mentioned above are said to form a stereo pair. We can say that stereo vision is the ability to use binocular disparity to determine whether objects are closer or further than the fixation point of the eyes.

Stereo imagery is growing in multimedia[8], tele-presence systems, entertainment, remote operations, robot navigation and machine vision. Among the benefits of these type of images, we can mention that stereo images improve the feeling of presence, and are useful in remote manipulation for feeling spatial relationships. A typical stereo system with a single left-right pair needs two times more data than a monoscopic system. Therefore, the attention is toward image compression methods with regards to the special characteristics of stereo pairs.

However, one of the problems existing in any imaging system is finding a compression method as close to optimum as possible, and stereo imaging is no exception.

Considering this issue, disparity estimation is almost an inevitable part of any stereo compression system. In this thesis the problem is tackled using a Rate-Distortion optimized disparity estimation.

There are basically two different techniques for compressing stereo pairs. The first technique, which is called disparity-compensated transform-domain predictive coding, concentrates on minimizing the mean-square error between the original stereo pair and the compressed stereo pair. The second technique, called mixed resolution coding, is a psychophysically justified technique that exploits known facts about human stereo vision to code stereo pairs in a subjectively acceptable manner[18]. The first of these approaches is used in this thesis. The quality of stereo pairs involves measuring the ability of an observer to understand the depth in coded stereo pairs. Usually, as mentioned in [18], the depth which is felt in coded stereo image pairs is greater than of original stereo image pairs.

1.1 Problem Statement

The problem of optimal bit allocation between different parameters in a stereo image compression system, including the disparity vector, the displaced frame difference, and the structure of the quadtree in the variable block size scheme is considered to be solved in a rate-distortion frame.

In a block matching method for disparity estimation, it is desired to choose disparity vectors for which the residual error or the displaced frame difference is minimum, while in a rate-distortion based frame the constraint of keeping the rate under a specific budget and simultaneously minimizing the distortion is also going to be added.

Combining the constraints above results in the problem of estimating the disparity using a rate-distortion criterion for compression of stereo image pairs.

1.2 Contributions

Although there has been a lot of work done in the field of motion estimation using different rate-distortion based methods or non rate-distortion based ones, as well as fixed block size or variable block size methods, the amount of corresponding work for stereo applications (the topic of this thesis) has been much less. Furthermore my work focuses on the idea of developing and implementing an algorithm based on mathematical theory and using the ideas and concepts of mathematical optimization like dynamic programming, which have not been applied to the problem of disparity estimation, targeting coding applications.

A paper based on this thesis has been accepted in the IEEE Canadian Conference on Electrical and Computer Engineering.

1.3 Organization of the Thesis

The thesis begins with a review of literature and background on disparity estimation, motion estimation, rate-distortion optimal schemes, problem formulation and notation in chapter 2.

Chapter 3 discusses the proposed rate-distortion optimal variable block size disparity estimation method in detail. Chapter 4 presents the experimental results to show the improvements achieved using this method. The conclusion and future work follows in chapter 5.

Chapter 2

Problem Formulation

In this chapter, we first present the basic concepts related to our problem and the specific problem formulation. A review of related work follows. Finally a summary of the chapter is presented.

The problem addressed in this thesis is disparity estimation in stereoscopic imagery in a rate-distortion framework using a variable-block-size method for the purpose of compression. The procedure as to how this problem is developed and then formulated to be solved is described step by step below.

First we introduce the nature of a stereo image and the idea of compressing a stereo pair. Here the idea of disparity compensation automatically follows, which basically involves the concept of coding the right image first by using some standard method such as the JPEG standard for a single image compression. Then it is the turn of left image to be predicted from the right image in order to reduce the amount of data that must be sent to the receiver. But, as is the characteristic of any prediction, there still remains some amount of prediction error, which if sent to the receiver adds to the quality of left image reconstruction. Again this error is coded to save even more bandwidth. At the decoder, the error is added back to the decoded right image in order to reconstruct the left image.

The elements for this technique include intraframe coding and disparity estimation and compensation. Therefore we need to discuss intraframe coding and disparity estimation. Then the fact that disparity is similar to motion in image sequences brings the idea of relating disparity to the concept of motion which is already a mature field in video coding and compression.

As opposed to the abstract ideal case, every real system is facing bandwidth or rate budget limitations and therefore automatically some amount of imperfection, referred to as distortion, is going to be introduced in the reconstruction phase and it becomes inevitable to balance rate and distortion for different parameters.

In this thesis, as a realization of intra-frame coding, the JPEG method is addressed and the Discrete Cosine Transform or DCT is used instead of other common transforms like Wavelet Transform. This is because in this research, the focus is on the disparity estimation aspect of the problem and furthermore, the DCT remains a very efficient technique.

In the disparity and motion estimation part of the chapter, the focus is on block based methods which involve block matching techniques which itself has the issues of performance and computation burden. One important subgroup of block based methods are variable block size motion estimation techniques.

Then the above mentioned concepts are put in a rate distortion optimal framework and since variable block size, rate-distortion optimal disparity estimation has not been reported for stereo coding these concepts together lead to the topic of this thesis.

2.1 Introduction to Stereo Imagery

In a three dimensional imagery system, a pair of spatially rotated and spaced apart cameras is used to give to the eyes two views from two different angles from a scene[1]. The benefits that these images have, in comparison with two dimensional imagery, makes them useful in areas like entertainment, robot vision, guidance of mobile robots, and some applications in virtual environments like remote inspection, co-ordinate measurement and telemanipulation. Of course for each application, the issues specific to that application must be taken into account.

Here we try to concentrate on some specific characteristics of a three-dimensional imaging system which are important and are decisive in the performance of these systems in different applications. These can be the extent of the comfortable stereo-region, the minimum detectable depth interval, and the image magnification. The overlapped fields of view of two cameras, which are spaced apart to view the same scene from different angles, form a volume, in three dimensional space which can be named the stereo-region. Every object point in the stereo region has two spatially separated images. Then if these pictures are presented properly to the right and left eyes, a feeling of a three dimensional picture can be achieved by the observer.

Fig 2.1 shows the geometry of a stereoscopic imaging system.

The right and left planes are coplanar. P_1 and P_2 are two arbitrary points in the image and the parameter f is the focal length which is an intrinsic parameter of the camera.

Given a point in the left image, the vector that connects this point to its matching point in the right image is called disparity. The disparity vector can have both positive and negative horizontal and vertical components and based on camera geometry.

The performance of stereoscopic systems depends on various parameters. Since the extent of the region of space which can be displayed in three dimensions without causing undue eye strain to the viewer, and also the minimum detectable depth interval, are related to system's geometrical and optical parameters, these parameters are important in controlling the characteristics of stereo systems. Distortion in three-dimensional images can sometimes be due to the difference between the depth magnification and the magnification of the xy-plane[1]. Therefore, in order to design a stereoscopic system for a specific application, the relationships between the performance of a three dimensional television system and the controlling parameters must be explored.

2.2 Stereo Pair Compression

There exist two main techniques for stereo pair data compression: disparity-compensated transform domain predictive coding and mixed-resolution coding. Disparity-Compensated Transform domain coding aims to minimize the error between the original stereo pair and the compressed stereo pair. Mixed resolution coding is a perceptually justified method that is suitable when the compressed stereo pair is viewed by a human. Here we deal with disparity- compensated transform domain predictive coding.

The basic operation of a Disparity-Compensated Transform Domain coding is as follows: First, the right picture is coded independently of the left picture using a desired technique[24], and it is then transmitted to the receiver. Second, the left picture is broken into subblocks, then for each subblock in the left picture, an approximating block is found in the encoded right picture, and side information identifying this block is transmitted to the receiver. In the next step, the transform of the left picture subblock is predicted based on the transform of its approximating subblock, and finally, after a suitable bit allocation, the transform-domain prediction error is quantized and transmitted to the receiver in order to enable the decoder to reconstruct the stereo image as closely as possible to the original one which was at the transmitter.

The disparity-compensated transform domain predictive coding technique for stereo pairs is similar to a standard technique for video sequence coding, which is DCT coding of a motion compensated prediction error. But one has to remember that, the nature of the displacement that exists and and must be compensated in a stereo pair is different from that in an image sequence: In a stereo pair, the image of every object in the left picture is displaced with respect to its image in the right picture, while in cases such as for example, video teleconferencing, from frame-to-frame only a few objects in the scene change their location. Furthermore, in a stereo pair produced by the ideal parallel-axes geometry, only horizontal unidirectional displacements are possible, while in an image sequence, the image of an object can have displacements vertically and horizontally in both of these directions, and also can have rotation.

Since the basic intraframe coding technique used in thesis is the block DCT method used in JPEG[21], we give a brief review of JPEG. The block diagram of a JPEG coder is shown in Fig 2.2.

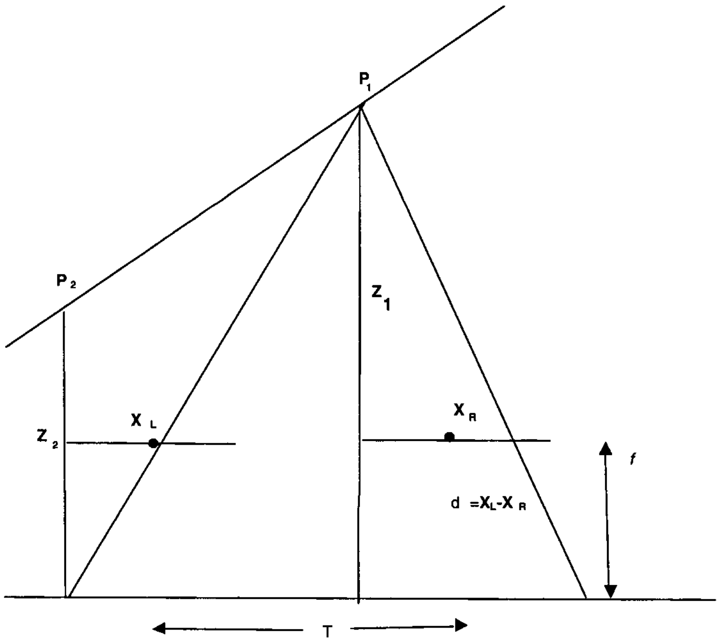

Figure 2.1: A stereoscopic imaging system

First we must mention that the concept of transform coding describes coding techniques where the source is decomposed using a linear transform and where each of the frequency components obtained from the decomposition is then quantized.

JPEG is an international image compression standard which is based on the transform technique. A typical transform-based image coder consists of the front-end linear transform and thereafter, a scalar quantizer and entropy coder. As a brief description of JPEG standard, the image is decomposed into 8 × 8 blocks for transforming, quantization and entropy coding. Blocks are processed in a raster scanning order and discrete cosine transform or DCT is done on each block independently from other blocks. In the next phase, each 8 × 8 block is quantized by a uniform scalar quantizer. The step sizes are defined in an 8 × 8 matrix for each of the 64 frequency components. These values are encoded in the header.

