

# Wavelet Transforms for Stereo Imaging

A thesis submitted to the University of Wales for the degree of

**Doctor of Philosophy**

By

**Fangmin Shi**

Mechatronics Research Centre
University of Wales College, Newport
August 2002

**To my parents and**
**Hui and Chuyi**

## Table of Contents

## Declaration / Statements

### DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... (candidate)

Date .....08 NOVEMBER 2002.....

### STATEMENT 1

This thesis is the results of my own investigations, except where otherwise stated.

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ..... (candidate)

Date .....08 NOVEMBER 2002.....

### STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate)

Date .....08 NOVEMBER 2002.....

## Acknowledgements

I have been lucky with people around me. I always think so. The three years at UWCN has proved this again.

The fact that I am able to complete the PhD research and present this thesis mainly contributes to Professor Geoff Roberts. I am very grateful for the study opportunity he offered me, which opened my eyes well beyond the PhD research. Without his support, encouragement and push, this thesis would not be possible.

I am also lucky to have Dr Neil Rothwell Hughes as the second supervisor. Each time our discussions or his comments on my reports, papers and draft thesis saw his understanding and scientific intuition.

Thinking back, I have had many good memories at Mechotronics Research Centre and at Allt-Yr-Yn Campus. I would like to thank all the students colleagues, Mrs Jane John and all the IT and library staff in particular Mrs Jennifer Coleman and Mrs Bronwen Stone. I will never forget Jennifer's flowers to me when I was so depressed after the first burglary in my life. I am also very grateful to many people who have made efforts to make my life easier, in particular at the most difficult earliest stage when I suffered from both the language difficulty and culture shock. I am thinking of Liren Wang, Viviane Bouchereau, Ioannis Akkizidis, Alex Mouzakitis and Eric Llewellyn. I still feel their friendship now as the feeling of many old Chinese friends. I would like to thank all of them not only for the research work but also for the happy time together.

My parents, the best parents in the world, always give me whole-hearted love. All my life cannot reward their sacrifice to the family and unselfish support to my education since my childhood. The love from all my three sisters and two brothers and their families is also surrounding me all the time. I am indebted to them all.

My deepest gratitude goes to my husband, Hui Yang, and our daughter, Chuyi Yang, with whom I share everything except wavelets.

## Summary

Stereo vision is a means of obtaining three-dimensional information by considering the same scene from two different positions. Stereo correspondence has long been and will continue to be the active research topic in computer vision. The requirement of dense disparity map output is great demand motivated by modern applications of stereo such as three-dimensional high-resolution object reconstruction and view synthesis, which require disparity estimates in all image regions.

Stereo correspondence algorithms usually require significant computation. The challenges are computational economy, accuracy and robustness. While a large number of algorithms for stereo matching have been developed, there still leaves the space for improvement especially when a new mathematical tool such as wavelet analysis becomes mature.

The aim of the thesis is to investigate the stereo matching approach using wavelet transform with a view to producing efficient and dense disparity map outputs. After the shift invariance property of various wavelet transforms is identified, the main contributions of the thesis are made in developing and evaluating two wavelet approaches (the dyadic wavelet transform and complex wavelet transform) for solving the standard correspondence problem. This comprises an analysis of the applicability of dyadic wavelet transform to disparity map computation, the definition of a wavelet-based similarity measure for matching, the combination of matching results from different scales based on the detectable minimum disparity at each scale and the application of complex wavelet transform to stereo matching.

The matching method using the dyadic wavelet transform is through SSD correlation comparison and is in particular detailed. A new measure using wavelet coefficients is defined for similarity comparison. The approach applying a dual tree of complex wavelet transform to stereo matching is formulated through phase information. A multi-scale matching scheme is applied for both the matching methods. Imaging testing has been made with various synthesised and real image pairs.

Experimental results with a variety of stereo image pairs exhibit a good agreement with ground truth data, where available, and are qualitatively similar to published results for other stereo matching approaches. Comparative results show that the dyadic wavelet transform-based matching method is superior in most cases to the other approaches considered.

# Nomenclature

## Notations

| | |
|---|---|
| $\mathbb{N}$ | Positive integers including 0 |
| $\mathbb{Z}$ | Integers |
| $\mathbb{R}$ | Real numbers |
| $\mathbb{C}$ | Complex numbers |
| $\mathcal{L}^2(\mathbb{R})$ | Finite energy function space |
| $\mathcal{H}$ | Hilbert space |
| $(X,Y,Z)$ | World co-ordinate of a three-dimensional point |
| $(x,y)$ | Two-dimensional image co-ordinate |
| $(u,v)$ | Two-dimensional image co-ordinate in pixels |

## Acronyms and abbreviations

| | |
|---|---|
| DyWT | Dyadic wavelet transform |
| CWT | Complex wavelet transform |
| DTCWT | Dual tree complex wavelet transform |
| SSD | Sum of the squared differences |
| MRA | Multi-resolution analysis |
| STFT | Short time Fourier transform |
| RDS | Random dots stereograms |
| 3D | Three-dimensional |
| 2D | Two-dimensional |

## List of Figures

## List of Tables

# 1 Introduction

The work on computer vision started thirty years ago. This is therefore a relatively new area where new applications appear all the time. It can be defined as a set of computational techniques aimed at estimating or making explicit geometric and dynamic properties of the three-dimensional (3D) world from digital images (Trucco and Verri, 1998). It gives solutions to various real world problems from industrial inspection, autonomous vehicles, robots to medical image analysis (Davies, 1997). Many of the applications require the perception of depth information, which is partially lost during the perspective projection. One method of depth recovery, known as stereo vision, requires the determination of corresponding points between two or more images. The recovered depth information can be used to plan a robot's path, position a robot arm, determine the pose of a CAD model or generate landscaping contours. As such tasks may be of critical importance, the depth recovery must be well characterised and precisely determined. Stereo matching provides the foundation for all of these in stereo vision. It is thus considered to be a central problem in 3D computer vision.

Mathematically, to infer information about the 3D world from mere 2D projections is an ill-posed problem (Bertero *et al*, 1988) in the sense that there is an infinite number of solutions. However, *a prioi* knowledge can be used as physical constraints that can limit these solutions. In stereo vision, additional matching constraints renders such an ill-posed problem soluble. If the relative position between the corresponding points, known as disparity, is known, the depth information and the original three-dimensional object structure can be recovered. Disparity information can be obtained by comparing

the similarity of a pair of stereo images using a correlation measure. This was the approach used in the early vision research (Marr, 1982).

## 1.1 Research Motivation

Stereo correspondence has long been and will continue to be the active research topic in computer vision. The requirement of dense disparity map output is great demand motivated by modern applications of stereo such as three-dimensional high resolution object reconstruction and view synthesis, which require disparity estimates in all image regions.

Stereo correspondence algorithms usually require significant computation. Conventional matching method compares the similarity of small patches of images by applying a correlation measure. Based on the result of psychophysical and physiological studies into human vision which indicates that the retinal images are likely to be processed in multiple frequency channels, an efficient matching algorithm using multiple scale approach was developed (Marr and Poggio, 1979). This works by applying a smoothing filter, e.g. Gaussian filter, to the original images at different resolutions and then determines the corresponding points by detecting the intensity extrema of the filtered images. The matching results at the coarser scale can be used as a guide to matching at the finer scales, which is known as coarse-to-fine strategy (Rosenfeld and Thurston, 1977).

Although many other matching methods (Grimson, 1981; Sanger, 1988; Barnard, 1989; Devleeschauwer, 1993; Yang, 1993) were developed later, the filtering method and the coarse-to-fine strategy have been widely used. In particular, Mallat (Mallat, 1989a; Mallat, 1989b) made a significant contribution in developing an efficient wavelet

multiresolution representation, by which some of the multi-scale ideas in stereo vision have been formalised and refined. Since then the wavelet analysis, as a useful tool, has found more and more applications.

The research this thesis was first motivated by improving the efficiency of existing matching algorithms. While investigating alternative approaches, the wavelet transform was considered to have potential in this area. This thesis concentrates on two kinds of wavelet transforms: dyadic wavelet transform and complex wavelet transform.

## 1.2 Aims and Objectives of the Thesis

The aim of the research is to make use of wavelet analytical tool to produce dense disparity map output. Suitable shift invariant wavelet transforms will be identified. Hierarchical wavelet-based stereo matching algorithms will be constructed capable of generating dense disparity maps as output. This will give rise to a computationally efficient implementation.

The specific objectives of the thesis are:

- Explore Mallat's wavelet multiresolution analysis to determine whether it is suitable for stereo matching
- Identify the existing wavelet transforms with a view of their suitability to stereo matching
- Formulate and implement wavelet-based matching methods
- Evaluate the performance of the algorithms proposed by comparison with each other and with a conventional matching approach.

## 1.3 Methodology

Conventional multi-scale stereo matching methods will first be reviewed. Wavelet theory will then be studied with a view to developing wavelet-based stereo matching methods. In terms of the existing matching categories, the investigation for the wavelet-based matching method will start with the consideration of generating dense disparity maps, i.e. using correlation measure and phase information.

## 1.4 Outline of the Thesis

There are seven chapters altogether in this thesis. Each chapter starts with an introductory section and is concluded with a summary of the chapter and references. Generally the thesis follows a structure of review, related theory, development and implementation, results and discussion. Specifically, the organisation from Chapter 2 is as follows:

Chapter 2 presents the background of stereo vision and an overview of stereo matching including stereo geometry, matching constraints and matching methods. Chapter 3 reviews wavelet theory and discusses the importance of wavelet shift invariance property to stereo imaging.

These are followed by the main investigations of this thesis in Chapters 4, 5 and 6. Chapter 4 identifies a shift invariant wavelet transform, i.e. Dyadic Wavelet Transform (DyWT), for stereo matching with a view to solving the windowing problem and to reducing the computational complexity. It begins with the formulation of the DyWT and its fast implementation. A DyWT based matching method is then developed. Based on the conventional correlation measure, a new measure of the Sum of Squared Differences (SSD) using DyWT coefficients is defined and called W-SSD. This

measure allows a coarse-to-fine multi-scale disparity estimate associated with the disparity due to the hierarchical structure of DyWT.

As disparity maps can also be extracted from local phase differences between two bandpass signals, Chapter 5 addresses a wavelet-based matching approach from the phase point of view. It starts with a comparative discussion between global Fourier phases and local Gabor phases. Then the conventional phase-based disparity computation method is described. This is followed by the presentation of the phase-based matching approach using Magarey and Kingsbury's dual tree complex wavelet transform (DTCWT) (Magarey and Kingsbury, 1998).

Implementation, imaging tests and the algorithm evaluation of applying the new wavelet-based methods to compute disparity maps are presented in Chapter 6. Various experiments are made with a range of stereo images: random dot stereograms, artificial images with ground truth data and real images naturally taken in the laboratory. As the development of W-SSD matching method is the main contribution of this thesis, the implementation procedure with this method is detailed in this chapter. An evaluation of its performance is first made in terms of different object structure and surfaces with stereograms and then made by comparison with other methods, i.e. the complex wavelet transform based, Gaussian pyramid based and standard correlation based matching results.

Finally, Chapter 7 presents a summary of the work covered by the thesis, lists and describes the contributions of the thesis and makes some suggestions for future work.

Three appendices are included at the end of the thesis. Appendix A describes the camera calibration method used for the work presented in Chapter 2. Appendix B gives

further information on some mathematical notations used in the thesis. Appendix C contains copies of the papers that have been published as the research work progressed.

### 1.4.1 Main Contributions

The main contributions of the work presented in this thesis are:

- Applicability of dyadic wavelet transform to disparity map computation
- Definition of a wavelet-based similarity measure for matching
- Combination of matching results from different scales based on the detectable minimum disparity at each scale
- Application of DTCWT to stereo matching

These are explained in Chapters 4, 5 and 6 and summarised in Chapter 7, section 7.3.

## 1.5 References

Barnard, S. T. 1989. Stochastic Stereo Matching over Scale. *International Journal of Computer Vision*, (3), pp. 17-32.

Bertero, M., Poggio, T. A. and Torre, V. 1988. Ill-Posed Problems in Early Vision. *Proceedings of the IEEE*, **76** (8), pp. 869-889.

Davies, E. R. 1997. *Machine Vision: Theory, Algorithms, Practicalities.* 2nd end. Academic Press.

Devleeschauwer, D. 1993. Intensity-Based, Coarse-to-Fine Approach to Reliably Measure Binocular Disparity. *Computer Vision Graphics & Image Processing*, **57** (2),

pp. 204-218.

Grimson, W. 1981. A Computer Implementation of a Theory of Human Stereo Vision. *Phil. Trans. Royal Soc. London*, **V292**, pp. 217-253.

Mallat, S. 1989a. Multifrequency Channel Decomposition of Images and Wavelet Models. *IEEE Trans ASSP*, **37** (12), pp. 2091-2110.

Mallat, S. 1989b. A Theory for Multiresolution Signal Decomposition: the Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11** (7), pp. 674-693.

Marr, D. 1982. *Vision.* New York: W. H. Freeman and Company.

Marr, D. and Poggio, T. 1979. A Computational Theory of Human Stereo Vision. *Proc. Royal Society of London*, **204**, pp. 301-328.

Rosenfeld, A. and Thurston, M. 1977. Coarse-fine Template Matching. *IEEE Trans. System, Man, and Cybernetics*, **7**, pp. 104-107.

Sanger, T. D. 1988. Stereo Disparity Computations Using Gabor Filter. *Biol. Cybern.*, **59**, pp. 405-418.

Trucco, E. and Verri, A. 1998. *Introductory Techniques for 3-D Computer Vision.* New Jersey: Prentice Hall.

Yang, Y. 1993. *Local, Global and Multilevel Stereo Matching.* IEEE Conf. on Computer Vision & Pattern Recognition. pp. 274-279,

# 2 Background of Stereo Vision

## 2.1 Introduction

The human visual system makes use of the two slightly different views from each eye to produce three-dimensional (3D) perception. In the industry domain, many applications require a machine to obtain 3D information about the environment. This technique is known as *stereo vision*. For example, stereo vision provides autonomous vehicles (Curwen *et al*, 1992) with 'sight', enabling them to understand their surroundings such as path planning by avoiding obstacles on their way or determining their locations. Such systems can be used for vehicle navigation, for the location and tracking of known objects and for vehicle surveillance. In medical surgery (Edwards *et al*, 1995; Skrinjar and Duncan, 1999; Davies, 2000) researchers have applied stereo-guided vision systems in the operating theatre for timely 3D information of the operative object and accurate location of a pointer or a tool.

This chapter addresses the background issues of stereo vision that are related to this thesis. This content starts with an introduction to stereo imaging. Of all the conventional matching methods reviewed in section 2.2, the correlation-based method is of particular interest in this thesis because it produces dense depth maps. Section 2.3 specially focuses on the discussion of its mathematical representation, its advantages and disadvantages, in particular the windowing problem, when it is applied to perform matching. The phase-based method is another approach to obtaining disparity maps and is outlined in section 2.4. To improve the algorithmic computational efficiency, a multi-

scale approach is applied in the thesis. Section 2.5 discusses how the conventional coarse-to-fine strategy works. Section 2.6 gives the summary of this chapter.

## 2.2 Stereo Imaging

Stereo vision works with two-dimensional (2D) images, which are taken simultaneously by two cameras at slightly different positions. A simple case of such a system assumes that the two cameras are parallel to each other. The geometry is illustrated in Figure 2.1.

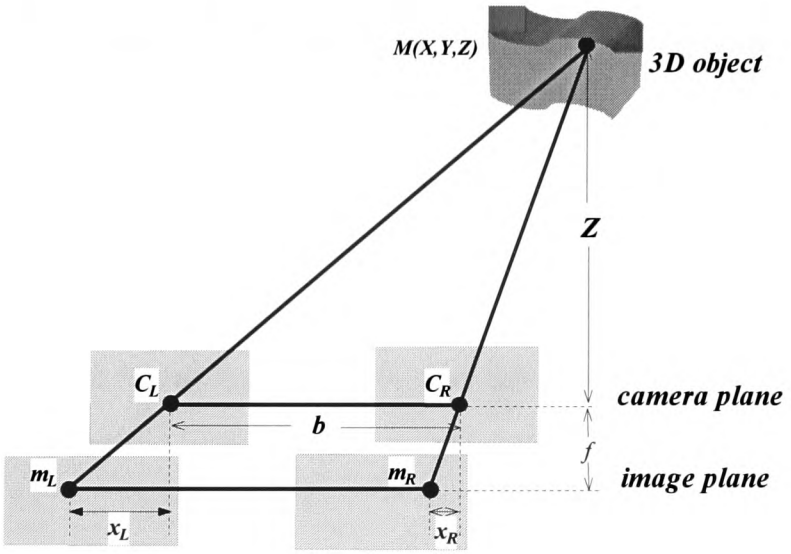



**Figure 2.1 Simple stereo geometry**

In Figure 2.1, the plane that is used for projection is called camera plane, whose optical centres are denoted by $C_L$ and $C_R$, respectively at each plane. The line joining the optical centres is called base line. The line that passes the optical centre and is perpendicular to the camera and image plane is called optical axis. A 3D physical point $M$ projects into two image planes by 2D image points $m_L$ and $m_R$ respectively. Due to the parallel cameras, the image co-ordinates have the same vertical values ($y$-axis) and have only a horizontal shift with reference to each image co-ordinate origins which will

be defined in the next paragraph. Focal length $f$ is defined as the distance between the base line and the image planes. The distance of the physical point $M$ relative to the base line is the depth that is of interest in stereo vision task. The collection of such depth values forms a *depth map*.

Four co-ordinate systems are to be used from now on. The system that represents a random point $M$ with respect to any origin in the 3D world is called *world co-ordinate system* $(X,Y,Z)$. *Camera co-ordinate system* uses the optical centre $C_L$ or $C_R$ as its origin, the base line as its $x$-axis and the optical axis as its $z$-axis. Then the co-ordinates of image points $m_L$ and $m_R$ in camera co-ordinate system are denoted by $(x_L, y_L, f)$ and $(x_R, y_R, f)$. These image points can also be expressed by pixel values $(u_L, v_L)$ and $(u_R, v_R)$, which are convenient for digital image storage and computation, in each image plane using 2D *image co-ordinate system* with respect to an origin in the top left corner of the image plane. If the middle point of the base line is defined to be the origin of the world co-ordinate system and all the three world co-ordinate axes parallel to those in the camera co-ordinate system, then such a system is described as *cyclopean co-ordinate system*.

According to the triangulation of perspective projection in Figure 2.1, consider a 3D point $M(X,Y,Z)$ in the cyclopean coordinate system, the following relationships hold:

$$\frac{x_L}{X+b/2}=\frac{f}{Z} \tag{2.1}$$

$$\frac{x_R}{X-b/2}=\frac{f}{Z} \tag{2.2}$$

Then the depth value can be derived straightaway:

$$Z = \frac{f \cdot b}{x_L - x_R} \tag{2.3}$$

In this equation, the focal length $f$ is one of the camera's intrinsic parameters, which can be obtained by camera calibration. The technique that is used for camera calibration and the experimental data of the camera system used in the thesis are attached in Appendix A. The parameter $b$ is the length of the base line, which is known once the stereo camera system is set up.

The result of $(x_L\text{-}x_R)$ is defined as disparity $d$, i.e. $d= x_L\text{-}x_R$, which is the difference of positions between the two image points corresponding to the same 3D physical point. To determine the disparity, it needs to be known which point in one image plane is the corresponding point to a given point in another image plane. This is known as *correspondence problem*. For example, in Figure 2.1, it is important to make sure that $m_L$ in the left image plane matches $m_R$ in the right plane, which are the respective projections of 3D point $M$ in each plane.

It is extremely difficult to solve the correspondence problem because finding correspondences is an ill-posed problem (Bertero *et al*, 1988). To remove the ambiguity (Marr and Poggio, 1976) in the correspondence, some constraints must be applied during the matching process:

- Constraint 1: *Similarity* (*Comparability*) (Grimson, 1981)

  The matching points must have similar intensity values or similar feature attribute values.

- Constraint 2: *Uniqueness* (Marr and Poggio, 1979)

One point in the left image plane should usually have no more than one match in the right image.

- Constraint 3: *Continuity* (Marr and Poggio, 1979)

  The disparity of the matched points should vary smoothly almost everywhere over the image.

- Constraint 4: *Ordering* (Baker and Binford, 1981)

  If $m_{L1}$ and $m_{L2}$ are points in $L$ and their corresponding points are $m_{R1}$ and $m_{R2}$ in $R$ respectively, then $m_{R1}$ and $m_{R2}$ should keep the same ordering as $m_{L1}$ and $m_{L2}$ and vice versa.

The geometry that is shown in Figure 2.1 is a simple case with horizontal disparity only. Generally, disparity may take place both horizontally and vertically. This means that given a point in one image plane, searching for its corresponding point in another plane is in two-dimensional space. However, under epipolar constraint that is described below, the searching space could be reduced to one dimension.

Figure 2.2 illustrates the epipolar geometry. Two cameras are placed with their optical axes angled inwards. As the base line goes through the two image planes $L$ and $R$, the intersections $E_L$ and $E_R$ are defined as *epipoles*. Joining each epipole and image point is an *epipolar line*, e.g. line $E_L m_L$ and $E_R m_R$. The plane that the base line, the 3D point $M$ and its two image points $m_L$ and $m_R$ form is called *epipolar plane*. Actually, epipolar lines are the intersections of the epipolar plane with the image planes.

The epipolar geometry assures a constraint that given a point $m_L$ in $L$, its corresponding point $m_R$ must lie on the corresponding epipolar line $E_R m_R$ in $R$. This is the epipolar constraint.

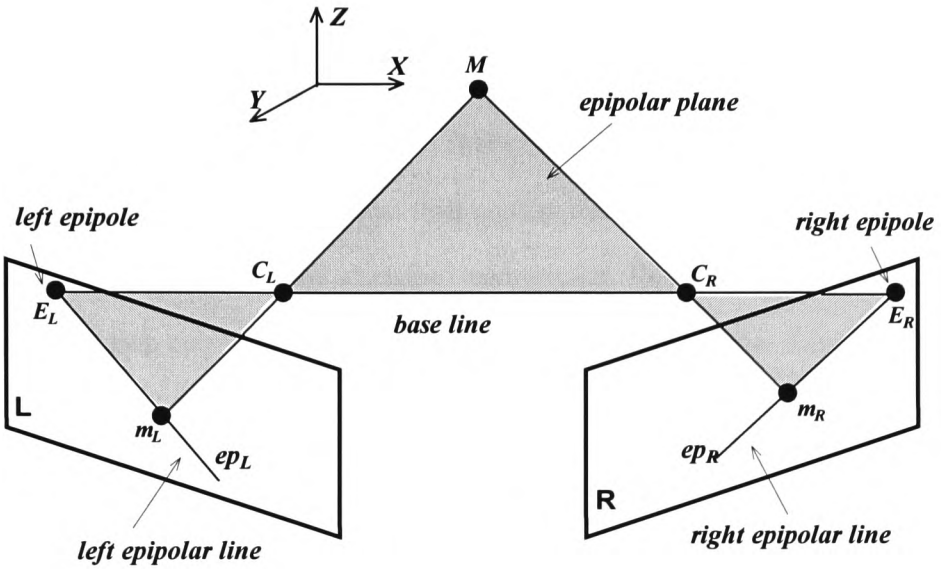



**Figure 2.2 Epipolar geometry**

After camera calibration and finding correspondences, according to equation (2.3) depth information can be obtained. The technique for camera calibration (Tsai, 1987) is relatively well understood. But how to find correspondences remains difficult in computer vision domain. Matching is the fundamental computational task. Therefore, the vision problem is often seen as correspondence problem. Successful matching is the key point.

The basic idea of performing stereo matching is to establish correspondence by comparing the similarity between the selected primitives. There are many matching methods. With different matching primitives, various matching algorithms use different matching criterion and strategy. Generally the proposed correspondence algorithms can be classified into two classes, correlation- and feature-based methods (Dhond and Aggarwal, 1989).

Feature-based method (Marr and Poggio, 1979; Ohta and Kanade, 1985; Ohta and Kanade, 1985; Pollard *et al*, 1985; Ayache and Faugeras, 1985; Grimson, 1985) uses image features as matching primitive. It first extracts a collection of features such as edges, corners, lines and curves and then applies matching by their attributes. Such features are stable to noise and change of environment. However, this method gives a sparse depth map output. Interpolation is necessary to provide a dense depth map.

Correlation-based matching is a basic matching method, which directly applies to image intensity values on the basis of similarity of windowed areas. The standard and refined implementations have been proposed in many papers such as (Moravec, 1977), (Gennery, 1980), (Okutomi and Kanade, 1992), (Faugeras *et al*, 1993), (Kanade and Okutomi, 1994) and (Wei *et al*, 1998). This method can produce sufficiently dense depth maps. More discussions will be presented in the next section as the work in this thesis has concentrated on the improvement of this method.

In recent years, a third approach to stereo matching, known as phase-based matching (Sanger, 1988; Weng, 1993; Zhong *et al*, 1994; Jenkin and Jepson, 1994; Zhou *et al*, 1996) has had much attention. In this method, the Fourier phase information of images is believed to contain useful information, enough to be used for matching. The difference between the phase information of the corresponding points is used to compute the stereo disparity. This method will be discussed later as another main contribution of the thesis is based on it.

## 2.3 Correlation-Based Matching

### 2.3.1 Theory

In correlation-based matching, small patches of windows are used as matching primitive and the correlation of the windows in two images is used as the similarity measure.

Let $L(i, j)$ and $R(i, j)$ be the intensity of a pair of stereo images, $2w+1$ and $2h+1$ the width and the height of the correlation window. Disparity values along the horizontal and vertical axes are denoted by $d_1$ and $d_2$. These parameters are illustrated in Figure 2.3.

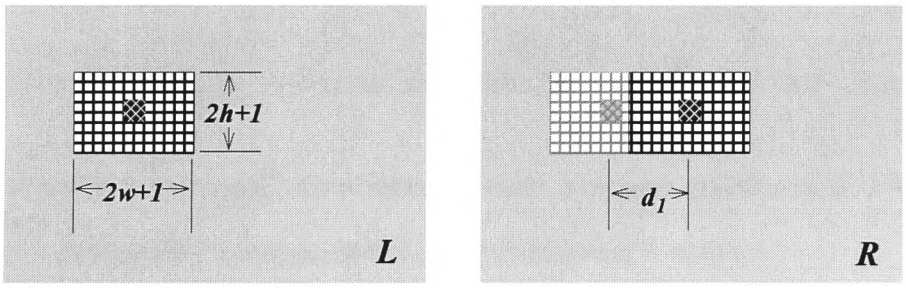



**Figure 2.3 Parameters for correlation-based matching**

The similarity measure using *SSD* (sum of squared differences) is:

$$ssd(d_1, d_2) = \sum_{j=d_2-h}^{d_2+h} \sum_{i=d_1-w}^{d_1+w} (L(i,j) - R(i,j))^2 \tag{2.4}$$

The *SSD* measure is widely applied in both existing matching research and applications. It is also used in the thesis. The process is to choose one point centred window in left image, using it as a template to move around in the right image until a corresponding point is found, i.e. the *ssd* value is minimized.

Using the epipolar constraint which can always reduce the search area from two dimensions to one, to simplify the mathematical representation, the methods described in the thesis only discuss one-dimensional case unless otherwise specified.

### 2.3.2 Advantages and Disadvantages

The main advantages of using SSD measure is that it is easy to implement (and debug) and produces dense disparity map output. This can be very helpful for the reconstruction of high resolution 3D surfaces and video tracking, which are in great demand by modern applications. However, it has three main disadvantages:

- The size of the correlation windows does affect the matching result. This is left to be discussed in detail in next section.
- Due to the point by point template comparison, the computation is time consuming.
- It directly uses intensity values for comparison, which are sensitive to the changes of the environment illumination and contrast and image distortions.

### 2.3.3 Windowing Problem

As fixed size of correlation window is applied to the whole stereo images, it is important to choose an appropriate window size of the correlation-based method especially when the stereo disparity contains both big and small values. If the window is too narrow, the windowed intensity variation would not be distinctive enough and false matches may take place. If it is too wide, resolution is lost, as neighbouring image areas with various disparities would be combined in the measurement.

## 2.4 Phase-Based Matching

The theoretical principle of phase-based matching is the Fourier shift theorem (Bracewell, 1986), which states that a signal's shift in time (or space) domain can be computed by the phase information in frequency (or spatial frequency) domain.

Suppose $x_1(t)$ and $F_1(f)$ be an original signal and its Fourier transform. Shifting $x_1(t)$ by $\tau$ forms $x_2(t)$, that is, $x_2(t)= x_1(t-\tau)$ and its Fourier transform is $F_2(f)$. Their mathematical representations are:

$$F_1(f) = \int_{-\infty}^{\infty} x_1(t) \cdot e^{-j2\pi ft} dt = |F_1(f)| \cdot e^{j\theta_1} \tag{2.5}$$

$$F_2(f) = \int_{-\infty}^{\infty} x_2(t) \cdot e^{-j2\pi ft} dt = |F_2(f)| \cdot e^{j\theta_2} \tag{2.6}$$

Where $|F_1(f)|$ and $|F_2(f)|$ are complex magnitudes and $\theta_1$ and $\theta_2$ are complex phase angles.

Substitute $x_2(t)$ by $x_1(t-\tau)$, then

$$\begin{aligned} F_2(f) &= \int_{-\infty}^{\infty} x_1(t-\tau) \cdot e^{-j2\pi f(t-\tau)} d(t-\tau) \cdot e^{-j2\pi f\tau} \\ &= F_1(f) \cdot e^{-j2\pi f\tau} \end{aligned} \tag{2.7}$$

The relationship between the two Fourier transforms is as follows:

$$|F_2(f)| = |F_1(f)|, \tag{2.8}$$

$$and\ \theta_2 = \theta_1 - 2\pi f\tau \tag{2.9}$$

The Fourier magnitude of the shifted signal remains unchanged but the difference between the Fourier phases is linear to the shift value. Hence the shift value τ can be obtained by:

$$\tau = \frac{\Delta\theta}{2\pi f}, \text{ where } \Delta\theta = \theta_2 - \theta_1 \tag{2.10}$$

In Fourier shift theorem, the parameter τ refers to the global shift value of a signal and the computation using equation (2.10) does not give any information in the spatial domain. Actually, a real pair of stereo signals can be intuitively assumed to be the shifted version of each other, whose disparity, the shifted value with respect to each pixel, is dependent on the pixel position. This requires the short-time Fourier Transform (STFT) that gives the Fourier transform in the joint time-frequency or position-frequency space.

Gabor (Gabor, 1946) proposed the windowed Fourier transform, e.g. STFT, for local spatial and frequency analysis:

$$STFT(\tau, f) = \int_{-M/2}^{M/2} [x(t) \cdot w(t-\tau)] \cdot e^{-j2\pi ft} dt \tag{2.11}$$

where *w(t)* is the window function and *M* is the window size. The idea is to segment the signal using a window function, followed by doing a Fourier analysis on these segmentations. A popular used window is Gaussian function, smooth and having minimum area of spatial and spectral width:

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/2\sigma^2} \tag{2.12}$$

With Gaussian function, equation (2.11) can also be represented by the form of convolution:

$$GT(\tau, f) = \int_{-M/2}^{M/2} [x(t) \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\tau)^2/2\sigma^2}] \cdot e^{-j2\pi ft} dt \\ = x(t) * gf(t, f) \tag{2.13}$$

where *gf(t, f)* is the Gabor filter:

$$gf(t, f) = \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/2\sigma^2} \cdot e^{j2\pi ft} \tag{2.14}$$

A set of such filters have each a different range of frequency sensitivity to span the whole spatial frequency spectrum, each of which can provide a local phase value after convolving with the signal.

The research using STFT-based method was pioneered by Sanger (Sanger, 1988). He called this *correspondenceless* approach because the disparity results from direct calculation. It can also produce dense depth map output under an efficient implementation. In addition, Fleet (Fleet *et al*, 1991) proposed disparity computation equation using the derivatives of the phases. Weng (Weng, 1993) developed similar filters forming the Windowed Fourier Phase (WFP) method. All of these methods were implemented after many iterations because the difficulty with the Gabor phase based method lies in the non-linearity of the local shift and local phase difference under STFT. The simple relationship, as shown in (2.10), does not hold any more for STFT.

STFT also has its disadvantage of the limit on its time-frequency resolution capability due to the Heisenberg uncertainty principle (Battle, 1988). The principle states that the multiplication of the width of the window function in time and frequency domains remains constant. Therefore, in the case of STFT, a narrow window gives short signals

poor frequency localization, whereas a wide window gives low frequency signals poor time resolution. More explanation can be found in (Feichtinger and Strohmer, 1997).

As discussed by Kaiser (Kaiser, 1994), the STFT represents an inaccurate and inefficient method of time-frequency localisation. The inaccuracy results from the aliasing problem of high- and low-frequency components that do not fall within the frequency range of the window. The inefficiency arises from the responding frequencies, which must be analysed at each time interval, regardless of the window size or the dominant frequencies present. In addition, several window lengths must usually be analysed to determine the most appropriate choice. For analyses where a predetermined scaling may not be appropriate because of a wide range of dominant frequencies, a method of time–frequency localisation that is scale independent, such as wavelet analysis, should be employed. This is one of the reasons for the creation of wavelet theory. The wavelet approach to disparity computation using phase-based method is the other task of the thesis. This method is described in detail in Chapter 5.

## 2.5 Multiple Scale Approach to Matching

Many image related algorithms are time consuming. Since pyramid structure (Burt, 1984) was applied to images to improve the computational efficiency, multiple resolution or multiple scale approach has been generally used in vision problems. Firstly, an image representation with multiple resolutions is constructed. Then matching starting from coarse resolution to fine resolution is performed. This coarse-to-fine strategy not only speeds up the matching process but also provides a better solution to the false-target problem (Barnard, 1989). Stereo matching (Rosenfeld and Thurston, 1977) and edge detection (Canny, 1986) are early examples of the application of the approach.

Conventionally Gaussian or Laplacian pyramids are employed for algorithm implementation (Marr, 1982; Rosenfeld, 1984). This method works by applying a smoothing filter to the original images first. This is because vision problems like stereo matching actually belong to the class of optimisation problems. It is much more rigorous to locate the extrema of the smoothed version of an image than its original, which can then give a good starting point to locate the extrema of the original image.

In order to smooth images and detect intensity changes, a good filter should be used. Marr and Hildreth's work (Marr and Poggio, 1976; Marr and Poggio, 1979) shows that the most satisfactory operator is the filter $\nabla^2 G$, where $\nabla^2$ is the Laplacian operator $(\partial^2/\partial x^2 + \partial^2/\partial y^2)$ and $G$ denotes the two-dimensional Gaussian distribution $G(x,y) = e^{\frac{x^2+y^2}{2\pi\sigma^2}}$, $\sigma$ is the standard deviation. Then $\nabla^2 G$ is:

$$\nabla^2 G(x,y) = \frac{-1}{\pi\sigma^4}(1 - \frac{x^2+y^2}{2\sigma^2})e^{\frac{-x^2-y^2}{2\sigma^2}} \tag{2.15}$$

There are some advantages of choosing $\nabla^2 G$:

- Different $\sigma$ forming large or small filters enable the smoothing of an image at multiple scales.
- Secondly, the Gaussian part of the expression effectively removes all structures at scales much smaller than the space constant $\sigma$ of the Gaussian.
- Thirdly, the derivative part significantly reduces the amount of computation. Figure 2.4 illustrates the spatial configuration in neurophysiological terms of the various first- and second-order differential operators (Marr, 1982). The first-order operators, i.e. $\partial/\partial x$ and $\partial/\partial y$, can be thought of as measuring the difference between the values at two neighbouring positions along the $x$- and $y$- axis. In this way, the

second-order operators, i.e. $\partial^2/\partial^2 x$, $\partial^2/\partial^2 y$, $\partial^2/\partial x \partial y$ and $\partial^2/\partial^2 x + \partial^2/\partial^2 y$ (the Laplacian operator, $\nabla^2$), are shown in (c)-(f), respectively. In particular, the Laplacian operator has the circularly symmetric form as appeared in (f).

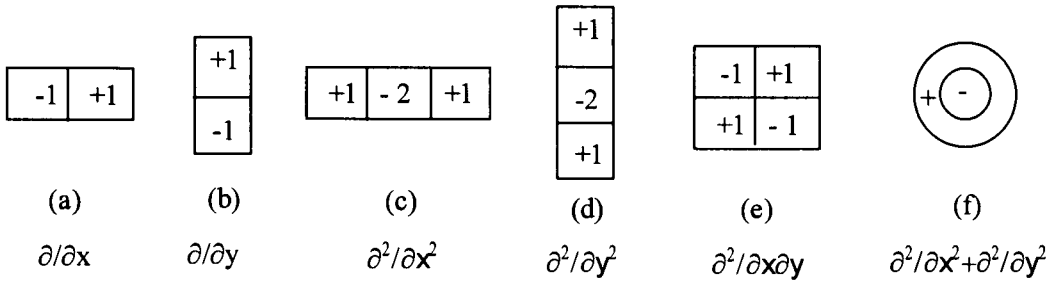



**Figure 2.4 The spatial configuration of first- and second-order differential operators**

To detect the intensity extrema, filtering an image with the operator $\nabla^2 G$ at different resolutions, reflected by the space constant $\sigma$ of the Gaussian, needs to be done first. Then some algorithm is applied to the filtered images. Using this way, Canny (Canny, 1986) put forward an effective edge detection approach.

In stereo matching, the search process starts at a coarse scale and the roughly matched results are used to guide searching at finer scales. It works as follows:

- Generating a pair of image pyramids from the original image pairs so that only a few and prominent features are present at the coarse levels. The original images are at the finest level of the image pyramids.
- Starting the matching process at the coarsest level.
- Using the matches obtained at the coarser level to guide the matching process gradually up to the finest level.

In 1989 Mallat combined the concept of wavelets and the multiscale vision method and put forward a complete theory of wavelet multiresolution analysis (Mallat, 1989). This theory provides a general hierarchical image decomposition method which facilitates improvements in coarse-to-fine strategy. The basic wavelet theory will be described in Chapter 3 and the proposed method using wavelet-based matching method will be formulated in Chapters 4&5.

## 2.6 Summary

This chapter has presented the theoretical issues of stereo vision and an overview of stereo matching, including the stereo geometry, matching constraints and matching methods. In particular, two conventional approaches e.g. the correlation- and phase-based matching methods have been outlined and the existing problems inherent in them have been discussed. From the computational point of view, multiple scale approach to vision has also been examined. These demonstrate the viability of applying other approaches to stereo matching. Wavelet analysis, which provides a general multi-resolution image representation, has the potential to overcome the problems. Therefore, relevant wavelet theory will be introduced in Chapter 3.

## 2.7 References

Ayache, N. and Faugeras, O. D. 1985. *Depth Maps Obtained by Passive Stereo.* Proc. of the 3rd Workshop on Computer Vision: Representation and Control. pp. 197-204, Bellaire.

Baker, H. H. and Binford, T. O. 1981. *Depth from Edge and Intensity Based Stereo.* International Joint Conference on Artificial Intelligence. pp. 631-636,

Barnard, S. T. 1989. Stochastic Stereo Matching over Scale. *International Journal of Computer Vision*, (3), pp. 17-32.

Battle, G. 1988. Heisenberg Proof of the Balian-Low Theorem. *Lett. Math. Phys.*, (15), pp. 175-177.

Bertero, M., Poggio, T. A. and Torre, V. 1988. Ill-Posed Problems in Early Vision. *Proceedings of the IEEE*, **76** (8), pp. 869-889.

Bracewell, R. N. 1986. *The Fourier Transform and its Applications.* McGraw-Hill Book Company.

Burt, P. J. 1984. The Pyramid as a Structure for Efficient Computation. *In:* ed. Multiresolution Image Processing and Analysis. pp. 6-35. Springer-Verlag

Canny, J. 1986. A Computational Approach to Edge Detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **8** (6), pp. 679-698.

Curwen, R., Blake, A. and Zisserman, A. 1992. *Real-Time Visual Tracking for Surveillance and Path Planning.* European Conference on Computer Vision, ECCV92. pp. 879-883

Davies, B. L. 2000. A Review of Robotics in Surgery. *J. Eng. in Medicine, Proc. H. of IMechE. Special Millennium Issue*, **214** (1), pp. 129-140.

Dhond, U. R. and Aggarwal, J. K. 1989. Structure from Stereo - A Review. *IEEE Transactions on Systems, Man and Cybernetics*, **19** (6), pp. 1489-1510.

Edwards, P. J., Hawkes, D. J. and Hill, D. L. G. 1995. Augmentation of Reality in the Stereo Operating Microscope for Otolaryngology and Neurosurgical Guidance. *Journal of Image Guided Surgery.*

Faugeras, O., Hotz, B., Mathieu, H., Vieville, T. and Zhang, Z. 1993. *Real-Time Correlation-Based Stereo: Algorithm, Implementations and Applications.* INRIA Sophia-Antipolis.

Feichtinger, H. G. and Strohmer, T. 1997. *Gabor Analysis and Algorithms.* Birkhauser Boston.

Fleet, D. J., Jepson, A. D. and Jenkin, M. R. M. 1991. Phase-Based Disparity Measurement. *Computer Vision Graphics and Image Processing: Image Understanding*, **53** (2), pp. 198-210.

Gabor, D. 1946. Theory of Communication. *Journal of IEE*, **93**, pp. 429-459.

Gennery, D. B. 1980. *Object Detection and Measurement Using Stereo Vision.* Proc. ARPA Image Understanding Workshop. pp. 161-167, College Park, MD.

Grimson, W. 1981. A Computer Implementation of a Theory of Human Stereo Vision. *Phil. Trans. Royal Soc. London*, **V292**, pp. 217-253.

Grimson, W. E. L. 1985 . Computational Experiments with a Feature-Based Stereo Algorithm. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **7** (1), pp. 17-34.

Jenkin, M. R. M. and Jepson, A. D. 1994. Recovering Local Surface Structure through Local Phase Difference Methods. *CVGIP*, **59**, pp. 72-93.

Kaiser, G. 1994. *A Friendly Guide to Wavelets.* Boston: Birkhäuser.

Kanade, T. and Okutomi, M. 1994. A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16** (9), pp. 920-932.

Mallat, S. 1989. A Theory for Multiresolution Signal Decomposition: the Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11** (7), pp. 674-693.

Marr, D. 1982. *Vision.* New York: W. H. Freeman and Company.

Marr, D. and Poggio, T. 1976. Cooperative Computation of Stereo Disparity. *Science*, **194**, pp. 283-287.

Marr, D. and Poggio, T. 1979. A Computational Theory of Human Stereo Vision. *Proc. Royal Society of London*, **204**, pp. 301-328.

Moravec, H. P. 1977. *Towards Automatic Visual Obstacle Avoidance.* Proc. 5th International Joint Conference on Artificial Intelligence. pp. 584

Ohta, Y. and Kanade, T. 1985. Stereo by Intra- and Inter-scanline Search using Dynamic Programming. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **7** (2), pp. 139-154.

Okutomi, M. and Kanade, T. 1992. A Locally Adaptive Window for Signal Processing. *International Journal of Computer Vision*, (7), pp. 143-162.

Pollard, S., Mayhew, J. and Frisby, J. 1985. PMF: A Stereo Correspondence Algorithm Using a Disparity Gradient Limit. *Perception*, **14** (4), pp. 449-470.

Rosenfeld, A. 1984. *Multiresolution Image Processing and Analysis.* Springer-Verlag.

Rosenfeld, A. and Thurston, M. 1977. Coarse-fine Template Matching. *IEEE Trans. System, Man, and Cybernetics*, **7**, pp. 104-107.

Sanger, T. D. 1988. Stereo Disparity Computations Using Gabor Filter. *Biol. Cybern.*,

**59**, pp. 405-418.

Skrinjar, O. and Duncan, J. 1999. *Real Time 3D Brain Shift Compensation*. Information Processing in Medical Imaging. pp. 42-55,

Tsai, R. Y. 1987. A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation*, **RA-3** (4), pp. 323-344.

Wei, G. Q., Brauer, W. and Hirzinger, G. 1998. Intensity- and Gradient-Based Stereo Matching Using Hierarchical Gaussian Basis Functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20** (11), pp. 1143-1160.

Weng, J. 1993. Image Matching Using the Windowed Fourier Phase. *International Journal of Computer Vision*, **3**, pp. 211-236.

Zhong, S., Shi, Q. Y. and Cheng, M. 1994. A Stereo Matching Method Based on Wavelet Transform. *Pattern Recognition and Artificial Intelligence*, **7** (1), pp. 27-33.

Zhou, J., Peng, J. X. and Ding, M. Y. 1996. Image Matching Based on Wavelet Features. *Pattern Recognition and Artificial Intelligence*, **9** (2), pp. 125-129.

# 3 Wavelets and Matching

## 3.1 Introduction

Modern wavelet theory was initially motivated by the search for a better time-frequency signal representation than that provided by the short time Fourier transform (STFT). The wavelet transform is considered the most recent solution (Valens, 1999) to overcome the drawbacks of STFT, as discussed in chapter 2.

This chapter provides the necessary wavelet theory required by the following chapters and highlights the shift invariant property of wavelet transforms for stereo matching. The following sections, starting with section 3.2, give a brief introduction to the wavelet and its properties. The advantages of using wavelets compared with STFT and an overview of the wavelet transform are presented in section 3.3. Section 3.4 explains why shift invariance of a wavelet transform is of vital importance to the image matching task. It is made clear that achieving shift invariance of a wavelet transform is a prerequisite for its application to matching. Some existing matching methods using shift invariant wavelet transforms are reviewed in section 3.5. The conclusion of this chapter is summarised in section 3.6.

## 3.2 Wavelets

The wavelet theory presented here comes from a comprehensive study of many references such as (Daubechies, 1992; Chui, 1992; Kaiser, 1994; Strang and Nguyen, 1997; Teolis, 1998; Mallat, 1998; Hubbard, 1998; Frazier, 1999).

The general wavelet transform (Cohen and Kovacevic, 1996), mathematically expressed in equation (3.1), has a similar form to the Fourier transform:

$$WT(b,a) = \int x(t) \cdot \varphi_{a,b}(t) dt \tag{3.1}$$

In contrast with the Fourier transform, which decomposes a signal into sinusoids of various waves of various frequencies, the wavelet transform represents a signal using a family of wavelets $\varphi_{a,b}(t)$ as basis functions:

$$\varphi_{a,b}(t) = \frac{1}{\sqrt{a}} \varphi(\frac{t-b}{a}), \; a>0, \; b \in \mathbb{R} \tag{3.2}$$

where $a$ is a scale parameter, $b$ is a translation parameter. They are formed by dilating and translating a mother wavelet $\varphi(t)$ which is a small oscillatory function with finite support.

It is important to note that the mother wavelet is not specified in (3.2). This is a difference between the wavelet transform and the Fourier transform. The theory of the wavelet transform just defines a framework within which appropriate wavelets can be designed according to a specific requirement.

To become a candidate as a wavelet function, the following admissibility condition, given in (Daubechies, 1992), must be satisfied:

$$\int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < +\infty \tag{3.3}$$

where $\Psi(\omega)$ is the Fourier transform of $\varphi(t)$. This property implies that $\Psi(\omega)$ vanishes at the zero frequency, i.e.

$$\left|\Psi(\omega)\right|^2\Big|_{\omega=0} = 0 \tag{3.4}$$

It is equivalent to say that a wavelet function must have a zero mean in the time domain:

$$\int_{-\infty}^{\infty} \varphi(t)\, dt = 0 \tag{3.5}$$

and therefore it must be oscillatory.

Figure 3.1 shows some typical mother wavelet functions (Daubechies, 1992). Figure 3.2 illustrates a few wavelets generated by shifting and dilating the Morlet wavelet function (Chui, 1992). It can be seen from Figure 3.2 that the lower scales correspond to the more compressed, i.e. higher frequency wavelets, and the higher scales correspond to the more stretched, i.e. lower frequency wavelets.

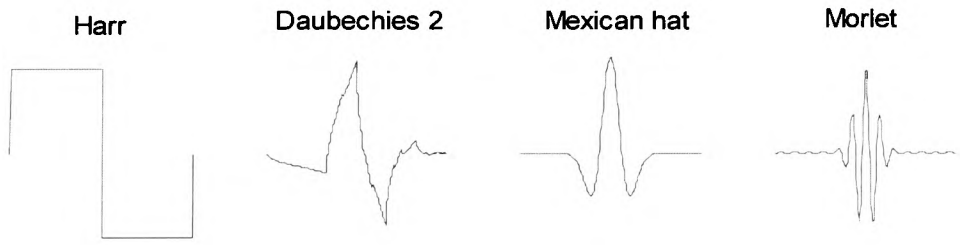



**Figure 3.1 Mother wavelets**

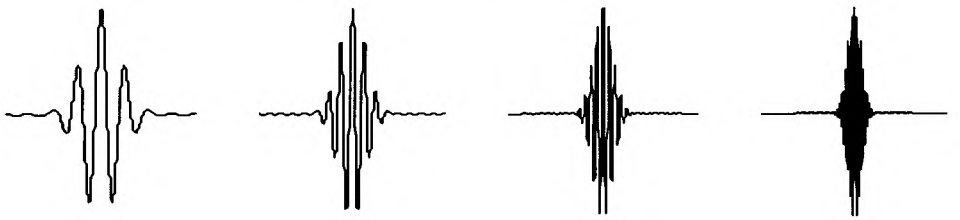



**Figure 3.2 Shifted and dilated wavelets**

In Figure 3.1 the different types of wavelets would result in different performances in different applications of wavelet transforms. For specific applications, the wavelet functions have to be carefully chosen. There are several factors that need considering:

- Width

As the definition of the wavelet indicates, expressed by equation (3.5), the support for a wavelet should be limited, i.e. compact support. For example, the support widths for the Harr and Daubechies *2* wavelets are the same, i.e. *1*. Strictly speaking, the Mexican hat and the Morlet have indefinite support. However, they both have an interval where the energy is mostly concentrated. Beyond the interval, the function values tend to be zero. This interval is called the effective support. For the Mexican hat, it is *[-5,5]* and for the Morlet, *[-4,4]*.

The balance between the width of a wavelet function in the time domain and in the frequency domain determines the resolution of the wavelet. A narrow function will have good time resolution but poor frequency resolution, whereas a wide window will have poor time resolution, yet good frequency resolution. This is an illustration of the Heisenberg uncertainty principle (Battle, 1988), discussed in chapter 2.

- Shape

An appropriate wavelet function should be chosen as a reflection of the type of features contained in the original signals. Boxcar-like wavelets such as the Harr, the first plot in Figure 3.1, are suitable for the sequences with sharp jumps or steps wavelet (Torrence and Compo, 1998). However, for smoothly varying signals, a smooth waveform, e.g. the Mexican hat wavelet shown in Figure 3.1, should be used.

As types of filters, smooth and symmetric wavelets are necessary for most vision applications (Daubechies, 1992; Prasad *et al*, 1997). The smoothness is measured by the moment defined by:

If

$$\int_{-\infty}^{\infty} t^k \varphi(t)\, dt = 0, \quad 0 \le k \le K, \quad k, K \in \mathbb{N} \tag{3.6}$$

then the wavelet is said to have $K$ vanishing moment. The higher the $K$ is, the smoother or the more regular the wavelet. The moments of the wavelets shown in Figure 3.1 are 1 for Harr, 2 for Daubechies 2, and indefinite for both Mexican hat and Morlet. It also can be seen that Harr, Mexican hat and Morlet are all symmetric. The symmetry is significant for image processing applications because imaging systems are more tolerant of symmetric errors than asymmetric ones (Daubechies, 1992).

- Continuous or discrete

This distinction depends on the parameters of scale and shift. If both of them are continuous, then the wavelet functions map a signal into a continuous series of transformations. In the case of both parameters being discrete, the wavelet functions form a discrete sequence of transformations. This will be discussed in more detail in section 3.3.

- Orthogonal or non-orthogonal

The wavelet transform can be considered as a signal representation using a set of wavelets as basis functions. The remarkable property that is achieved by many wavelets is orthogonality. A wavelet basis set, containing of all the dilations and translations, is orthogonal when their inner products are zero:

$$\int_{-\infty}^{\infty} \varphi_{a_1,b_1}(t)\varphi_{a_2,b_2}(t)dt = \int_{-\infty}^{\infty} \frac{1}{\sqrt{a_1}}\varphi(\frac{t-b_1}{a_1})\frac{1}{\sqrt{a_2}}\varphi(\frac{t-b_2}{a_2})dt = 0, \tag{3.7}$$

$$a_1, a_2>0, b_1,b2\in\mathbb{R}$$

Orthogonal wavelets are able to produce a full and complete representation of signals without any information redundancy. The number of wavelet coefficients at each scale is the most compact in this case. On the contrary, non-orthogonal wavelets are redundant. The redundancy tends to get severe when the scale gets large because the wavelet spectrum between adjacent time intervals at large scales is highly correlated. However, the non-orthogonal wavelet transform is useful for those analyses where smooth and continuous variations of wavelet coefficients are expected. The redundancy also helps improve the ability to reject noise. Section 3.3.4 will cover a further discussion of this.

- Complex or real

Complex wavelet functions, used as the kernel in equation (3.1), can return complex transformed coefficients of both amplitude and phase. Real wavelet functions only return a single component.

- Shift invariant or shift variant

A transform operator $\boldsymbol{H}$ is shift-invariant when a delay of the input $x$ produces a delay of the output $y=\boldsymbol{H}x$ (Chui, 1992). The Fourier transform is shift-invariant because the Fourier transform of a delayed signal in time domain has a delayed phase in the frequency domain. However, not all types of the wavelet transforms possess this property. A wavelet transform is shift invariant if the transformed energy distribution is

maintained at all scales (Kingsbury, 2000a). The importance of wavelet shift invariance will be demonstrated in section 3.4

As discussed above, the wavelet transform can be classified from many points of view according to the different types of wavelets, for example, continuous or discrete, orthogonal or non-orthogonal, complex or real, and shift invariant or shift variant. Considering the characteristics of the vision task and the issue of computational efficiency, a discrete, orthogonal, real and shift invariant wavelet transform would be ideal. Various wavelet transforms with their different properties will be identified in section 3.3. In particular, shift invariance and information redundancy are major considerations discussed for each type of wavelet transforms in section 3.4. A specific wavelet transform will be chosen in Chapter 4 and its completeness will be discussed in section 4.2.

## 3.3 Wavelet Transforms

In wavelet analysis a scalable and shiftable function is used as the kernel window, i.e. the mother wavelet. The spectrum is calculated for every position as the window is shifted along the signal. Such calculation is repeated every time with a slightly shorter (or longer) window. The collection of the results is a time-scale representation of the original signal, all with different resolutions.

In contrast with the STFT that uses a constant sized window, the wavelet transform analyses a signal at multiple resolutions and the product of time interval and frequency interval is constant at all scales (Cohen and Kovacevic, 1996). If the frequency interval increases by a scale factor, the time interval decreases by the same factor. The comparison of the time-scale plane between the STFT and the wavelet transform is illustrated in Figure 3.3.

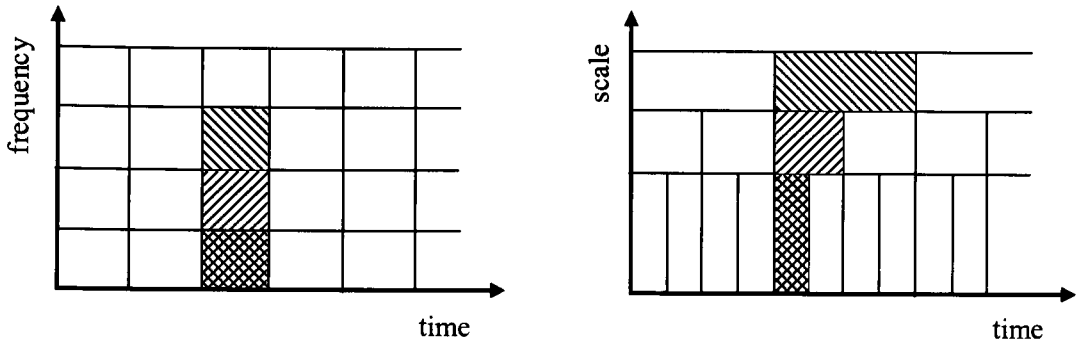



**Figure 3.3 Time-frequency/scale planes**

***Left*: STFT and *Right*: the wavelet transform**

Therefore, the wavelet transform is able to analyse high-frequency components using small windows and low-frequency components using large windows. This property is ideal when dealing with non-stationary signals that contain both short high-frequency components and long low-frequency components. A signal is non-stationary if its frequency components do not appear at all times (Teolis, 1998). Images are typical non-stationary signals, which consist of a slowly changing background corresponding to low-frequency components and rapidly changing details corresponding to high-frequency components.

In practice, the commonly used wavelet transforms can be categorized as follows.

### 3.3.1 Continuous Wavelet Transform

The form of equation (3.1) refers to the continuous wavelet transform. Parameters $a$ and $b$ are continuous values in this case. $\{\varphi_{a,b}(x)\}$ constitutes an overcomplete representation. The information is highly redundant with the continuous wavelet representation but it is shift invariant (Teolis, 1998).

### 3.3.2 Discrete Wavelet Transform

The redundancy with the continuous wavelet transform can be reduced by discretizing $a$ and $b$: $a = a_0^j$, $b = na_0^j b_0$, $a_0 > 0$, $b_0 > 0$, and $n, j \in \mathbb{Z}$. This uniformly sampled formation is called the discrete wavelet transform. The redundancy still exists but the amount of the redundancy depends on the choice of $a_0$ and $b_0$, as discussed below.

If the sampling interval $\tau = a_0^j b_0$ tends to be very small (minimum is zero), the discrete wavelet transform is close to the continuous wavelet transform above.

If the sampling interval $\tau$ increases and becomes large relative to the rate of variation of the wavelet coefficients, then the redundancy is greatly reduced. This observation is particularly prominent when $a_0 = 2$ and $b_0 = 1$. Daubechies (Daubechies, 1992) found that when $a=2^n$, $b=k2^n$ and $n, k \in \mathbb{Z}$, the basis functions $\{\varphi_{kn}(x) = 2^{-n/2}\varphi(2^{-n}x - k)\}$ are orthogonal for certain choices of wavelets, which means there is no redundancy in the representation. This is efficiently implemented by Mallat (Mallat, 1989b) using two channel filter banks and is known as Multi-Resolution Analysis (MRA). Section 3.3.3 will discuss it in more detail.

### 3.3.3 Wavelet Multiresolution Analysis

Equation (3.4), which shows the Fourier transform of a wavelet, has a zero value at the zero frequency, thus such a function can be seen as a band-pass filter. A series of dilated wavelets $\varphi_a(t) = \frac{1}{a}\varphi(t)$ can be seen as a filter bank and their Fourier transform $\Psi_a(\omega)$ must have following property according to Fourier theory:

$$\Psi_a(\omega) = |a|\Psi(a\omega) \tag{3.8}$$

This means that dilating a signal in time by a factor of two will result in compression in the Fourier spectrum by a factor of two and also shift all frequency components down by a factor of two. This observation can be used to decompose a signal into a series of wavelets. However, in order to cover the whole spectrum of a signal, a low-pass filter associated with a scaling function $\phi(t)$ was introduced by Mallat (Mallat, 1989b). The scaling function must satisfy:

$$\int_{-\infty}^{\infty} \phi(t)dt = 1 \tag{3.9}$$

As a wavelet can be considered as a band-pass filter and a scaling function as a low-pass filter, thus a series of dilated and the scaled wavelet function can be represented by a filter bank (Strang and Nguyen, 1997). The decomposition of MRA is to pass the signal through this filter bank. It is implemented by iteratively applying the band-pass and low-pass filters to split the signal spectrum into two equal parts followed by downsampling the outputs by a factor of two, which simply removes every other sampling component of a sequence. Figure 3.4 shows the process, where *Hp* and *Lp* are the band-pass and low-pass filters, and $\boxed{\downarrow 2}$ represents downsampling by two. After downsampling, the signal will then have half the number of the previous samples. The scale of the signal is doubled after the half band filtering operation. The output of *Hp* contains the high-frequency components reflecting the detailed content of the original signal and thus is called the *Detail Part*, represented by *D* in Figure 3.4. The output of *Lp* by contrast contains the low-frequency content of the original signal. As it assembles the original signal, it is thus called the *Approximation Part*, represented by *A* in Figure 3.4. The Approximation Part still contains some details that are of interest and

therefore it can be split again in the same way as for the original signal. In this way, a hierarchical decomposition at multiple levels is created. The decomposition can stop at any level. Usually the maximum number of the levels is dependent on the length of the sampling sequence of the original signal. In Figure 3.4 the structure starts with the original signal represented by $A_0$ until the *jth* level.

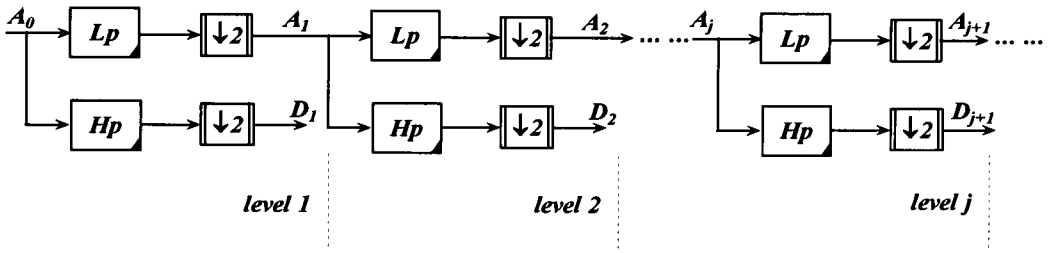



**Figure 3.4 Hierarchical MRA decomposition structure**

Although the concept of scalable windows, i.e. wavelets, has existed for nearly one hundred years, the applications of the wavelet transform have not been so popular until a fast implementation of MRA was proposed (Chui, 1992). Its computational efficiency results from the hierarchical filter bank architecture. It can be proved that the associated wavelets underlining the filters must be orthogonal so that MRA provide a complete signal representation (Mallat, 1989b).

Unfortunately, the use of the downsampling causes MRA to lack shift invariance. For example, assume a discrete system is represented by $y(n) = \sum h(n)x(n-k)$, where *h(n)* is the system impulse response, *x(n)* and *y(n)* are the input and the output. When the input is a unit impulse signal *x(0)* = 1, the output is *y(n)* = *h(n)*. If the impulse is delayed one sample in time to *x(0)* = 1, then the output is equally delayed one sample to *y(n)* = *h(n-1)*. This system, summarised in table Table 3.1, is thus shift invariant.

**Table 3.1 A shift invariant impulse response system**

| | |
|---|---|
| impulse 1: $x_1(n) = [1, 0, 0, \ldots]$ | $\longrightarrow$ response 1: $y_1(n) = [h(0), h(1), h(2), \ldots]$ |
| impulse 2: $x_2(n) = [0, 1, 0, 0, \ldots]$ | $\longrightarrow$ response 2: $y_2(n) = [0, h(0), h(1), h(2), \ldots]$ |

Now add the downsampling operation for the input and output sequences, which is shown in Table 3.2. The downsampling only keeps the odd-numbered components.

**Table 3.2 Illustration of shift variance due to downsampling**

| | |
|---|---|
| impulse 1: $\boxed{\downarrow 2}\, x_1(n) = [1, 0, 0, \ldots]$ | $\rightarrow$ response 1: $\boxed{\downarrow 2}\, y_1(n) = [h(0), h(2), \ldots]$ |
| impulse 2: $\boxed{\downarrow 2}\, x_2(n) = [0, 0, 0, 0, \ldots]$ | $\rightarrow$ response 2: $\boxed{\downarrow 2}\, y_2(n) = [0, h(1), h(3), \ldots]$ |

The result in Table 3.2 indicates that the responses after downsampling are completely different because of the delay (shift in time).

This issue was discussed by Strang (Strang and Nguyen, 1997) in detail and the shift-invariance problem was considered to be the main drawback of MRA. Therefore, it cannot be directly used for tasks such as pattern recognition and computer vision that require shift-invariant transforms.

### 3.3.4 Dyadic Wavelet Transform

Shift variance is one of the disadvantages of the general discrete wavelet transform. If orthogonality is not emphasised in the signal representation, information redundancy is

introduced. However, the wavelets need not be orthogonal and in some applications the redundancy can help to reduce the sensitivity to noise (Lang *et al*, 1995) or improve the shift invariance of the transform (Teolis, 1998).

For the purpose of shift-invariance, the dyadic wavelet transform (*DyWT*) is formulated by discretising the scale parameter $a$ along a dyadic sequence, e.g. $a=2^j$ ($j\in\mathbb{Z}$), but the translation parameter $b$ remains continuous:

$$DyWT(b,2^j) = \int x(t)\cdot 2^{-j/2}\varphi(2^{-j}(t-b))dt\,, j\in \mathbb{Z},\ b\in \mathbb{R} \tag{3.10}$$

Mallat (Mallat, 1998) proved that if the Fourier transform, $\hat{\varphi}(2^j\omega)$, of the dyadic wavelets satisfies

$$A \le \sum \left|\hat{\varphi}(2^j\omega)\right|^2 \le B \tag{3.11}$$

for two constants $0 \le A \le B$, then the transform with such dyadic wavelet bases defines a complete and stable representation. It is complete because $< x(t),\varphi_{2^j,b}(t) > = 0$ if and only if $\varphi_{2^j,b}(t) = 0$ (Teolis, 1998). Equation (3.11) guarantees its stability because the dyadic wavelets are integrable (Mallat, 1998).

The algorithmic efficiency of the dyadic wavelet transform is greatly improved compared with the continuous wavelet transform although the information is still redundant along the position axis due to the continuity of the translation parameter. However, the redundancy has distinct benefits to offer in practical signal processing applications such as image matching (Teolis, 1998). Firstly, a redundant system has an inherent degree of noise robustness that is proportional to the degree of redundancy.

The larger the degree of the redundancy, the more tolerant the transform is to perturbation and the less sensitive it is to noise. Secondly, the freedom to choose a wide selection of wavelets is gained without the requirement of non-redundancy. Many non-orthogonal wavelets with arbitrary joint time-frequency localization may be used. This makes it easier to choose a wavelet that resembles the original signal to achieve better processing as a wavelet transform actually implements correlation operation. In addition, the advantages of dyadic wavelet transform also include the property of shift invariance, which is necessary for image matching and will be discussed in section 3.4. Although its main drawback is the increased computational complexity and storage requirement as compared to MRA, its reasonable computational speed as well as all of the above advantages has made it attractive for many vision tasks. Table 3.3 lists the advantages and disadvantages of the dyadic wavelet transform. Its application to stereo matching will be investigated in chapter 4.

**Table 3.3 Advantages and disadvantages of dyadic wavelet transform**

| Advantages | Disadvantages |
|---|---|
| ▪ Shift invariance<br>▪ Inherent degree of noise robustness<br>▪ Freedom in the choice of wavelets<br>▪ Reasonable computational speed | ▪ Increased computational complexity and storage requirements compared with MRA |

### 3.3.5 Complex Wavelet Transform

All the above discussions are based on real wavelets. But for real wavelets, orthogonality (non-redundancy) and shift invariance are not compatible (Daubechies, 1992). Complex wavelets are thus introduced as mother wavelets to construct complex wavelet transform for a trade-off between orthogonality and shift invariance. Although complex wavelet bases cannot be made orthogonal, such carefully created functions (Kingsbury, 2000a) may result in limited information redundancy. Use of complex wavelets is another way to achieve shift-invariance.

There are different ways to generate complex wavelets. A review of these will be given in section 3.5.

## 3.4 Importance of Wavelet Shift-Invariance to Stereo Matching

Under the definition of shift-invariance in section 3.2, if a wavelet transform is shift-invariant, the transformed energy distribution should be maintained at all scales. As indicated in section 3.3, the efficient MRA is not shift invariant. Examples can be found in (Strang and Nguyen, 1997) and (Teolis, 1998) to show that the MRA coefficients of one signal and a version shifted by one sample give significantly different results both within and between subbands. The following paragraphs illustrate how it affects correspondence computation in stereo matching.

To compute stereo disparity, one image can be intuitively assumed to be the shifted version of the other. The value of the shift with respect to each pixel is called the disparity, which is dependent on the pixel position.

As an example of the problem introduced by shift-variance, Figure 3.5 shows two epipolar lines (Faugeras, 1993) from two stereo images. Figure 3.6 displays the wavelet

decomposition results of 5 level approximations (*a1~a5*) and details (*d1~d5*), using the MRA approach. Level 0 refers to the original signals. By comparing these decomposed plots, for instance, between left: d2 and right: d2, it can be seen that the details are not shifted according to the shifting of the original signals because the values of the sequence after downsampling depends on whether the shift value is odd or even. This demonstrates that MRA is a shift-variant transform. Such a transform does not therefore give robust results when applied to stereo matching.