The DC coefficient which is the first quantized coefficient, is predicted from the previously encoded block to give more compression and represents the average sample value in a block. The remaining 63 coefficients are encoded using data from that specific block. Then the entropy coder, which is lossless, provides a mapping from each of the various quantization indices to given codes. The AC coefficients are processed in a zig-zag manner. Run length codes represent the sequence of quantized coefficients as *(run,value)* pairs. Run indicates the number of zero coefficients between the current non-zero coefficient and the previous nonzero coefficient and Value is for the value of the current coefficient. The sequence of Runs and Values are compressed using arithmetic coding or Huffman coding.

We are using the discrete cosine transform or DCT, because it is better than other linear transforms for image coding which are non-adaptive. The right picture can be transform coded using the DCT on 8×8 subblocks, followed by quantization and varaiable length coding.

2.3 Disparity and Motion Estimation

In this section, we present a brief review related to disparity estimation and motion estimation techniques and in the following section, we review the rate-distortion optimal schemes which are applied to image compression.

A good disparity or motion estimation algorithm can result in useful reduction of the redundancies with a fast implementation, which is necessary if one is needing real time coding applications[12].

Block-matching algorithms are established on the basis of the matching of blocks between two images[24]. For each block, a motion vector is evaluated by matching the block in one image within a search area in the other image using the Mean Absolute Error (MAE) criterion for example. Among all the block-matching algorithms, full search gives best match. The full search method exhaustively matches all the possible candidates within the search window, where the candidate with the least MAE would be introduced as the best matched block, and the associated motion vector is calculated. However Full Search needs a lot of heavy computations which makes it difficult for real-time implementation.

In block-based methods, the disparity or motion vector for one block is fixed and represents the disparity vector for all the pixels in that block. The block-matching technique is very widely used in video coding systems like H.261, H.263 and MPEG. It tries to find a block that is most similar to a current block within a pre-defined search area in a reference frame. To achieve this, full search is computationally intensive[23], and may not be practical for high-quality video coding applications such as MPEG or High Definition Television, because these applications need large search areas for increasing efficiency. Hence, many fast Block Matching Algorithms have been developed to lighten the heavy computations of the full-search block matching algorithm. In the basic full search method, there is a region in which a match to the block can be found specifically by its width and height, called the search area.

For conventional block matching, an image frame is segmented into two-dimensional small blocks having the size of $N \times N$ pixels. For each block, one has to search for the displacement which produces the best match among the candidate blocks in the preceding frame. Usually, the search is performed within a maximum range of $(-N_D, N_D)$ pixels in horizontal and vertical directions which take the position of the reference block as the center. For each block in the search area, one has to calculate how much similarity with the reference block exists. There are many similarity criteria, i.e., the Cross-Correlation Function, Mean Square Error (MSE), Mean Absolute Error (MAE), etc. Among them mean absolute error is the most widely used alternative because its computation procedure is very simple[4]. The MAE criterion for a block of size $N \times N$ pixels is

$$\mathrm{MAE}(m_1, m_2) = \sum_{n_1=1}^{N} \sum_{n_2=1}^{N} |x_t(n_1, n_2) - x_{t-1}(n_1 + m_1, n_2 + m_2)| \tag{2.1}$$

$N_D \leq m_1, m_2 \leq N_D$

where x_t is the reference block in the current image and x_{t-1} is a candidate block within the search area in previous image. The displacement $(m1, m2)$ which minimizes MAE of (m_1, m_2) is chosen as the motion or displacement vector.

The Full Search method gives the best match, but since in this technique, the error has to be evaluated at all positions, this method is very expensive from the viewpoint of computational cost and more efficient methods have been developed to reduce the computational requirement.

One method [18] for approximating subblocks using disparity compensation is considering that a subblock is located in the left picture, then this subblock is placed above the identically positioned subblock from the right picture, and the error between the two subblocks is computed. Next, the left picture subblock is displaced one pixel to the left with respect to the right picture, and again the error between the left picture subblock and the subblock over which it was then situated is computed and saved. This process is repeated until the left picture subblock has been shifted over a range larger than the disparity associated with the closest object imaged in the stereo pair. The best match is the one that has the least error. Usually it is better that some information about the application is known beforehand in order to determine how large the shifting range must be. Among the parameters that affect the appropriate shifting range are the separation between the left and the right cameras and the distance to the nearest object being imaged. The method of finding matches already explained is called the block-matching method for disparity estimation. The problem of finding matches for subblocks at the left edge is very similar to two problems that exist in the coding of image sequences or video: finding matches for background areas that are uncovered by moving objects, and finding matches for objects that move into the image area between two frames.

Some disparity compensation algorithms find left-picture/right-picture matches with integer-pixel accuracy. At the expense of increased side information and computational complexity, disparities could be estimated to subpixel accuracy. This can improve the performance of a Disparity Compensated Transform Domain Predictive coder.

Using larger subblocks in disparity-compensated prediction requires less overhead information, but the resulting predictions are not as good as the case of small subblocks.

Since relatively little work on block-based disparity estimation has been published, we now address the related issue of motion estimation. Recent developments in video compression have many applications such as multimedia transmission, videophones, teleconferencing, CD-ROM storage. The main idea of compression techniques is to remove spatial and temporal redundancies existing in video sequences. The most popular technique is motion-compensated predictive coding. In video coding applications like video conferencing and for example High Definition Television, some coding techniques such as motion compensation and spatial transformation are widely used. To achieve high data compression in video coding, it is necessary to employ motion estimation. Whenever the motion estimation is more accurate, the efficiency of compression gets better, therefore it is an important tool to get better compression. It is well known that motion estimation needs intensive computation, but in video coding it is necessary.

We assume that the reader is familiar with the basic topics of motion estimation, therefore here we briefly touch on the fast block matching methods developed by researchers.

In order to reduce the computations in the full search method, several fast searching algorithms such as the two-dimensional logarithmic search, three-step search, conjugated direction search, etc have been developed so far. These fast algorithms reduce the number of positions to be searched based on the assumption that the matching error increases in a monotonic manner when the search position moves farther from the position of minimum error. However, this assumption cannot be always true in real world applications, and therefore the motion vector which has been estimated can converge to a local minimum of the error instead of the global minimum.

In [5] a motion estimation algorithm using an Adaptive Search Center predicted from its adjacent blocks, and a Nonlinear Center Biased Search Point Pattern is presented which does not have the problem of being trapped by the local minimum, and its main characteristic is finding the majority motion vector in one step. In comparison with other block-based search algorithms, like the full-search and three-step-search, the mentioned algorithm has an average Peak Signal to Noise ratio or PSNR very close to the PSNR of full search algorithm, and also is faster than the three step-search in the sense of average search time. Various other search algorithms have been proposed, like for example Three Step Search, Cross Search, One at a Time Search and Four Step Search. In most of these algorithms the effort is toward reducing the computation amount by decreasing the number of search points, or changing the pattern adopted for search. Moreover some algorithms are based on the assumption that the motion vectors of adjacent macro blocks in the frame are related to a large degree. Using this assumption, the number of search points can still be reduced.

One algorithm of this type is called the Predictive Search Algorithm, which profits from the linear weighting of the motion vector of the three adjacent macro blocks to achieve a predicted motion vector. Of course the results show that for achieving this computation reduction, a reduction in accuracy of matching is produced which result to lower Peak Signal to Noise Ratio or PSNR.

The algorithms mentioned above have proven to be able to reduce the computational requirements to a considerable extent. Three Step Search has been the most popular one, because it is relatively simple. However it has two main problems: firstly, its Peak Signal to Noise Ratio is basically lower that that of Full search and secondly, it can give misleading results, because it can easily be trapped in a non-optimum solution. Instead, the Adaptive Center Non-linear Three Step Search [5] which is a modified version of three step search works faster, has a PSNR near to that of full search and does not have the trouble of being trapped in a non optimum solution. It is by nature different from three step search in three aspects: firstly, its search center is predicted from the left and top adjacent macro block based on the achieved model.

Secondly, it uses a search pattern in the form of a non-linear center biased search point pattern which goes outward in similar to a spiral movement; thirdly, it stops faster, if the minimum position is closer to the search center.

2.4 Rate-Distortion Optimality

In this section we review the issue of rate distortion optimality by first giving an overview of the classical rate distortion theory and after that considering some practical ways through which this theory can be used in different applications[2],[11].

As is going to be discussed later in this chapter, classical rate-distortion theory, if applied to sources like images, shows limitations in practice and therefore it is necessary to consider rate distortion optimization scenarios which are mostly limited to special cases.

The theory of rate distortion, which is a part of source coding, deals with using the redundancy which may exist in any source, in order to compress the source, which means trying to represent the source with the fewest possible bits and meanwhile trying to maintain the quality up to a specific level.

In order to do that, first the different characteristics of different types of sources like image, video, audio and data must be studied to get the best possible result. Then comes the issue of lossy compression, which means a relative amount of quality is sacrificed for the purpose of getting more compression.

For image and video coding, lossy techniques are preferred because of the properties of the human visual system that allows the decoded image or video to look almost indistinguishable from the original ones[17]. The reproduction quality or the amount of error or similar concepts are related to distortion and these two values have a reciprocal relationship.

The main goal here is to overview how these rate-distortion tradeoffs are used in practical image coding. We are considering rate-distortion in the transform domain, as is used in standards related to this area.

Transform coding includes decomposing a source into its frequency components using block transforms such as the discrete cosine transform. After decomposing the source into its frequency components, quantization is applied. If we consider a pair of stereo images, we need disparity estimation as a tool for achieving better compression. The encoder computes the disparity parameters and the decoder uses this for reconstruction.

The encoder has some degrees of freedom in selecting levels of rate-distortion performance. For example, in JPEG, it is possible to select different rates for an image and get different distortions in the reconstructed image.

Hence in other terms the rate-distortion optimality problem is to design the encoder of an image in the best manner, such that some rate-distortion criterion is satisfied. If the source model is a deterministic model, the goal would be to optimize the parameter selection strategy for a specific input which comes to the encoder. In standards like JPEG, the coder is designed to be adjusted for every input in the desired way, which gives more exact results than considering a collection of inputs together.

By being familiar with the mathematical theory of optimization, it goes without saying that, by the term "optimal' solution, we mean the one giving the best objective function among all possible other solutions.

So far, we have talked generally about Rate-Distortion optimization, but in practice there are some issues that must be considered. For example one issue that must be taken into account is what is most suitable to be chosen as the basic coding unit. Depending on the application, the coding unit may be different. For example, it is possible to consider video frames as the basic coding units for video coding, and measure rate and distortion for each frame in the sequence, and then decide about the optimal solution. In applications like our topic, the technique that we use for this part operates at a finer level and considers coding choices for a single block. More details about this technique are presented in Chapter 3.

It is clear that we can use more complex algorithms for encoding in off-line cases but in live or real-time encoding there are limits for the time and complexity of the algorithm. Also if coding is done only once, instead of several times, the algorithm need not be very simple, because usually in these cases the quality is important. Standards like JPEG provide a common decoding framework and it is possible to develop encoders for a range of applications.