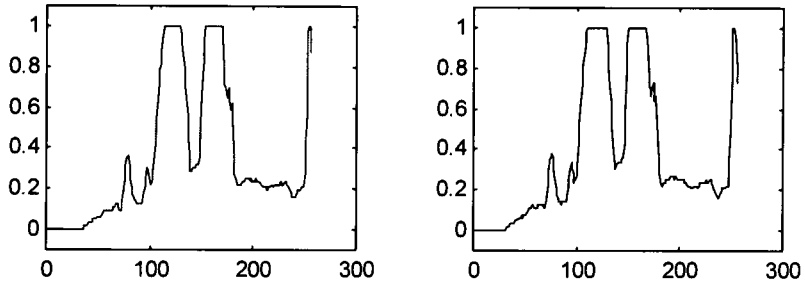



**Figure 3.5 Scan lines from two stereo images**

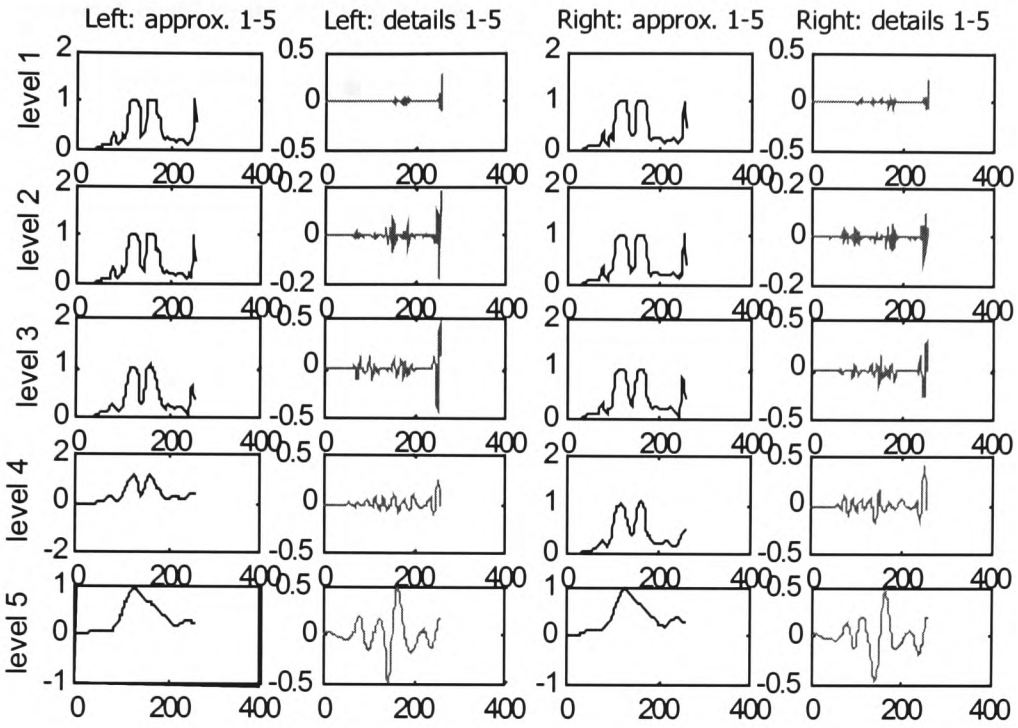



**Figure 3.6 Wavelet 1D decompositions at 5 levels**

## 3.5 Existing Wavelet-Based Matching Methods

As indicated in section 3.3.3, the development of Mallat's MRA (Mallat, 1989a) stimulated a great deal of research work because of its fast implementation but the lack of shift invariance is its main drawback. There is no doubt that vision tasks such as stereo matching require shift invariant transform as illustrated in section 3.4. Conversion of shift variance to shift invariance is of vital importance to solve the correspondence problem. This section reviews the main existing wavelet-based matching methods.

In 1991 Mallat (Mallat, 1991) studied the conventional signal representation using the positions of multiscale zero-crossings and pointed out that such a representation is not stable. By adding a measure of the length between neighbouring zero-crossings based

on wavelet coefficients, he built a new stabilized and complete wavelet-based zero-crossing representation. Importantly, it is shift invariant. As the zero-crossings of a wavelet transform reflect the locations of discontinuities in the signal, therefore, the disadvantage of this method is the limited coded input for the case of smooth images with sparse singularities.

In 1992, Simoncelli *et al.* (Simoncelli *et al*, 1992) introduced a new term, *shiftability,* which means that a transform treats its input signal in a uniform manner regardless of the relative alignment of the input, by exploring shiftable wavelet transforms in terms of spatial position, scale and orientation (in the two-dimensional case). They proposed the concept of *jointly shiftable* and developed a one-dimensional wavelet transform that is jointly shiftable in translation and scale, and a two-dimensional transform that is jointly shiftable in translation and orientation. They implemented this by designing a steerable pyramid. Stereo matching was performed using a coarse-to-fine least-squares gradient-based disparity estimator. However, although the method does give shiftability, its drawback is lower speed.

In order to achieve shift invariance, some researchers turned their attention to complex wavelet transforms that offer a number of potentially advantageous properties such as shift invariance, symmetry and approximate linear phase. The linear phase property can be used in phase-based matching as discussed in Chapter 2.

Magarey and Kingsbury (Kingsbury and Magarey, 1996;(Magarey and Kingsbury, 1995) designed a complex wavelet transform, which is approximately shift invariant, to estimate motion and displacement. It was hierarchically generated by complex Gabor-like wavelet and scaling filters. Its architecture resembles the Mallat's MRA analysis, but with two parallel trees corresponding to real and imaginary parts respectively.

Under this structure, the magnitude of the wavelet coefficients remains approximately unchanged while the phase follows the linear behaviour of the associated wavelet filter. With specially designed wavelets, such a transform can be used to detect phase shift produced by small displacement values provided that the original image is reasonably flat in the spectral domain.

Modified versions of such filter architecture designs were also developed later in (Kingsbury and Magarey, 1996; Magarey and Dick, 1998; Kingsbury, 2000b). The most recent design was to generate complex coefficients by using a dual tree of wavelet filters by Kingsbury (Kingsbury, 2000c). It was claimed that the advantages of this structure include approximate shift invariance, limited redundancy (*2:1* for *1*-dimension, $2^m$*:1* for *m*-dimension), good directional selectivity in two or higher dimensions and perfect reconstruction.

Spaendonck *et al.* (Spaendonck *et al*, 2000) challenged the Kingsbury's dual tree structure and claimed that neither tree genuinely corresponds to a wavelet transform. He then put forward a new complex wavelet transform with no redundancy, superior directional selectivity and perfect reconstruction. Nevertheless, the shift invariance property and the application to stereo matching were not examined in the paper.

Other matching approaches using complex wavelets, e.g. complex Daubechies wavelet, can be found in (Pan, 1996).

A comparison of the four wavelet-based methods discussed above that have been involved in stereo matching applications is given in Table 3.4. Both the advantages and the disadvantages are listed in the table.

**Table 3.4 A comparison of four wavelet-based matching methods**

| | Mallat's Zero-Crossings | Simoncelli's Shiftability | Magarey's Complex Trees | Kingsbury's Dual Trees |
|---|---|---|---|---|
| ***Advantages*** | ▪ Shift invariant<br>▪ Efficient computation | ▪ Shift invariant<br>▪ Jointly invariant with scale etc. | ▪ Approximately shift invariant<br>▪ Simple structure and implementation | ▪ Approximately shift invariant<br>▪ Limited redundancy |
| ***Disadvantages*** | ▪ Reduced performance with smooth images | ▪ High computational complexity | ▪ Only small disparity values can be detected | ▪ Complicated filter bank architecture |
| ***Common Points*** | ▪ With information redundancy<br>▪ Applying coarse to fine matching strategy | | | |

This section has given an overview of wavelet-based matching methods that have been reported in the literature. The advantages and the disadvantages suggest that there still exists some opportunities to investigate a simple way to better achieve the matching task. Chapter 4 will propose such a matching approach using a dyadic wavelet transform. The benefits of applying dyadic wavelet transform have been discussed in section 3.3.4. In addition, the implementation structure will be adjusted in order to improve computational efficiency. The matching results of applying the dyadic WT will be presented in Chapter 6.

## 3.6 Summary

This chapter has presented the basic wavelet theory necessary for the stereo matching task. The categories of wavelet transforms and their properties have been reviewed. The

issue of shift invariance of various wavelet transforms has been highlighted. The importance of the wavelet shift invariance to the specific stereo matching task has been identified and approaches to achieve shift invariance have been discussed. Some existing stereo matching methods using shift invariant wavelet transforms have been outlined and four of them, Mallat's Zero-Crossings, Simoncelli's Shiftability, Magarey's Complex Trees and Kingsbury's Dual Trees, have been compared in terms of their advantages and disadvantages.

It can be seen that seeking a balance among the properties of shift invariance, minimum redundancy and fast and accurate matching is still an open problem. A new approach using a shift invariant wavelet transform to stereo matching is developed in this thesis and is discussed in detail in Chapter 4.

## 3.7 References

Battle, G. 1988. Heisenberg Proof of the Balian-Low Theorem. *Lett. Math. Phys.*, (15), pp. 175-177.

Chui, C. K. 1992. *Wavelets: A Tutorial in Theory and Applications.* San Diego, USA: Academic Press, Inc.

Cohen, A. and Kovacevic, J. 1996. Wavelets: The Mathematical Background. *Proceedings of the IEEE*, **84** (4), pp. 514-522.

Daubechies, I. 1992. *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics.

Faugeras, O. 1993. *Three Dimensional Computer Vision: a Geometric Viewpoint.* UK: The MIT Press.

Frazier, M. 1999. *Introduction to Wavelets through Linear Algebra.* Springer-Verlag.

Hubbard, B. B. 1998. *The World According to Wavelets: The Story of a Mathematical Technique in the Making.* A K Peters Ltd.

Kaiser, G. 1994. *A Friendly Guide to Wavelets.* Boston: Birkhäuser.

Kingsbury, N. G. 2000a . *Complex Wavelets and Shift Invariance.* Proc IEE Colloquium on Time-Scale and Time-Frequency Analysis and Applications, IEE. London.

Kingsbury, N. G. 2000b . Complex Wavelets for Shift Invariant Analysis and Filtering of Signals. *Journal of Applied Computation and Harmonic Analysis,*

Kingsbury, N. G. 2000c . *A Dual-Tree Complex Wavelet Transform with Improved Orthogonality and Symmetry Properties.* IEEE International Conference on Image Processing. pp. 375-378, Vancouver, Canada.

Kingsbury, N. G. and Magarey, J. 1996. *Wavelets in Image Analysis: Motion and Displacement Estimation.* Proc. Irish DSP and Control Conference. pp. 199-217, Dublin.

Lang, M., Guo, H., Odegard, J. E., Burrus, C. S. and Wells, R. O. 1995. *Nonlinear Processing of a Shift Invariant DWT for Noise Reduction.* Proc. Wavelet Application II, SPIE. pp. 64 0-651 Orlando, FL.

Magarey, J. and Dick, A. 1998. *Multiresolution Stereo Image Matching Using Complex Wavelets.* International Conference on Pattern Recognition 1998. pp. 4-7,

Magarey, J. 1997. *Motion Estimation Using Complex Wavelets.* PhD thesis. Department of Engineering, Cambridge University.

Kingsbury, N. G. and Magarey, J. 1996. *Wavelets in Image Analysis: Motion and Displacement Estimation.* Proc. Irish DSP and Control Conference. pp. 199-217, Dublin.

Mallat, S. 1989a. Multifrequency Channel Decomposition of Images and Wavelet Models. *IEEE Trans ASSP*, **37** (12), pp. 2091-2110.

Mallat, S. 1989b. A Theory for Multiresolution Signal Decomposition: the Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11** (7), pp. 674-693.

Mallat, S. 1991. Zero-Crossings of a Wavelet Transform. *IEEE Transactions on Information Theory*, **37** (4), pp. 1019-1033.

Mallat, S. 1998. *A Wavelet Tour of Signal Processing.* Academic Press.

Pan, H. P. 1996. *General Stereo Image Matching Using Symmetric Complex Wavelets.* SPIE Proceedings.

Prasad, L., Iyengar, S. S. and Ayengar, S. S. 1997. *Wavelet Analysis with Applications to Image Processing.* CRC Press.

Simoncelli, E. P., Freeman, W. T., Adelson, E. H. and Heeger, D. J. 1992. Shiftable Multiscale Transforms. *IEEE Trans. Information Theory*, **38** (2), pp. 587-607.

Spaendonck, R., Fernandes, F. C. A., Coates, M. and Burrus, C. S. 2000. *Non-Redundant, Directionally Selective, Complex Wavelets.* IEEE International Conference on Image Processing. pp. 379-382,

Strang, G. and Nguyen, T. 1997. *Wavelets and Filter Banks.* Revised end. Wellesley-Cambridge Press.

Teolis, A. 1998. *Computational Signal Processing with Wavelets.* Birkhauser Boston.

Torrence, C. and Compo, G. P. 1998. A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, **79** (1), pp. 61-78.

Valens, C. 1999. A Really Friendly Guide to Wavelets. *http://perso.wanadoo.fr/polyvalens/clemens/wavelets/wavelets.html*,

# 4 Stereo Matching by Dyadic Wavelet Transform

## 4.1 Introduction

The purpose of this chapter is to identify a shift invariant wavelet transform for stereo matching with a view to solving the windowing problem and reducing the computational complexity compared to the conventional correlation-based matching method discussed in chapter 2.

In terms of the properties of shift invariance and computational complexity, the Dyadic Wavelet Transform (DyWT) possesses the practical advantages over other wavelet transforms that have been presented in chapter 3. The discretisation along the dyadic scales of DyWT enables considerably efficient computation. However, the DyWT results in redundancy along the translation axis, but this redundancy cannot be avoided because a continuous translation variable is necessary in order to generate a dense disparity map output. This chapter begins with the formulation of the DyWT and its fast implementation (Holschneider *et al*, 1989). A DyWT based matching method is then developed. Based on the conventional correlation measure, the Sum of Squared Differences (SSD) as discussed in chapter 2, a new measure using DyWT coefficients is defined and called W-SSD. This measure allows a coarse-to-fine multi-scale disparity estimate associated with the disparity due to the hierarchical structure of DyWT. The next chapter, Chapter 5, presents the experiments and results using the new DyWT based matching approach proposed in this chapter.

## 4.2 The 1-D Dyadic Wavelet Transform

This section covers the development of a stable and complete signal representation using dyadic wavelets.

### 4.2.1 Wavelet Frames

A signal *x(t)* can be represented by a set of orthogonal basis functions. Such a representation has no information redundancy. More generally, a non-orthogonal set of basis functions (i.e. not linearly independent) can also be used provided that some redundancy is allowed. Such a family of basis functions can be generalised into the concept of frames (Appendix B) defined below.

A sequence of functions $\{\alpha_n\}_{n\in\mathbb{N}}$ in a Hilbert space $\mathcal{H}$ is called a frame (Daubechies, 1992) if there exist two constants $A > 0$ and $B < \infty$ so that for any $x \in \mathcal{H}$,

$$A\|x\|^2 \le \sum_n |< x,\alpha_n >|^2 \le B\|x\|^2 \tag{4.1}$$

$A$ and $B$ are called the frame bounds. When $A = B$, the frame is considered to be tight. Usually frames, even tight frames, are not orthogonal bases. A frame constitutes an orthogonal basis if and only if $A = B = 1$. If $A > 1$, the frame is redundant. $A$ gives the redundancy ratio and is called the minimum redundancy factor.

As discussed in chapter 3, the dyadic wavelet bases are mathematically represented by:

$$\varphi_{2^j b}(t) = \frac{1}{\sqrt{2^j}}\varphi(\frac{t-b}{2^j}),\ j \in \mathbb{Z}, b\in\mathbb{R} \tag{4.2}$$

Although $b$ is a continuous parameter in DyWT in theory, it has to be discretised when the implementation is conducted in software. Without loss of generality, let $b = n2^j b_0, j \in \mathbb{Z}$ and $b_0 \in \mathbb{R}$. If the sampling interval $2^j b_0$ is small enough, $b$ can be very close to continuous. Then equation (4.2) becomes:

$$\varphi_{2^j n}(t) = \frac{1}{\sqrt{2^j}} \varphi\left(\frac{t - n2^j b_0}{2^j}\right) \tag{4.3}$$

As stated in chapter 3, to be eligible for a wavelet, a function must satisfy the admissibility condition (Grossmann and Morlet, 1984):

$$C_\Psi = \int \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < +\infty \tag{4.4}$$

Daubechies (Daubechies, 1992) proved that if $\{\varphi_{2^j n}\}$ constitutes a frame in $\mathcal{H}$ then the frame bounds must satisfy:

$$A \le \frac{C_\Psi}{b_0 \log_e 2} \le B \tag{4.5}$$

This is equivalent to the following inequality (Daubechies, 1992; Chui and Shi, 1993; Chui and Shi, 1993):

$$A \le \frac{1}{b_0} \sum \left|\Psi(2^j \omega)\right|^2 \le B \tag{4.6}$$

In particular, the condition (4.6) under which the collection $\{\varphi_{2^j n}\}$ or $\{\varphi_{2^j b}\}$ constitute a frame requires that condition (4.5) holds for $\varphi$ (Daubechies, 1992).

### 4.2.2 Completeness and Stability of DyWT

Consider the DyWT that consists of the wavelet frame $\{\varphi_{2^j b}\}$:

$$DyWT(b,2^j) = \int_{-\infty}^{\infty} x(t) \cdot \frac{1}{\sqrt{2^j}} \varphi(\frac{t-b}{2^j}) dt \tag{4.7}$$

Using the format of the inner product, it can also be written as:

$$DyWT(b,2^j) = < x(t) \cdot \varphi_{2^j b}(t) > \tag{4.8}$$

To prove that this representation is complete and stable is equivalent to answering the question whether the wavelet coefficients $< x \cdot \varphi_{2^j b} >$ completely characterize $x$ or, in other words, whether $x$ can be reconstructed in a numerically stable way from the coefficients.

Having studied the connection between wavelet frames and numerically stable and complete signal representations, Daubechies (Daubechies, 1992) showed that for any appropriate wavelet $\varphi$ and sampling constant $b_0$, there exists a dual wavelet frame $\tilde{\varphi}_{2^j b}$, with which $x$ can be reconstructed by:

$$x = \sum_{j,b} < x \cdot \varphi_{2^j b} > \tilde{\varphi}_{2^j b} \tag{4.9}$$

The calculation of $\tilde{\varphi}_{2^j b}$ is not discussed (Mallat, 1998) as signal processing is the target here rather than signal reconstruction.

Equation (4.9) implies that in order for the DyWT to be complete and stable, it is required that $\{\varphi_{2^j b}\}$ constitute a frame. Conditions (4.4) and (4.6) guarantees such a frame.

### 4.2.3 Wavelets to Be Used

Any functions that meet the conditions represented by equations (4.4) and (4.6) are eligible as wavelet frames. Two of them (Daubechies, 1992) are given below as examples and are to be used later in chapter 5.

- Mexican hat wavelet

This is the normalised second derivative of the Gaussian function, $\exp(-t^2/2)$:

$$\varphi(t) = \frac{2}{\sqrt[4]{9\pi}}(1-t^2)e^{-\frac{t^2}{2}} \tag{4.10}$$

Its waveform is plotted in Figure 3.1. Its Fourier transform is:

$$\Psi(\omega) = -\frac{\sqrt[4]{64\pi}\,\omega^2}{\sqrt{3}}e^{-\frac{\omega^2}{2}} \tag{4.11}$$

For $\{\varphi_{2^j b}\}$, the frame is nearly tight when $A \approx B = \frac{2}{b_0} C_\Psi \log_2 e$ for all $b_0 \leq 0.75$ (Daubechies, 1992) (Mallat, 1998). The frame bound is inversely proportional to $b_0$, which measures the redundancy of the frame. When $b_0$ is halved, the redundancy doubles.

- Morlet wavelet

This is a sinusoid-modulated Gaussian function. Its mathematical expression in the time $\varphi(t)$ and frequency domain $\hat{\varphi}(\omega)$, respectively, are:

$$\varphi(t) = \frac{1}{\sqrt[4]{\pi}}(\cos \varepsilon_0 t - e^{-\frac{\varepsilon_0^2}{2}})e^{-\frac{t^2}{2}} \tag{4.12}$$

$$\hat{\varphi}(\omega) = \frac{1}{\sqrt[4]{\pi}}(e^{-\frac{(\omega-\varepsilon_0)^2}{2}} - e^{-\frac{(\omega^2+\varepsilon_0^2)}{2}}) \tag{4.13}$$

where $\varepsilon_0$ is a constant usually chosen to be 5. For this value of $\varepsilon_0$, the second component in equation (4.12) becomes so small that it can be neglected in practice. In contrast with the Mexican hat wavelet, the sampling constant, $b_0$, of Morlet wavelet needs to be greater than 2.5 in order for $\{\varphi_{2^j b}\}$ to be a tight frame (Daubechies, 1992).

### 4.2.4 "Algorithme à Trous"

A fast calculation for DyWT has been developed and is called in French the *Algorithme à Trous* (Holschneider *et al*, 1989). It is computed with a fast filter bank algorithm. The structure, shown in Figure 4.1, is similar to MRA of Figure 3.4 but without subsampling.

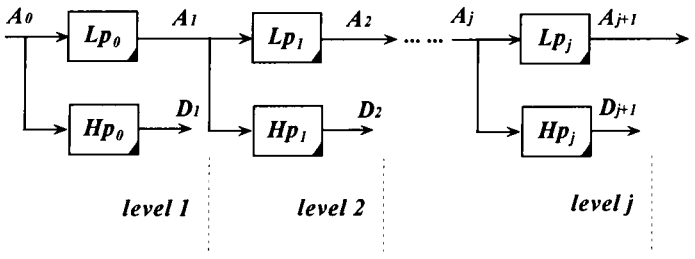



**Figure 4.1 Structure of algorithme à Trous**

Suppose that the scaling and wavelet function $\phi$ and $\varphi$ are designed with a pair of finite impulse response low-pass and high-pass filters *Lp* and *Hp*, whose Fourier transforms are:

$$\Phi(\omega) = \frac{1}{\sqrt{2}} \hat{L}p(\frac{\omega}{2})\Phi(\frac{\omega}{2}) \tag{4.14}$$

$$\Psi(\omega) = \frac{1}{\sqrt{2}} \hat{H}p(\frac{\omega}{2})\Phi(\frac{\omega}{2}) \tag{4.15}$$

where $\Phi(\omega), \Psi(\omega)$, $\hat{L}p(\omega)$ *and* $\hat{H}p(\omega)$ are Fourier transforms of the scaling function, the wavelet, the low-pass filter and the high-pass filter.

The algorithm is implemented by cascading convolutions with dilated filter banks. That is:

for any $j \geq 0$,

$$A_{j+1}(t) = A_j(t) * Lp_j(t),\ D_{j+1}(t) = A_j(t) * Hp_j(t),$$

$$\text{and } A_j(t) = \frac{1}{2}(A_{j+1}(t) * Lp_j(t) + D_{j+1}(t) * Hp_j(t)) \tag{4.16}$$

## 4.3 The Dyadic Wavelet Transform on Images for Matching

When working on images, the two-dimensional wavelet transform is usually applied. However, due to the specialty of the stereo matching task, one-dimensional searching can be used on images by applying the epipolar constraint (Faugeras, 1993), which was illustrated in chapter 2. This thesis makes use of the one-dimensional wavelet transform along epipolar lines.

### 4.3.1 Working on Images under the Epipolar Constraint

As shown in Figure 2.2 of Chapter 2, corresponding points must lie on the associate epipolar lines. Finding the right epipolar line on which the search starts is the first step for finding the corresponding points.

In most of the cases, stereo cameras are placed parallel to each other, i.e. the optical axes of the cameras are parallel, as illustrated in Figure 2.1 of chapter 2. This case is re-plotted in Figure 4.2,where the extension of one row in the left image plane ***L*** coincides with its epipolar line at the same line in the right image plane ***R***. This implies that there is only horizontal displacement. In other words, to locate the point corresponding to $m_L(j_L, n_L)$ of the left image, it would be enough to search along the scanline $n_R = n_L$ in the right image (Trucco and Verri, 1998).

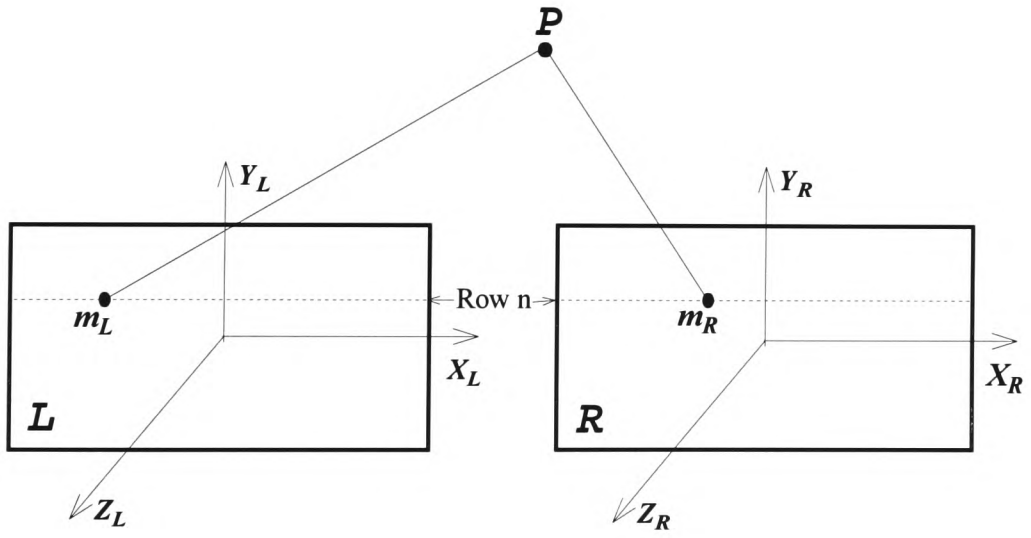



**Figure 4.2 Parallel cameras**

If the stereo cameras are not parallel to each other, both horizontal and vertical disparities would exist. The disparity values are represented by a two-dimensional

vector. The general case has been shown in Figure 2.2 of chapter 2. It is simplified in Figure 4.3. The epipolar lines no longer coincide with the image rows. Instead, they are determined by the intersection between the epipolar plane and the image planes, which has been explained in chapter 2.

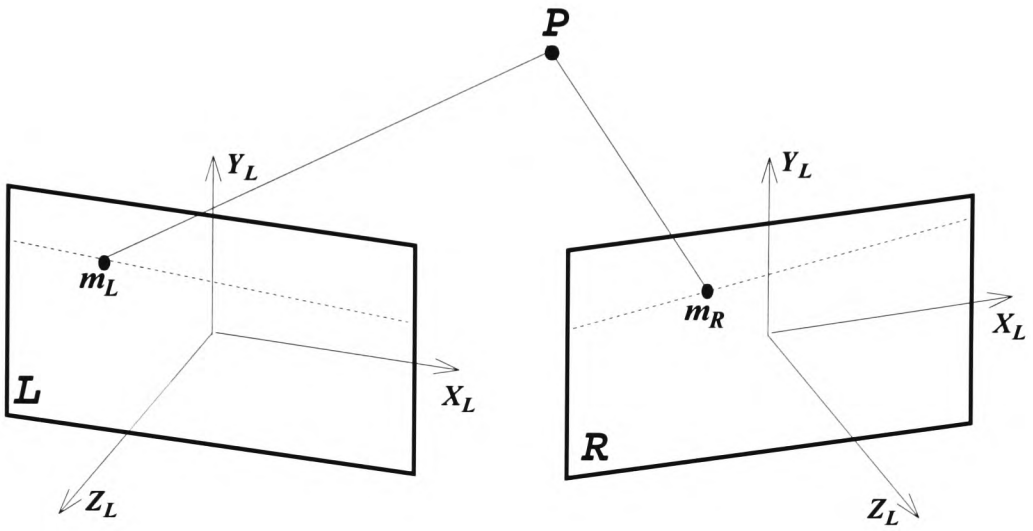



**Figure 4.3 Non-parallel cameras**

### *4.3.2* Wavelet-Based SSD Measure

Principally, disparity analysis is the evaluation of the existing geometrical differences between a pair of similar patches of images. A conventional similarity measure adopts the sum of squared differences (SSD) between the small patches within matching windows, as defined in section 2.3 of chapter 2.

In this thesis, a new SSD measure using the coefficients of the dyadic wavelet transform, presented in chapter 3, rather than the image intensity values is defined. Let $epL(x_L)$ and $epR(x_R)$ be two epipolar lines of the left and right stereo images, respectively. Their dyadic wavelet transforms at scale $2^j$ are represented by $DyWT_l$

$(2^j,x_L)$ and $DyWT_r(2^j,x_R)$, where $j \in \mathbb{N}$. For each point $x_L$ of the left image, if its corresponding point in the right image is $x_R$, then the disparity is $d = x_L - x_R$. Given $x_L$, the method used in this thesis to find $x_R$ is by minimizing the following wavelet-based SSD (W-SSD) measure, denoted by $wssd(2^j,d)$:

$$wssd\ (2^j,d) = \sum_{\tau=-2^j\sigma}^{2^j\sigma} \left| DyWT_{\ l}(2^j,x_L+\tau) - DyWT_{\ r}(2^j,x_L-d+\tau) \right|^2 \qquad (4.17)$$

where $\tau$ is the index of the comparative window and $2^j\sigma$ is the half window width. These parameters at $r(x_R)$ are illustrated in Figure 4.4.

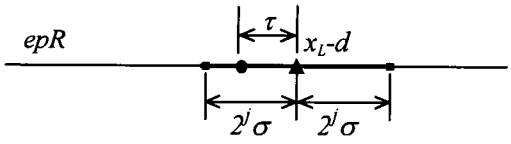



**Figure 4.4 W-SSD comparative window**

The corresponding point is assigned to the pixel that has the minimum *wssd* value in the comparative window, i.e. the disparity of $x_L$ at one scale is

$$d(2^j,x_L) = \arg\min_d \{wssd\ (2^j,d)\},\ d \in [-2^j\sigma, 2^j\sigma] \qquad (4.18)$$

### 4.3.3 Performance of the W-SSD

The performance of the W-SSD depends on the following factors:

i. The choice of $\sigma$

The characteristic of a wavelet function is its locality in both spatial and spectral domains. As discussed in chapter 3, wavelets have either a compact support or an

effective support (Chui, 1992). The value of $\sigma$ can be chosen to be half the support width of a mother wavelet. For example, with the Mexican hat wavelet, $\sigma$ is 5 and with the Morlet wavelet, $\sigma$ is 4. $2\sigma$ covers the whole period of a mother wavelet. It is therefore reasonable to use $2\sigma$ as the comparison window. For different scale, this window size should be $2^j*2\sigma$. In this way, an automatic windowing for matching is obtained which is superior to the conventional SSD-based matching approach.

ii. The type of wavelets

Based on the discussion in section 3.2, smooth and symmetric wavelets such as the Mexican hat and Morlet wavelets may be more suitable for visual applications. The results of applying different wavelets to stereo matching will be compared in chapter 5.

iii. The method of combining the *wssd* values at various scales

In contrast to the conventional SSD, which is only a function of the pixel positions, the W-SSD method provides comparative results at various scales when it is applied to the same pair of epipolar lines. The method used to combine the *wssd* values from each scale determines the precision of the matching results. The coarse-to-fine strategy discussed in chapter 2 will be modified and used for W-SSD based matching in section 4.3.4.

### 4.3.4 Hierarchical Matching Process

The matching process starts with the selection of a mother wavelet and suitable parameter $\sigma$. Next, given a point $m_L$ in the left image, a set of *wssd* values within the neighbourhood around the position of $m_L$ using equation (4.17) at the coarsest scale are computed. Of all the *wssd* values, a scale-dependant $m_R$ is determined so that the

$wssd(2^j,d)$ has the minimum value. Additionally, in the process of the computation, the constraints discussed in chapter 2, e.g. similarity, uniqueness, ordering and continuity are all used to guide the matching. The use of these constraints can be found in a wide range of conventional matching methods (Marr and Poggio, 1976; Grimson, 1985; Kanade and Okutomi, 1994).

For example, Figure 4.5 illustrates how the ordering constraint is used in this thesis. In this figure, $ep_L$ and $ep_R$ are epipolar lines; $x_{L1}$ and $x_{R1}$ are corresponding points; $x_{L2}$ is the point right next to $x_{L1}$ in $ep_L$ and also the centre point of the original searching area in $ep_R$.; $\Pi$ is the effective searching area; $x_{R2}$ is the searching result, i.e. the corresponding point of $x_{R1}$. Assume $x_{L1}$ and $x_{R1}$ have been found to be the corresponding points between $ep_L$ and $ep_R$. Considering the next point $x_{L2}$ to $x_{L1}$ on the line $ep_L$, the searching area for its corresponding point $x_{R2}$ without the ordering constraint should be $(x_{L1}-2^j\sigma, x_{L1}+2^j\sigma]$ by default. Under the ordering constraint, the possible position of $x_{R2}$ must be at the right side of $x_{R1}$ because $x_{L2}$ lies to the right of $x_{L1}$. This observation means that the searching area for $x_{R2}$ could start with the position of $x_{R1}$ instead of $x_{L2}-2^j\sigma$ in the right image. In Figure 4.5, the actual searching length is the highlighted section in $ep_R$. Thus, the computational expense is reduced under the ordering constraint. When matching is performed from left to right, the effective searching area $(x_{R1}, x_{L2}+2^j\sigma]$, denoted by $\Pi$, is actually applied.

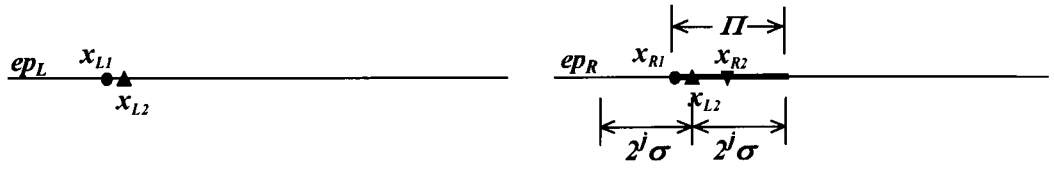



**Figure 4.5 Matching under the ordering constraint**

The ordering constraint is without doubt powerful. However, it fails due to self-occlusion when pole-like objects are in the foreground as shown in Figure 4.6, where point $X_2$ falls into the region known as the forbidden zone (Baker and Binford, 1981): the shaded area. In the left epipolar line $ep_L$, $x_{L1}$ is to the right of $x_{L2}$, whereas in the right epipolar line $ep_R$, this ordering is reversed. To comply with the ordering constraint, the points which provide information about the object front surface must lie entirely within the non-shaded area in Figure 4.6 (Davies, 1997).

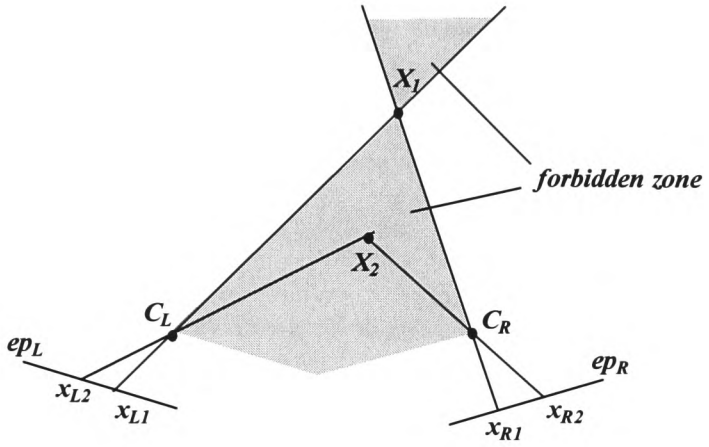



**Figure 4.6 An exception of ordering of object points**

Occlusion is a difficult phenomenon to be dealt with by stereo implementation. Special treatment must be given with complex scenes containing occlusion. As the focus of this thesis is on the applicability of wavelet-based stereo matching development, occlusion is not considered at this stage and recommended as part of the future work as discussed in section 7.4.

If the original image size is 64*64, then the maximum scale number can be $2^6$. Following the above computational process, each pixel $x$ would have 6 disparity values from $d(2^1,x)$...to $d(2^6,x)$. Each of them may correspond to a different disparity value.

The matching constraint of uniqueness specifies no more than one corresponding point. Therefore, same method of counting all of these values is needed. The DyWT based coarse-to-fine combination strategy used in this thesis is described below.