To formulate the concept in mathematical terms, consider U coding units where each coding unit has P different operating points; for each coding unit i its rate is r_{ij} and distortion d_{ij} and j is the quantizer in which starting from $j = 1$ to $j = P$ the quality gets worse and worse. Of course we can decrease distortion as much as we wish when the rate is not limited to a specific amount, but it is important to try to achieve the best quality for some constraints on the rate. Regarding our application, the technique that we are using is assuming a constraint on the total rate budget, among other classes of formulations that might consider the delay as is the case for real time video coding, but we give review of them here.

The type of our problem is a budget constrained problem, i.e., the rate is limited by some restriction on the maximum total number of bits that can be used. This total number of bits available, or budget R_T, has to be distributed among the different coding units in a way that the overall distortion is minimized. This problem is finding the optimum operating point, $x(i)$, for each coding unit i such that

$$\sum_{i=1}^{U} r_{ij} \leq R_T \tag{2.2}$$

and a function related to distortion $f(d_{1x(1)}, d_{2x(2)}, \ldots d_{Ux(U)})$ is minimized. This function can be the average distortion:

$$f(d_{1x(1)}, d_{2x(2)}, \ldots d_{Ux(U)}) = \sum_{i=1}^{U} d_{ix(i)} \tag{2.3}$$

Another method for doing this is the technique of comparing and sorting, in which the distortions of two potential solutions are compared and sorted until the one which has the minimum amount of distortion is achieved. As is the property in typical similar problems, the best solution shows some kind of equality of distribution among different coding units. Of course the minimum average distortion problem is more used. However, sometimes this kind of problem with limited rate budget does not simply contain a limit on the total budget but some limitations for some groups of coding units.

In rate-distortion optimal schemes there is also the case where there is a constraint on the buffer size and the time or delay, which is the problem when one deals with a series of video sequences (which is not the issue of our research), but in order to have a thorough review of rate-distortion optimal schemes, the details of it are explained below.

As an example, consider the case when a video encoder compresses 30 frames per second and the system operates with an end-to-end delay of 3 seconds. Then the decoder will wait three seconds to decompress and display the first frame and at any given time there will be $= 3\frac{1}{\frac{1}{30}} = 90$ video frames in the system which are either stored in the encoder or decoder buffers or being transmitted. It is up to the encoder to guarantee that the rate budget for each frame is selected in a way that the frames don't arrive too late to the decoder. Hence, this problem can be stated as trying to select the optimal set of quantizers for which there is no delay in the transmission path from encoder to decoder and meanwhile the total distortion is minimized. Of course if there is no bandwidth constraint which is practically impossible, this problem would be easy to solve. Therefore here comes the issue of investigating channel characteristics in terms of constant or time varying channel delays or constant or time variant bit rate budget or other characteristics, the details of which are beyond the discussion here, because here we intend to do source coding and not error control or channel coding. Nevertheless, we give a general introduction to this case.

If the characteristics of the channel are not known exactly, we cannot deterministically know what the future rates will be, but if we have enough information to model the channel, of course the problem becomes easier.

The solution to this optimization problem is basically founded on on the discrete Lagrangian Optimization. This approach was first used in source coding, following the framework being explained in tree pruning and entropy-constrained allocation problems.

As the basic idea of this technique we can assume introducing a non negative real number as Lagrange multiplier $\lambda \geq 0$, and let us consider the Lagrangian cost $d_{ij}(\lambda)+ \lambda.\mathrm{r}_{ij}$.

Fig. 2.3 shows a graphical interpretation of the Lagrangian cost.

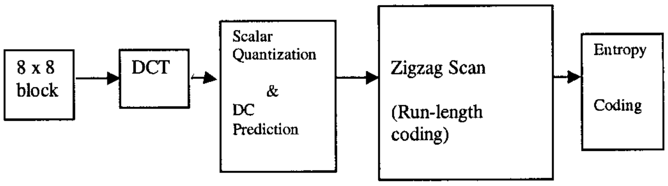

Figure 2.2: JPEG Block Diagram

The figure shows the rate distortion curve and various slopes which give different solutions.

As the quantization index j increases, we have an increase in the rate and proportionally a decrease in distortion. The role of Lagrange multiplier here is to allow us to select different rate -distortion points on the curve. Minimizing the Lagrangian cost $d_{ij}+\lambda.\mathrm{r}_{ij}$ when $\lambda = 0.$, is the same as the distortion minimization because hereby a point closer to the vertical axis is selected.

It means if a particular set of quantizers $x'(i)$ for each coding unit minimizes:

$$\sum_{i=1}^{C} d_{ix(i)} + \lambda.r_{ix(i)} \tag{2.4}$$

then it is also the optimal solution to the budget-constrained problem which was mentioned before, for the special case where the budget is:

$$R_T = R(\lambda) = \sum_{i=1}^{C} r_{ix'(i)} \tag{2.5}$$

so that

$$D(\lambda) = \sum_{i=1}^{C} d_{ix'(i)} \leq \sum_{i=1}^{C} d_{ix(i)} \tag{2.6}$$

for any x satisfying

$$\sum_{i=1}^{U} r_{ij} \leq R_T \tag{2.7}$$

with R being equal to

$$R_T = R(\lambda) = \sum_{i=1}^{C} r_{ix(i)} \tag{2.8}$$

as mentioned above.

We can write

$$\min(\sum_{i=1}^{C} d_{ix(i)} + \lambda r_{ix(i)}) = \sum_{i=1}^{C} \min(d_{ix(i)} + \lambda . r_{ix(i)}) \tag{2.9}$$

because for a given λ the budget constraint has been removed.

Using this, the minimum can be computed independently for each coding unit. For each coding unit i, the point on the rate-distortion characteristic that minimizes $d_{ix(i)}+\lambda.\mathrm{r}_{ix(i)}$ is that point at which the line of absolute slope λ is tangent to the convex hull of the rate-distortion characteristic as in Fig.2.3. For this reason λ is called the slope, and referring to the fact that, λ is the same for every coding unit on the sequence, this optimization algorithm is called the constant slope optimization[16].

This is more used in optimization problems where the cost and objective functions are continuous and differentiable and also for discrete optimization problems as well.

Using the Lagrange multiplier method, finding the best quantizer for a given λ is easy. It is necessary to to find the correct λ for reaching to the best solution at the required rate which means λ must be found such that $R(\lambda)$ as defined above is close or equal to the budget.

In order to reduce the computational burden, if we don't want the rate to be exactly equal to the budget rate, the number of iterations required in searching for λ can reduced. Moreover, in some cases such as video coding applications, where we may be performing allocations on successive frames which have similar characteristics, as a heuristic one can initialize the Lagrange multiplier for a frame with the values at which a good solution was achieved for previous frames, which will reduce the number of iterations. In other words, finding smart ways to guess the initial value of λ can give us a fast way to reach the answer.

Optimization problems with several constraints can be solved using Lagrangian techniques. These approaches are based on generalized Lagrangian relaxation methods in which a different Lagrange multiplier is needed for each constraint and therefore we will have a vector form of lagrange multiplier, but we have to be aware that the search in a multidimensional space is not as simple and fast as it is for just one single Lagrange multiplier.

Usually the practical problems show some specific characteristics that give us some illustrations of the structure of the constraints, and that can be useful to reduce the computation for the search for the vector Lagrange multiplier. Therefore it is very good, because in cases like this, we have at least an initial idea of how to search for finding the optimal vector Lagrange multiplier Λ in an iterative manner.

One significant disadvantage of Lagrangian techniques is that through these techniques it is not possible to have operating points that are not located on the convex hull of the Rate-Distortion characteristics. As another formulation is therefore to formulate the allocations as deterministic dynamic programming problem.

In this case, a tree that represents all possible solutions is considered. Each stage of the tree corresponds to one of the coding units j and each node of the tree at a given stage shows one possible cumulative rate. Fig 2.4 shows how two new nodes are produced with the appropriate accumulated rate by adding the rate for each possible quantizer to the accumulated rate at block $i-1$.

The cost related to each branch is the distortion corresponding to the particular quantizer, and therefore if one travels on the tree from the root to the leaves, one can add up the accumulated distortions for each of the solutions. In the optimization theory and the related graph theory, this is always a good way for representing all possible solutions.

If two different solutions result to the same rate, it is obvious that the solution having higher distortion up to that point must be eliminated or in other terms pruned

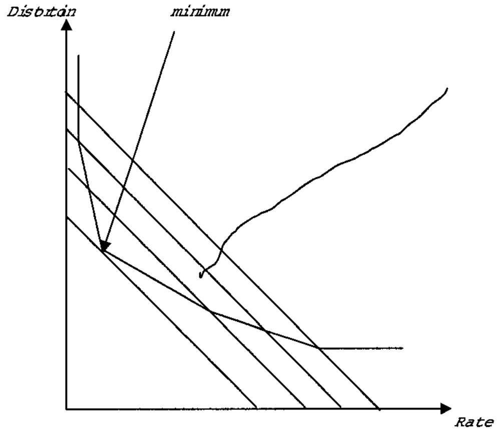

Figure 2.3: The graphical form of Lagrangian cost

from the tree. The Viterbi algorithm or Dykstra algorithm, is suitable to find the shortest path in a graph.

Dynamic programming can also be used for optimal tree pruning applications to prune trees in for example wavelet-packet optimization, or tree-structured vector quantization.

It is also not very difficult to use additional constraints to the tree growth so that our two main problems, namely, budget constraint problem and delay or buffer constraint problems can be solved. For example, to introduce an overall budget constraint, it is enough to prune the branches that exceed the desired total rate allocation or in other words the tree can not grow higher than the already given budget constraint. Similarly, if a buffering constraint exists, then it is necessary to prune the branches that require more than the maximum buffer size at a given stage.

Again to make the long story short about this algorithm, at stage i, for all the nodes which have not been pruned, add the branches corresponding to all quantization choices available at that stage. The rate r_{ij} is decisive in recognizing the end node and the distortion d_{ij} is added to the path distortion. Then the branches that have rates above the rate constraint are pruned and then, for each remaining node at stage $i + 1$, only the branches which have the lowest cost are kept.

The complexity of dynamic programming grows exponentially with the number of coding units but the complexity of Lagrange Multiplier has a linear proportion with the number of coding units. The Lagrangian optimization is good in terms of complexity. But, the Lagrangian approach has one drawback in that only points in the convex hull of the global operational rate-distortion characteristic are accessible. To overcome this problem, one idea might be to try to have as many points as possible on the curve.

Now we are considering the problems that are to some extent dependent. Until now, the discussion was based on the assumption that selection of the coding mode for each unit can be made independently without influencing the other units.

It must be remembered that sometimes this assumption is not a correct assumption. This is usually the case in coding schemes using some sort of prediction. For example, assume that each coding unit i is predicted from the preceding coding unit $i-1$. The predictor is constructed using the past quantized data, and thus we code $X_i - P(\widehat{X}_{i-1})$;i.e., the prediction error. When quantization is used, the prediction error and accordingly the rate-distortion operating points which can be potential solutions for i depend on the previous decisions for quantizer for $i-1$. Every different selection of $x(i-1)$ causes a different characteristic.