The proposed new method of combining disparities from different scales is based on the fact that at any given scale, $j$, the maximum detectable disparity is $\pm 2^j\sigma$. This is because at each scale the searching area is $[-2^j\sigma, 2^j\sigma]$ relative to the central point of the comparative window. Therefore, for the above 6 scaled example, if the $\sigma$ of the wavelet being used is 2 then the maximum detectable disparities at consecutive scales are in turn 4, 8,16, 32 and 64. In other words, different scales correspond to different preferential ranges of disparity. The optimised disparity areas should be [0, 4) for scale 1, [4, 8) for scale 2, [8, 16) for scale 3 and so on.

Starting with the coarsest scales from pixel to pixel, if the disparity at the coarsest scale belongs to its scale-preferred disparity range, then this disparity is accepted as the final output for the pixel. If the disparity is less than the range, then the consideration for the pixel disparity is passed to the next finer scale. The process is repeated until the finest scale is obtained.

Consider the case of one pixel matching. Let the final result, the disparity of pixel $x_L$, denoted by $d(x_L)$, and the maximum scale number be $2^J$. Assume the scale number is big enough to cover the maximum real disparity, which is less than $2^J\sigma$. The selective procedure for the right scale hence the right disparity is shown as a flow chart in Figure 4.7. If no suitable disparity and scale are detected, then the output is assigned as *Null*, which means that pixel does not have a corresponding point at all or some error occurs (Marr, 1982).

As an example, if the disparities at pixel $x_L$=20 at different scales respectively are

$d(6,20) = 1$, $d(5,20) = 3$, $d(4,20) = 5$, $d(3,20) = 4$, $d(2,20) = 6$, $d(1,20) = 4$,

then according to Figure 4.7 the computational result of $d(2,20) = 6$ should be saved as the final disparity for pixel 20, i.e. $\boldsymbol{d}(20)=6$. Actually the decision can be made when the detected scale reaches 2. Therefore, the computation of $d(1,20)$ is not needed.

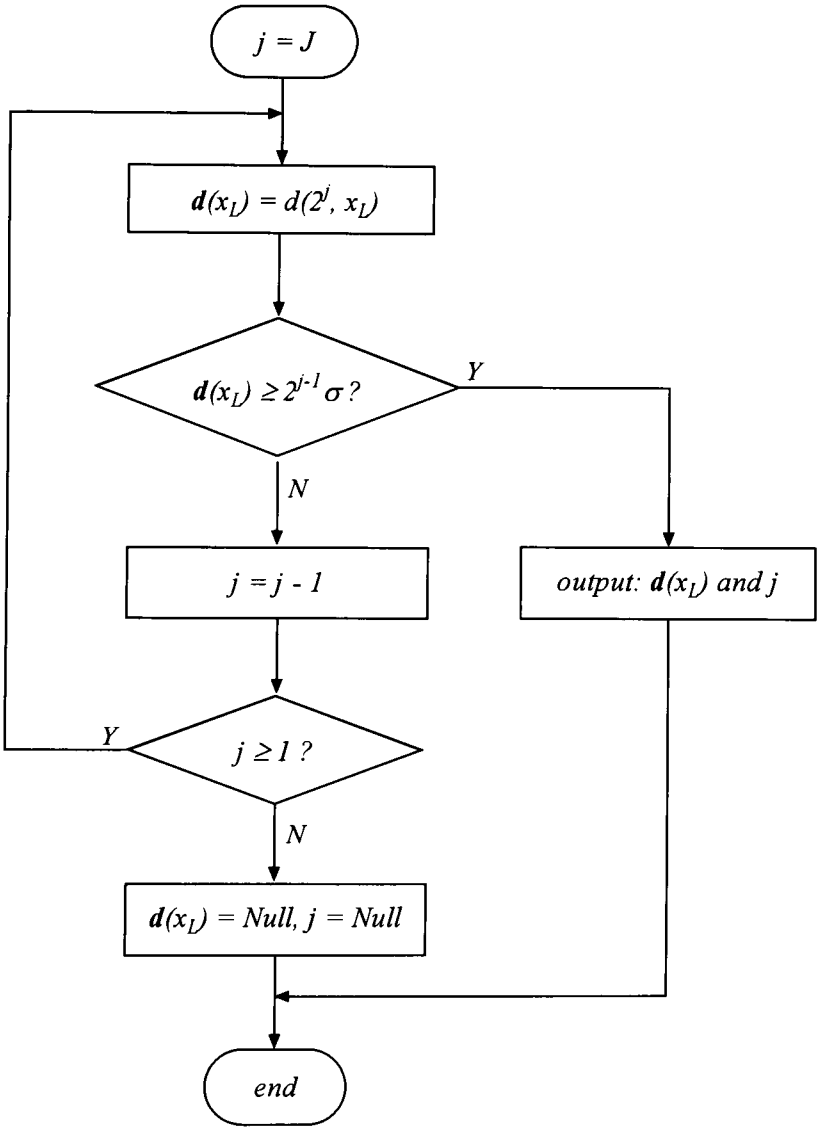



**Figure 4.7 Flow chart of W-SSD coarse-to-fine disparity combination for one pixel**

## 4.4 Implementation Structure

The proposed W-SSD method described above can be summarized as the following steps. The inputs of the method are two epipolar lines *epL* and *epR*, and the output is the disparity vector ***d(x)*** relative to the pixels of *epL*.

i. *EpL* and *epR* are decomposed into the wavelet representations at dyadic scales according to equation (4.7).

ii. At the coarsest scale, the disparity values are computed for each pixel at *epL* using the W-SSD measure defined in (4.17). They are examined to see whether they belong to the disparity range for that scale. For those points that get the disparities, save them and mark the positions.

iii. For the unmarked pixels, pass to the next lower scale. Repeat step ii until the determination of the disparity at the finest scale is complete.

This structure is illustrated in Figure 4.8.

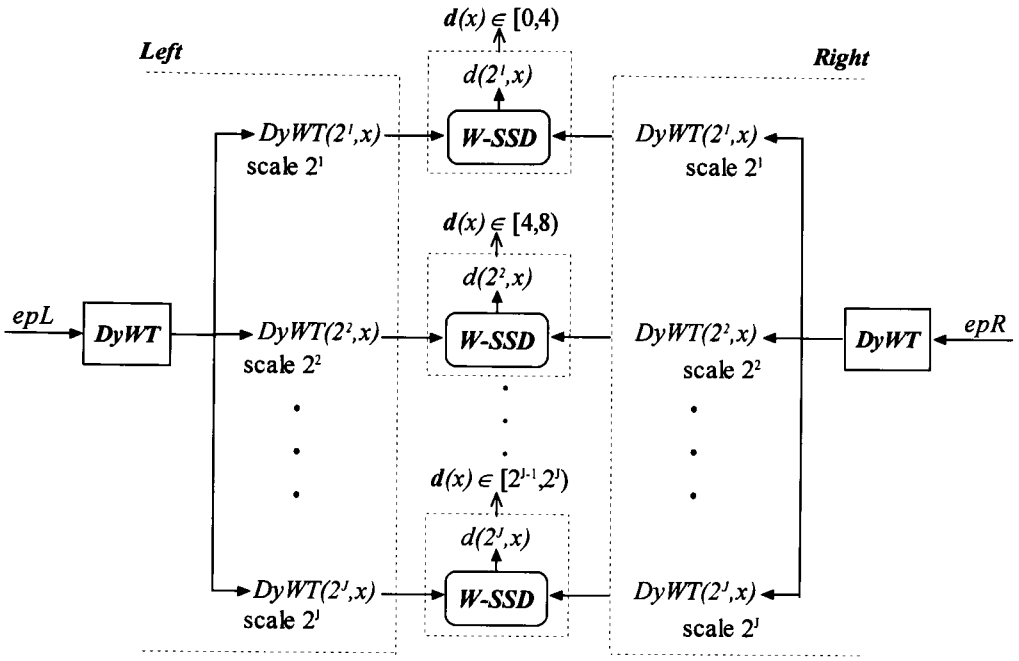



**Figure 4.8 The W-SSD multi-scale stereo matching method**

## 4.5 Summary

This chapter has presented a new DyWT based approach to computing stereo disparity maps. The DyWT has been discussed in detail in the one-dimensional case. As a signal representation, the completeness and stability of the DyWT have been discussed using the theory of frames. Two appropriate wavelet frames, the Mexican hat and Morlet wavelets have been formulated as they are used for the stereo matching application in later chapters. The fast implementation of DyWT, *Algorithme à Trous*, has also been discussed. In the two-dimensional image case, the epipolar constraint has been reviewed as it is used to allow the reduction of searching from two dimensions to one dimension. Based on the DyWT coefficients, a novel matching approach using W-SSD measure for similarity comparison has been proposed in this chapter.

The main contributions of this chapter are the definition of the multi-scale W-SSD measure using the coefficients of the DyWT and the development of combining the matching results from different scales based on the detectable minimum disparity at each scale. To test the proposed approach, experiments with various images will be made in chapter 6.

## 4.6 References

Baker, H. H. and Binford, T. O. 1981. *Depth from Edge and Intensity Based Stereo.* International Joint Conference on Artificial Intelligence. pp. 631-636,

Chui, C. K. 1992. *Wavelets: A Tutorial in Theory and Applications.* San Diego, USA: Academic Press, Inc.

Chui, C. K. and Shi, X. 1993. Inequalities of Littlewood-Paley Type for Frames and Wavelets. *SIAM J. Math. Anal.*, **24** (1), pp. 263-277.

Daubechies, I. 1992. *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics.

Davies, E. R. 1997. *Machine Vision: Theory, Algorithms, Practicalities.* 2nd end. Academic Press.

Faugeras, O. 1993. *Three Dimensional Computer Vision: a Geometric Viewpoint.* UK: The MIT Press.

Grimson, W. E. L. 1985. Computational Experiments with a Feature-Based Stereo Algorithm. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **7** (1), pp. 17-34.

Grossmann, A. and Morlet, J. 1984. Decomposition of Hardy Functions into Square

Intergrable Wavelets of Constant Shape. *SIAM J. of Math. Anal.*, **15** (4), pp. 723-736.

Holschneider, M., Kronland-Martimet, R., Morlet, J. and Tchamitchian, P. 1989. A Real-Time Algorithm for Signal Analysis with the Help of the Wavelet Transform. *In:* Anonymous. *Wavelets, Time-Frequency Methods and Phase Space.* pp. 289-297. Springer-Verlag, Berlin

Kanade, T. and Okutomi, M. 1994. A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16** (9), pp. 920-932.

Mallat, S. 1998. *A Wavelet Tour of Signal Processing.* Academic Press.

Marr, D. 1982. *Vision.* New York: W. H. Freeman and Company.

Marr, D. and Poggio, T. 1976. Cooperative Computation of Stereo Disparity. *Science*, **194**, pp. 283-287.

Trucco, E. and Verri, A. 1998. *Introductory Techniques for 3-D Computer Vision.* New Jersey: Prentice Hall.

# 5 Disparity Computation Using Wavelet Phases

## 5.1 Introduction

The previous chapter described a novel wavelet-based stereo matching approach which uses the W-SSD measure to compute disparity maps for stereo images. As described in Chapter 2 section 2.4, disparity maps can also be extracted from local phase differences between two bandpass signals. This method has generated interest mainly because of its potential for fast parallel computation and because of its applicability to theories of stereopsis in the human visual cortex (Fleet *et al*, 1991). This chapter addresses the issues in the phase-based matching approach. It begins with a comparative discussion between global Fourier phases and local Gabor phases. Then the conventional phase-based disparity computation method is described. This is followed by the presentation of the phase-based matching approach using Magarey and Kingsbury's complex wavelet transform (Magarey and Kingsbury, 1998).

## 5.2 Fourier Phases and Global Shift vs Gabor Phases and Local Shift

If there are two signals, one of which is obtained by shifting the other, assuming their Fourier transforms are known, then the shift value can be computed from the difference between the two Fourier phases according to the Fourier shift theorem (Bracewell, 1986), as discussed in Chapter 2. Rewriting equation (2.10):

$$\tau = \frac{\Delta\theta}{2\pi f}, \text{ where } \Delta\theta = \theta_2 - \theta_1 \tag{5.1}$$

where τ is called *global shift*, and θ is Fourier phase. The global shift refers to the displacement of every sample of the signal. The Fourier phases are contained everywhere in the signal, but where in the signal the phases occur are not known. This is unacceptable in stereo matching. As indicated in Chapter 2, a pair of stereo images can be thought of shifted versions of each other, whereas the shift value between the correspondence points, i.e. the disparity, is dependent on the pixel position. Such displacement is known as *local shift*.

The earliest study of local phase representation was the Gabor transform (Gabor, 1946) which uses complex Gabor filters $g(x,\sigma,\omega)$:

$$g(x,\sigma,\omega) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} e^{j\omega x} \tag{5.2}$$

where ω is the spatial frequency of the filter and σ is the Gaussian parameter expressed as a fraction of the wavelength in order to preserve a constant relative bandwidth. The Gabor transform of convolving the left and right epipolar lines, *L(x)* and *R(x)*, with the Gabor filters is therefore:

$$GT_L(x) = L(x) * g(x,\sigma,\omega) = \frac{1}{\sqrt{2\pi}\sigma} \int L(x') e^{-(x-x')^2/2\sigma^2} e^{j\omega_L(x-x')} dx' \tag{5.3}$$

$$GT_R(x) = R(x) * g(x,\sigma,\omega) = \frac{1}{\sqrt{2\pi}\sigma} \int R(x') e^{-(x-x')^2/2\sigma^2} e^{j\omega_R(x-x')} dx' \tag{5.4}$$

As $GT_L(x)$ and $GT_R(x)$ are complex values, they can be written in polar form and be expanded into real and imaginary components:

$$GT_L(x) = |GT_L(x)| e^{j\phi_L(x)} = GT_{L,r}(x) + jGT_{L,i}(x) \tag{5.5}$$

$$GT_R(x) = |GT_R(x)| e^{j\phi_R(x)} = GT_{R,r}(x) + jGT_{R,i}(x) \tag{5.6}$$

The Gabor phases can be calculated by:

$$\phi_L(x) = \tan^{-1}\left[\frac{GT_{L,r}(x)}{GT_{L,i}(x)}\right] \tag{5.7}$$

$$\phi_R(x) = \tan^{-1}\left[\frac{GT_{R,r}(x)}{GT_{R,i}(x)}\right] \tag{5.8}$$

Gabor transform is used to locate the joint local information on position and frequency. It can be plotted as a scalogram. Figure 5.2 shows a scalogram of an original signal $L(x) = \cos(2\pi f_0 \frac{x}{100}), f_0 = 2, 0 \le x \le 200$, shown in Figure 5.1*. The left plot of Figure 5.2 shows the magnitude and the right plot shows the phase value, both with reference to the horizontal pixel position and the vertical frequency.

* The labels of nearly all the figures in this thesis are omitted. However, the meaning of the axes can be easily seen either from the context or from the equation from which they are generated.

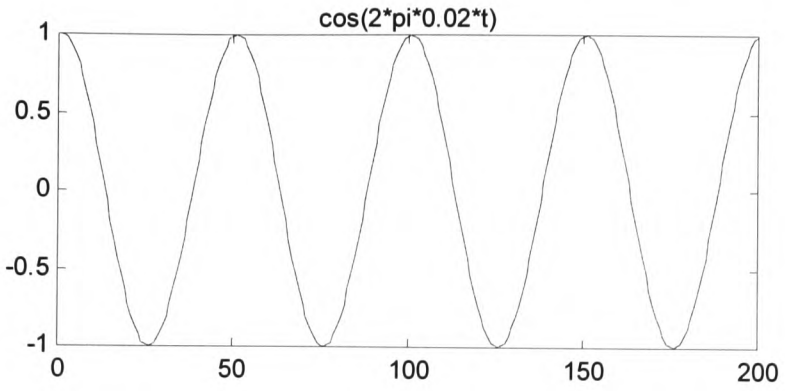



**Figure 5.1 A sinusoid signal**

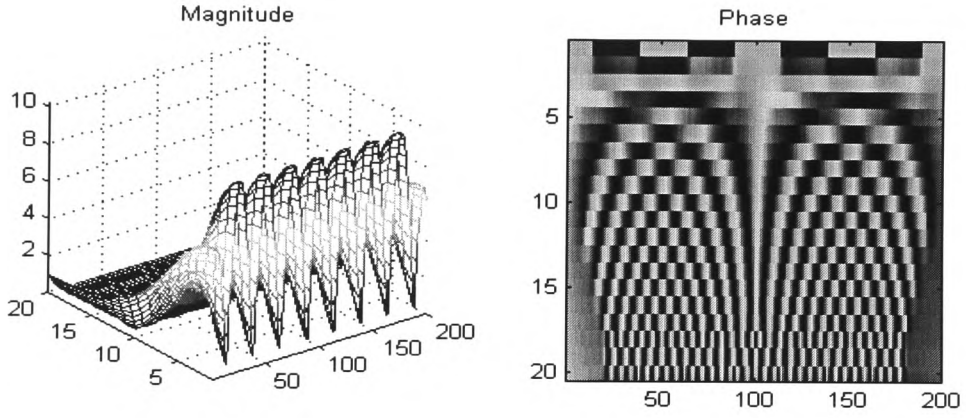



**Figure 5.2 Scalogram of *L(x)* (Left: Magnitude, Right: Phase)**

## 5.3 Conventional Phase-Based Disparity Computation

In contrast to the Fourier transform, the advantage of Gabor transform is the ability to see where in the position the frequencies, i.e. local frequencies, occur in the original signal. This observation is pretty clear when an original signal is non-stationary. This is very useful for stereo matching. In particular, disparity measurement can be directly infered from the local phase difference (Sanger, 1988; Fleet *et al*, 1991; Maimone and Shafer, 1995):

$$d(x) = (\phi_L(x) - \phi_R(x))/\varpi \tag{5.9}$$

where $\varpi$ is the average local spatial frequency. In most cases, the band-pass filter frequencies are the same for left and right images, $\omega_L = \omega_R$ (Fleet, 1993). Sanger (Sanger, 1988) approximated the average of the filter centre pass frequencies from left and right images:

$$\varpi = \frac{1}{2}(\omega_L + \omega_R) \tag{5.10}$$

Another estimate of the average frequency (Jenkin and Jepson, 1994) is made by the average of phase derivatives:

$$\varpi(x) = \frac{1}{2}(\phi_L'(x) + \phi_R'(x)) \tag{5.11}$$

The phase-based matching method makes use of the fact that the disparity from band-pass signals is equivalent to the local phase difference between them. This approach has generated some interest mainly due to its potential for fast parallel computation and its applicability to theories of stereopsis in the human visual cortex (Fleet *et al*, 1991; Jenkin and Jepson, 1994).

As an illustration of this method, two shifted versions of signals are given below:

$$\textit{Original signal: } L(x) = \begin{cases} \cos(2\pi \frac{f_0}{100} x_1) & 0 \le x_1 < 100 \\ \cos(2\pi \frac{f_1}{100} x_2) & 100 \le x_2 < 200 \end{cases}, f_0 = 2, f_1 = 5;$$

*Shifted version*: $R(x) = \begin{cases} \cos(2\pi \frac{f_0}{100}(x_1 - \tau_1)) & 0 \le x_1 < 100 \\ \cos(2\pi \frac{f_1}{100}(x_2 - \tau_2)) & 100 \le t_2 < 200 \end{cases}, \tau_1 = 2, \tau_1 = 6$, two shifts occur at different intervals.

The two signals are plotted in Figure 5.3. Their respective Gabor transforms are shown in Figure 5.4 with respect to magnitude and phase parts.

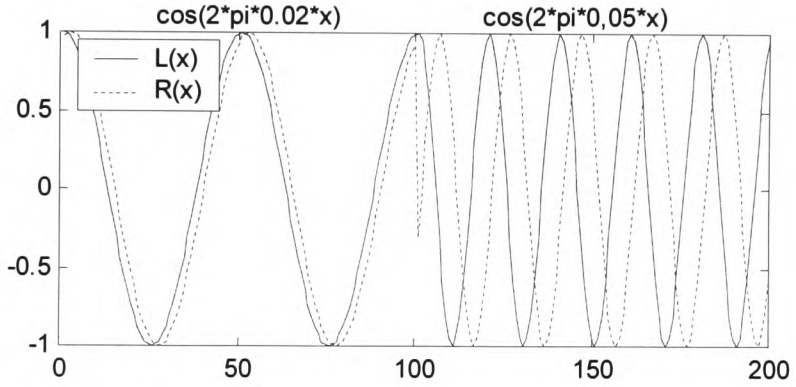



**Figure 5.3 Non-Stationary signals with shift at different intervals**

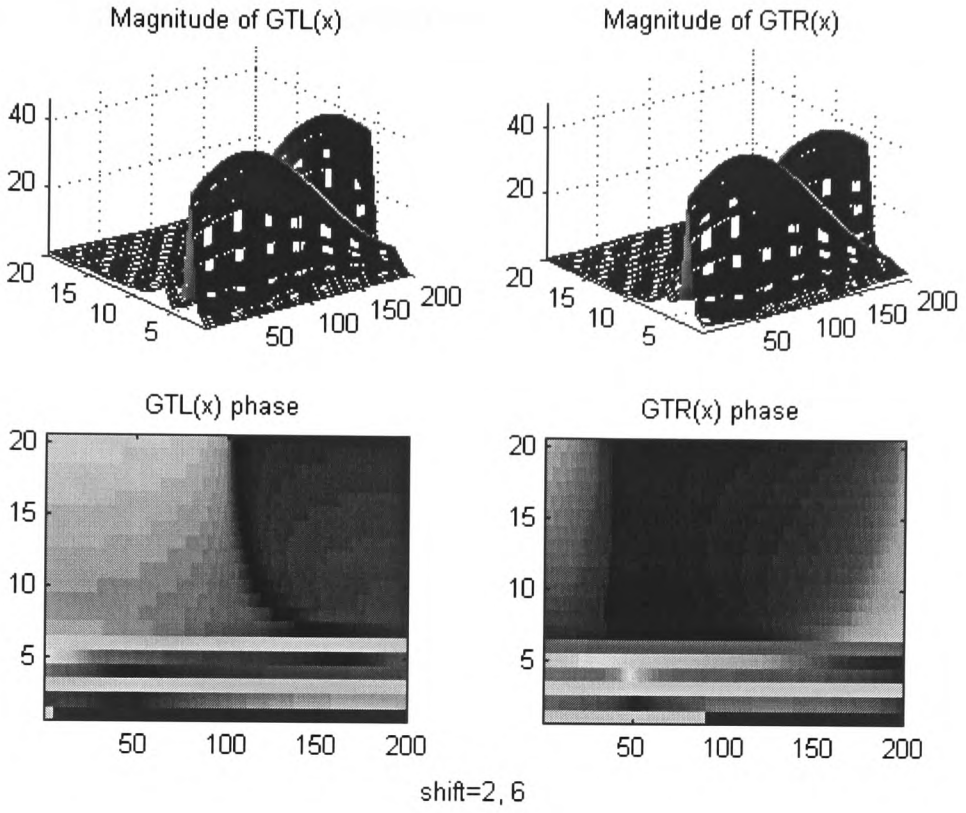



**Figure 5.4 Scalograms of two shifted signals**

Now assume only the two Gabor transforms as shown in Figure 5.4 are known. With a prioi knowledge that the one original signal is obtained by shifting the other, determining the shift value for each position, i.e. the disparity, can be carried out using equation (5.9). The computed disparity in this case is shown in Figure 5.5.

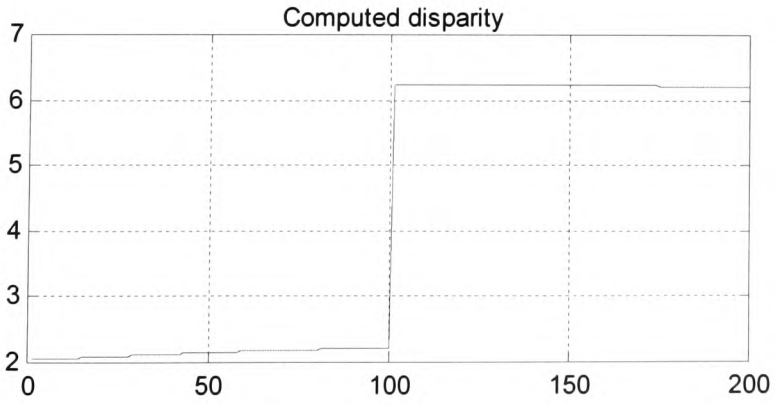



**Figure 5.5 Computed disparity**

Figure 5.5 shows two clearly separated ranges of shift as pre-assigned in the original signal pairs displayed in Figure 5.3. As the disparity directly results from the calculation using equation (5.9), this method also gives a means of generating disparity to subpixel accuracy.

An issue in the phase-based disparity computation is the so-called *Phase Wrapping* (Openheim and Lim, 1981) - the periodicity of the phase causes problems in the recovery of disparity from phase known as "wrapping " effects. If two signals are band-pass filtered at some frequency, then the maximum disparity which can be obtained from the phase difference is $d_{\max} = \pm\pi/\omega$. If the disparity is larger than this half wavelength, then the phase difference will wrap around, and equation (5.9) will become $\Delta\phi(x) = \omega \cdot d(x) - 2\pi f$, where $\Delta\phi(x) = \phi_L(x) - \phi_R(x)$. To avoid phase wrapping, the maximum possible frequency $\omega_{\max}$ and the maximum disparity $d_{\max}$ in the band-passed signals must satisfy (Fleet *et al*, 1991; Maimone and Shafer, 1995):

$$\frac{\pi}{\omega_{\max} + \sigma^{-1}} > |d_{\max}| \tag{5.12}$$

It is difficult to estimate disparity when only a single band-pass filter is used. Sanger (Sanger, 1988), Fleet (Fleet *et al*, 1991), Weng (Weng, 1993) and Maimone (Maimone and Shafer, 1995) used different averaging methods by calculating the disparities from the signals filtered at several spatial frequencies. The selection of the centre frequency of filters, i.e. the effective window of the filters, raises the same windowing problem as correlation-based matching methods.

## 5.4 Complex Wavelet Phases for Stereo Matching

### 5.4.1 Complex Wavelet Transform

With a view to achieving shift invariance and good directional selectivity of wavelet transform, Kingsbury and Magarey (Kingsbury and Magarey, 1996) developed the complex wavelet transform (CWT). Its structure is shown in Figure 5.6, in which each block is a complex low- or high-pass filter (*Lo* or *Hi*) and includes downsampling by two at its output. The CWT filters have complex coefficients and generate complex output samples. Compared with the decimated MRA tree of Figure 3.4 in section 3.3.3, the output sampling rates remain unchanged, but each sample contains a real ($r$) and imaginary ($i$) part.

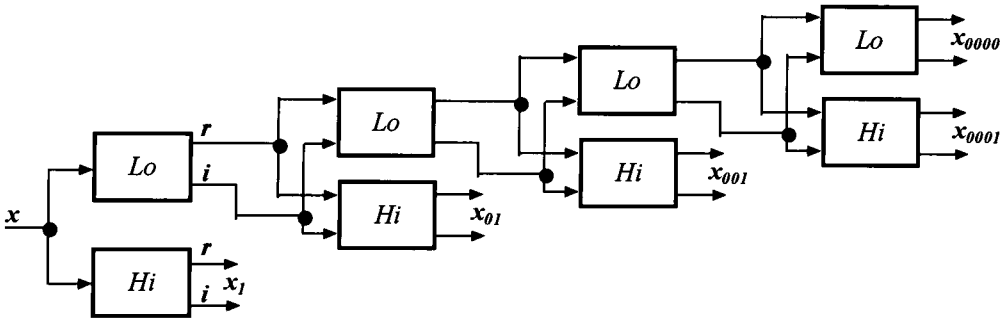



**Figure 5.6 Four levels of complex wavelet scheme for a 1D input signal**

As discussed in section 3.2, the energy of the shift invariant wavelet transform is expected to be invariant to any shifts of the input in time or space (Simoncelli *et al*, 1992). Consider a step function input signal, analysed with the CWT using the complex filters. The energy at each level, shown in Figure 5.7, is proved to be approximately constant by Kingsbury (Kingsbury, 1999; Kingsbury, 2000a).

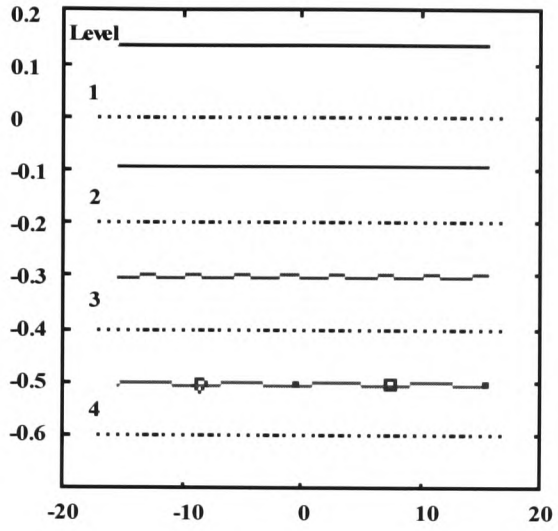



**Figure 5.7 CWT energy at levels 1 to 4**

The complex filters are designed such that the magnitudes of their step responses vary slowly with input shift while the phases vary rapidly. The approximate linearity of the complex phases with input shift (as with Fourier coefficients) is one of the main properties of the CWT that can be used for matching. Disparity estimation based on the measurement of phase shift is possible as discussed in section 2.4.

In the case of one dimension, such filters are designed to have a Gabor function form. The impulse response of *Lo* and *Hi* filters in Figure 5.6 are:

$$\begin{aligned} h_0 &= \frac{1}{10}[1-j, 4-j, 4+j, 1+j] \\ h_1 &= \frac{1}{48}[-3-8j, 15+8j, -15+8j, 3-8j] \end{aligned} \tag{5.13}$$

Extending this to two-dimensions, the parallel column and row path use the same filtering block except that the row filtering is performed using the conjugates of $h_0$ and $h_1$. At each level $m$, the CWT produces six complex bandpass subimages $\{D^{(n,\,m)}, n=1,$

…, 6} and two lowpass subimages $\{A^{(1,m)}, A^{(2,m)}\}$. Figure 5.8 shows the CWT structure for images over two levels.

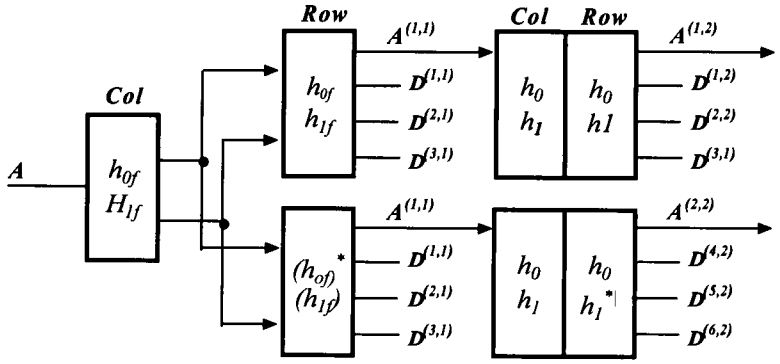



**Figure 5.8 Two-dimensional CWT structure at two levels**

If the original images have a size of $r*c$, then the subimages are of dimensions $r/2^m$ * $c/2^m$. Therefore the overall redundancy of the transform is 4 to 1. Each subband has a corresponding equivalent spatial filter, which has a characteristic spatial frequency $\Omega^{(n,m)}$. This frequency is computable from $h_o$ and $h_1$ (Magarey and Kingsbury, 1995).

As an improvement to the algorithm in terms of perfect reconstruction, a dual tree of wavelet filters was suggested by Kingsbury (Kingsbury, 2000c). The structure of the Dual-Tree Complex Wavelet Transform (DTCWT) is shown in Figure 5.9. Compared with the decimated MRA tree of Figure 3.4 in section 3.3.3, the DTCWT has two parallel, fully decimated trees, tree *a* and *b* in Figure 5.9. The requirement for the filter design is that the delay of filters $H_{ob}$ and $H_{1b}$ should be one sample offset from the delay of $H_{0a}$ and $H_{1a}$. This ensures that the level 1 downsampled values in tree *b* pick up the opposite samples to these in tree *a*. This is equivalent to doubling all the sampling rates in a conventional wavelet tree as shown in Figure 3.4 of section 3.3.3 by eliminating the downsampling by multiplier of two after the level 1filters, $H_{0a}$ and $H_{1a}$.

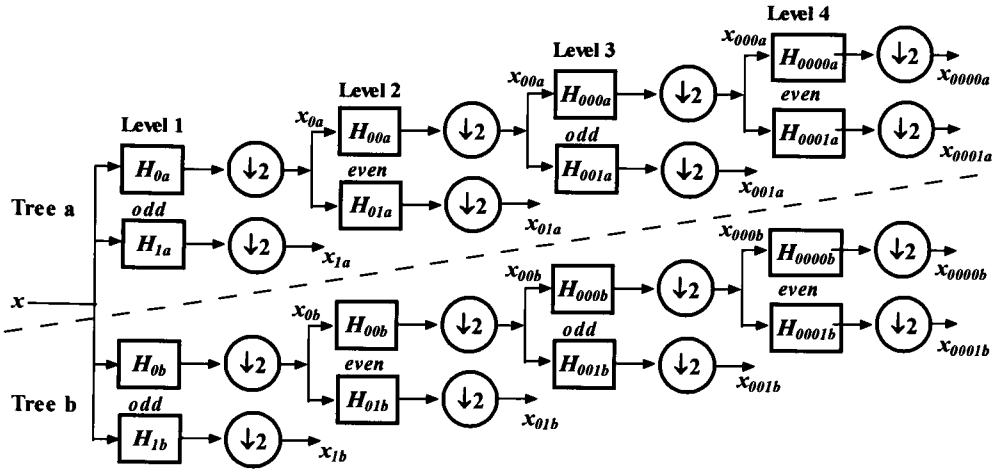



**Figure 5.9 Structure of dual-tree of CWT (DTCWT)**

As an illustration of calculating the DTCWT coefficients, an example is given below using the signal that is used in section 5.2 and shown in Figure 5.1: $L(x) = \cos(2\pi f_0 \frac{x}{100}), f_0 = 2, 0 \le x \le 200$. The first level filters used in this example are plotted in Figure 5.10. The computed DTCWT is displayed in Figure 5.11, which contains three levels of decomposition result. The first two columns show the output from tree a&b, in which the solid lines represent the approximation parts, i.e. the output from the low-pass filters and the dotted lines denote the detail parts, i.e. the output from the high-pass filters. The third and fourth columns show the phase values for approximation and detail part respectively extracted from the complex decomposition components.