Techniques that consider the two frames independently would select (for the given slope λ)quantizer 2 for both frames;i.e.,it would incur a cost $J_1(2)$for some λ then, given that quantizer 2 was selected for frame 1, would choose the minimum among all $J_2(2,x)$, which turns out to be $J_2(2,2)$. However, in this particular example, the greedy approach, allocating first for frame 1 and then for frame 2, can be outperformed. The better overall performance can be achieved when quantizer 1 is used for the first frame and quantizer 2 is used for the second. Even through $J_1(2) \ < \ J_1(1)$ we have that $J_1(2) + J_2(2,2) \ > \ J_1(1) + J_2(1,2)$.[1]

Here we consider as examples which shows some kind of dependency problem which is incorporated in MPEG video coding to make the matter more clear. Here are introduced briefly the tree based dependency and remind that trellis based dependency is also a topic in these sorts of problems. The selection of macroblock-level quantization in an MPEG video stream is a dependent problem because the rate r_{ij} for macroblock i and quantizer j depends on the quantizer chosen for macroblock $i-1$. This is because predictive entropy coding of the quantization indices is used to increase the coding efficiency. In this situation, it is possible to represent the possible selections as a trellis where each state represents one quantizer selection for a given macroblock, with each stage of the trellis corresponding to one macroblock. Dynamic programming can then be used to find the minimal cost path in this trellis, where the branch cost is typically defined as the lagrangian cost introduced above. Taking the dependency into account avoids greedy selection of coding parameters, where the quantizer assignment is optimized for the current coding unit alone.

In general, trellis-based dependencies arise in cases where the underlying structure is such that the memory in the system is finite i.e., coding choices for i depend only on a finite set of previous coding units and the number of possible cases is also finite. In other words, in this case, the available coding parameters for a given coding unit depend on the state of the system or the finite set of parameters that completely determine achievable values. For those type of dependencies one can use a dynamic programming approach, where the state corresponds to the state of the system, and branches (each corresponding to a choice of quantization) have associated a Lagrangian cost that combines the rate and distortion for the given parameter choices.

The second example to be considered for this sort of problem is the tree based dependency problem and can be seen when the effect of motion compensation in an MPEG framework is analyzed. After motion compensation, the encoder transmits the difference between the previously decoded frame and the current frame which is then compressed and sent to decoder to tell of how much error exists in the prediction and based on this information, the decoder can retrieve the frame at the decoding stage. This residual frame will depend on how the quantization parameters for all previous frames since the last INTRA-frame are chosen. All possible combinations generated by successive quantizer choices can be demonstrated as a tree for which the number of branches growing exponentially with the number of levels of dependency, which is in this case the number of frames since the last INTRA-frame.

2.4.1 Application to Practical Image Compression

Now that we have explained enough of the mathematical theories and concepts in an abstract form, it is time to investigate how the above mentioned theories can be adapted and applied to the practical issue that we are dealing with, which is stereo image compression. This is very similar to the ways this theory has been applied before to image and video compression and therefore having a good knowledge of the applications of these concepts to image and video compression, can give us some insight and ideas for stereo image compression.

For the first type of problems that we considered that were budget constrained optimization problems, in the context of image and video coding, we consider transform coding, which was introduced earlier related to DCT coding.

Again we explain that the budget-constrained optimization problem, which is a topic in source coding theory, tries to minimize the quantization distortion subject to a bit-rate constraint. To formulate the problem in a very general form, it is the problem of selecting the best combination of transform, quantizer, and entropy coder in a rate distortion framework. Depending on the real case, one or all of the above parameters can be jointly optimized. Sometimes for example, one of the parameters like the transform is fixed, and the two other parameters, for example quantization and entropy coder, are jointly optimized. The quantization modes can be in the form of scalar quantizers or vector quantizers. Scalar quantizers are simpler; vector quantizers are more complex, but if they are designed well, for example in the case of our proposed algorithm for quantizing the disparity vectors which have two (vertical and horizontal) components, it can cause to reduction of bit rate very well. One method for rate-distortion optimization is to keep one parameter fixed and changing the other ones, but joint optimization is a more developed method. An example for this can be coding standards like JPEG, where the quantizer choice (8×8 quantizer matrix for the image)and the entropy coding choice(Huffman table)can be optimized for each image separately. A more detailed description of JPEG was presented previously in this chapter.

In a similar manner, optimization techniques can be applied with very good performance for video coding frameworks such as MPEG. Of course, it must be mentioned that in video coding the temporal redundancy plays a much more important role than the spatial redundancy and therefore here we have little to do with this, because here in stereo image pair coding, the focus of our attention is naturally the spatial redundancy existing between the two right and left frames. There has been a great deal of research done in the field of video coding on rate-distortion based techniques for variable-block-size motion compensation and also on variable bit-rate motion vector encoding based on DPCM.

Further, rate-distortion optimization techniques can be applied to shape coding, where trade-offs between the fidelity of the shape representation versus the bit rate needed to represent the shape can be optimized. These techniques are likely to be used in standards such as MPEG-4, which have support for video objects, which is not the topic of this research since we use the frames and blocks, not objects, as our basic unit. On the contrary to fixed transform based rate-distortion optimization, there is also adaptive transform based rate-distortion optimization.

2.5 Summary

Fast Block Matching Algorithms can fall into different groups based on their characteristics: 1)Block Matching Algorithms with unimodal error surface assumption such as Logarithmic Search and Three Step Search, 2)Block Matching algorithms based on pixel-sampling, 3)Hierarchical or Multiresolution Block Matching Algorithm and 4)block matching algorithms using spatial and temporal correlation. The first group of Block Matching Algorithms can be trapped in a local minimum with a large amount of matching errors because their basic assumption(unimodal error surface) often does not hold in practical video sequences. The second group of block matching algorithms has a limitation in computations reduction, because excessive sub-sampling may cause a local minimum phenomenon which leads to incorrect results. Instead, this group of Block matching algorithms is often exploited as a good tool that can be incorporated in several existing fast algorithms to obtain an additional complexity reduction. The third group of block matching algorithms works relatively well and provides fast computation. Usually, it only uses motion vector information from coarser levels for motion vector refinement at finer levels without considering other useful information such as spatial/temporal correlations. However these Block Matching Algorithms are not always very reliable. In the fourth group, these methods predict a starting point estimate by using spatial or temporal correlations, and also have reliable performance. In general, these methods predict an initial estimate by using spatial or temporal correlations at the first stage, and then refine the estimate within a smaller search area in the next stages. However, since the schemes assume that motion vector field always has strong spatial/temporal correlations, if this assumption is not true, the fall in their quality has bad effects on video frames. Moreover, due to their irregular data-flow, they are not suitable for hardware implementation.

Regarding motion estimation, block matching algorithms that could compensate the drawbacks in both full search and fast block matching algorithms were described. The methods were shown to enhance the complexity of computation and the local minimum problems.

Afterwards an efficient method for searching full resolution motion vectors without using motion vector decimation at all was introduced. Very briefly said, after the search at the coarsest level two or three initial motion vector candidates are chosen for the next level. At the following levels, the motion vector candidates are defined within smaller search ares. This technique was also proved to be faster than Full Search but had a slightly smaller PSNR. It was also good to be implemented by hardware.

Then a review of rate-distortion optimal schemes was presented, especially considering the practical usage of this theory to image coding. First the relationship between these techniques and rate-distortion theory which is a part of information theory was presented. Standards-based image codings can use optimization techniques as far as the encoder has to optimize the choice of its coding parameters. Following this the resource allocation problem in the general form was presented together with two of its examples which were budget-constrained allocation and delay-constrained allocation. After that Lagrangian optimization and dynamic programming techniques were explained, as useful and good tools for solving the allocation problems. In [6] an efficient motion estimation algorithm has been proposed which minimizes the rate subject to a constraint on overall distortion

The most useful concept that can be understood from investigating different sorts of dependent coding problems with an application to an MPEG framework is the fact that the number of branches grows exponentially. But the promising point is that through using good heuristics, it is possible to make some approximations in order to reduce the search load for the optimal solution. For example, heuristics that consider the fact that the finer the quantizer, the smaller the error or greedy approaches where, for example, only a few quantization choices are kept at any given stage. One other technique to overcome this problem is by using models of the dependent Rate-Distortion characteristics so that not all the operating points in the tree need to be explicitly computed, or by considering models of the rate and assuming a good estimate of quality being the quantizer characteristics.

Chapter 3

Rate-Distortion Optimal Variable Block Size Disparity Estimation Method

In this chapter, a detailed description of the method proposed in this thesis will be presented.

Several methods and algorithms which also have many similarities have been developed for motion estimation. Of course many of these methods are based on the practical experiences of their developers regarding some special applications. The approach that has been adapted in this thesis, is to formulate the problem mathematically as an optimization problem and then trying to solve it using the techniques existing in optimization theory instead of just trying to follow the less fruitful method of only doing simulations on every work that had been done in this area.

This method considers jointly rate-distortion optimal selection of coding parameters in a general disparity compensated coder. The general coder uses variable block size disparity estimation and residual coding. This is essentially the optimal bit allocation problem for a frame at a given rate constraint. This method not only gives the general formulation and solution using the Lagrange multiplier method and dynamic programming, but also shows how the general theory can be adapted and applied to a JPEG like coder. Experimental results are presented in the next chapter. The technique can also be applied to JPEG compliant coders with fixed block size disparity estimation. This approach can also be applied to distortion constrained coding, and therefore allows a fine tuning of either the rate or distortion.

3.1 An Introduction to the Technique

The block diagram of the method is shown in Fig.3.1.

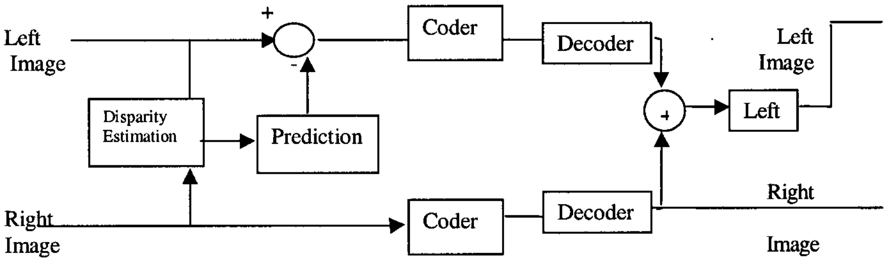

Figure 3.1: The general scheme for disparity compensated stereo image pair compression

A disparity compensated predictive coder consists of disparity estimation, disparity compensation and transform coding modules. This provides spatial redundancy reduction between the two frames. For simplicity and efficiency, this is usually performed on blocks rather than on individual pixels. The image being encoded is divided into either fixed or variable size blocks and then the disparity vector for each block is estimated using block matching. Compared to using fixed block size disparity estimation, variable block size disparity estimation is more general and shows a better adaptation to disparity discontinuities. Fixed block size disparity estimation can be considered as a special case of variable block size disparity estimation.

JPEG-like standards use a fixed block size method. After this phase, the displaced frame difference or DFD is formed as the residual of the current and predicted image. Then transform coding uses spatial redundancy that exists in the DFD. We are using DCT and then the coefficients are quantized and variable length coding is done on them. Each block in this method is labelled as corresponding to one of several modes which are going to be explained later in detail and encoded according to that. Then quantization and coding are performed differently for each block considering its mode.

According to [27] regardless of the specifics of disparity estimation, disparity compensation, transform coding or mode selection, the performance of the coder in the sense of rate-distortion depends on the complete set of parameters from the modules mentioned above. For example among these parameters we can mention the variable block size disparity estimation structure, the disparity vectors and the mode selection for each block. Parameters are chosen based on distortion. For example the disparity vectors are found using MAE as the matching criterion; a buffer constraint based rate control which does not depend on disparity estimation or disparity compensation can be used to select the quantization scaling factor for scaling the DCT coefficient quantization matrix; and the mode selection for each block is based on distortion.

Because the goal of the coder is to minimize the distortion given a rate constraint or to minimize the rate given a distortion constraint, it is inevitable to do rate-distortion optimization to have a good coding performance. But selection of one of the parameters influences the others and therefore it is not very efficient to optimize the parameters separately and in this proposed method the parameters are optimized jointly. Here the method seeks optimal bit allocation for an image. Considering the corresponding methods for motion estimation, the method of Lagrange multiplier and dynamic programming is the most commonly use approach.

In this thesis a rate-distortion optimized approach is adapted for choosing the coding parameters in a joint manner. The structure of quadtree is used for the variable block size disparity estimation and the parameters of each node can be the mode and the displaced frame difference and the disparity vector. This implementation has the benefit of compatibility with JPEG like coding. It is also to be mentioned that fixed block size is a special case of variable block size in which there is no splitting.

Another reason for using variable block size disparity estimation is the tradeoff that it presents between the bits considered for disparity vectors and the bits considered for the displaced frame difference or DFD and hence is suitable for the optimization of parameters in a joint manner. In this method the concept of rate-distortion which is a topic of information theory is used. It other words, this is a parametric decision problem where different sets of parameters which are decisive in the rate-distortion function can be selected in a joint manner.

The problem is formulated by the Lagrange multiplier method which converts the constrained optimization problem to an unconstrained one and it is solved by the technique of dynamic programming which is a successful method for solving similar resource allocation problems in optimization theory. This method is applied to variable-block-size disparity estimation with DCT DFD coding in order to get the practical results which are presented in what follows.

3.2 General Problem Formulation

The structure of a quadtree is used for variable block size disparity estimation . Each block in the quadtree is a *tree node.* A block is called a *parent node* if it is further divided into four nodes, which are named *child nodes.* A block is a *leaf node* if it is not further subdivided. In the general theory, the tree can be represented by a series of bits that indicate ending by a leaf with 0 and by being a parent node with 1. A frame is split into 32 × 32 superblocks and then each superblock can be segmented into 16 × 16 and 8 × 8 blocks. The reason for choosing this structure is the good tradeoff that it presents for prediction and overhead information.

Assume T is a particular tree structure for a frame which belongs to the set of all possible quadtree structures $\mathbf{T}$. Let $N(T)$ be the number of tree nodes in T.

Assume there are K parameters which affect the rate and distortion for a given tree node and P is the parameter set associated with all the tree nodes in T. For example a disparity vector, a label for mode and quantization scaling factor. Each block also has a rate and distortion associated with those parameters.

3.2.1 Lagrange Formulation

Using the Lagrange cost function

$$J(T,P) = D(T,P) + \lambda R(T,P) \tag{3.1}$$

the above constrained optimization problem can be converted to an unconstrained problem using the Lagrange multiplier method and becomes

$$\min_{\mathbf{T}} \min_{\mathbf{P}} J(T,P) = \min_{\mathbf{T}} \min_{\mathbf{P}} (D(T,P) + \lambda R(T,P)) \tag{3.2}$$

The solution to equation (3.2) named as (T^*,P^*) is also a solution to (3.1) for the case of $R(T^*,P^*) = R_{budget}$.

The distortion of the reconstructed frame is the sum of the distortions of individual nodes.

3.2.2 Dynamic Programming Optimization

The above optimization problem can be solved using dynamic programming. In this problem once the parameter set s_n for the nth block is known, the distortion and the DFD coding rate of the block depend entirely on s_n. The one dimensional dependency is introduced only by the differential coding of the s_n. It can be shown that a dynamic programming recursion formula can be established and a trellis can be constructed where states of the trellis represent all the possible choices of s_n. By relating each arc between trellis states with the Lagrange cost, the deterministic finite-state optimization problem is converted to the shortest path finding problem. The special case of forward Dynamic programming, known as the Viterbi Algorithm (VA), can be used to find the optimal solution. Using the VA, only one incoming path (the minimal cost path) is kept for each trellis node at each stage. The accumulated cost and the path are recorded. The shortest path can be found at the end of the trellis along with the optimal trellis states which are back traced.

The intuitive approach for solving (3.2) by the Viterbi Algorithm is to first grow the trellis for a given tree T and find the shortest path for T. Then the optimal solution is found by performing the minimization for all trees within the possible quadtree set $\mathbf{T}$. For a given tree, let's consider the ith stage in the trellis. At each stage, there are M_s trellis states, each representing a set of control parameters s for the corresponding block.

Since the superblock structure is used in here, an alternative approach is adopted which forms a multilevel trellis by incorporating the possible subtree structures into the trellis. This is illustrated in Fig.3.1. The black circles inside the rectangle are the trellis nodes where each node represents one particular choice of s (for example, a combination of disparity vector, mode, and quantization step size which is used to encode the block. Each node also contains the resulting distortion and rate. The distortion and DFD coding rate are dependent only on the current s while the overhead rate for coding these depends on that of the preceding trellis node too.

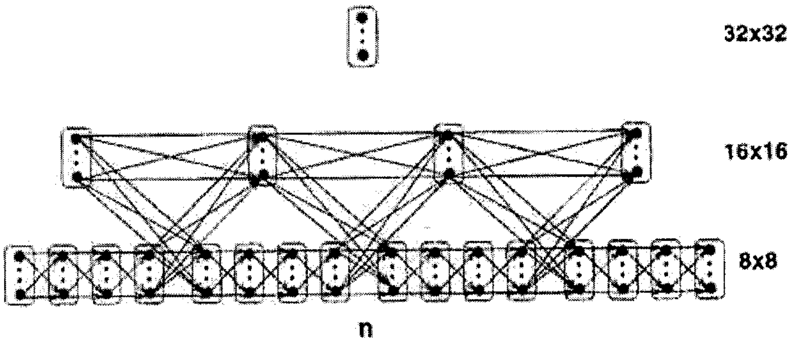

Figure 3.2: The trellis structures for an individual superblocks from [27]

In this figure, the top structure shows the multilevel trellis for an individual superblock. It consists of trellis nodes of all the possible 3 level quadtree decomposition from the 32×32 level to any of the 16×16 or the 8×8 levels in the same superblock. Thus the optimal path is forced to select only valid quadtree structures. The bottom one shows the trellis structure for three consecutive superblocks in a scanning order.

After drawing the trellis structure, in order to find the shortest path for a given λ

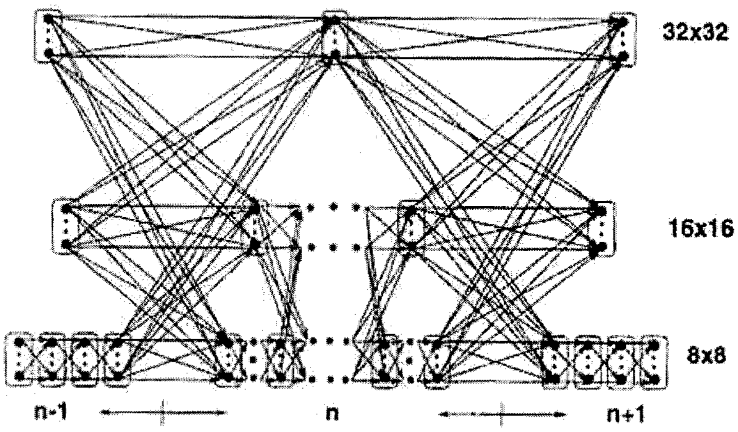

Figure 3.3: The trellis structures for three consecutive superblocks from [27]

the Viterbi Algorithm is used. By iterating through λ, we can get the optimum solution which for the specified bit constraint, gives the minimum amount of distortion.

The superiority of this method is that it is controlled by λ but there are some other strategies which don't follow a strong mathematically proven basis and just use some ad hoc techniques for controlling for example the step size of the quantizer which is just a single parameter. But following the above formulation and solution, gives us a general rate-distortion optimization performance framework which to some extent guarantees the optimal solution as is the typical characteristic of every mathematical based modelling and solution method.

Of course if we have a constraint on distortion then the problem can be formulated accordingly and the same procedure can be followed.

3.3 Application to JPEG-Like Coder

In this section the main goal is to show how the above mentioned theory can be applied to a JPEG-Like coder including the selection of parameters and calculating the rate and distortion for constructing the trellis.

JPEG-Like coding is introduced in Chapter 2. We note here that in this thesis the method is applied to luminance only because for the chrominance components the application would be the same. Therefore we change the color image first to a black and white image and apply the technique on that.

The block diagram of the method is shown in Fig.3.1.

A disparity vector for each block is determined by block matching using MAE. The disparity vector is quantized. Of course in developing the whole algorithm the special intrinsic characteristics of stereo image pairs have been used. As an example considering the camera alignment and the nature of producing stereo images it is quite reasonable to assume that the horizontal component of disparity is greater than the vertical component. Though the amount of vertical disparity is much smaller than the horizontal disparity, the search area is considered to have a ratio of 2:1 for width to height. The components of the disparity vectors can either be positive or negative and hence the last bit is the sign bit.

In this method, determining the disparity vector and choosing the mode is based on distortion. The data from the two images is saved in the buffer of each image. Note that the image is saved as a multiple of 32 in height and 32 in width and the rest of the data is discarded. This has been done for the sake of simplicity, because the image is divided into 32×32 blocks and the processing is done on these blocks. Of course for an application, the program can do the necessary adaptations.

The Coder in this software runs the Rate-Distortion Algorithm and saves the coded image and also saves the decoded image. The main modules that constitute the whole program are the ones to compute disparity vectors and Mean Average Error for each block, the optimal quadtree structure selector for a frame according to the Lagrangian cost, and the DFD's arranger for blocks in the block array. The decoder program decodes the left image from right image, the disparity vectors and DFD's and saves it to the "output.bmp" file. It also calculates the distortion between the left image and the decoded image and returns it. Another function calculates the bit stream for DFD's. If skip mode is selected, disparity vectors and mode are saved. Otherwise the DCT for each DFD is computed and quantized and the coefficients are arranged in semi zigzag order and coded with variable length coding(VLC). The method here is Huffman coding. The DFD's are added to the bit stream as unsigned characters and extracted in the decoder. For the disparity vectors, the signed character that is considered, is added to the bitstream and then at the decoder extracted from the bit stream.