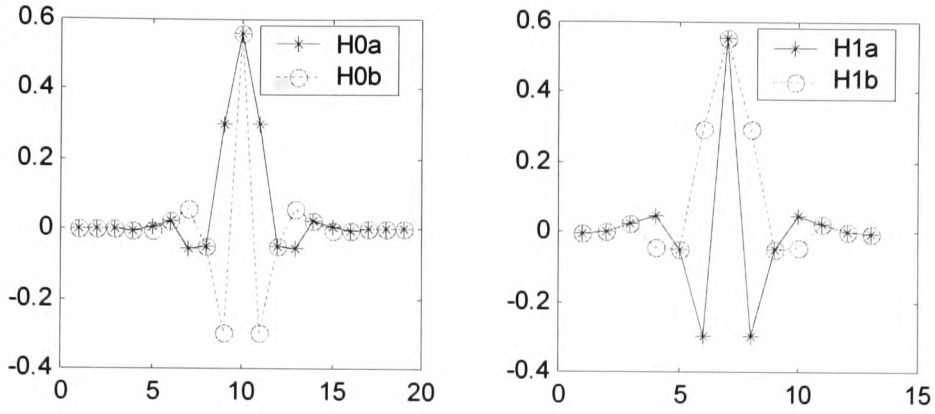



**Figure 5.10 Kingsbury filters at level 1**

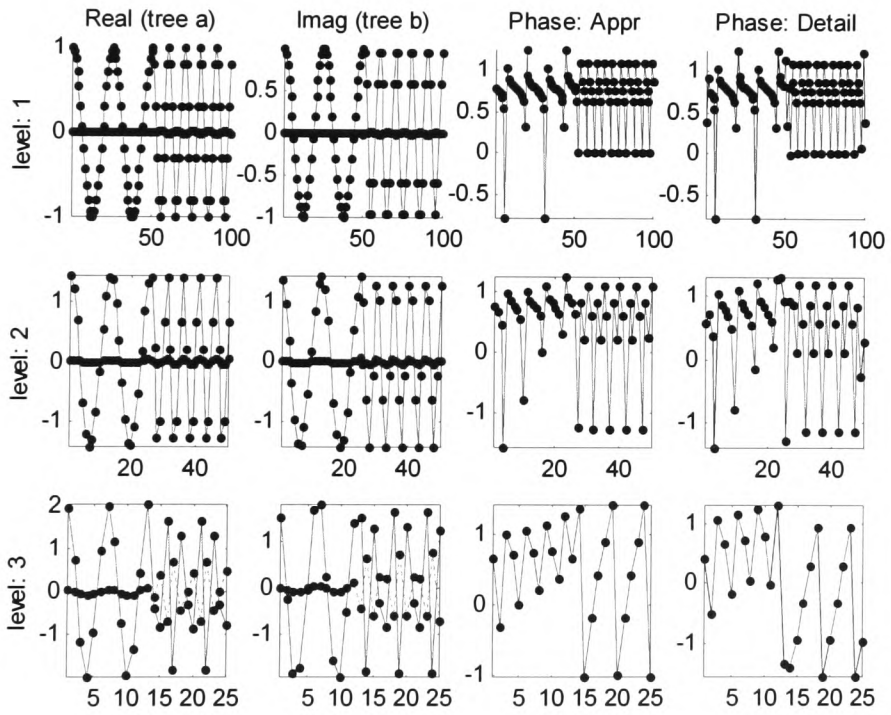



**Figure 5.11 DTCWT coefficients of a sinusoid signal**

### 5.4.2 Complex Phase-Based Matching Using Wavelets

Magarey (Magarey, 1997) presented a detailed method for motion estimation using the complex wavelet transform as shown in Figure 5.8. Since the new dual-tree structure was designed (Kingsbury, 2000b), its performance with stereo matching has not been reported. This section formulates the phased-based disparity computation using DTCWT as shown in Figure 5.9.

In the case of two-dimensional images, apart from the difference of filter design, CWT and DTCWT have got the same output structure at each level, i.e. six bandpass coefficients and two lowband coefficients. Let a pixel of an image at row $r$ and column $c$ be denoted by vector $\mathbf{p}$ that $\mathbf{p} = [r\ c]^T$. The disparity value relative to $\mathbf{p}$ is $\mathbf{d}$, where $\mathbf{d} = [d_1,\ d_2]^T$. Assume $D_l^{(n,m)}(\mathbf{p})$ and $D_r^{(n,m)}(\mathbf{p}+\mathbf{d})$, where $n$=1,...,6 and $m$ is the scale number, represent the band-pass DTCWT coefficients of the corresponding points at each subband $(n,m)$. The following relationship as a model of phase behaviour around an integer-indexed DTCWT subband coefficient can be derived from the DTCWT structure (Magarey and Kingsbury, 1998):

$$D_r^{(n,m)}(\mathbf{p}+\mathbf{d}) \approx D_l^{(n,m)}(\mathbf{p})\, e^{j\theta(\mathbf{d})} \tag{5.14}$$

$$\text{where } \theta(\mathbf{d}) = 2^m \Omega^{(n,m)} \mathbf{d} \tag{5.15}$$

and the limit for each disparity $\mathbf{d}$ is set to $\mathbf{d} \in [-0.5, 0.5]*[-0.5, 0.5]$

The idea to estimate the disparity is by minimising the *subband squared difference*, $SD^{(m)}(\mathbf{p},\mathbf{d})$, which is defined as a summation over the six oriented subbands, $SD^{(n,m)}(\mathbf{p},\mathbf{d})$, as follows:

$$SD^{(m)}(\mathbf{p},\mathbf{d}) = \sum_{n=1}^{6} SD^{(n,m)}(\mathbf{p},\mathbf{d}) \tag{5.16}$$

$$\text{where } SD^{(n,m)}(\mathbf{p},\mathbf{d}) = \left| D_l^{(n,m)}(\mathbf{p}) - D_r^{(n,m)}(\mathbf{p}+\mathbf{d}) \right|^2 \tag{5.17}$$

Expanding equation (5.16) with (5.17) and (5.14), $SD^{(m)}(\mathbf{p},\mathbf{d})$ can be approximated as a quadratic surface (Magarey, 1997):

$$SD^{(m)}(\mathbf{p},\mathbf{d}) = Ad_1^2 + Bd_2^2 + Cd_1d_2 + Dd_1 + Ed_2 + G \tag{5.18}$$

where the coefficients $\{A, B, C, D, E, F, G\}$ are:

$$\begin{aligned}
A &= \sum_{n=1}^{6} \left| D_l^{(n,m)}(\mathbf{p}) D_r^{(n,m)}(\mathbf{p}) \right| (\Omega_1^{(n,m)})^2 \\
B &= \sum_{n=1}^{6} \left| D_l^{(n,m)}(\mathbf{p}) D_r^{(n,m)}(\mathbf{p}) \right| (\Omega_2^{(n,m)})^2 \\
C &= \sum_{n=1}^{6} \left| D_l^{(n,m)}(\mathbf{p}) D_r^{(n,m)}(\mathbf{p}) \right| 2\Omega_1^{(n,m)}\Omega_2^{(n,m)} \\
D &= \sum_{n=1}^{6} \left| D_l^{(n,m)}(\mathbf{p}) D_r^{(n,m)}(\mathbf{p}) \right| (-2)\Omega_1^{(n,m)}\theta^{(n,m)}(\mathbf{p}) \\
E &= \sum_{n=1}^{6} \left| D_l^{(n,m)}(\mathbf{p}) D_r^{(n,m)}(\mathbf{p}) \right| (-2)\Omega_2^{(n,m)}\theta^{(n,m)}(\mathbf{p}) \\
G &= \sum_{n=1}^{6} \left( \left| D_l^{(n,m)}(\mathbf{p}) \right| - \left| D_r^{(n,m)}(\mathbf{p}) \right| \right)^2 + \sum_{n=1}^{6} \left( \left| D_l^{(n,m)}(\mathbf{p}) D_r^{(n,m)}(\mathbf{p}) \right| \right) \left( \theta^{(n,m)}(\mathbf{p}) \right)^2
\end{aligned} \tag{5.19}$$

Equation (5.18) can be further transformed to the following $\{\mathbf{d}, \alpha, \beta\}$ form:

$$SD^{(m)}(\mathbf{p},\mathbf{d}) \approx \alpha(d_1 - d_{10})^2 + \beta(d_2 - d_{20})^2 + \gamma(d_1 - d_{10})(d_2 - d_{20}) + \delta \qquad (5.20)$$

The surface minimum coordinates $\mathbf{d_0} = [d_{10}\ d_{20}]$ in (5.20) are extracted as:

$$\mathbf{d}_0 = \begin{bmatrix} \frac{2BD - CE}{C^2 - 4AB} & \frac{2AE - CD}{C^2 - 4AB} \end{bmatrix} \qquad (5.21)$$

This value $\mathbf{d_0}$ therefore is used to determine the disparity estimate at pixel $\mathbf{p}$ of level $m$.

In equation (5.20), the curvature parameters $\alpha$, $\beta$, $\gamma$ = $A$, $B$, $C$, and the surface minimum $\delta$ is:

$$\delta = G - Ad_{10}^2 - Bd_{20}^2 - Cd_{10}d_{20} \qquad (5.22)$$

which indicates the closeness of the match between $D_l^{(n,m)}(\mathbf{p})$ and $D_r^{(n,m)}(\mathbf{p}+\mathbf{d})$.

Equation (5.21) is the disparity estimation at one level. Next the disparity field needs to be refined using the information at the next finer level *m-1*. As the subimage at level *m-1* is twice the size of that at level *m*, interpolation is needed for the surface $SD^{(m)}$ to have the same size as $SD^{(m-1)}$.

Fleet and Jepson (Fleet, 1992) have developed a method for interpolating the downsampled outputs of a complex band-pass filter by modulating a low-pass interpolating kernel to the centre frequency of the band-pass filter and convolving with the kernel:

$$D^{(n,m)}(\mathbf{p}+\mathbf{d}) = \sum_{\mathbf{k}} H_{\mathbf{d}}^{(n,m)}(\mathbf{k}) D^{(n,m)}(\mathbf{p}-\mathbf{k}) \qquad (5.23)$$

where $H_{\mathbf{d}}^{(n,m)}(\mathbf{k}) = H_{\mathbf{d}}(\mathbf{k}) e^{j2^m \Omega^{(n,m)} \cdot (\mathbf{k}+\mathbf{d})}$, $H_{\mathbf{d}}(\mathbf{k})$ is the 2D low-pass interpolating kernel e.g. sinc kernel.

Substitute the interpolated band-pass output to (5.16) to form an update surface $SD'^{(m)}$ and solve (5.17) and (5.20) for an update set of parameters $\{\mathbf{d}'_0, \alpha', \beta', \gamma', \delta'\}$. The matching strategy is to allow $\mathbf{d}'_0$ acting as a starting point for the disparity estimation at level *m-1*. Then combine the surface $SD^{(m-1)}$ with $SD^{(m)}$ to form the cumulative squared difference surface $CSD^{(m-1)}$:

$$CSD^{(m-1)}(\mathbf{p}, \mathbf{d}) = SD'^{(m)}(\mathbf{p}, \mathbf{d}) + SD^{(m-1)}(\mathbf{p}, \mathbf{d}) \tag{5.24}$$

Like its components, $CSD^{(m-1)}$ is quadratic as equation (5.20) and its solution parameters provide the disparity estimate at level *m-1*. This procedure can continue to refine the estimate by incorporating finer levels until the finest level is reached (Magarey, 1997).

Motion estimation using initial complex wavelet transform as shown in Figure 5.8 is discussed in detail in (Magarey and Kingsbury, 1998). However, its performance with stereo matching and applying the new dual-tree filter design structure have not been found in the literature. The application of DTCWT to stereo matching will be carried out in Chapter 6 and the results will be compared with the method using W-SSD.

## 5.5 Summary

This chapter has reviewed a different approach to stereo matching using local phases. A method for computing local Gabor phases in a band-passed signal was examined. To avoid the difficulty in combining the disparities from the filtered signal using various centre frequencies, an approach applying the complex wavelet transform has been presented. A dual-tree structure of the filter design and its shift invariance property has been discussed. The complex filters are designed such that the magnitudes of their step responses vary slowly with input shift while the phases vary rapidly. The chapter has

shown that the complex wavelet phases are approximately linear with the input shift. Based on this observation, the relationship between left and right integer-indexed DTCWT subband coefficients is derived. Disparity is thus estimated by minimising the subband squared difference.

The main contribution of this chapter is the formulation of the phase-based matching using DTCWT coefficients. This phase-based matching method will be tested in the next chapter, Chapter 6 and mainly used for comparing with the W-SSD based matching using DyWT as presented in Chapter 4.

## 5.6 References

Bracewell, R. N. 1986. *The Fourier Transform and its Applications.* McGraw-Hill Book Company.

Fleet, D. J. 1992. *Measurement of Image Velocity.* Kluwer Academic Publishers.

Fleet, D. J. 1993. Stability of Phase Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15** (12), pp. 1253-1268.

Fleet, D. J., Jepson, A. D. and Jenkin, M. R. M. 1991. Phase-Based Disparity Measurement. *Computer Vision Graphics and Image Processing: Image Understanding*, **53** (2), pp. 198-210.

Gabor, D. 1946. Theory of Communication. *Journal of IEE*, **93**, pp. 429-459.

Jenkin, M. R. M. and Jepson, A. D. 1994. Recovering Local Surface Structure through Local Phase Difference Methods. *CVGIP*, **59**, pp. 72-93.

Kingsbury, N. G. 1999. *Shift Invariant Properties of the Dual-Tree Complex Wavelet*

*Transform* . IEEE International Conference on Acoustics, Speech, and Signal Processing. pp. 1221-1224, Phoenix, AZ, USA.

Kingsbury, N. G. 2000a . *Complex Wavelets and Shift Invariance.* Proc IEE Colloquium on Time-Scale and Time-Frequency Analysis and Applications, IEE. London.

Kingsbury, N. G. 2000b . Complex Wavelets for Shift Invariant Analysis and Filtering of Signals. *Journal of Applied Computation and Harmonic Analysis*,

Kingsbury, N. G. 2000c . *A Dual-Tree Complex Wavelet Transform with Improved Orthogonality and Symmetry Properties.* IEEE International Conference on Image Processing. pp. 375-378, Vancouver, Canada.

Kingsbury, N. G. and Magarey, J. 1996. *Wavelets in Image Analysis: Motion and Displacement Estimation.* Proc. Irish DSP and Control Conference. pp. 199-217, Dublin.

Magarey, J. 1997. *Motion Estimation Using Complex Wavelets.* PhD thesis. Department of Engineering, Cambridge University.

Magarey, J. and Kingsbury, N. 1998. Motion Estimation Using a Complex-Valued Wavelet Transform. *IEEE Transactions on Signal Processing*, **46** (4), pp. 1069-1084.

Maimone, M. W. and Shafer, S. A. 1995. *Modeling Foreshortening in Stereo Vision Using Local Spatial Frequency.* Report CMU-CS-95-104, Carnegie Mellon University.

Openheim, A. V. and Lim, J. S. 1981. The Importance of Phase in Signals. *Proceedings of the IEEE*, **69**, pp. 529-541.

Sanger, T. D. 1988. Stereo Disparity Computations Using Gabor Filter. *Biol. Cybern.*,

**59**, pp. 405-418.

Simoncelli, E. P., Freeman, W. T., Adelson, E. H. and Heeger, D. J. 1992. Shiftable Multiscale Transforms. *IEEE Trans. Information Theory*, **38** (2), pp. 587-607.

Weng, J. 1993. Image Matching Using the Windowed Fourier Phase. *International Journal of Computer Vision*, **3**, pp. 211-236.

# 6 Testing and Evaluation of Proposed Approaches

## 6.1 Introduction

The previous two chapters, Chapter 4&5, have described two different wavelet-based matching methods. One of them is based on correlation using the dyadic wavelet transform, whereas the other makes use of local phases by complex wavelet transform. Implementation of these two methods will be described in this chapter. Various experiments will be made with a range of stereo images: random dot stereograms, artificial images with ground truth data and real images naturally taken in the laboratory. As the development of the W-SSD matching method is the main contribution of this thesis, the implementation procedure will be detailed in this chapter. Experimental results with both methods will be presented. A comparison between the DYWT and the Complex WT as well as with a conventional pyramid matching and a standard SSD method will be made.

## 6.2 Test Data

Three different categories of images are used in this chapter to test the performance of the DyWT-based matching approach and to compare it with other matching methods.

### 6.2.1 Random Dot Stereogram

The matching approach was applied initially to the so-called *Random Dot Stereograms* (RDS), which Julesz (Julesz, 1971) constructed to demonstrate that the human visual system can recover the three-dimensional depth of a scene from the two-dimensional retinal fields without using any other visual cues. The small positional differences, i.e. disparities, of the images of a scene on the two spatially separated retinas are used to precisely locate the depths of objects. This can be easily modelled by constructing two images, called stereograms, which are composed of randomly placed dots in the background in a different colour and are identical except for two central squares with two different pixel displacement between them. By viewing the stereograms with crossed eyes or through a stereoscope, a vivid sensation of depth arises, which is the result of binocular disparities alone.

Due to the known structure, testing of a matching method with stereograms can be quantitatively evaluated. In this thesis three types of Random Dot Stereograms are generated depending on the grey values of the dots.

Figure 6.1(a) shows a simple case, where the background and two squares have only one single grey value, respectively. The displacements between the two central squares are four and eight pixels. The scan lines at row sixty from the left and right images are plotted in Figure 6.1(b). For simplicity, the axis labels are omitted in this and following scan line plots, which are the grey values against the pixel positions.

Figure 6.2(a) and Figure 6.3(a) show two similar stereograms, each of which has randomly distributed dots and two central squares in the right images have four and eight pixels displacement. The distinction is the grey values of the black dots. The pixel values in Figure 6.2(a) are either 0 or 1, i.e. they are a pair of binary images, whereas in

Figure 6.3(c) the random dots have a range of grey values from 0 to 1. Both of the cases can be demonstrated by one of their scan lines at row sixty shown in Figure 6.2(b) and Figure 6.3(b).

Note that all of the above three pairs of RDSs are shown as the size of 128*128. As they have the same shifted pixels in the central squares, they possess the same disparity map, the 2D and 3D formats of which are plotted in Figure 6.4.

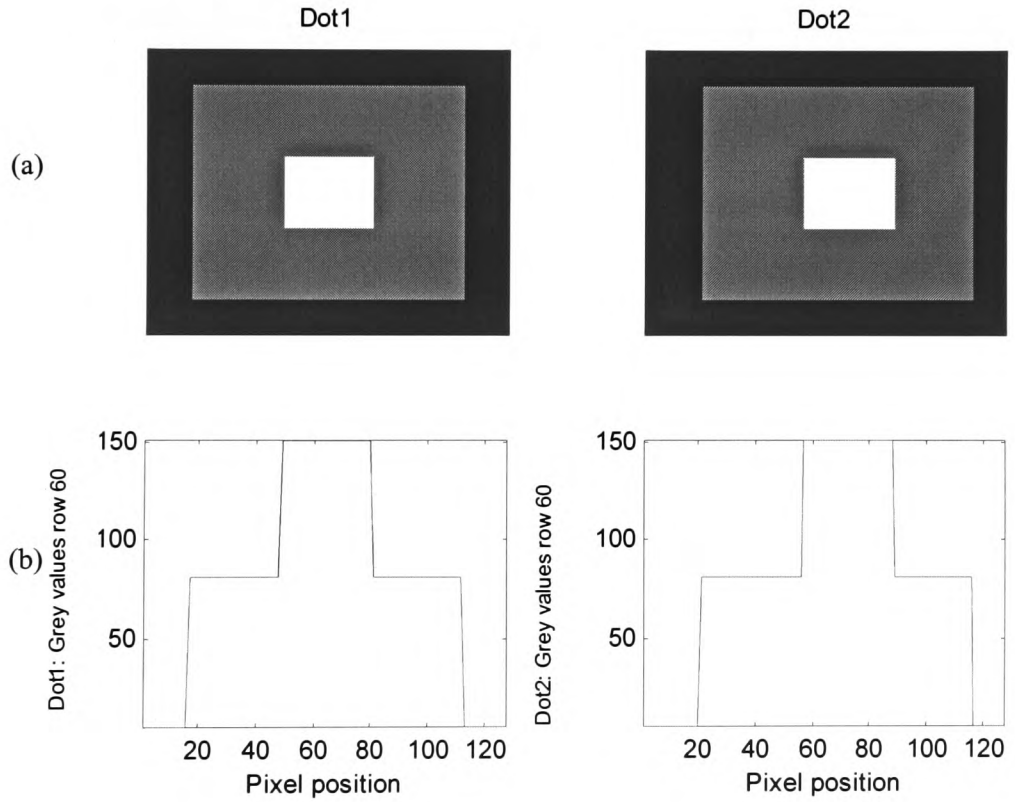
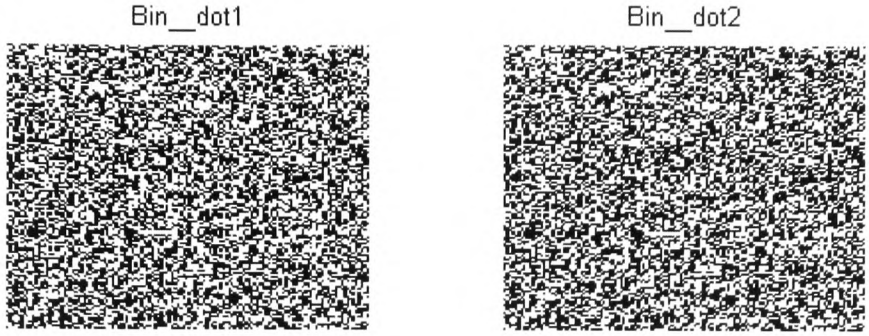



**Figure 6.1 Stereo pair 1: *Dots***

(a)




(b)

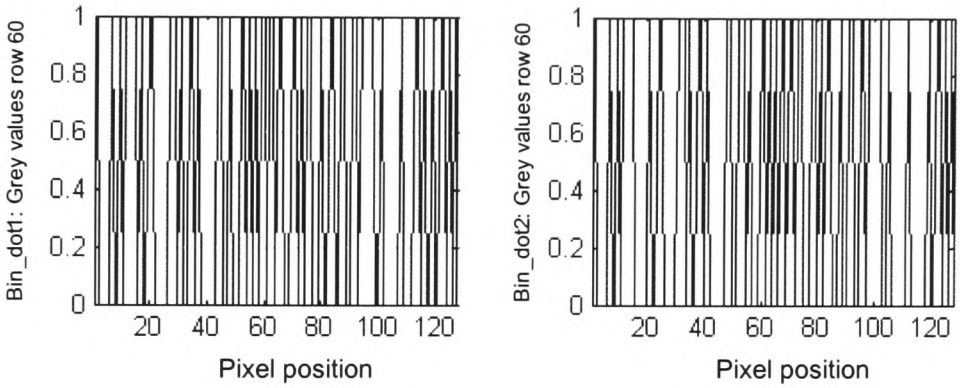



**Figure 6.2 Stereo pair 2: *Bin_ dots****

---

* The steps in the plots of the two scan lines are due to the printer resolution. The grey values are actually either 0 or 1.

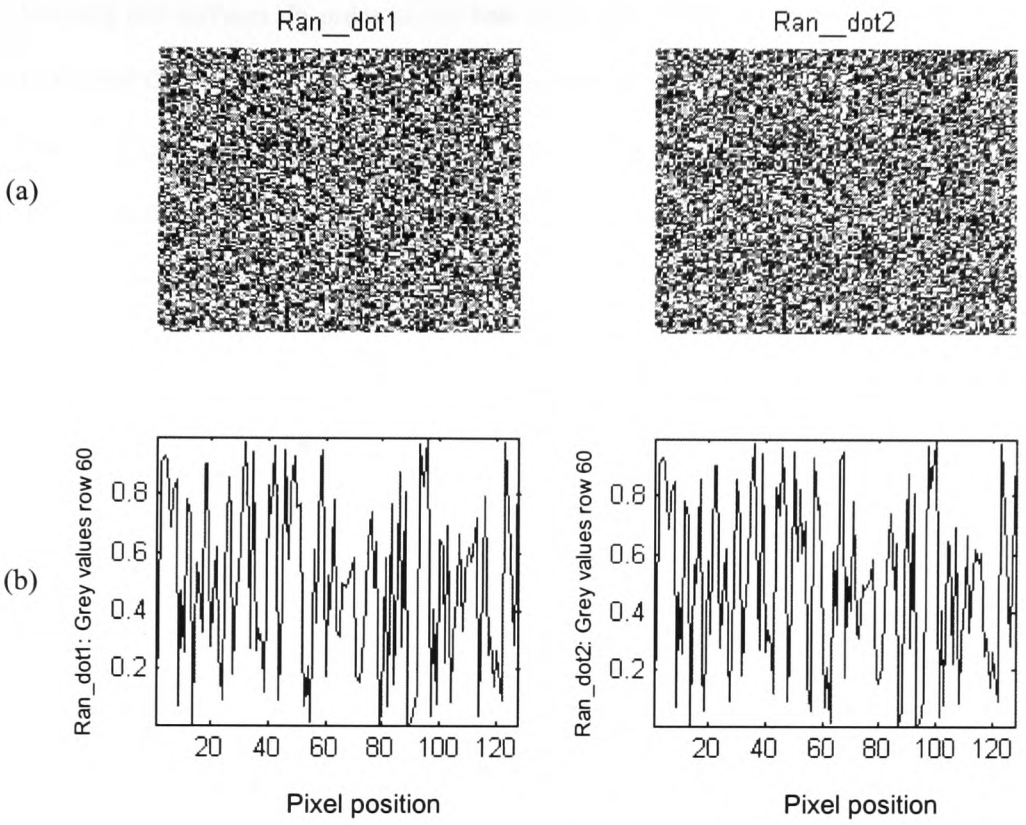



**Figure 6.3 Stereo pair 3: *Ran_ dots***

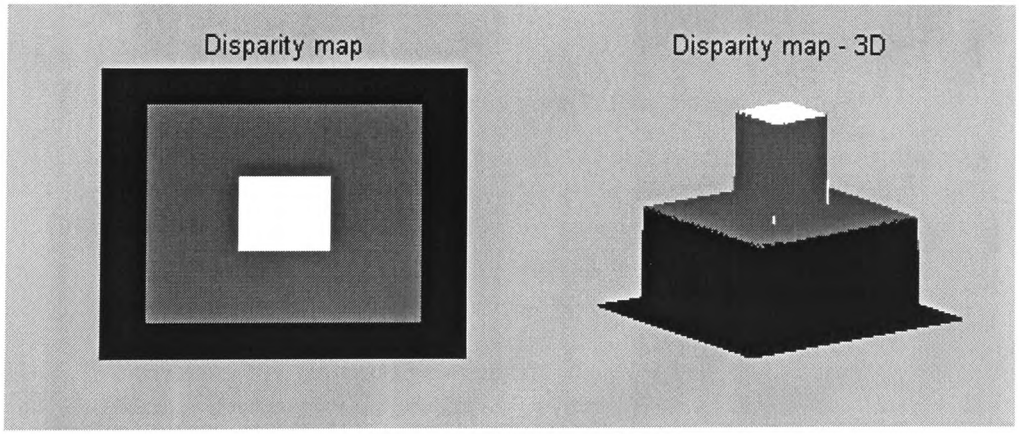



**Figure 6.4 RDS: true disparity map**

Although the structure of the above three pairs of RDS is different, they have the same disparity map which consists of two parallel squares. This means each square's corresponding points in the 3D world have the same distance to the camera plane, hence

forming flat surfaces. In order to test how the proposed DyWT method would respond to slanted flat objects and rounded objects, two additional pairs of RDS are constructed. Figure 6.5 gives the stereo RDS, *Ran_ramp*, representing a smooth ramp in the image as shown in Figure 6.6. The maximum disparity of the ramp is 16 pixels. Figure 6.7 shows a pair of RDS, *Ran_ball*, whose underlining structure is a hemisphere. Its true disparity map is shown in Figure 6.8. The maximum disparity value is 32 pixels, which is the radius of the ball. All the images generated in Figure 6.5 and Figure 6.7 are of size 128*128 pixels with random grey values from 0 to 1.

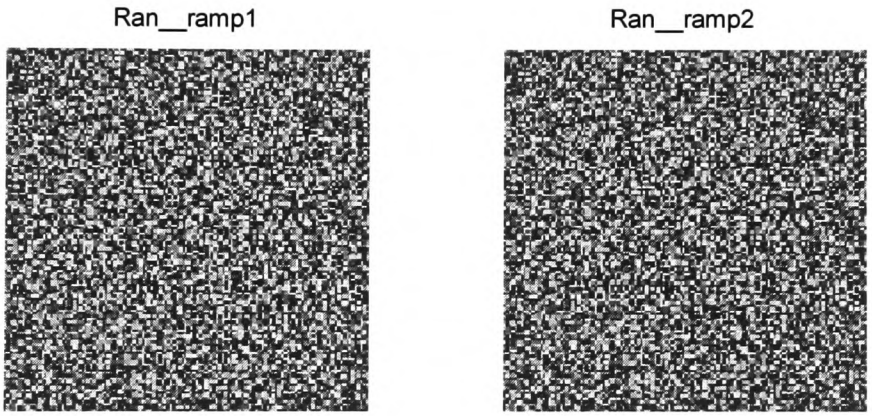



**Figure 6.5 Stereo pair 5: *Ran_ramp***

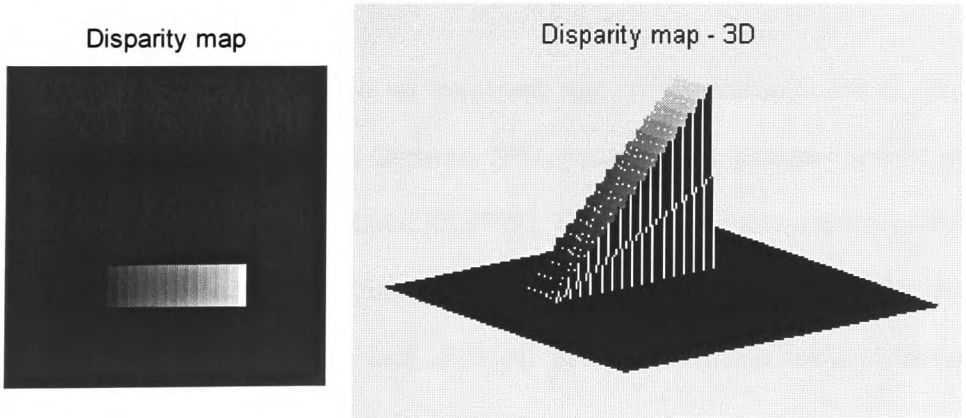



**Figure 6.6 Disparity map: *Ran_ramp***

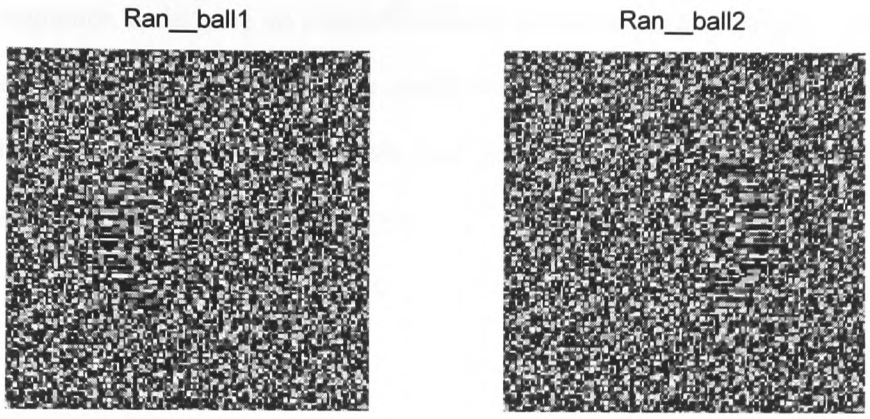



**Figure 6.7 Stereo pair 4: *Ran_ball***

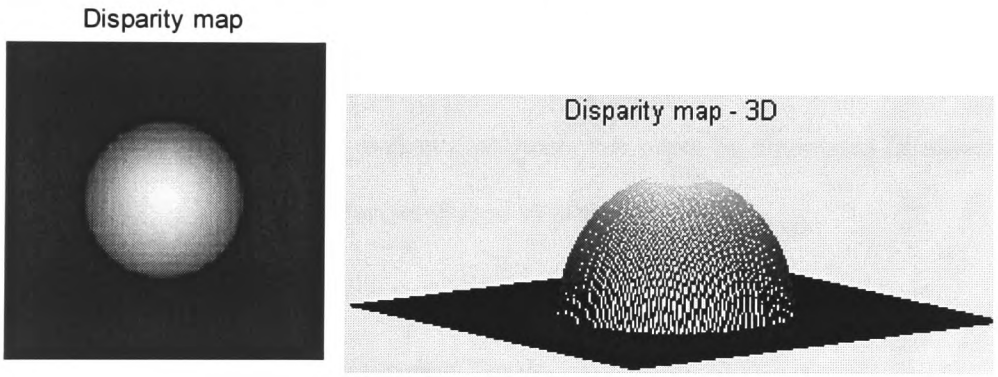



**Figure 6.8 Disparity map: *Ran_ball***

The RDS has historically played an important role in the investigation of stereo matching (Marr and Poggio, 1976; Grimson, 1981; Sanger, 1988; Kanade and Okutomi, 1994; VALENTINOTTI and TARAGLIO, 1999). This is because it represents a simple and controlled method of the formation of stereo image pairs. However, these computer-generated RDS would look different in the real stereo image formation process, ignoring in particular distortions due to the perspective projection.

An assumption underlying all correlation-based stereo matching methods is that the perspective distortions are relatively small over the window used to carry out the correlation, except at finite many depth discontinuities (Dhond and Aggarwal, 1989). The ability to tolerate local distortions due to perspective is a measure of the robustness of a matching algorithm. In the limit of extreme perspective distortion or occlusion even robust but sparse matching techniques like feature matching will fail (Kanade and Okutomi, 1994). This is at the root of the ill-posed nature of stereo matching (Bertero *et al*, 1988), and the errors due to perspective distortion can be viewed as a form of noise in the measurement process. Nearly all new matching techniques represent an attempt to ameliorate this problem and thus aims to minimise the inevitable post matching filtering step, but the need for this filtering or the application of some other equivalent constraints can never be eliminated (Rothwell Hughes, 1999). This will be further discussed in section 6.9.

### 6.2.2 Artificial Images with Ground Truth

This thesis adopts the synthesised stereo images of real scenes, such as the face, the lamp and the table constructed by the University of Tsukuba. Figure 6.9 shows the image pair, each with size 284*386 and the ground truth disparity map is shown in Figure 6.10. This imagery was obtained from the Microsoft website (Szeliski, 2000).


**Figure 6.9 Stereo pair 4: *Tsukuba* images**

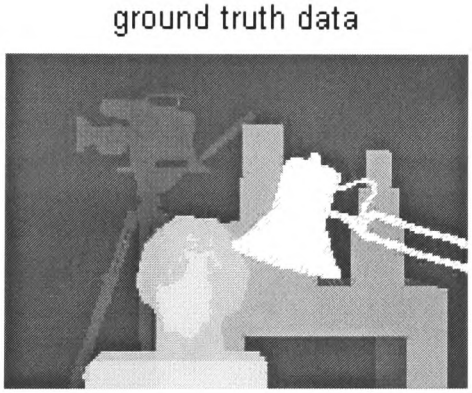



**Figure 6.10 Disparity map of *Tsukuba* images**

### 6.2.3 Real Images

Two pairs of stereo images obtained in the laboratory (Rothwell Hughes, 1999) were taken in a natural scene and used for evaluation, as shown in Figure 6.11, *Boxes* and Figure 6.12, *Toys*. The images were taken using a camera at two close positions with only a slight horizontal displacement. Although there are no ground truth data for the real images, they are included here to get an intuitive view of stereo disparity and for comparative purposes later in this chapter and the following chapter. These real images

contain not only perspective distortions and occlusions, but also changes due to directional and non-constant illumination, which was included sunlight.