The left image is divided into 32×32 blocks as superblocks. The superblock is divided into blocks of 16×16 and 8×8 and for each block, by doing the comparison in the right image, the disparity vectors and the average error are calculated. For calculation of the disparity vector, a block of the left image is compared image with similar blocks and shifted block in four directions. This shift amount has been found by choosing different possible values and selecting the most suitable one. The comparison is done by finding MAE between the two blocks and the selected disparity vector is the one where the MAE is minimum. In implementation, the edges of the image had to be considered separately and this special case was considered. In the next step which we denote as the third step, the proper combination for each superblock must be selected. The criterion of selection is MAE of the blocks. We have seven types of blocks: In the first type, the MAE of the superblock is negligible (negligibility is selective and in rate-distortion is chosen by a parameter). In this case the DFD is not saved and the block takes number seven. If the MAE is not negligible, the MAE of the superblock is compared with the MAE of the sum of the sixteen subblocks; if it is less than their sum added to a variable, which can be chosen and is effective in calculation of the Lagrangian cost, the block is assigned number three and its DFD is saved. The third type is blocks of size 16×16 such that their MAE is ignorable; they are identified with number six and their DFD is not saved. The fourth type are the blocks of size 16×16 such that their MAE is not ignorable and is compared with MAE of the sum of the four subblocks. If it is less than their sum added to a special variable, which is selectible and is effective in calculating the Lagrangian cost, the block is assigned number two and its DFD is saved. The fifth type consists of blocks of size 8×8 such that their MAE is ignorable that are characterized by number five and the DFD is not saved. The sixth type includes blocks of size 8 × 8 such that their MAE is not ignorable and that are identified with number one, and their DFD is saved and anothe type is the default type. This is done for all superblocks of the left image. In the fourth phase, coding is done. If the blocks are of type five, six or seven, three bits for them identify the type of block, six bits for the disparity vector in horizon and five bits for the disparity vector in height is considered which in other terms implies the before mentioned ratio of 200 percent. In our method for the purpose of generalization in case of imperfections in camera alignment or other nonideal conditions, the disparity vector components can take on either negative or positive values and therefore a sign bit is allocated to each component. If the blocks are of type one, two or three, in addition to the above coding, the DFD is also coded such that first, its DCT is calculated and then the coefficients are quantized and are arranged in a semi-zigzag order and VLC is done upon them. In this way, the coding is accomplished and the coding rate is calculated.

In the fifth step or optimizing the rate-distortion, the image is retrieved and its distortion is calculated in comparison with the left image. If the optimum condition is not satisfied the steps three to five are repeated.

For the disparity vectors, the signed character that is considered, is added to the bitstream and then at the decoder extracted from the bit stream. Among other parts of the program is the one than extracts the DCT coefficient from bit stream and saves them.

This method has been applied to several stereo pairs and the experimental results are presented in the next chapter.

Chapter 4

Experimental results

In this chapter, results of compression of the left image in a stereo pair are presented, because the right image is compressed using intraframe coding.

The procedure for illustrating the results is by depicting the PSNR versus rate curves, which are obtained by varying parameters affecting the compression algorithm such as the quantizer step size.

As a benchmark coding of left image is implemented using MATLAB, using imwrite function with different values of the quality parameter. For the disparity compensated predictive coding, which was implemented in software using Visual C++ programming language on a Pentium Computer under Windows, the cases of fixed and variable block size disparity estimation are considered. The relationship for calculating the PSNR was $20\log_{10}(255/\mathrm{RMSE})$ where RMSR is the root mean squared error. The elements which add up to form the total rate were explained in detail in Chapter 3.

The experimental results, as mentioned previously, were performed on the luminance component of the image obtained from the RGB image using the coefficients 0.299, 0.5876 and 0.114. The stereo pairs on which the simulations were done have been depicted in the Appendix.

The PSNR as a function of rate for the test images using JPEG, fixed block size and variable block size disparity estimation have been depicted in Figs 4.1, 4.2, 4.5, 4.8, 4.9 and 4.10. All the results show a significant improvement resulting from disparity compensation over intra frame coding, as expected. Illustration of coded images with JPEG and variable block size method with similar bit rates are shown in Fig 4.3, 4.4, 4.6 and 4.7 which also show very impressive results. The comparison of fixed block size (32 × 32) and variable block size using the rate-distortion technique is shown in more detail.

As the curves show more in details, the quality is improved by using the method of variable block size, in comparison with simply using a fixed block size scheme. Of course based on different reasons like for example the spectral characteristics of each image, the amount of compression and quality improvement differ as is also quite an expected result introduced by applying intra frame JPEG coding.

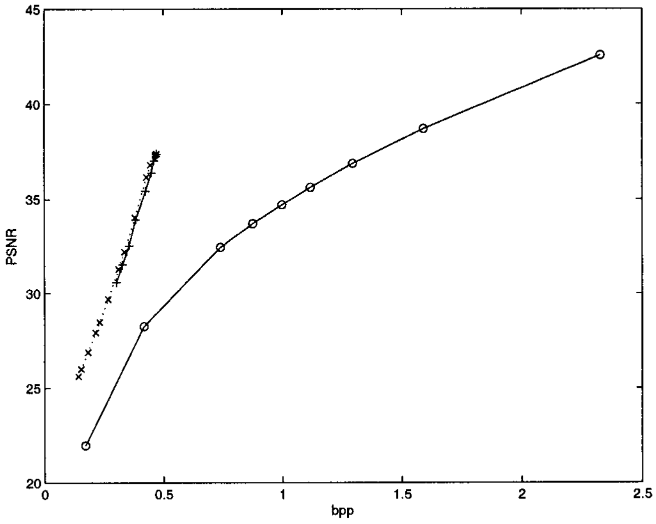

Figure 4.1: Result of applying the method using fixed(+) and variable(x) block size cases to “manege” stereo pair in comparison with JPEG coding

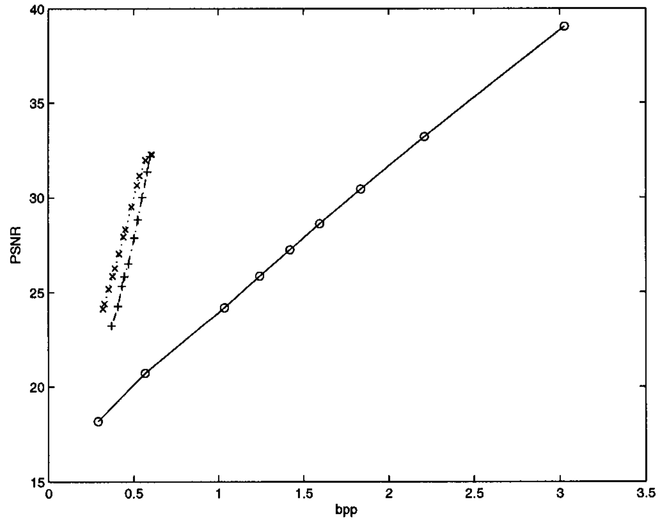

Figure 4.2: Result of applying the method using fixed(+) and variable(x) block size cases to "redcar" stereo pair in comparison with JPEG coding

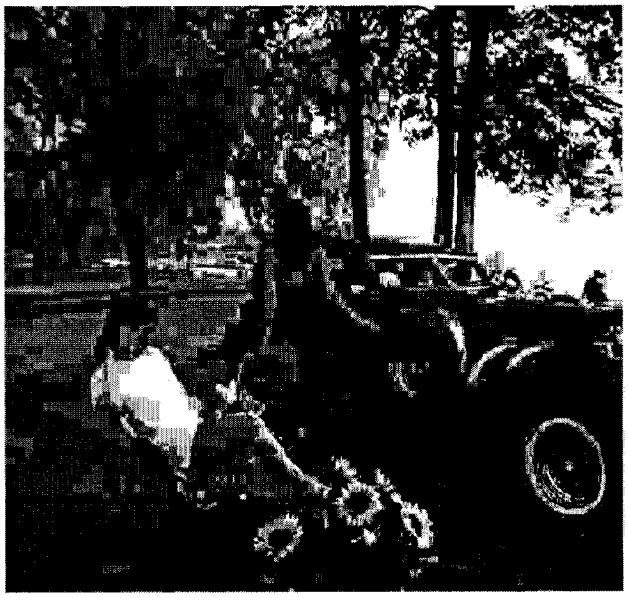

Figure 4.3: Coded image with JPEG for "redcar" with 0.1933 bpp

Figure 4.4: Coded image with variable block size method for "redcar" with 0.1153 bpp

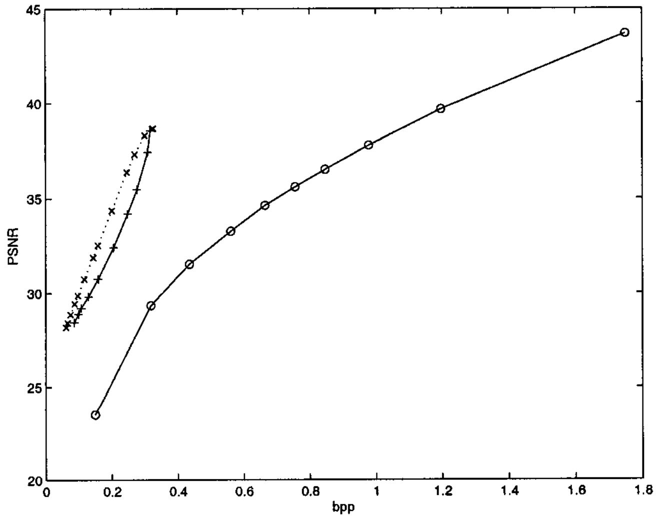

Figure 4.5: Result of applying the method using fixed(+) and variable(x) block size cases to "trapeze" stereo pair in comparison with JPEG coding

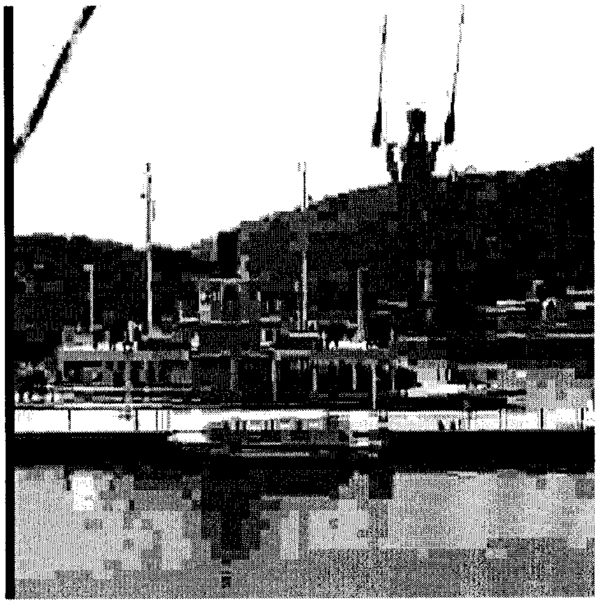

Figure 4.6: Coded image with JPEG for “trapeze” with 0.1496 bpp

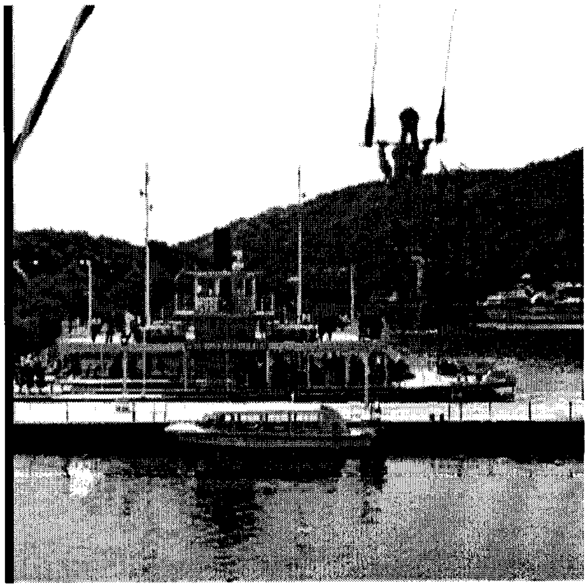

Figure 4.7: Coded image with JPEG for “trapeze” with 0.0616 bpp

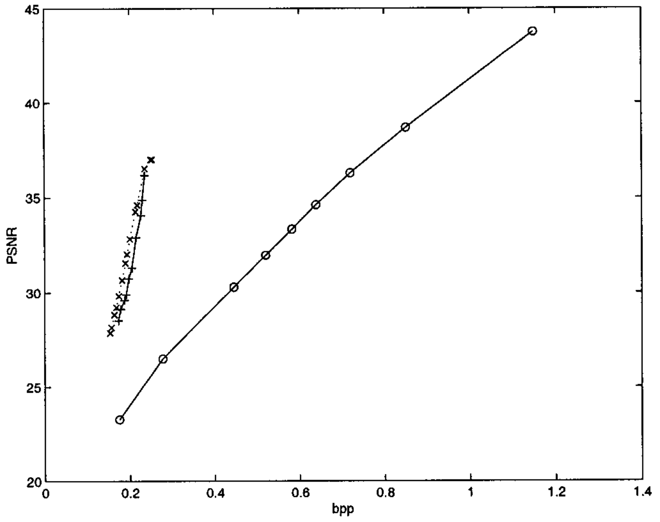

Figure 4.8: Result of applying the method using fixed(+) and variable(x) block size cases to "lab" stereo pair in comparison with JPEG coding

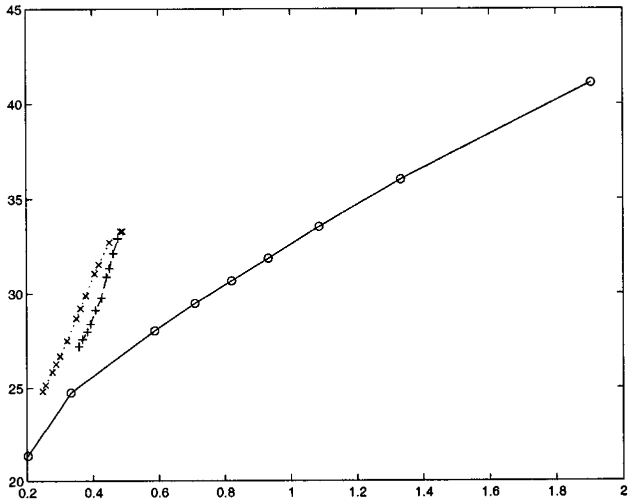

Figure 4.9: Result of applying the method using fixed(+) and variable (x) block size cases to "book" stereo pair in comparison with JPEG coding

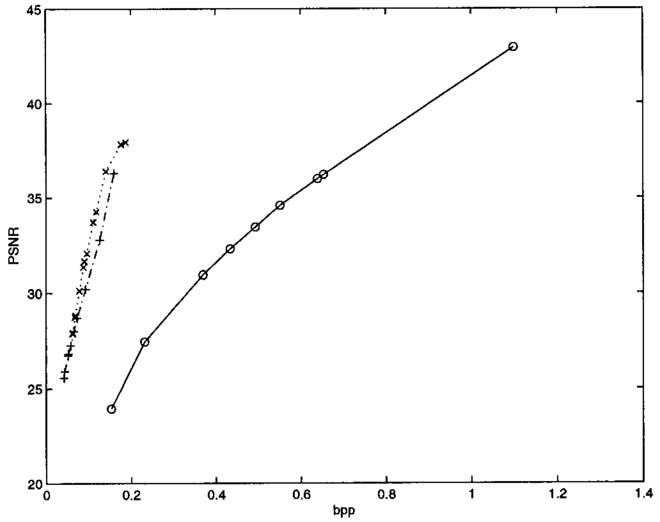

Figure 4.10: Result of applying the method using fixed(+) and variable(x) block size cases to "ball" stereo pair in comparison with JPEG coding

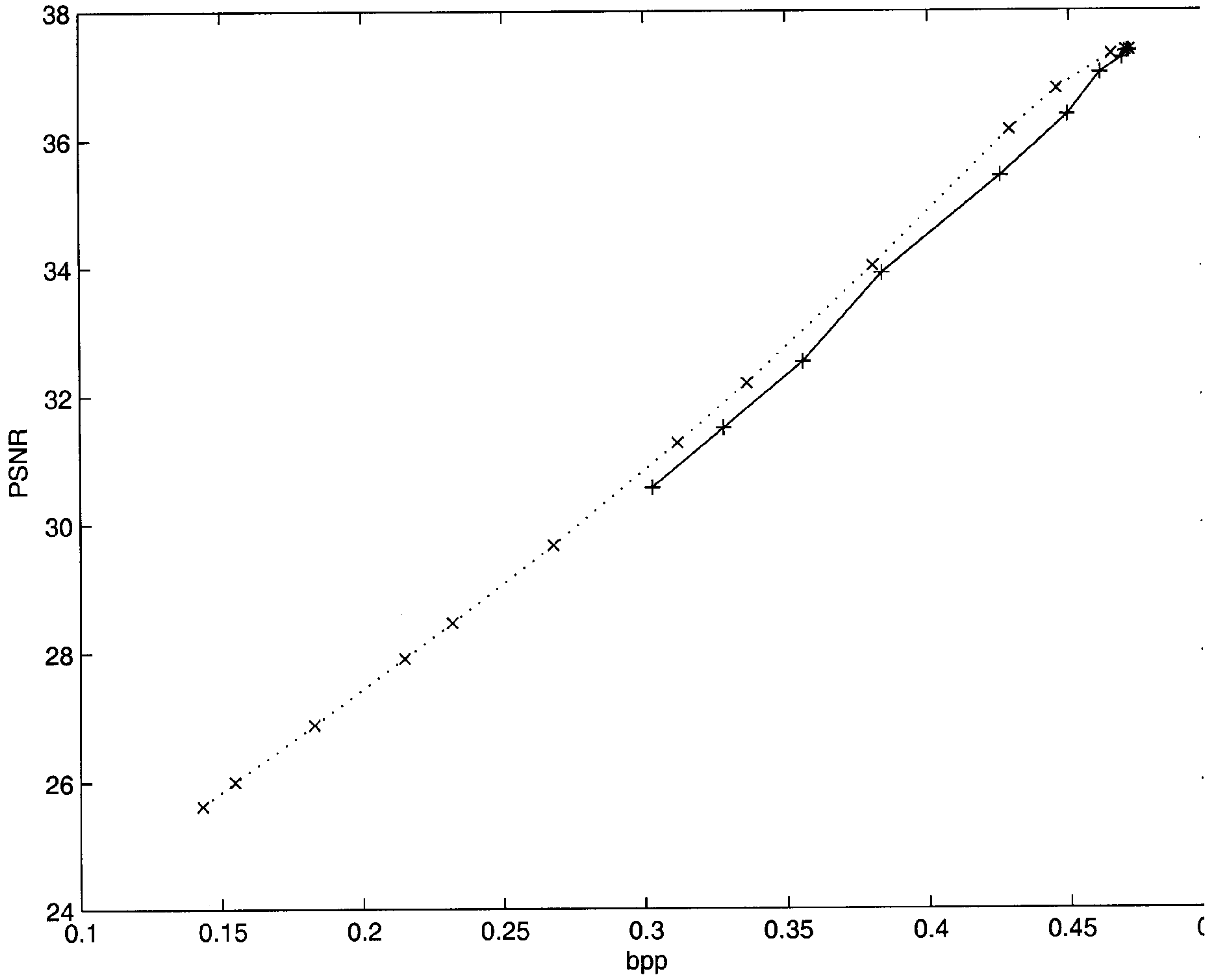

Figure 4.11: Result of applying the method using fixed(+) and variable(x) block size cases to “manege” stereo pair

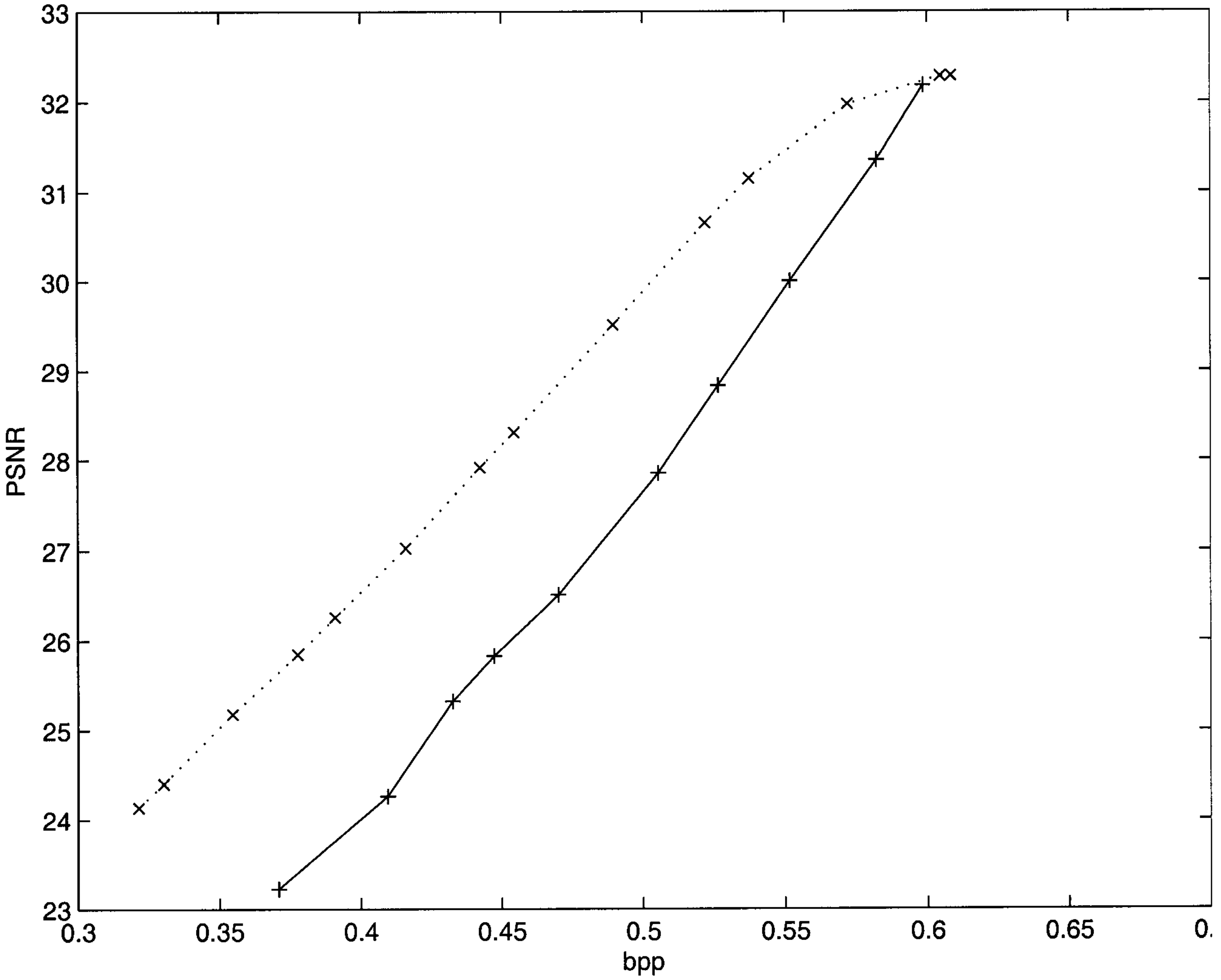

Figure 4.12: Result of applying the method using fixed(+) and variable(x) block size cases to “redcar” stereo pair