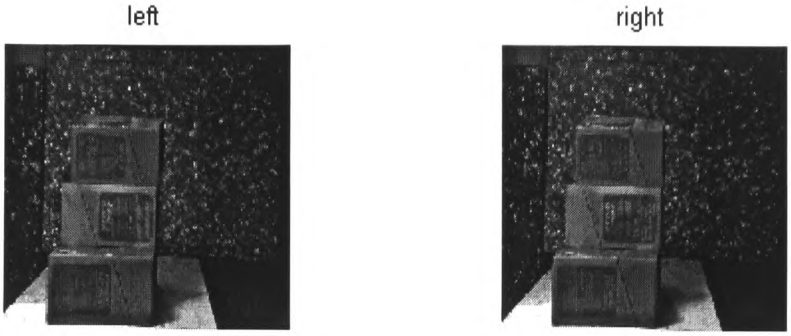



**Figure 6.11 Stereo pair 5: *Boxes***


**Figure 6.12 Stereo pair 6: *Toys***

### 6.2.4 Use of the test images

Five pairs of random dot stereograms with the same size (128*128 pixels) were generated, three of which, *Dots*, *Bin_dots* and *Ran_dots*, use the same square shaped disparity map output but a different intensity structure, and two of which, *Ran_ramp* and *Ran_ball* have a slope and round shaped disparity map. As the structure is exactly known, the new proposed DyWT method will be tested with all of them. In addition,

one pair of synthesised images of known ground truth data, the *Tsukuba*, and two pairs of real images with no ground truth information, *Boxes* and *Toys*, are also presented. All of these image pairs will be used to test the new DyWT matching method. For comparative purpose, *Ran-dots*, *Tsukuba* and *Boxes* are used to evaluate the performance of the DyWT method by comparison with other matching methods. All the images used in the thesis are summarised in Table 6.1, which includes a reference to the figure numbers depicting the images, a brief description of their characteristics and the truth data if applicable, and a list of the algorithms with which they will be used in the later sections.

As a demonstration of the approach proposed in section 4.3, sections 6.4.1 to 6.4.3 elaborate the process illustrated in Figure 4.6 with one of the above test pairs, *Bin_dots*. *Bin_dots* is chosen because its binary structure is not as simple as *Dots* and not as complicated as *Ran_dots*. Section 6.3 defines some parameters used to evaluate the computational results. Results using other test data are given in section 6.4.4.

**Table 6.1 A summary of images used**

| Image name | Shown in | Brief description | Truth data | Algorithm tested | Tested in sections |
|---|---|---|---|---|---|
| *Dots* | Figure 6.1 | Simple RDS with three grey levels | Layered squares with disparity values being 4 and 8 | DyWT | 6.4.4 |
| *Bin_dots* | Figure 6.2 | Binary RDS | | Explicit description of implementation by DyWT | 6.4.1 to 6.4.3 |
| *Ran_dots* | Figure 6.3 | 0 to 1 grey-scale RDS | | DyWT | 6.4.4 |
| | | | | DTCWT | 6.5 |
| | | | | Pyramid | 6.6 |
| | | | | Standard SSD | 6.7 |
| *Ran_ramp* | Figure 6.5 | 0 to 1 grey-scale RDS | A underlining ramp | DyWT | 6.4.4 |
| *Ran_ball* | Figure 6.7 | 0 to 1 grey-scale RDS | A floating half ball | DyWT | 6.4.4 |
| *Tsukuba* | Figure 6.9 | Synthesised images | Available | DyWT | 6.4.4 |
| | | | | DTCWT | 6.5 |
| | | | | Pyramid | 6.6 |
| | | | | Standard SSD | 6.7 |
| *Boxes* | Figure 6.11 | Real images of layered boxes with simple geometry | Not available | DyWT | 6.4.4 |
| | | | | DTCWT | 6.5 |
| | | | | Pyramid | 6.6 |
| | | | | Standard SSD | 6.7 |
| *Toys* | Figure 6.12 | Real images of child's toys with more varied shapes | Not available | DyWT | 6.4.4 |

## 6.3 Performance Measurement

In terms of the complexity and the accuracy evaluation of the approach, three measures are used for each computed result.

1. Computational complexity

Normally big O-notation (Borodin and Munro, 1975) is used as a means for describing an algorithm's performance. In terms of the different matching approaches use in the thesis, this will be discussed in the comparative section 6.8. In addition, the program running time is also used, which can be directly obtained in MATLAB while a program is running.

2. Matched rate

It is calculated as a ratio of correctly matched pixels to all the pixels of the image wherever the ground truth data are applicable. The matched pixels refer to those whose computed disparity values are the same as in the ground truth image at the same positions.

3. Standard deviation of the difference image

In order for the accuracy estimation, a measure, denoted by *std_d*, is defined as the standard deviation of the difference image between the ground truth data and the computed disparity map:

$$std\_d = std\ (std\ (ground\ truth\ image - computed\ disparity\ map)) \tag{6.1}$$

## 6.4 Implementation 1: by DyWT

### 6.4.1 Coarsest Level Estimation

The size of each of the "Bin_dots" is 128*128. As there are only horizontal displacement values in the stereo images, matching is performed along the corresponding rows. This section illustrates the coarsest level disparity estimation

taking the epipolar lines, *epL* and *epR*, respectively from row sixty of the "Bin_dots" pair as an example.

#### *6.4.1.1 DyWT Decomposition*

A plot of the lines is shown in Figure 6.13. According to the theory of section 4.3, the maximum scale level of the DyWT decomposition is $2^7$. Figure 6.14 presents the decomposition result at seven dyadic scales, which are the input data for the W-SSD computation using equation (4.17). Note that the solid lines from now on are associated with the left line *epL* or the result based on it and the dashed lines are associated with the right line *epR* or the result based on it unless otherwise stated.

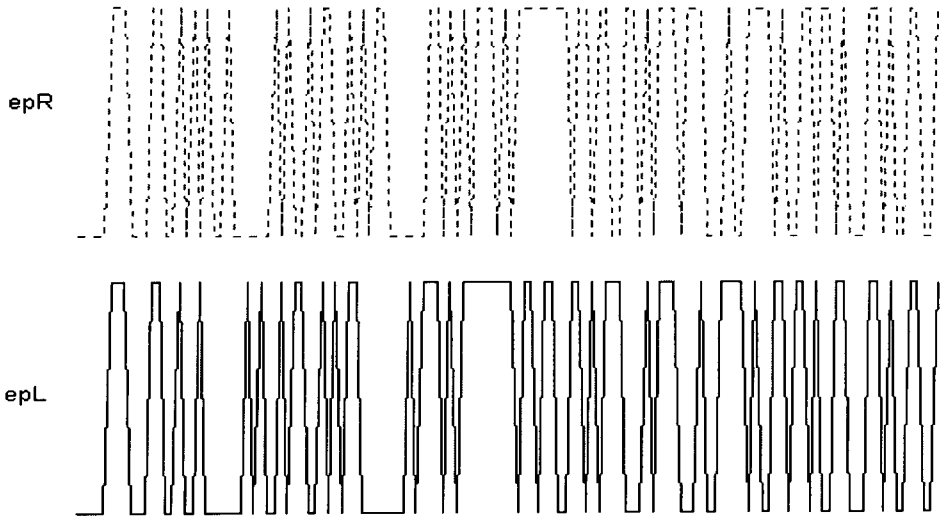



**Figure 6.13 Epipolar lines in one plot**

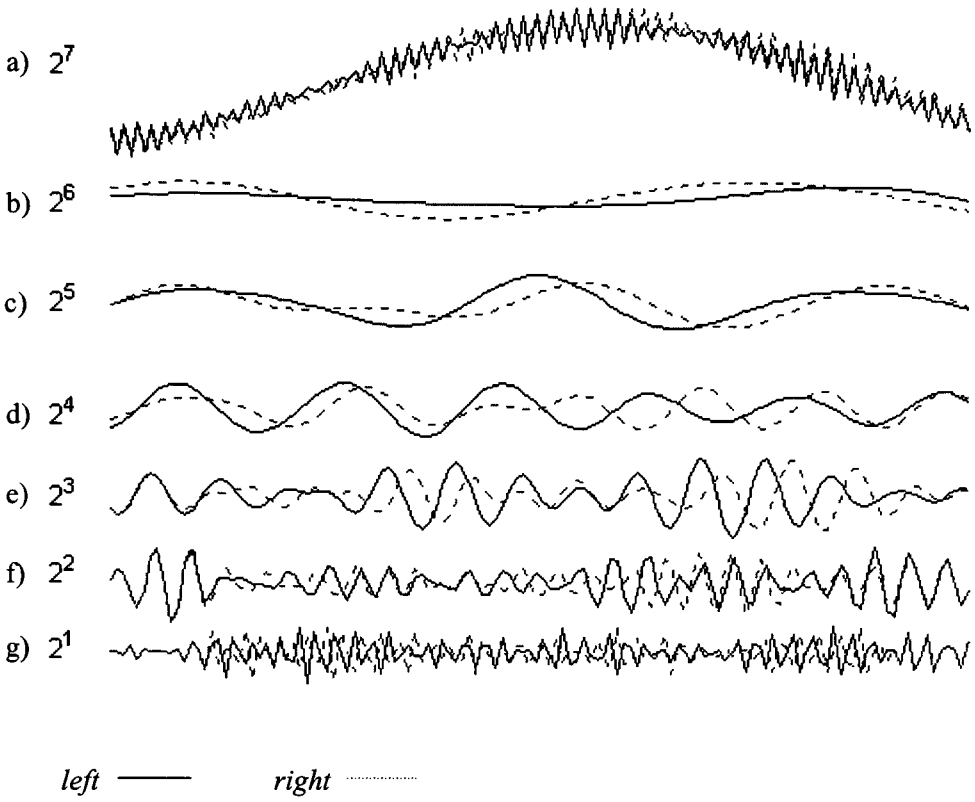



**Figure 6.14 DyWT of epipolar lines**

*6.4.1.2 Choosing the Coarsest Level*

It can be seen from Figure 6.14 that more detail is revealed at finer scales. As the scale gets coarser, the DyWT gets flatter. This can be explained from the filtering point of view.

For the filtering view of the DyWT, equation (3.10) can be rewritten as the following convolution form:

$$\begin{aligned} DyWT(b,2^j) &= \int_{-\infty}^{\infty} x(t) \cdot \frac{1}{\sqrt{2^j}} \varphi(\frac{t-b}{2^j}) dt \\ &= x(t) * \varphi(-\frac{t}{2^j}) \end{aligned} \tag{6.2}$$

which means the DyWT can be viewed as the output of a filtering operation where the wavelet function $\varphi(-\frac{t}{2^j})$ plays the role of the analysis filter and acts as a band-pass filter centered around its scale-dependant frequency. The filtering window, which is [$-2^j\sigma$, $2^j\sigma$] as discussed in chapter 4, becomes bigger as the scale $j$ increases. Just as in a conventional filtering operation, a filter with a larger window passes a slowly changing signal and that with smaller window does vice versa.

In the case of the Morlet wavelet, $\sigma = 4$, the window size at each scale is [$-2^{j+2}$, $2^{j+2}$]. By default the coarsest level should be the maximum decomposed level, i.e. $2^7$ in this example. Assume $j$ = 7, 6, or 5, the half length of the wavelet window is 512, 256 or 128, respectively. This is illustrated in Figure 6.15, where (a), (b) and (c) show the wavelet decomposition of the same signal (any signal with length 128) but with a different size of Morlet window. It can be seen that when the Morlet wavelet is moving along the signal as a filtering function, all of the original signal sample at scale $2^7$ and $2^6$ and the majority of the original signal sample at scale $2^5$ is always involved in the convolution. The results therefore cannot contain sufficient detail to be used as inputs for W-SSD matching as shown in equation (4.17). For example, the Morlet window size at scale $2^5$ is 128 as it ranges between [-64, 64]. When it is convolved with a signal of length 128, the number of summation / multiplication operations are always less than 128 except at the centre point with full length 128, which means some detailed information is lost. Such scales do not have any meaning for filtering and are therefore not used as the starting scale for W-SSD estimation. In general, the levels that yield sufficient detail for W-SSD matching should be chosen according to the inequality (6.3):

$$1 \leq 2 \cdot 2^{j} \sigma < L \tag{6.3}$$

where $L$ is the length of the original signal. For example, if $L$=128, $\sigma$=4, then $j$ can be 1, 2 and 3. Therefore the coarsest level with sufficient detail for matching in this example is $2^3$.

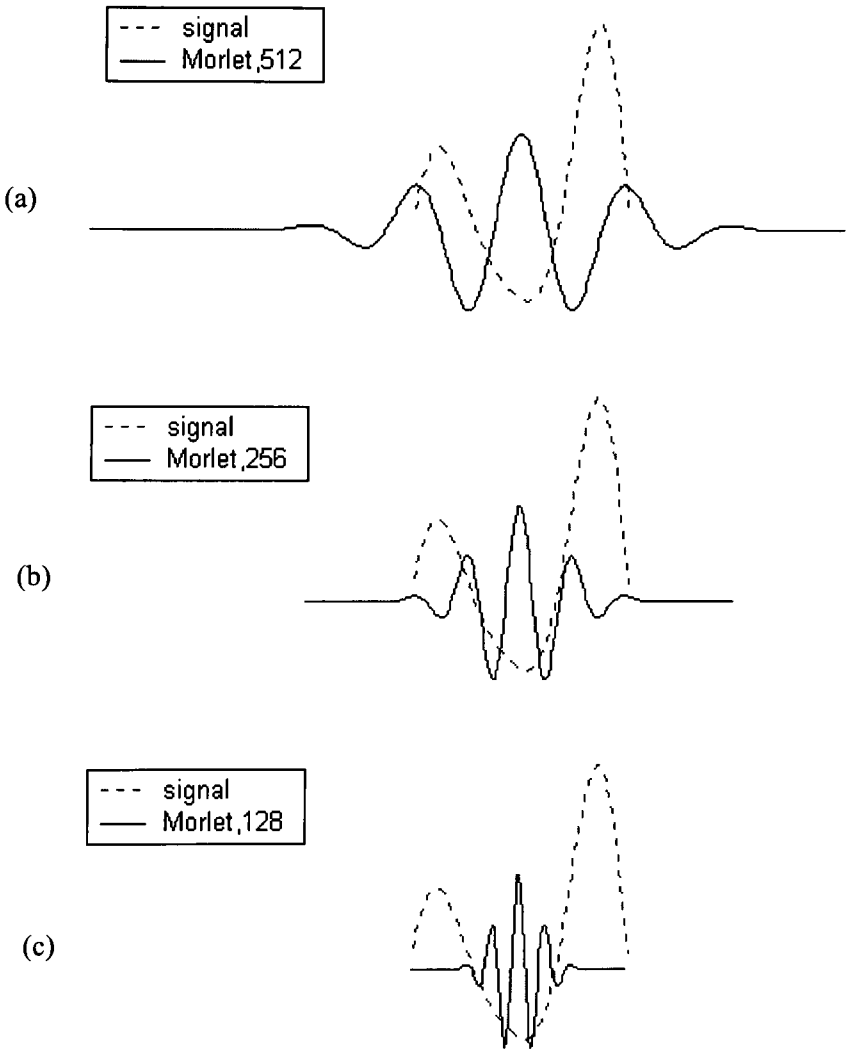



**Figure 6.15 A comparison of the sizes between a signal and dilated wavelets at coarser scales**

*6.4.1.3 Disparity Computation at the Coarsest Level*

At level $2^3$, the inputs for disparity computation are $DyWT_l(2^3, x)$ and $DyWT_r(2^3, x)$, which are plotted in Figure 6.14(e). As discussed in section 4.3, the W-SSD measure for each point at this scale is:

$$wssd\ (2^3, d) = \sum_{\tau=-32}^{32} \left| DyWT_{\ l}(2^3, x_L + \tau) - DyWT_{\ r}(2^3, x_L - d + \tau) \right|^2 \qquad (6.4)$$

And the disparity is:

$$d(2^3, x_L) = \arg\min_d \{wssd(x_L, 2^3, d)\},\ d \in [-32, 32] \qquad (6.5)$$

The computed results are plotted in Figure 6.16.

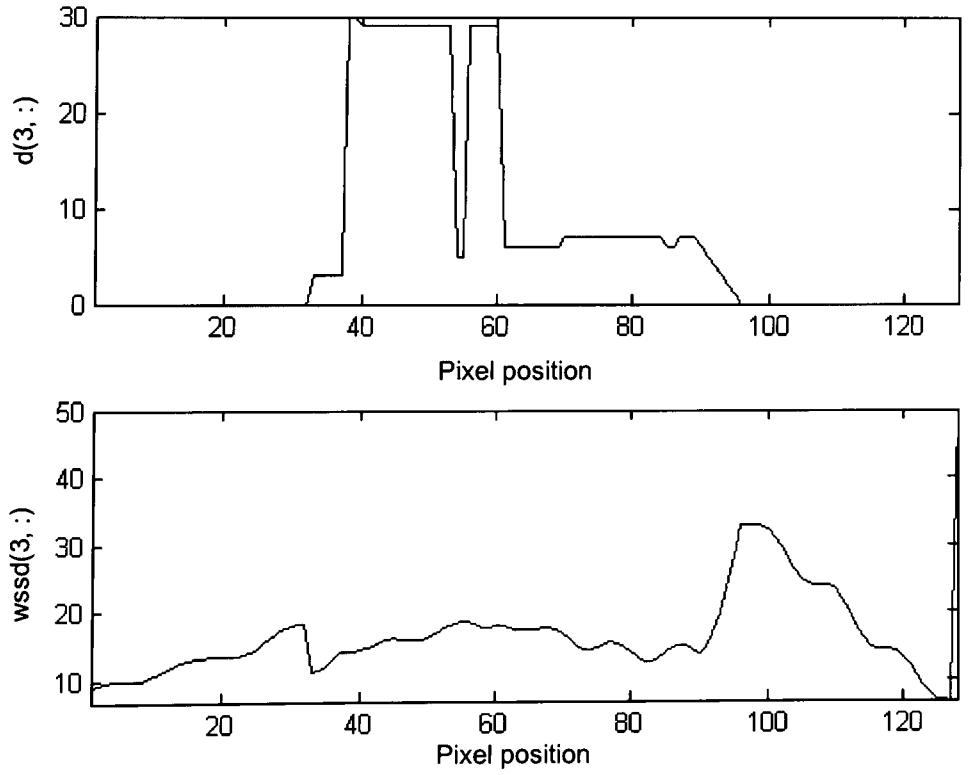



**Figure 6.16 Disparity and *wssd* values at scale $2^3$**

In principle, this method chooses corresponding points by their similarity, i.e. minimising the *wssd* measure in the windows around the center of each point. Such minimum can always be obtained and assigned to each point. However, the differences among the minimum wssd values along the pixel positions are sometimes very large. For example, the minimum *wssd* is 6.6 for some pixels, whereas it is more than 30 for some other pixels. It is assumed in this thesis that the pixels whose minimum *wssd* values are bigger than a threshold are considered as the unmatched positions at that level. The disparities and *wssd* at these positions are assigned big values, say 50, to distinguish them from other positions. The modified *d(3,x)* and *wssd(3,x)* are plotted in Figure 6.17.These positions as well as those that have the disparity of outside of 32 and 16 are then passed to the next level for further determination of disparity. As discussed in section 4.3, the positions with disparity between 32 and 16 are saved as part of the final result.

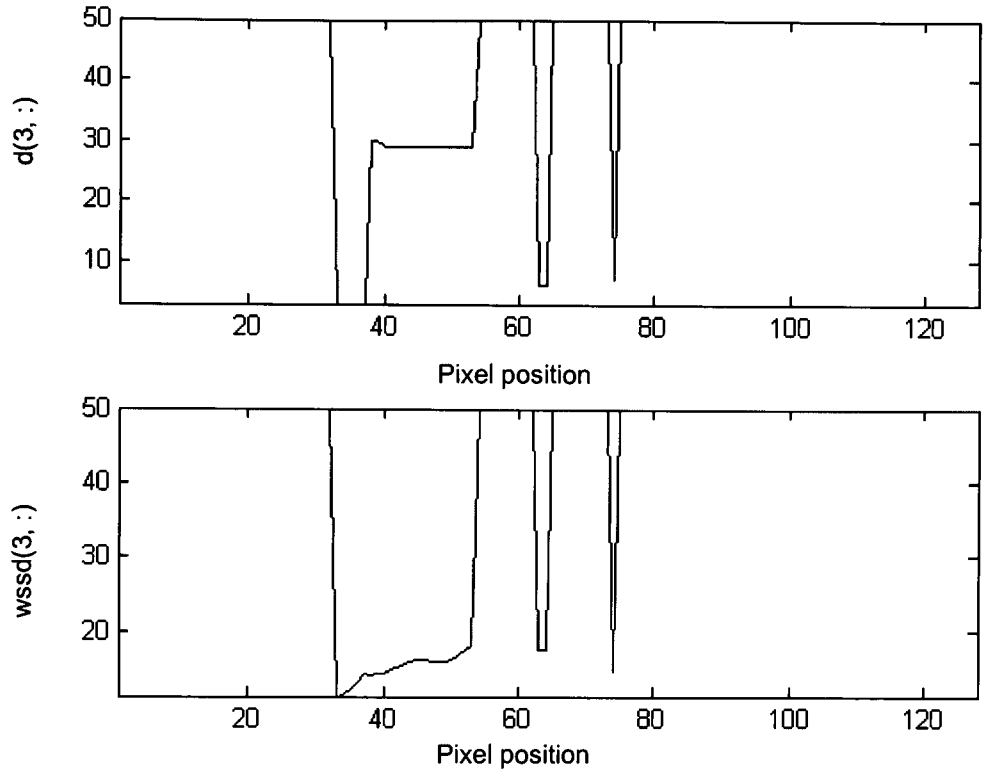



**Figure 6.17 Disparity and *ssd* values at scale $2^3$ after thresholding**

In order to choose a threshold, the standard deviation of the *wssd*, denoted by *std*, is taken into account. When *wssd(x)* > *std*, let *wssd(x)* = 50 and *d(x)* = 50. The pixels with disparities of 50 are effectively marked for the next level of matching.

### 6.4.2 Coarse-to-Fine Estimation

Those positions that are not picked up at the coarsest level and those that have the *wssd* values of 50 will be passed to the next level for further disparity determination. Such refinement is repeated until the finest level is reached. The results at level $2^2$ and $2^1$ of the last example are shown in Figure 6.18 and Figure 6.19, respectively. This procedure was illustrated in the flow graph in Figure 4.6.

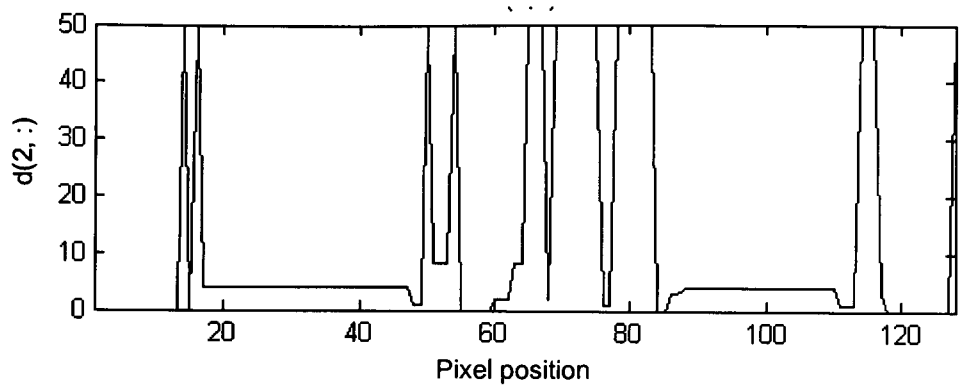



**Figure 6.18 Disparity value at level 2**

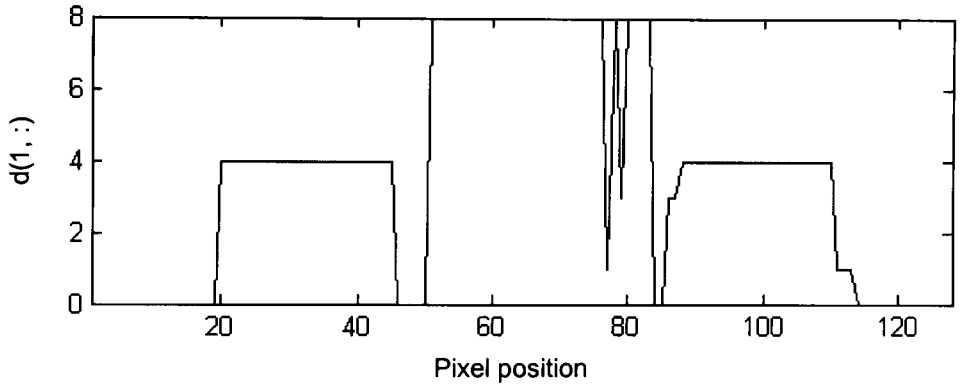



**Figure 6.19 Disparity value at level 1**

### 6.4.3 Disparity Map on Images

Figure 6.19 is the final output of disparity using wavelet decomposition for the stereo lines shown in Figure 6.13. These are a pair of stereo lines from the pair of stereo images shown in Figure 6.2. The disparity computation on both image pairs is made line by line following the same procedure described in section 6.4 and 6.4.2. The results in the form of a disparity map and a 3D depth map are displayed in Figure 6.20.

Considering the evaluation measurement discussed in section 6.3, the quantitative values are as follows:

- Computational time: 2.29s

- Matched rate: 99.2%

- *std_d*: 0.2038.

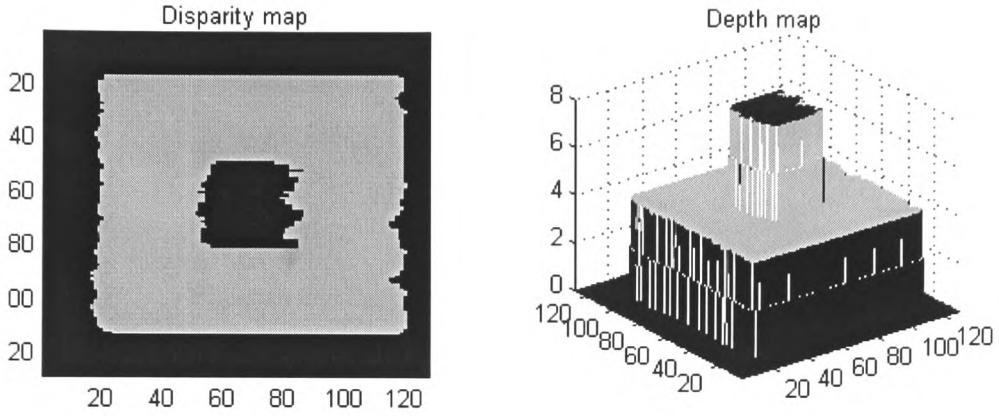



**Figure 6.20 Disparity map and depth map: *Bin_dots***

### 6.4.4 Results with Various Images

The process for computing disparity maps using DyWT, described in this section, has been applied to a number of different stereo images. The computed results on the test images shown in section 6.2 are displayed below as from Figure 6.21 to Figure 6.26. The computation used an Intel Pentium 800, 128M RAM, Microsoft Windows 2000 operated PC and run in MATLAB 6.0 with compiled M files.

Figure 6.21 shows the computed disparity map for the test stereo pair *Dots* shown in Figure 6.1. Compared with the true disparity map shown in Figure 6.4, Figure 6.21 contains 256 incorrect matching pixels. The matching rate is therefore 98.3%. The computed *std_d* is 0.1852.

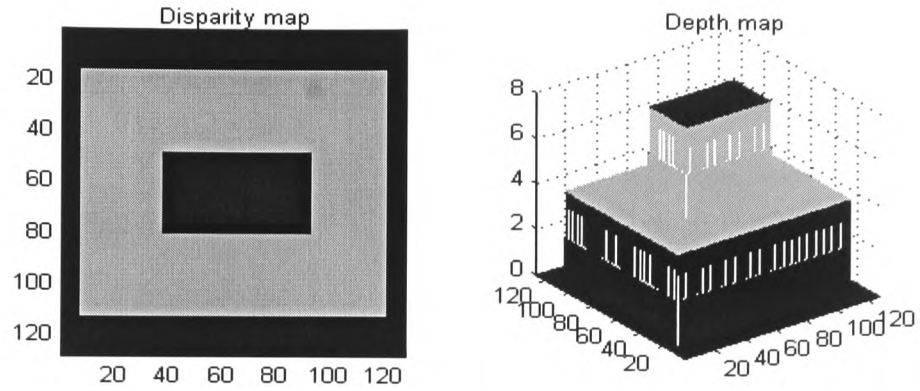



**Figure 6.21 Computed disparity map of the *Dots* pair**

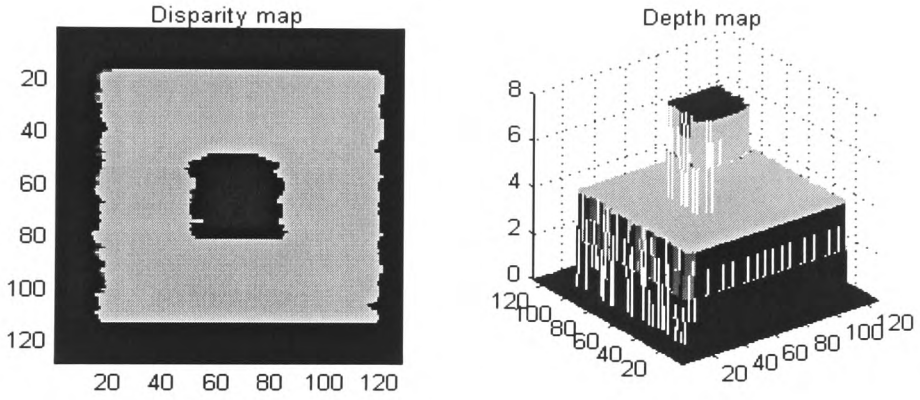



**Figure 6.22 Computed disparity map of the *Ran_dots* pair**

Figure 6.22 is the result for the stereo pair *Ran_dots* of Figure 6.3. The matching rate with this is 98.1% and *std*_d is 0.4107.

Further tests using *Ran_ramp* and *Ran_ball* of Figure 6.5 and Figure 6.7 were carried out. As discussed in section 6.2.1, these images are designed to test the capability of matching slanted flat and rounded surfaces. The disparity maps obtained from DyWT method for these images are presented in Figure 6.23 and Figure 6.24, respectively, in which the ramp with clear slope as shown in Figure 6.6 and the ball shape similar to Figure 6.8 can be seen. For these images, the disparity maps show more deviation as the

disparity values get bigger. The overall matching rate and *std_d* are 96.2%, 1.5794 and 93.0%, 5.8924 for *Ran_ramp* and *Ran_ball* respectively.

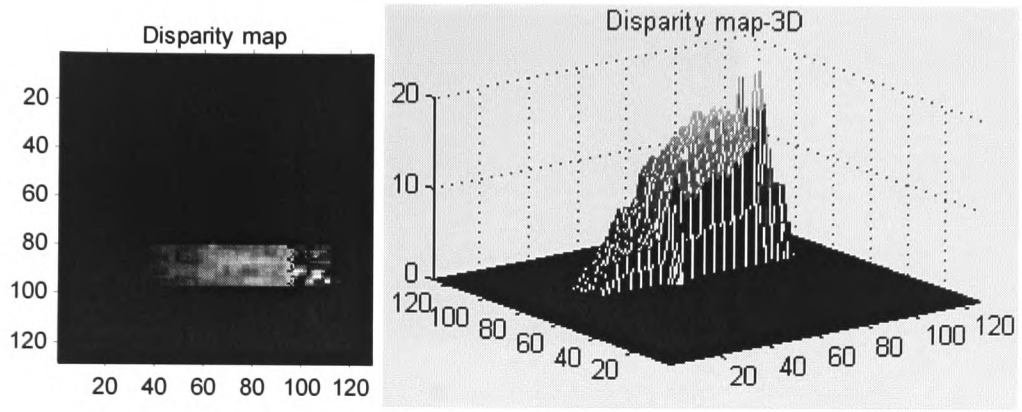



**Figure 6.23 Computed disparity map of the *Ran_ramp* pair**

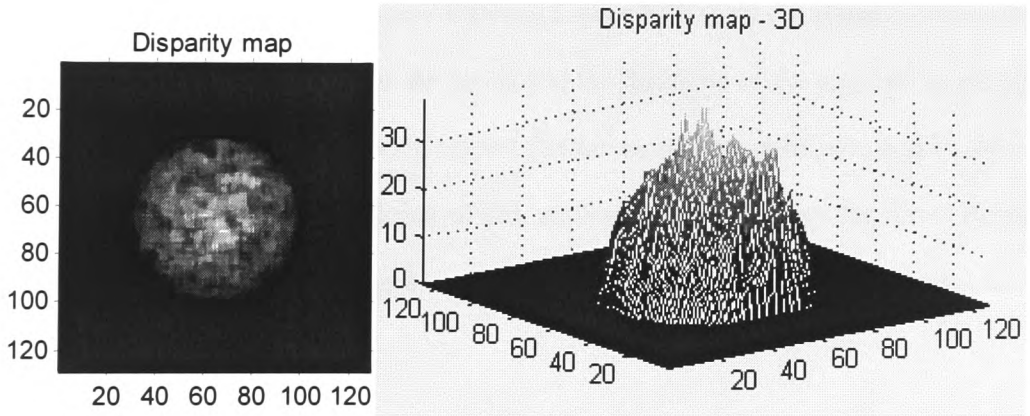



**Figure 6.24 Computed disparity map of the *Ran_Ball* pair**

Figure 6.25 shows the result of Tsukuba pairs of Figure 6.9. It can be seen that the overall disparity map is good. It picks up the details in the feature area such as around the face. The results for areas such as the face (10 & 11 pixels), the lamp (14 pixels) and the table (8 pixels) are in close agreement with the ground truth data. Compared with the ground truth image shown in Figure 6.10, the accurately matched pixels takes

up to 78.3%, whereas the standard deviation of the difference image of the two (*std_d*) is 5.4028.

**Figure 6.25 Computed disparity map of the *Tsukuba* pair**

The approach is also executed with natural images of Figure 6.11 and Figure 6.12 used in (Rothwell Hughes, 1999). Figure 6.26 and Figure 6.27 show the computed disparity map, respectively. The layers of the boxes and the locations of the toys can be clearly seen in the result. The computation takes 4.93s for each of the pairs, which is based on the PC referred to at the beginning of this section and results from the direct output from MATLAB. As these images are uncalibrated, no comparison with ground truth data is possible.

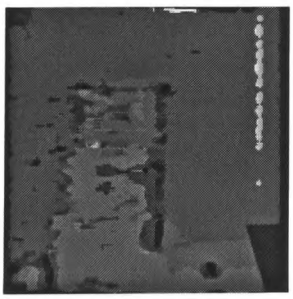


**Figure 6.26 Computed disparity map of the *Boxes***

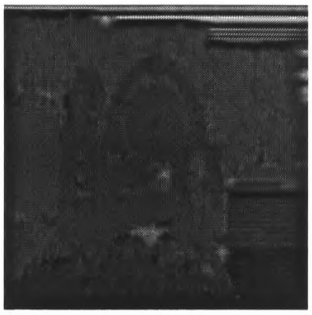


**Figure 6.27 Computed disparity map of the *Toys***

It is worth noticing that the above computed disparity maps show blurred or serrated edges and some matched feature areas look wider than they really are. There are two main reasons for this. Firstly this can be attributed to the smoothing effect of wavelet transform i.e. because wavelet basis functions perform as filtering functions. This means the smoothing effect cannot be avoided. Secondly, the SSD method itself can blur the image edges because the resolution is inherently ±1 pixel. It is possible to apply mean or median filters to eliminate the serrations. As discussed in section 6.2.1, generalised post-processing methods such as using fuzzy filtering approach (Rothwell Hughes, 1999) can be applied to reduce the streakiness while retaining the sharpness of the edges. Further discussion on this can be found in section 6.9.

Other mismatches are mainly due to noise characteristics, which is one of the big problems of SSD matching. A suitable threshold should be chosen to remove the points with higher SSD value. Hard thresholding and soft thresholding (Chambolle *et al*, 1998) are commonly used. The standard deviation is adopted as a soft threshold in this chapter to reduce the SSD noise.

## 6.5 Implementation 2: by DTCWT

The computational method for disparity estimation using DTCWT is described in section 5.4. It computes the disparity at coarsest level, and then refines it at finer levels until the finest level is reached. For initial disparity estimates $\mathbf{d}\in[-0.5, 0.5]$, the procedure for hierarchical refinement used in his thesis is as follows:

Step 1. Compute $m$ levels DTCWT coefficients $D_l^{(n,m)}(\mathbf{p})$ and $D_r^{(n,m)}(\mathbf{p})$, n=1,...,6, using the decomposition structure shown in Figure 5.9.

Step 2. Form the subband squared difference $SD^{(m)}(\mathbf{p},\mathbf{d})$ using equations (5.16) to (5.20) at the coarsest level $m$.

Step 3. Estimate the disparity $\mathbf{d}_0$ by equation (5.21) for level $m$.

Step 4. Apply interpolation equation (5.22) to form a update surface $\mathcal{SD}^{(m)}$

Step 5. Form surface $SD^{(m-1)}(\mathbf{p},\mathbf{d})$ using equations (5.16) to (5.20) at next finer level $m-1$.

Step 6. Add $\mathcal{SD}^{(m)}$ and $SD^{(m-1)}(\mathbf{p},\mathbf{d})$ to form the cumulative squared difference $CSD^{(m-1)}$ using (5.24).

Step 7. Solve (5.24) for $\mathbf{d}$ for $m-1$ level estimate.

Step 8. Go back to Step 4 to continue the refinement procedure until the finest level is reached.

Imaging testing with this method has been carried out using three of the image pairs given in section 6.2, *Ran_Dots, Tsukuba* and *Boxes*. Figure 6.28 to Figure 6.30 show the corresponding computed disparity maps for the DTCWT method. The measures of

the computing time, matched rate and deviation for each image pair are given under each caption.

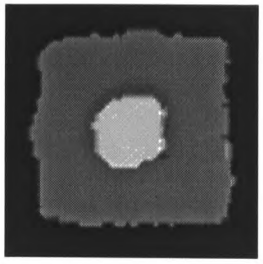


**Figure 6.28 Computed disparity map using DTCWT: *Ran_dots***

**Computing time: 3.14, matched rate: 98.0%, deviation: 0.4156**

**Figure 6.29 Computed disparity map using DTCWT: *Tsukuba***

**Computing time: 9.01, matched rate: 69.8%, deviation: 8.28**

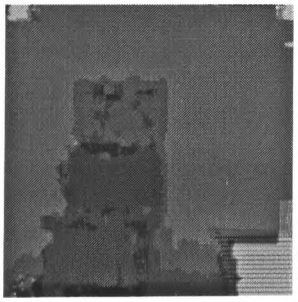


**Figure 6.30 Computed disparity map using DTCWT: *Boxes***

**Computing time: 8.15, matched rate: -, deviation: -**

## 6.6 Implementation 3: by Gaussian Pyramid

For comparison purposes, matching results with the above test image pairs using Gaussian pyramid are presented in this section. This is one of the conventional hierarchical matching method and is used by many researchers (Marr and Poggio, 1976; Grimson, 1981; Barnard and Fishler, 1982; Rosenfeld, 1984; Cantoni *et al*, 1989; Yang *et al*, 1993; Kumar and Desai, 1994; Rojas *et al*, 1997; Szeliski and Scharstein, 1998)

The process of stereo matching using the Gaussian pyramid approach is described in section 2.5. The computed disparity maps are given in Figure 6.31 to Figure 6.33 using the image pairs of *Ran_dots, Tsukuba* and *Boxes*, respectively. The measures of the Computing time, matched rate and deviation for each image pair are given under each caption.

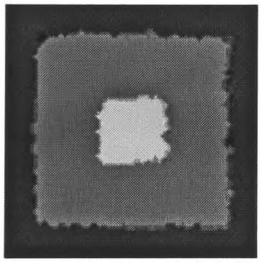


**Figure 6.31 Computed disparity map with Gaussian pyramids: *Ran_dots***

**Computing time: 2.36, matched rate: 98.7%, deviation: 0.3827**

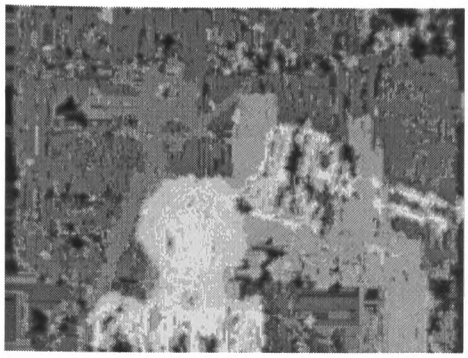


**Figure 6.32 Computed disparity map with Gaussian pyramids: *Tsukuba***

**Computing time: 6.29, matched rate: 72.6%, deviation: 8.1795**

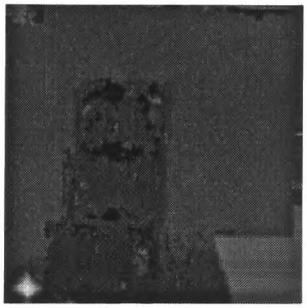


**Figure 6.33 Computed disparity map with Gaussian pyramids: *Boxes***

**Computing time: 5.09, matched rate: -, deviation: -**

Figure 6.31 to Figure 6.33 show the results of applying the Gaussian Pyramid for *Ran_Dots, Tsukuba* and *Boxes images*, respectively.

## 6.7 Implementation 4: standard SSD

This implementation provides a baseline performance for a standard correlation-based matching using SSD measure as discussed in (Trucco and Verri, 1998). The disparity maps are computed with the above three stereo images and are shown in Figure 6.34 to

Figure 6.36. The standard correlation method uses the SSD measure discussed in section 2.3.1. As is well known (Marr, 1982); (Trucco and Verri, 1998), different sizes of searching area and correlation window affect the matching result. Specific parameters for the three images are labelled below each caption. The measures of computing time, matched rate and deviation for each image pair also follow each caption.

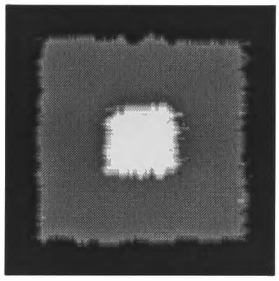


**Figure 6.34 Computed disparity map with standard SSD: *Ran_dots***

**Searching area: 9, window size: 5*3**

**Computing time: 4.27, matched rate: 98.5%, deviation: 0.3856**

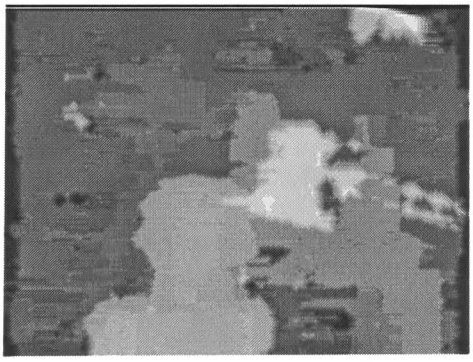


**Figure 6.35 Computed disparity map with standard SSD: *Tsukuba***

**Searching area: 17, window size: 9*9**

**Computing time: 17.20, matched rate: 73%, deviation: 7.2836**

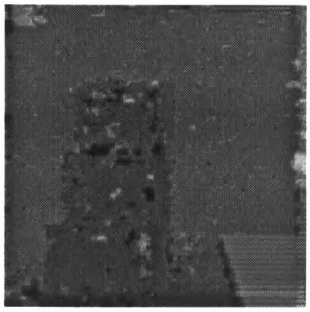


**Figure 6.36 Computed disparity map with standard SSD: *Boxes***

**Searching area: 21, window size: 9*9**

**Computing time: 11.73, matched rate: -, deviation: -**

## 6.8 Comparative Evaluation

This section is summarises the computed results and presents an evaluation of DyWT-based matching method in comparison with other methods. Random dot stereograms do not consider the perspective distortions, but they are used in this thesis to allow a check of the performance of the proposed method for different image textures and scene structure. A comparison with the five types of RDS discussed in section 6.2.1 is first made in section 6.8.1. Following this, section 6.8.2 compares the new DyWT method with three other matching approaches discussed in sections 6.5 to 6.7, using three stereo pairs: *Ran_dots, Tsukuba* and *Boxes* given in section 6.2.2 and 6.2.3.

With respect to the computational complexity, it is difficult to determine the O-notation for the three algorithms. The DyWT- and pyramid-based method performs image matching by applying the epipolar constraint, which reduces the search routines from two dimensions to one dimension. For one-dimensional matching between epipolar lines, the computational complexity using Dyadic wavelet transform and Gaussian

pyramid is $O(N)$ for $N$ inputs. As far as the DTCWT-based method is concerned, it is claimed (Kingsbury, 2000) that the redundancy with DTCWT is 4:1 compared with Mallat's multiresolution tree (Mallat, 1989) as shown in Figure 3.4 of Chapter 3. The computational complexity for the latter is $O(NlogN)$. Therefore, the complexity measure for DTCWT can be considered as $O(NlogN)$.

As big $O$-notation is defined (Horowitz and Shani, 1978), it is used to express an upper bound for the computational time. In some cases, the algorithm efficiency is not as normally suggested such as $O(N) < O(NlogN) < O(N^2) < O(2^N)$. For example, if there exist two algorithms which perform the same task on $N$ inputs, and the first has a computing time which is $O(N)$ and the second $O(N^2)$, the normal judgment is that the first algorithm is faster than the second. Assume the case that the actual computing times for the two algorithms are $10^4N$ and $N^2$, then it is easy to see that algorithm two is faster for all $N<10^4$. Only when $N>10^4$, is algorithm one faster.

It is difficult to determine the real computing time for the implementation of the two wavelet-based matching methods as so many loops are involved. Alternatively, it is possible to make sure all of the methods programmed by the same person using the same programming style, implemented under the same computer condition, as given in section 6.4.4 and then measure the running time from the beginning to the end by calling the MATLAB function.

### 6.8.1 Comparison of Results from Stereograms Using DyWT Method

Five types of random dot stereograms have been generated altogether. *Dots*, *Bin_dots* and *Ran_dots* reflect 3D objects with flat surfaces whose textures vary from simple to random. *Ran_ramp* and *Ran_ball* are designed to represent slanted flat and rounded surfaces. As shown by the computed disparity results presented in section 6.4.3 and

6.4.4, the DyWT method with various stereograms achieved a successful matched rate and low standard deviation, std_d. These are summarised in Table 6.2. It can be seen that DyWT-based matching approach has a better performance for flat objects than slanted flat and rounded objects.

**Table 6.2 Test of DyWT with stereograms**

| | *Dots* | *Bin_dots* | *Ran_dots* | *Ramp* | *Ball* |
|---|---|---|---|---|---|
| rate | 98.3% | 99.2% | 98.1% | 96.2% | 93.0% |
| std_d | 0.1852 | 0.2038 | 0.4107 | 1.5794 | 5.8924 |

As the five stereograms have the same size of 128*128, the computing time using DyWT matching is the same at 2.29 seconds.

### 6.8.2 Comparison of Results Between Four Approaches

Table 6.3 shows the comparative results of DyWT matching method with the other three methods, DTCWT-, Pyramid and standard SSD as discussed in sections 6.5 to 6.7, for computing time, matched rate and the standard deviation. One pair of test images from each test data category (stereograms, synthesised and real images as shown in section 6.2) is chosen for evaluation. For comparative reasons, *Ran_dot*, *Tsukuba* and *Boxes* are used. It can be seen from Table 6.3 that of the four methods, DyWT-based matching takes a similar time as the Pyramid-based matching for all the three pairs of images, both of which are quicker than standard SSD. The *std_d* measure reflects the deviation of the computed disparity map from its corresponding ground truth data. In the case of DyWT matching, this deviation remains small. The matched rate with the DyWT method is reasonably high although sometimes it is not the highest. DyWT is

5% better than standard SSD for *Tsukuba*. However, for random dot stereograms the standard SSD and Pyramid are better than DyWT. This suggests that DyWT be better at handling perspective distortions than SSD. Of the four matching algorithms, DTCWT shows the lowest matched rate and most *std_d*, but it is quicker than standard SSD. In general phase-based matching approaches can hardly compete with the standard SSD matching (Sanger, 1988). The reasons lie in the assumption of transferring the Fourier shift theorem to the linear relationship between the phase difference and shift values using the joint position-spatial frequency representation. To improve the accuracy of DTCWT, equation (5.1) depicting this assumption could be refined in future work.

**Table 6.3 Comparison of DyWT-based matching with the other three approaches**

| Method | *Ran_dots (128*128)* | | | *Tsukuba (284*386)* | | | *Boxes (256*256)* | | |
|---|---|---|---|---|---|---|---|---|---|
| | time(s) | rate | std_d | time(s) | rate | std_d | time(s) | rate | std_d |
| DyWT | 2.29 | 98.1% | 0.4107 | 6.03 | 78.3% | 5.4028 | 4.93 | - | - |
| DTCWT | 3.14 | 98.0% | 0.4516 | 9.01 | 69.8% | 8.2846 | 8.15 | - | - |
| Pyramid | 2.36 | 98.7% | 0.3827 | 6.29 | 72.6% | 8.1795 | 5.09 | - | - |
| Standard SSD | 4.27 | 98.5% | 0.3856 | 17.20 | 73.0% | 7.2836 | 11.74 | - | - |

## 6.9 Discussion

The proposed DyWT-based matching method has achieved good results with synthesised and real images. Some issues such as the blurring effect and the limitations of the method are discussed below.

- **The possibility of eliminating streakiness in the disparity maps by post processing**

As discussed in section 6.2.1 and 6.4.4, either the measurement noise due to perspective distortion or the streakiness caused by convolution operation can be reduced by some post filtering techniques. Median filters are effective in smoothing impulsive points whist preserving discontinuities but less effective in a mean squared error sense when used on areas of images which are smooth and corrupted with Gaussian noise (Kwan and Cai, 1993). Kalman filters can be used to reduce to the uncertainty in the measurement of disparity as they can weight the importance of each pixel in a spatial filter window (MATTHIES *et al*, 1989). In (Rothwell Hughes, 1999), fuzzy filters was designed to regularise the depth maps from a sequence of images. More work on post matching filtering can be seen in (Taguchi *et al*, 1994); (Takashima *et al*, 1995); (Trucco *et al*, 1996).

When filtering is used, the streaking tends to be tidied up, but the parts of disparity maps representing objects are more spread out than the original objects. This is an important and separate body of research which could be tackled in future work. However, the focus of the thesis is not on the filtering stage but on the stereo matching stage of the problem.

- **Limitations of the methods**

As the DyWT indirectly makes use of the intensity values, it has the similar limitation as the conventional correlation-based matching method, e.g. its sensitivity to the differences in foreshortening. In particular, as analysis and results shown in section 6.4, it yields better results with flat objects than slanted flat or rounded objects. However, DyWT appears to be more robust to perspective distortions than standard SSD as shown by the result for *Tsukuba*.

Ordering constraint is one of the important matching constraints used in the development of DyWT matching. However, in the case where the ordering constraint fails, as discussed in section 4.3.4, this method cannot produce expected results.

As much effort in this research has been focused on the understanding and development of the wavelet-based algorithm, the important issues of discontinuities and occlusions in stereo vision have not been focused on in the thesis. In general, discontinuities in disparity and occlusions near intensity edges often occur. In this case, one-to-one matching between the stereo images is invalid. The discontinuous and occluded areas need to be specially treated. Future work should consider this.

Compared with feature-based matching approaches with the disadvantage of yielding only sparse disparity maps, the proposed W-SSD method produces dense disparity maps. Many real applications such as 3D object reconstruction and video tracking have benefited from this advantage (Scharstein and Szeliski, 2002). It is possible to combine the advantages of W-SSD matching and other matching strategy to give dense disparity map output.

- **Are there any better disparity maps in the literature?**

A large number of algorithms for stereo correspondence in the literature have been developed, however, there is no benchmark for stereo correspondence and relatively

little work has been done on uniformly characterizing their performance. A recent review paper (Scharstein and Szeliski, 2002) presented a computed disparity map of *Tsukuba* image using standard SSD plus post-filtering, as shown in Figure 6.37. It can be seen from Figure 6.37 that serrated edges are also present and are similar to those observed in Figure 6.25.

**Figure 6.37 A disparity map using SSD (Scharstein and Szeliski, 2002)**

## 6.10 Summary

This chapter has described the implementation of the two wavelet-based stereo matching methods using DyWT and DTCWT described in Chapter 4&5. In particular, the disparity estimation from the coarsest level to applying coarse-to-fine strategy using DyWT has been detailed. Imaging experimental results with a variety of stereo image pairs have been carried out and discussed, which exhibit a good agreement with ground truth data, where available, and are qualitatively similar to published results for other stereo matching approaches (Grimson, 1985; Szeliski, 1999).

This chapter has also presented an evaluation of the proposed DyWT-based matching method. Five kinds of random dot stereograms have been tested in terms of flat, slanted flat and rounded surfaces, which shows that DyWT-based matching approach has a

better performance for flat objects than slanted flat and rounded objects. The evaluation is also made by comparison with DTCWT, pyramid and standard SSD methods. The comparison has been made in terms of the computational complexity, matched pixel rate and standard deviation of the difference image between the computed disparity map and the ground truth data. Three categories of stereo images have been used for the evaluation. The comparative results have shown that the DyWT-based matching method is superior in most cases to the other three approaches.

The main contributions of this chapter are the implementations of the two wavelet-based stereo matching methods using DyWT and DTCWT and the evaluation of the approaches. There are indeed some aspects of the wavelet-based matching methods that need further consideration. This will be discussed in the conclusion chapter, Chapter 7.

## 6.11 References

Barnard, S. and Fishler, M. 1982. Computational Stereo. *ACM Computing Surveys*, **14** (4), pp. 553-572.

Bertero, M., Poggio, T. A. and Torre, V. 1988. Ill-Posed Problems in Early Vision. *Proceedings of the IEEE*, **76** (8), pp. 869-889.

Borodin, A. and Munro, I. 1975. *The Computational Complexity of Algebraic and Numeric Problems*. New York: Elsevier.

Cantoni, V., Griffini, A. and Lombard, L. 1989. *Stereo Vision in Multi-resolution.* International Conference on Image Analysis and Processing. pp. 706-713,

Chambolle, A., Devore, R. A., Lee, N. Y. and Lucier, B. J. 1998. Nonlinear Wavelet Image Processing: Variational Problems, Compression, and Noise Removal Through

Wavelet Shrinkage. *IEEE Transactions on Image Processing*, **7** (3), pp. 319-334.

Davies, E. R. 1997. *Machine Vision: Theory, Algorithms, Practicalities.* 2nd end. Academic Press.

Dhond, U. R. and Aggarwal, J. K. 1989. Structure from Stereo - A Review. *IEEE Transactions on Systems, Man and Cybernetics*, **19** (6), pp. 1489-1510.

Grimson, W. 1981. A Computer Implementation of a Theory of Human Stereo Vision. *Phil. Trans. Royal Soc. London*, **V292**, pp. 217-253.

Grimson, W. E. L. 1985. Computational Experiments with a Feature-Based Stereo Algorithm. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **7** (1), pp. 17-34.

Horowitz, E. and Shani, S. 1978. *Fundamentals of Computer Algorithms.* Computer Science Press.

Julesz, B. 1971. *Foundations of Cyclopean Perception.* Chicago: University of Chicago Press.

Kanade, T. and Okutomi, M. 1994. A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16** (9), pp. 920-932.

Kingsbury, N. G. 2000. *A Dual-Tree Complex Wavelet Transform with Improved Orthogonality and Symmetry Properties.* IEEE International Conference on Image Processing. pp. 375-378, Vancouver, Canada.

Kumar, K. S. and Desai, U. B. 1994. *Integrated Stereo Vision - a Multiresolution Approach.* International Conference on Pattern Recognition. pp. 714-716,

Kwan, H. K. and Cai, Y. 1993. *Median Filtering Using Fuzzy Concept.* Proc IEEE Symp Circs and Sys. 824-827

Mallat, S. 1989. A Theory for Multiresolution Signal Decomposition: the Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11** (7), pp. 674-693.

Marr, D. 1982. *Vision.* New York: W. H. Freeman and Company.

Marr, D. and Poggio, T. 1976. Cooperative Computation of Stereo Disparity. *Science*, **194**, pp. 283-287.

Matthies, L., Kanade, T. and Szeliski, R. 1989. Kalman Filter-based Algorithms for Estimating Depth from Image Sequences. *International Journal of Computer Vision*, **3**, pp. 209-238.

Pollefeys, M., Koch, R. and Van Gool, L. 1999. Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters. *International Journal of Computer Vision*, **32** (1), pp. 7-25.

Rojas, A., Calvo, A. and Munoz, J. 1997. A Dense Disparity Map of Stereo Images. *Pattern Recognition Letters*, (18), pp. 385-393.

Rosenfeld, A. 1984. *Multiresolution Image Processing and Analysis.* Springer-Verlag.

Rothwell Hughes, N. 1999. *Fuzzy Filters for Depth Map Smoothing.* PhD Thesis, University of Wales.

Sanger, T. D. 1988. Stereo Disparity Computations Using Gabor Filter. *Biol. Cybern.*, **59**, pp. 405-418.

Scharstein, D. and Szeliski, R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, **47** (1), pp. 7-42.

Szeliski, R. 2000. http://www.research.microsoft.com/~szeleski/stereo

Szeliski, R. 1999. *Stereo Algorithms and Representations for Image-Based Rendering.* British Machine Vision Conference (BMVC'99). pp. 314-328, Nottingham, England.

Szeliski, R. and Scharstein, D. 1998. Stereo Matching with Nonlinear Diffusion. *International Journal of Computer Vision*, **28** (2), pp. 155-174.

Scharstein, D. and Szeliski, R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, **47** (1), pp. 7-42.

Taguchi, A., Takashima, H. and Murarta, Y. 1994. *Fuzzy Filters for Image Smoothing.* SPIE Conf Nonfinear Image Processing. pp. 332-339,

Takashima, H., Taguchi, A. and Murarta, Y. 1995. Edge Preserving Smoothing Using the Fuzzy Control Technique. *Elec Comm J Pr III*, **78** (1), pp. 43-72.

Trucco, E. and Verri, A. 1998. *Introductory Techniques for 3-D Computer Vision.* New Jersey: Prentice Hall.

Trucco, E., Roberto, V., Tinonim, S. and Corbatto, M. 1996. *SSD Disparity Estimation for Dynamic Stereo.* Proc British Machine Vision Conference. pp. 343-352,

Valentinotti, F. and Taraglio, S. 1999. A Hybrid Approach for Stereo Disparity Computation. *Machine Vision and Applications*, (11), pp. 161-170.

Yang, Y., Yuillet, A. and Lu, J. 1993. *Local, Global and Multilevel Stereo Matching.* IEEE Conference on Computer Vision & Pattern Recognition. pp. 274-279.

# 7 Conclusions and Further Work

## 7.1 Introduction

This thesis has shown and suggested two wavelet approaches to solving the standard correspondence problem. The shift invariance property of various wavelet transforms is firstly identified. The dyadic wavelet transform is in particular detailed to compute disparity maps through correlation comparison. Based on the conventional correlation-based matching method, a new W-SSD measure is defined for similarity comparison. Application of dual tree of complex wavelet transform to stereo matching is also formulated. Multi-scale matching scheme is applied for both the matching methods. Imaging tests have been made with various synthesised and real image pairs. Compared with the conventional matching method using Gaussian pyramid and standard SSD, the results have shown that wavelet-based matching methods are challenging approaches to standard stereo matching.

As a summary of the whole thesis, this chapter reviews the structure of the thesis, presents the discussions on the results obtained and highlights the contributions made in the thesis. Suggestions are also proposed for further investigation.

## 7.2 Review of the thesis

The objective of the thesis was to develop two wavelet approaches to solving the correspondence problem. The content of the chapters before the current one have been organised as follows:

Chapter 1 is an introductory chapter, which gives an outline of the thesis.

Chapter 2 and 3 are two review chapters with specific focus on stereo vision and wavelets, respectively.

Chapter 2 has presented background of stereo vision and an overview of the stereo matching including stereo geometry, matching constraints and matching methods. Two conventional correlation- and phase-based matching methods as well as multiple scale approach have been outlined. The discussion on their disadvantages has led to Chapter 3.

Chapter 3 has reviewed the wavelet theory and discussed the importance of wavelet shift invariance property to stereo imaging. The obstacle to applying wavelets to matching is then made specific, which is the shift variance problem of wavelet transforms. The literature survey showed that despite the growing ways to achieve shift invariance of wavelet transforms their applications to stereo matching often have disadvantages. The gap was specifically noted in terms of minimum information redundancy, computational expense and accuracy.

In Chapter 4 an approach using dyadic wavelet transform to compute stereo disparity maps is developed. As a signal representation, the one-dimensional dyadic wavelet transform has been discussed under the theory of frames. Two appropriate wavelet frames, the Mexican hat and Morlet wavelets have been formulated as they are used for the stereo matching application in later chapters. The fast implementation of DyWT, *Algorithme à Trous*, has also been discussed. In the two-dimensional image case, the epipolar constraint has been reviewed as it is used to allow the reduction of searching from two dimensions to one dimension. Based on the DyWT coefficients, a novel

matching approach using W-SSD measure for similarity comparison has been formulated in this chapter.

Another wavelet-based matching method using complex wavelet transform via phase information is described in Chapter 5. Fourier phases and local phases are discussed and a method for computing local Gabor phases in a band-passed signal is examined. The difficulty in combining the disparities from the filtered signal using various centre frequencies is discussed and a dual-tree structure of the filter design and its shift invariance property is developed and applied to disparity computation.

Implementation, imaging tests and the algorithm evaluation of applying the new wavelet-based methods to compute disparity maps are presented in Chapter 6. In particular, the disparity estimation from the coarsest level to applying coarse-to-fine strategy using DyWT has been discussed in detail. Imaging experimental results with a variety of stereo image pairs have exhibited a good agreement with ground truth data, where available, and are qualitatively similar to published results for other stereo matching approaches.

Chapter 6 has also presented an evaluation of the proposed DyWT-based matching method. Five kinds of random dot stereograms have been tested in terms of flat, slanted flat and rounded surfaces, which shows that DyWT-based matching approach has a better response to flat objects than slanted flat and rounded objects. The evaluation is also made by a comparison with DTCWT, pyramid and standard SSD methods. The comparison has been made in terms of the computational complexity, matched pixel rate and standard deviation of the difference image between the computed disparity map and the ground truth data. Three categories of stereo images have been used for the

evaluation. The comparative results have shown that the DyWT-based matching method is superior in most cases to the other three approaches.

Overall, it can be concluded that the objectives of the thesis as set out in in Chapter 1 have been met, i.e. to:

√ *Explore Mallat's wavelet multiresolution analysis to determine whether it is suitable for stereo matching*

Mallat's wavelet multiresolution analysis was explored in section 3.3.3. As its representation is not shift invariant, it is not suitable for stereo matching.

√ *Identify the existing wavelet transforms with a view of their suitability to stereo matching*

Various wavelet transforms were identified in section 3. As a result, the dyadic wavelet transform and the complex wavelet transform were chosen to be used to develop new wavelet-based matching methods because they are both shift invariant and offer efficient implementation.

√ *Formulate and implement wavelet-based matching methods*

Two wavelet-based matching methods were developed in the thesis. The method using DyWT through correlation measure was detailed in Chapter 4. And the method using DTCWT through phase difference was discussed in Chapter 5.

√ *Evaluate the performance of the algorithms proposed by comparison with each other and with a conventional matching approach.*

This is done in Chapter 6. The comparison has showed that the proposed approaches are good alternative to conventional matching methods.

## 7.3 Contributions

The main contributions that have been achieved in this thesis are summarised as follows:

- **Applicability of dyadic wavelet transform to disparity map computation**

Applying dyadic wavelet transform to compute disparity maps has not been previously reported. The thesis shows that it is feasible for dyadic wavelet transform to be used for stereo matching.

As reviewed in Chapter 3, not many researchers have used the wavelet approach to the stereo matching problem. It is clear that the most difficulty in applying wavelets to matching is to solve the shift variance problem of Mallat's wavelet multiresolution analysis as it is one of the most important milestones in wavelet history. This thesis chooses Dyadic Wavelet Transform to compute disparity maps by examining its properties in terms of shift invariance and computational complexity. It has been found that the discretisation along the dyadic scales enables considerably efficient computation whereas the continuity along the translation makes a dense disparity map output possible. Thus, it has the practical advantages over other wavelet transforms as discussed in Chapter 3.

Through a definition of the similarity measure using DyWT coefficients and a specially developed coarse-to-fine matching strategy, disparity maps are computed from various stereo image pairs. The results are in good agreement with those by other approaches as compared in Chapter 6. As the DyWT indirectly makes use of the intensity values, it has the similar limitation as the conventional correlation-based matching method, e.g. its sensitivity to the differences in foreshortening. In particular, as analysis and results

shown in sections 6.4 and 6.8, it yields better results with flat objects than slanted flat or rounded objects. In comparison with the standard SSD algorithm, it has better performance for synthesised *Tsukaba*. However, for random dot stereograms, standard SSD performs better. This suggests that DyWT-based matching method be better at handling perspective distortions than SSD.

In history, before wavelets were formulated, multi-scale image processing has been established by many vision researchers such as Burt and Adelson (Burt and Adelson, 1983), Marr (Marr, 1982) and Rosenfeld (Rosenfeld, 1984). Some of the ideas have later been formalised and refined by the wavelet theory. This makes the further study of application of wavelet transform to image analysis, e.g. the contribution discussed in this section of significant importance.

- **Definition of a wavelet-based similarity measure for matching**

After the properties of dyadic wavelet transform are examined in Chapter 3, a new similarity measure using DyWT coefficients, i.e. W-SSD, are defined in Chapter 4 based on conventional SSD measure. In contrast to the conventional SSD, the new W-SSD allows a coarse-to-fine multi-scale disparity estimate due to the hierarchical structure of DyWT. Once a specific mother wavelet is chosen, for a pair of fixed sized images, the comparative window at each scale in a coarse-to-fine procedure is correspondingly determined. The final disparity values can be picked up from the sub-results at each scale using the method described in Chapters 4 & 5. The windowing problem inherent in the conventional correlation-based matching is thus alleviated.

- **Combination of matching results from different scales based on the detectable minimum disparity at each scale**

This is a small contribution made in this thesis as it has a little difference from conventional coarse-to fine matching. The approach to doing this is described in detail in Chapter 4 and the implementation is discussed in Chapter 5. The maximum detectable disparity at each scale is dependent on the wavelet window. As the selected disparity values are set to be in a range of values, which are determined by the scaled window size, the procedure for keeping the optimal disparity values is simplified compared with the well-known conventional coarse-to fine approach.

- **Application of DTCWT to stereo matching**

Complex wavelet transform was firstly designed by Magarey and Kingsbury (Magarey and Kingsbury, 1995). Its application of complex wavelet transform to motion estimation has been reported by Magarey (Magarey, 1997). Since the new dual-tree structure, as an update to the original complex wavelet transform, was designed (Kingsbury, 2000), its performance with stereo matching has not been reported. The application of DTCWT for phased-based disparity computation is formulated in Chapter 5 and implemented in Chapter 6. The computed disparity maps have picked up the content and structure from original stereo pairs. However, the comparative result presented in section 6.8.2 has shown that its performance cannot compete with that of standard SSD due to the fundamental linear assumption made by equation (5.1).

## 7.4 Suggestions for Further Investigation

In order to further understand the performance of applying wavelet transforms to stereo matching, further investigation is suggested through the following ways.

- Consideration of discontinuities and occlusions

As discussed in Chapter 6, section 6.9, much effort in this research is focused on the understanding and development of the wavelet-based algorithm, the important issues of discontinuities and occlusions in stereo vision have not been considered in this thesis. In general, discontinuities in disparity and occlusions near intensity edges often occur. In this case, one-to-one matching between the stereo images is invalid and the discontinuous and occluded areas need to be specially treated.

Recent research has been reported on how to deal with the outliers and occlusions when using SSD measure. For example, a bi-directional matching process was discussed in (Pan and Magarey, 1998), where the matching process is carried out from the left to the right and from the right to the left image in two separate but identical processes. A multi-view approach has been proposed at the Microsoft Research (Szeliski, 1999), which links up corresponding image points over multiple viewpoints by correspondence tracking over adjacent image pairs. It works with an image sequence and integrates the disparity results obtained from several pairs by a correspondence linking algorithm. In addition, other methods can be found in (Szeliski and Scharstein, 1998) and (Clark, 2002) using non-linear diffusion and maximum likelihood respectively.

- Investigation for algorithm efficiency improvement

Mallat's wavelet multiresolution analysis is the most efficient implementation and has found many applications but it cannot be used for stereo matching as it lacks shift invariance. The orthogonality of wavelet bases is a necessity for efficient implementation. However, it is contradictory to the wavelet shift invariance. The new wavelet-based method developed in this thesis is efficient along the dyadic scale axis but there is much redundancy along the continuous translation axis. Although the

redundancy is useful for dense disparity map output, the price for it could be lower by investigating a more efficient implementation in the future work and at the same time a balance between the algorithmic efficiency and the dense disparity map output is remained. This could be achieved by studying the relationship between the discretisation of the translation parameter and the shift invariance of the corresponding wavelet transform.

- Investigation for other shift invariant wavelet transforms

Two wavelet-based matching approaches have been discussed in the thesis using correlation- and phase-based methods respectively. With a view to expanding the application of wavelet transform to stereo matching, future work may try the zero-crossings of wavelet transform, whose representation is described in (Mallat, 1998). That would result in the use of feature-based matching method, which will generate sparse disparity maps.

The purpose for investigating various wavelet-based matching method is to fully understand the characteristics of wavelet transforms to do effective stereo matching. The advantages from different methods may be integrated into further novel wavelet-based matching techniques.