Chapter 5

Conclusion

5.1 Summary of the Thesis

In this thesis, an introduction to the problem of disparity compensated compression of stereo pairs was presented, then the problem was formulated and basic and fundamental concepts regarding this problem were introduced. Different methods for disparity estimation, motion estimation and rate-distortion based methods, were reviewed. Then the proposed method was introduced and the experimental results were analyzed.

5.2 Thesis Contribution

In this thesis, the method of variable block-size motion estimation in rate-distortion framework was extended to the problem of disparity estimation.

A simple and efficient algorithm based on an overlap of different theories like information theory, optimization theory specifically the Lagrange multiplier and dynamic programming was developed and implemented as a software package in Visual C++ version 6 programming language and was applied to several stereo image pairs.

From the experimental results we found that basically, especially at lower rates, if we compare the method of fixed block size disparity estimation to standard JPEG coding, the method of fixed block size shows a superiority to usual JPEG coding in terms of PSNR or in other words distortion. Furthermore for our method of variable-block size there is even further improvement in terms of peak signal to noise ratio. Therefore the experimental results show that an enhancement in terms of distortion versus rate has been achieved by this method.

5.3 Future Work

Although, motion estimation for video coding is almost a mature field, and there has been a lot of work done in this area based on different techniques like rate-distortion criterion or fixed or variable block size or object based methods, disparity estimation in stereoscopic imagery is still a field that has to be explored using different techniques like block based methods and different criteria like rate-distortion constraints.

Among the other techniques that still need to be analyzed and investigated, we can mention very briefly the overlapped block disparity estimation[25], working on the issue of scalability [9],[22], reduction of blocking effects that may exist due to DCT-based compression[7], applying the method for wavelet coders[26] and performing the DFD coding, considering the characteristics of disparity-compensated stereo image pair residuals [14].

Bibliography

[1] A.M. Ariyaeeinia, "Analysis and design of Stereoscopic television systems," *Signal Processing: Image Communication,* Vol.13, pp.201-208, 1998.

[2] M.C. Chen and A.N. Willson, "Rate-Distortion Optimal Motion Estimation Algorithms for Motion Compensated Transform Video Coding," *IEEE. Trans. Circuits, Syst. Video Technol.,* vol. 8, NO. 2, pp. 147-158, April 1998.

[3] R.J. Chen and B.C. Chieu, "A fully Adaptive DCT based color image sequence coder," *Signal Processing: Image Communication,* vol.6, pp.289-301, 1994.

[4] K.W. Chun, J.B. Ra, "An improved block matching algorithm based on successive refinement of motion vector candidates," *Signal Processing: Image Communication,* vol.6, pp.115-122, 1994.

[5] H.Y. Chung, P.Y.S. Cheung and N.H.C Yung, "Adaptive Search Center Non-linear Three Step Search," *IEEE Trans. Image Processing,* vol.9, no.6, pp.982-993, 2000.

[6] W.C.Chung, F.Kossentini and M.J.T. Smith, "An efficient motion estimation technique based on a rate-distortion criterion," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* , vol.4, pp.1555-1559, 1996.

[7] F.Coudoux, M.Gazalet and P.Corlay, "Reduction of blocking effect in DCT-coded images based on a visual perception criterion ", *Signal Processing: Image Communication,* vol.11, p179-186, 1998.

[8] R.A Earnshow and J.A. Vince, *Multimedia Systems and Applications,* Academic Press, 1995.

[9] S.S.Hemami, "Distortion Analyses for Temporal Scalability Coding Techniques," In *Proceedings IEEE International Conference on Image Processing,* vol.3 pp.349-353, 1999.

[10] J. Lee, "Joint Optimization of Block Size and Quantization for Quadtree Based Motion Estimation," *IEEE Transaction on Image Processing,* vol.7, pp.909-912, June 1998.

[11] C.W.Lim, N.C.Kim, S. C. Jun and C.S.Jung, "Rate-Distortion Based Image Segmentation Using Recursive merging," *IEEE Trans. Circuits, Syst. Video Technol.,* vol.10, No. 7, pp.1121-1134, October 2000.

[12] B.Macq, M.P.Quelz and B.Simon, "Very low bit-rate image coding on adaptive multigrids," *Signal Processing:Image Communication*, vol.7, pp.313-331, 1995.

[13] H.B. Mitchell, N. Zilverberg, M. Avraham, "A Comparison of different block truncation coding algorithms for image compression," *Signal Processing: Image Communication,* vol.6, pp.77-82, 1994.

[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, Jan.1998.

[15] A. N. Netravali and B. G. Haskell, *Digital Pictures Representation, Compression, and Standards,* New york: Plenum Press, 1995

[16] A.Ortega and K.Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine,* vol.15, pp.23-50, November 1998.

[17] A.Ortega and K. Ramchandran, "From Rate-Distortion Theory to Commercial Image and Video Compression Technology," *IEEE Signal Processing Magazine,* vol.15, pp.20-22, November 1998.

[18] M.G. Perkins, "Data Compression of Stereo Pairs," *IEEE Transactions on Communications,* vol.40, N. 4, pp.684-696, April 1992.

[19] E.A. Riskin and R.M. Gray, "A Greedy Tree Growing Algorithm for the Design of Variable Rate Vector Quantizers," *IEEE Transactions on Signal Processing.,* vol.39, pp.2500-2506, November 1991.

[20] K. Sayood. *Introduction to Data Compression.* CA: Academic Press, 2000.

[21] S. Senbel, H. Abdel-Wahab, "Scalable and robust image compression using quadtrees,"*Signal Processing:Image Communication,* vol.14, pp.425-442, 1999.

[22] B.C. Song, J.B. Ra, "A fast multi-resolution block matching Algorithm for motion estimation,"*Signal Processing: Image Communication,* vol.15, pp.799-809, 2000.

[23] C. Stiller and J. Konrad, "Estimating Motion in image sequences," *IEEE Signal Processing Magazine,* vol.16, pp.70-91, July 1999.

[24] W.Woo and A. Ortega, "Stereo image compression based on Disparity Field Segmentation," *in Proc. of Visual Communic. and Image Proc., VCIP '97, San Jose, CA,* Feb. 1997.

[25] W. Woo, A. Ortega and Y. Iwadate, "Stereo Image Coding Using Hierarchical MRF Model and Selective Overlapped Block Disparity Compensation," In*IEEE International Conference on Image Processing,* vol.2, pp.467-471, 1999.

[26] Y.Yang and S.S.Hemami, "A rate distortion optimized motion compensated wavelet video coder," in *Picture Coding Symposium,* Portland, Oregon, April 1999.

[27] Y. Yang and S. Hemami, "Generalized Rate-Distortion Optimization for motion compensated wavelet video coders," in *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 10, No. 6, September 2000, pp.942-955.

[28] Y. Yang, S.S. Hemami, "Rate-Distortion Based Combined Motion Estimation and Segmentation," In *Proceedings IEEE International Conference on Image Processing,* vol.3, pp.920-924, 1998.

Appendix A

Original Stereo pairs

Some stereo image pairs used in Chapter 4 are shown here with the left image on the left and the right image on the right.

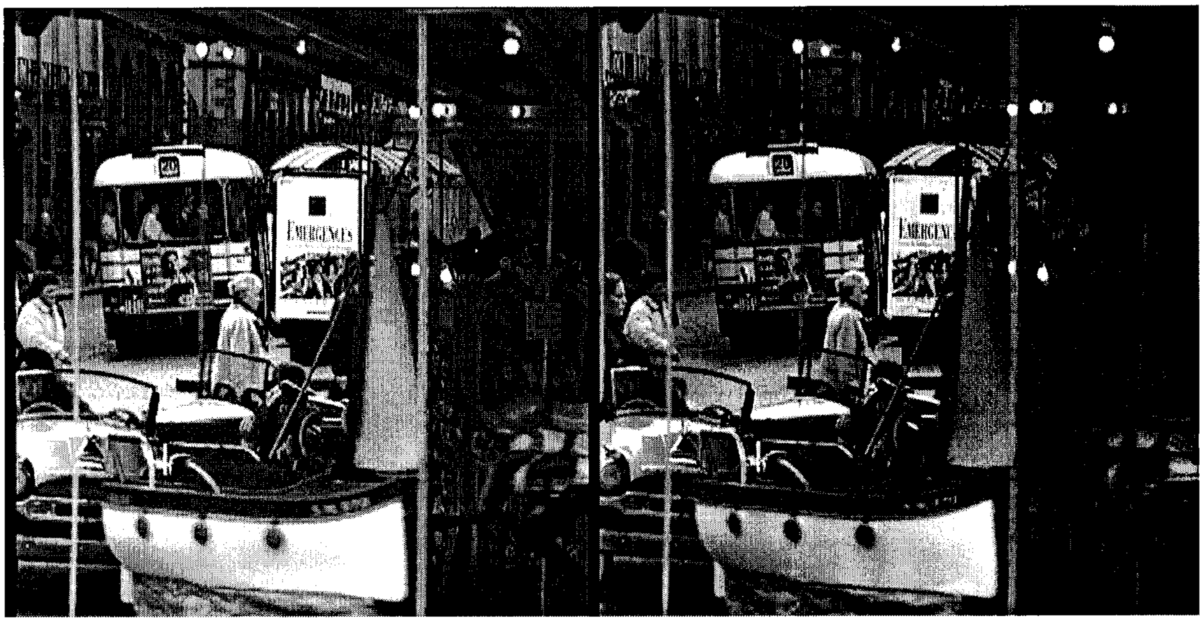

Figure A.1: Manege

Figure A.2: Redcar

Figure A.3: Trapeze