## 7.5 Potential Applications

Researchers have been investigating methods to acquire 3D information from objects and scenes for many years. As the visual quality becomes one of the main points of attention, not only the position of a small number of points have to be measured with high accuracy as a result of feature-based methods, but the geometry and appearance of all points of the surface have to be measured. This gives rise to the requirement for

dense disparity maps. The proposed W-SSD method not only produces dense disparity maps but also has the computational efficiency by naturally applying coarse-to-fine matching strategy due to shiftable and scalable wavelets. It can therefore be used in many real applications such as high-resolution 3D model reconstruction and motion estimation. However, due to the ill-posed nature of the matching problem, more future work as discussed in the previous section is needed.

## 7.6 References

Burt, P. and Adelson, E. H. 1983. The Laplacian Pyramid as a Compact Image Code. *IEEE Trans. Communications*, **31** (4), pp. 532-540.

Clark, F. O. 2002. Maximum-Likelihood Image Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24** (6), pp. 853-857.

Kingsbury, N. G. 2000. Complex Wavelets for Shift Invariant Analysis and Filtering of Signals. *Journal of Applied Computation and Harmonic Analysis*,

Magarey, J. 1997. *Motion Estimation Using Complex* Wavelets. PhD thesis. Department of Engineering, Cambridge University.

Kingsbury, N. G. and Magarey, J. 1996. *Wavelets in Image Analysis: Motion and Displacement Estimation*. Proc. Irish DSP and Control Conference. pp. 199-217, Dublin.

Mallat, S. 1998. *A Wavelet Tour of Signal Processing.* Academic Press.

Marr, D. 1982. *Vision.* New York: W. H. Freeman and Company.

Pan, H. P. and Magarey, J. 1998. Multiresolution Phase-Based Bidirectional Stereo.

*Digital Signal Processing*, **8** (4), pp. 255-266.

Rosenfeld, A. 1984. *Multiresolution Image Processing and Analysis.* Springer-Verlag.

Szeliski, R. 1999. *A Multi-View Approach to Motion and Stereo.* IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 157-163, Fort Collins.

Szeliski, R. and Scharstein, D. 1998. Stereo Matching with Nonlinear Diffusion. *International Journal of Computer Vision*, **28** (2), pp. 155-174.

# Appendix A Camera Calibration

## A.1 Introduction

The purpose of camera calibration is to quantify the relationship between what appears on the image plane and where it is located in the three-dimensional world. A simple camera model has four intrinsic and six extrinsic parameters to determine, which is discussed in section 2.

## A.2 Pinhole Camera Geometry

In stereo vision, a camera is modelled as a pinhole camera (Faugeras, 1993b). An image is formed in an image plane through an optical centre by 3D-2D perspective projection. Figure A.1 illustrates such a model system. The co-ordinate systems presented in section 2.2 apply to this figure. $\boldsymbol{R}$ is an image plane, $\boldsymbol{F}$ is the camera plane, $C$ is the camera optical center and c is its image in $\boldsymbol{R}$. $M$ and $m$ denote a physical point and its projective image. $(X,Y,Z)$, $(x,y)$ and $(u,v)$ stand for world, camera and pixel co-ordinate system, respectively, as discussed in section 2.2. $f$ is the focal length.

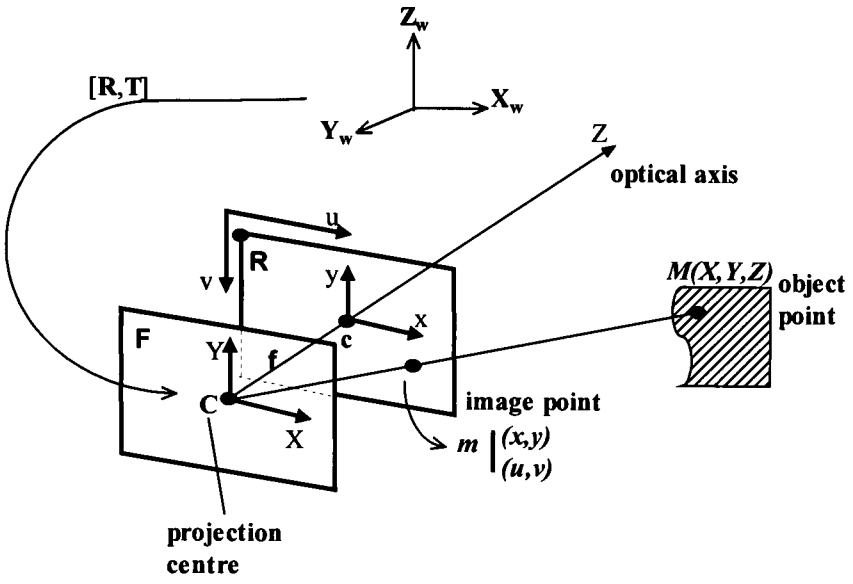



**Figure A.1 A pinhole camera model system**

Assume that the origin of the world co-ordinate system is at the center of the camera and the its *Z*-axis lies along the optical axis, as shown in Figure A.1, which is called the standard co-ordinate system. According to the camera model geometry (Faugeras, 1993b), the projection transformation from the 3D space into the 2D space can be described by a linear relationship:

$$\tilde{\mathbf{m}} = \mathbf{P} \cdot \tilde{\mathbf{M}} \tag{A.1}$$

where $\tilde{\mathbf{M}}$ and $\tilde{\mathbf{m}}$ represent the homogeneous co-ordinates of a 3*D* point and its 2*D* image, respectively, $\tilde{\mathbf{M}} = [X\ Y\ Z\ 1]$, $\tilde{\mathbf{m}} = [su\ sv\ s]^T$, $s \neq 0$ is a scale factor. The equation is equivalent to:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} \dfrac{f}{pixel\ width} & 0 & u_c & 0 \\ 0 & \dfrac{f}{pixel\ height} & v_c & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{A.2}$$

where $(u_c, v_c)$, the image co-ordinate of the optical centre, the ratios $\alpha_u = f / (pixel\ width)$ and $\alpha_v = f / (pixel\ height)$, and $f$, the focal length, are called the intrinsic parameters that depend only on the camera itself (Trucco and Verri, 1998).

More generally, the origin and the $Z$-axis of a three-dimensional world co-ordinate system can be any point and any line, not necessarily as the case shown above. A change of co-ordinates from any other frame to the standard co-ordinate system is needed. Introducing a homogeneous transformation matrix **K** (Trucco and Verri, 1998):

$$\mathbf{K} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0_3^T & 1 \end{bmatrix}$$

where **R** is a 3×3 rotation matrix, in the form of three row vectors, $\mathbf{R}=[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3]^T$, and **T** is a translation vector, $\mathbf{T}=[t_x\ t_y\ t_z]^T$, the following equation is obtained:

$$\tilde{\mathbf{m}} = \mathbf{P} \cdot \mathbf{K} \cdot \tilde{\mathbf{M}} = \mathbf{C} \cdot \tilde{\mathbf{M}} \tag{A.3}$$

The matrix **K** has six degrees of freedom, three for the orientation and three for the translation of the camera. These parameters are known as the extrinsic camera parameters (Trucco and Verri, 1998).

Now the general form of 3×4 matrix **C**, called the camera calibration matrix that links the camera intrinsic and extrinsic parameters, can be written as:

$$\mathbf{C}=\begin{bmatrix} \alpha_u \mathbf{r_1}+u_c\mathbf{r_3} & \alpha_u t_x+u_c t_z \\ \alpha_v \mathbf{r_2}+v_c\mathbf{r_3} & \alpha_v t_y+v_c t_z \\ \mathbf{r_3} & t_z \end{bmatrix} \tag{A.4}$$

Calibration is therefore the process of estimating the intrinsic and extrinsic parameters of the camera. It can be thought of as a two-stage process:

i. Estimating the matrix **C**, and

ii. Estimating the intrinsic and extrinsic parameters from **C.**

## A.3 Linear Calibration Method

Rewrite the calibration matrix **C** of (A.4) as:

$$\mathbf{C}=\begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{bmatrix} \tag{A.5}$$

where $q_{34}$=1 is set because of an arbitrary scale factor s in **C**. The linear relationship between the image points $m_i$ and the 3D reference points $M_i$ in homogeneous co-ordinate is:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix}=\begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & 1 \end{bmatrix}\bullet\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{A.6}$$

Solving this equation yields:

$$u = \frac{Xq_{11} + Yq_{12} + Zq_{13} + q_{14}}{Xq_{31} + Yq_{32} + Zq_{33} + 1}$$
$$v = \frac{Xq_{21} + Yq_{22} + Zq_{23} + q_{24}}{Xq_{31} + Yq_{32} + Zq_{33} + 1} \tag{A.7}$$

Alternatively,

$$\begin{aligned} Xq_{11} + Yq_{12} + Zq_{13} + q_{14} - uXq_{31} - uYq_{32} - uZq_{33} &= u \\ Xq_{21} + Yq_{22} + Zq_{23} + q_{24} - vXq_{31} - vYq_{32} - vZq_{33} &= v \end{aligned} \tag{A.8}$$

Therefore, given a set of $N$ 3D world points and their image co-ordinates, the following matrix equation can be built up:

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & & & & & & & & \\ \cdot & & & & & & & & & & \\ \cdot & & & & & & & & & & \\ \cdot & & & & & & & & & & \\ \cdot & & & & & & & & & & \\ \cdot & & & & & & & & & & \\ \cdot & & & & & & & & & & \\ \cdot & & & & & & & & & & \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -u_NX_N & -u_NY_N & -u_NZ_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -v_NX_N & -v_NY_N & -v_NZ_N \end{bmatrix} \bullet \begin{bmatrix} q_{11} \\ q_{12} \\ q_{13} \\ q_{14} \\ q_{21} \\ q_{22} \\ q_{23} \\ q_{24} \\ q_{31} \\ q_{32} \\ q_{33} \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ u_N \\ v_N \end{bmatrix} \tag{A.9}$$

where the matrix of the known is $2N \times 11$.

With 11 unknowns and each point providing two constraints, at least six points are needed to solve equation (A.9).

If equation (A.9) is written as:

$$[\mathbf{B}]\cdot[\mathbf{C}]=[\mathbf{UV}] \tag{A.10}$$

Then

$$[\mathbf{C}]=[\mathbf{B}]^{+}[\mathbf{UV}] \tag{A.11}$$

where $[\mathbf{B}]^{+}=\left[\mathbf{B}^{T}\mathbf{B}\right]^{-1}\mathbf{B}^{T}$

To calculate the intrinsic and extrinsic parameters, write **C** of (A.5) in the following form:

$$\mathbf{C}=\begin{bmatrix}\mathbf{q}_1^{\mathrm{T}} & q_{14}\\ \mathbf{q}_2^{\mathrm{T}} & q_{24}\\ \mathbf{q}_3^{\mathrm{T}} & q_{34}\end{bmatrix} \tag{A.12}$$

Combining (A.12) and the other form of **C** in (A.4) yields the parameters (Faugeras, 1993a):

$$\begin{aligned}
&t_z = q_{34}\\
&\mathbf{r}_3 = \mathbf{q}_3^{\mathrm{T}}\\
&u_c = \mathbf{q}_1^{\mathrm{T}}\mathbf{q}_3,\ v_c = \mathbf{q}_2^{\mathrm{T}}\mathbf{q}_3\\
&\alpha_u = \sqrt{\mathbf{q}_1^{\mathrm{T}}\mathbf{q}_1 - u_c^2},\ \alpha_v = \sqrt{\mathbf{q}_2^{\mathrm{T}}\mathbf{q}_2 - v_c^2}\\
&\mathbf{r}_1 = (\mathbf{q}_1^{\mathrm{T}} - u_c\mathbf{q}_3^{\mathrm{T}})/\alpha_u\\
&\mathbf{r}_2 = (\mathbf{q}_2^{\mathrm{T}} - v_c\mathbf{q}_3^{\mathrm{T}})/\alpha_v\\
&t_x = (q_{14} - u_c t_z)/\alpha_u\\
&t_y = (q_{24} - v_c t_z)/\alpha_v
\end{aligned} \tag{A.13}$$

An algorithm is developed to compute the calibration matrix **C** and the intrinsic and extrinsic parameters based on some known points. Known points mean their 3*D* positions $M_i(X_i, Y_i, Z_i)$ in the world co-ordinate system and their corresponding 2*D* image

positions $m_i(u_i, v_i)$ in the pixel co-ordinate system are known. As discussed above, at least six such points are needed because there are 11 unknown variables in **C**.

To locate the required points, a calibration pattern of known geometry is made, as illustrated in Figure A.2, which consists of two planar grids of black squares on a white background. The two planes are perpendicular to each other. The origin of the 3*D* world co-ordinate system in this frame is assumed at the intersection *O* of the two lines connecting the bottom sides of last row squares in each plane. The world co-ordinate system along with the axis directions at *X, Y and Z* is also shown in Figure A.2. For simplicity and accuracy, the corners of the squares are considered as the points of interest. Their 3*D* positions are to be measured with reference to this co-ordinate system. In Figure A.2, the corner points are indexed by the paralleled row and column edges. For example, $M_{cg}$ denotes the intersection point of row *c* and column *g*.

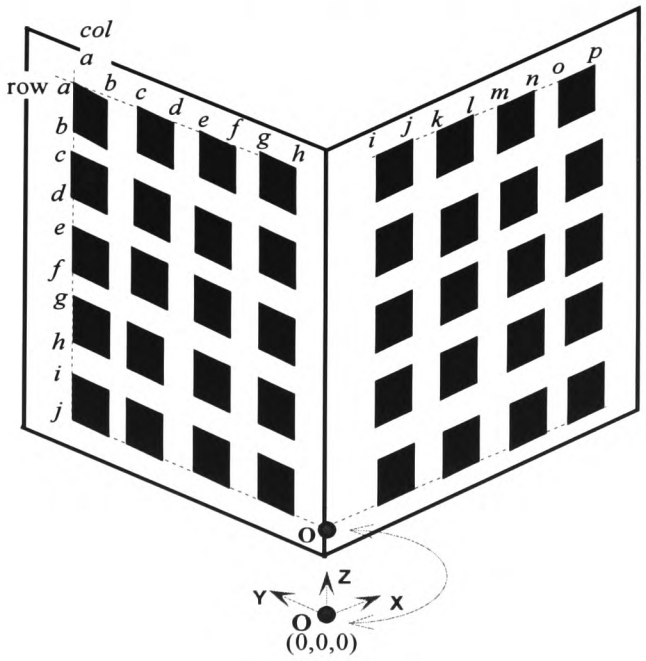



**Figure A.2 Illustration of a calibration pattern frame and the world co-ordinate system**

After the image of such a frame is taken, the 2D image positions, with reference to the origin of the image co-ordinate system at the top left corner of the image, are obtained by corner detection algorithm, which will be discussed later.

A model that was made for the project is shown in Figure A.3 and called *Pattern 1*.

## A.4 Algorithm Description

The algorithm for calibration matrix computation is described below. The process was implemented using MATLAB and the Image Processing Toolbox

1) Carefully measure the 3*D* co-ordinates of the corners in the calibration frame. As illustrated in Figure A.2, all the corner points of *Pattern 1* from $M_{aa}$ to $M_{jp}$ have been measured.

2) Take an image of *Pattern 1* and save it with size 256*256, see Figure A.3.

3) Detect the corner positions of the image by the corner extraction algorithm:

- Clear the background: to keep only the black grids in the white background, see Figure A.4

- Obtaining edges: apply Canny edge detection (Canny, 1986) algorithm to extract the edges of the squares, see Figure A.5.

- Getting corners:

As the feature of the image is very simple, there is no need to use complicated corner detection algorithm. In this thesis, corners are simply obtained by two intersecting edges of each square. Figure A.6 shows the image with the marked corners from $m_{aa}$ to $m_{hp}$.

4) Select a group of points (at least six points) at a time. Take their world and image co-ordinations as the inputs. Solve the Equation (A.9) using (A.11) for **C**.

5) Compute the intrinsic and extrinsic parameters from **C**.

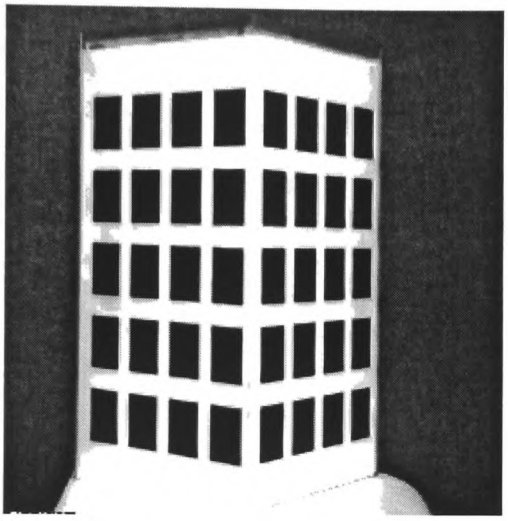


**Figure A.3 A calibration pattern:**

***Pattern 1***

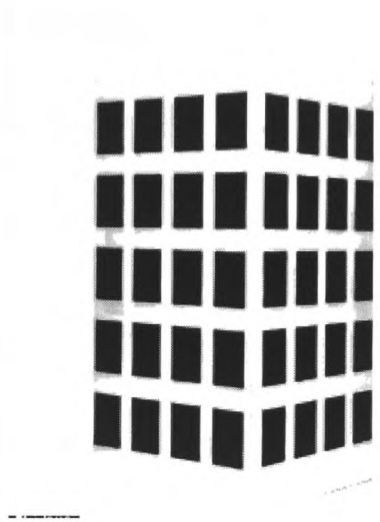


**Figure A.4 Clear pattern**

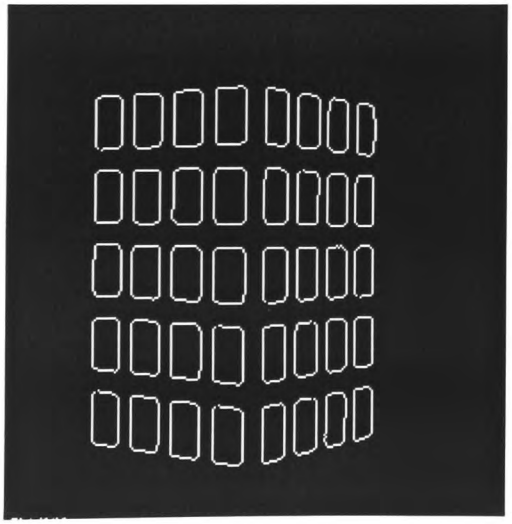


**Figure A.5 Getting edges**

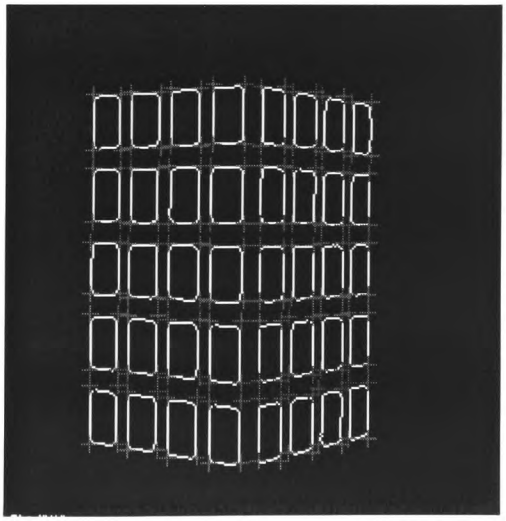


**Figure A.6 Getting corners**

## A.5 Results

Twenty points are randomly selected from the 8 × 16 = 128 corners each time. Five such groups are formed by the following components:

Group 1: $M_{aa}$; $M_{ac}$; $M_{bf}$,$M_{ci}$; $M_{co}$; $M_{ee}$; $M_{em}$; $M_{fd}$; $M_{gg}$; $M_{gk}$;

Group 2: $M_{aa}$; $M_{ag}$; $M_{bc}$; $M_{ch}$; $M_{dl}$; $M_{ee}$; $M_{eo}$; $M_{gc}$; $M_{fk}$; $M_{gm}$;

Group 3: $M_{ac}$; $M_{ak}$; $M_{bo}$; $M_{ce}$; $M_{ch}$; $M_{di}$; $M_{eg}$; $M_{fk}$; $M_{gm}$; $M_{ho}$;

Group 4: $M_{bb}$; $M_{be}$; $M_{cg}$; $M_{dp}$; $M_{ea}$; $M_{eh}$; $M_{fe}$; $M_{gg}$; $M_{gm}$; $M_{ha}$;

Group 5: $M_{ag}$; $M_{am}$; $M_{cc}$; $M_{df}$; $M_{dk}$; $M_{eb}$; $M_{eh}$; $M_{fc}$; $M_{he}$; $M_{hp}$;

Calibration matrix **C** is computed respectively for each group of data using equation (A.11), the result of which, denoted by **C1** to **C5** is displayed in Table A.1. The average value (**Cmean**) and the variance (**Cvariance**) from the set of the result are also computed and shown in the table. It can be seen that the computational result of **C** using this linear method is very much stable.

**Table A. 1 Computational result of calibration matrix C**

| | C1 | C2 | C3 | C4 | C5 | Cmean | Cvariance |
|---|---|---|---|---|---|---|---|
| $q_{11}$ | 0.7556 | 0.7421 | 0.7141 | 0.7007 | 0.7192 | 0.7264 | 0.0222 |
| $q_{12}$ | -0.7038 | -0.7061 | -0.7106 | -0.7184 | -0.7065 | -0.7091 | 0.0058 |
| $q_{13}$ | -0.0003 | 0.0015 | 0.0011 | 0.0044 | -0.0021 | 0.0009 | 0.0024 |
| $q_{14}$ | 158.8639 | 158.7368 | 159.2188 | 158.8099 | 157.9852 | 158.7229 | 0.4522 |
| $q_{21}$ | 0.2135 | 0.2125 | 0.1910 | 0.1904 | 0.1944 | 0.2004 | 0.0116 |
| $q_{22}$ | 0.0757 | 0.0749 | 0.0691 | 0.0628 | 0.0597 | 0.0684 | 0.0071 |
| $q_{23}$ | -1.3679 | -1.3571 | -1.3570 | -1.3412 | -1.3560 | -1.3558 | 0.0095 |
| $q_{24}$ | 234.1155 | 232.6446 | 233.2626 | 231.0087 | 232.7802 | 232.7623 | 1.1370 |
| $q_{31}$ | 0.0020 | 0.0019 | 0.0018 | 0.0018 | 0.0018 | 0.0019 | 0.0001 |
| $q_{32}$ | 0.0010 | 0.0009 | 0.0009 | 0.0008 | 0.0008 | 0.0009 | 0.0001 |
| $q_{33}$ | -0.0001 | -0.0000 | -0.0001 | -0.0001 | -0.0001 | -0.0001 | 0.0000 |

**Cmean** in Table A.1 is taken as the result of calibration matrix,

$$\mathbf{C} = \begin{bmatrix} 0.7264 & -0.7091 & 0.0009 & 158.7229 \\ 0.2004 & 0.0684 & -1.3558 & 232.7623 \\ 0.0019 & 0.0009 & -0.0001 & 1.0000 \end{bmatrix}$$

Intrinsic and extrinsic calibration parameters are computed using equation (A.13). The results are:

$(u_c, v_c) =$ (7.5790e-004, 5.1988e-004);

$(\alpha_u, \alpha_v) =$ (1.0151, 1.3723);

$$\boldsymbol{R} = \begin{bmatrix} 0.7156 & -0.6985 & 0.0009 \\ 0.1460 & 0.0499 & -0.9880 \\ 0.0019 & 0.0009 & 0.0001 \end{bmatrix};$$

$$\boldsymbol{T} = \begin{bmatrix} 0.6896 \\ -0.5182 \\ 1.0000 \end{bmatrix}.$$

## A.6 Test and Verification

Another algorithm is developed to test the accuracy of the calibration results. The test is performed by comparing two sets of 2D corner co-ordinates:

- initial image co-ordinates, $\{m_{xy} \mid xy = aa \sim jp\}$, of the pattern corners computed by the corner detection method described in section A.4.

- predicated image co-ordinates, $\{m'_{xy} \mid xy = aa \sim jp\}$, of the corners using the above obtained calibration matrix **C** according to equation (A.3), which takes the calibration matrix **C** and the 3D point co-ordinates as the input and gives 2D point co-ordinates as output.

Instead of displaying large number of the co-ordinate data here, Figure A.7 shows the two sets of points using * for the detected and + for the predicated. Intuitively, the two sets of data are very much close. The difference between them, $Diff = \{m_{xy} - m'_{xy} \mid xy = aa \sim jp\}$, is also computed point by point. The maximum, minimum and average differences are as follows:

$MaxDiff = 4.2807$, where $MaxDiff = max\ \{Diff\}$;

MinDiff = 0.0691, where $MinDiff = min\ \{Diff\}$;

MeanDiff = 0.7065, where *MeanDiff* = *mean* {*Diff*}.

It can be seen that the maximum error is no more than 5 pixels and the average difference is less than 1 pixel. Actually, this is the worst case with the maximum error compared with other computations using other data sets or other patterns. Apart from this case, three more cases are considered in order to make the comparison. The second case uses the above computed C1 as the calibration matrix for pattern 1. Only the top left corner point of each square is computed for comparison and the result is shown in Figure A.8. The other two patterns were also used for comparison purpose. The comparative results are shown in Figure A.9 and Figure A.10, respectively. The maximum, minimum and average differences in the above four cases are summarised in Table A.2.

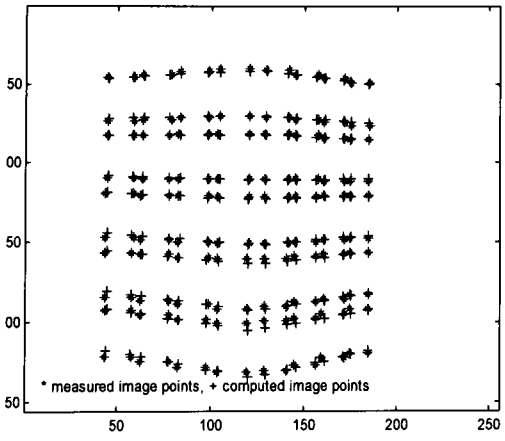



**Figure A.7 Comparison: case 1**

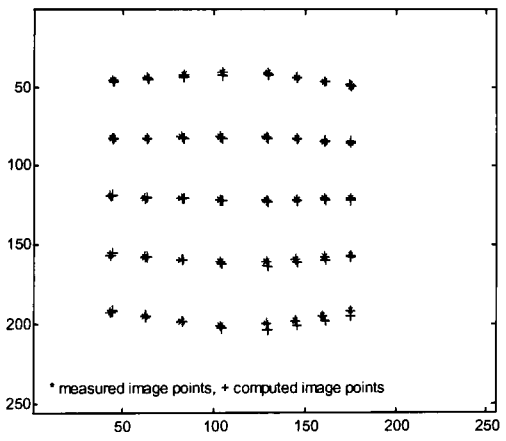



**Figure A.8 Comparison: case 2**

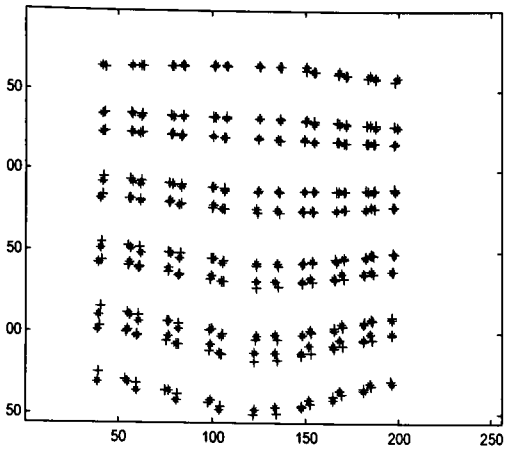


**Figure A.9 Comparison: case 3**

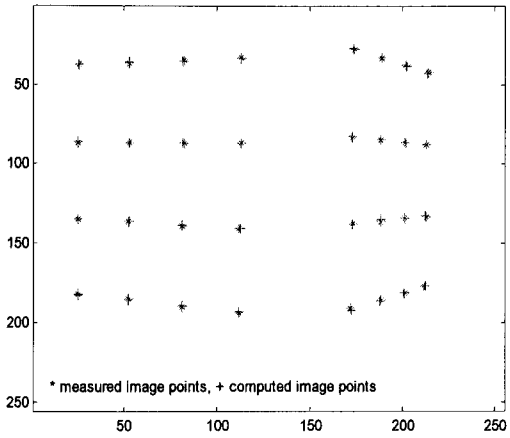



**Figure A.10 Comparison: case 4**

**Table A.2 Calibration matrix and image error data at different cases**

| Case | Calibration matrix | MaxDiff | MinDiff | MeanDiff |
|---|---|---|---|---|
| Case 2<br>*Pattern 1* | $\begin{bmatrix} 0.7556 & -0.7038 & 0.0003 & 158.8639 \\ 0.2135 & 0.0757 & -1.3679 & 234.1155 \\ 0.0020 & 0.0010 & -0.0001 & 1 \end{bmatrix}$ | 1.6416 | 0.0190 | 0.4567 |
| Case 3:<br>*Pattern 2* | $\begin{bmatrix} 0.4374 & -0.3818 & 0.0072 & 123.5388 \\ 0.0794 & 0.0528 & -0.8202 & 258.6395 \\ 0.0007 & 0.0007 & 0.0000 & 1 \end{bmatrix}$ | 4.3478 | 0.0251 | 1.1225 |
| Case 4:<br>*Pattern 3* | $\begin{bmatrix} 0.3626 & -0.2989 & 0.0149 & 126.7589 \\ 0.0569 & 0.0257 & -0.6569 & 284.3723 \\ 0.0006 & 0.0005 & 0.0000 & 1 \end{bmatrix}$ | 3.7127 | 0.0206 | 1.3308 |

Due to the principle of the calibration approach discussed in this thesis, the accuracy of the results depends on the accuracy of the raw data of object points. Therefore an accurate pattern and good corner detection algorithm are very important to the calibration.

## A.7 References

Canny, J. 1986. A Computational Approach to Edge Detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **8** (6), pp. 679-698.

Faugeras, O. 1993. *Three Dimensional Computer Vision: a Geometric Viewpoint.* UK: The MIT Press.

Trucco, E. and Verri, A. 1998. *Introductory Techniques for 3-D Computer Vision.* New Jersey: Prentice Hall.

# Appendix B Mathematical Explanation

## B.1 Hilbert Space

Hilbert space (Teolis, 1998) is the main space of interest in this thesis. Mathematically, a Hilbert space $\mathcal{H}$ is a normed vector space that has an inner product.

A norm $\|\cdot\|$ is a nonnegative real number that assigns to any vector $f \in \mathcal{H}$. A norm satisfies the following properties:

- Nonnegativity:

$$\|f\| \geq 0 \quad and \quad \|f\| = 0 \Leftrightarrow f = 0, \tag{B.1}$$

- Scaling:

$$\forall \lambda \in \mathbb{R}, \ \|\lambda f\| = |\lambda| \cdot \|f\|, \tag{B.2}$$

- Triangle inequality:

$$\forall f, g \in \mathcal{H}, \ \|f + g\| \leq \|f\| + \|g\| \tag{B.3}$$

An inner product $\langle \cdot, \cdot \rangle$ of two vectors $f, g \in \mathcal{H}$, must satisfy the conditions:

1) Linearity:

$$\forall \lambda_1, \lambda_2 \in \mathbb{R}, < \lambda_1 f_1 + \lambda_2, g >= \lambda_1 < f_1, g > + \lambda_2 < f_2, g > \tag{B.4}$$

2) Symmetry:

$$< f, g >= < g, f >^* \tag{B.5}$$

3) Self inner product:

$$< f, f > = \|f\|^2 \tag{B.6}$$

If $< f, g >= 0$, then $f$ and $g$ are assumed to be orthogonal.

As a consequence of these properties, an inner product in $\mathcal{H}$ satisfies the Cauchy-Schwarz inequality:

$$|< f, g >| \leq \|f\| \cdot \|g\|, \tag{B.7}$$

which is an equality if and only if $f$ and $g$ are linearly dependent.

An inner product of discrete signals $f[n]$ and $g[n]$ can therefore be defined by

$$< f, g >= \sum_{n=-\infty}^{\infty} f[n] g^*[n] \tag{B.8}$$

## B.2 Frames

A family of functions $\{\varphi_n\}_{n \in \mathbb{N}}$ in a Hilbert space $\mathcal{H}$ is called a frame if there exist two constants $A > 0$ and $B < \infty$ so that for all $f \in \mathcal{H}$,

$$A\|f\|^2 \le \sum_j |< f,\varphi_j >|^2 \le B\|f\|^2 \tag{B.9}$$

$A$ and $B$ are called the frame bounds.

If $A = B$, the frame is considered to be tight. In a tight frame, the following equation holds:

$$\sum_j |< f,\varphi_j >|^2 = A\|f\|^2 \tag{B.10}$$

which yields

$$f = A^{-1}\sum_j < f,\varphi_j > \varphi_j \tag{B.11}$$

However, usually frames, even tight frames, are not orthogonal bases. A frame constitutes an orthogonal basis if and only if $A = B = 1$. If $A > 1$, the frame is redundant. $A$ gives the redundancy ratio and is called the minimum redundancy factor.

Equation (B.11) gives a way to recover $f$ from the inner vector $< f,\varphi_j >$, if the frame is tight. For general frames, a frame operator $\boldsymbol{F}$ is defined as a linear operator so that:

$$(\boldsymbol{F}f)_j =< f,\varphi_j > \tag{B.12}$$

Daubechies (Daubechies, 1992) has proved that $\boldsymbol{F}^*\boldsymbol{F}$, where $\boldsymbol{F}^*$ is the adjoint of $\boldsymbol{F}$, is invertible.

## B.3 References

Daubechies, I. 1992. *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics.

Teolis, A. 1998. *Computational Signal Processing with Wavelets.* Birkhauser Boston.

# Appendix C Published Papers

# From Coarse to Fine Strategy to Wavelet Multiresolution Analysis

Fangmin Shi, Neil Rothwell Hughes and Geoff Roberts
*Mechatronics Research Centre*
*University of Wales College, Newport*
*Allt-Yr-Yn Campus*
*PO Box 180*
*Newport NP20 5XR, UK*
*fangmin.shi@newport.ac.uk*

## Abstract

*Coarse to fine searching is a very popular strategy in computer vision. Conventionally, it first pre-processes images at multiple scales with differential Laplacian operators and then applies algorithms from coarse to fine scales. The results at coarser scales can guide the search at finer scales. However, image wavelet multiresolution Analysis represents and analyses images by choosing wavelets that possess good local properties. It is an ideal extension of the conventional coarse to fine strategy. This paper briefly reviews the techniques for the two cases and points out the future directions for wavelet-based stereo matching research.*

***Keywords:*** *computer vision, coarse to fine strategy, stereo matching, wavelets*

## 1. Introduction

Coarse to fine strategy has been widely used in computer vision research since the 1970s. Stereo matching [11] and edge detection [2] are early examples of the use of the approach. Images are processed in advance at different scales and then the search or detection process starts at a coarse scale. The rough results are used to guide the hierarchical operations at finer scales. This strategy not only speeds up the matching process but also provides a better solution to the false-target problem [1]. Gaussian or Laplacian pyramids [9] are normally employed for algorithm implementation.

It was in 1988 when Stephane Mallat created a complete wavelet multiresolution theory [5] by combining wavelets and the multiscale computer vision method, that research on hierarchical vision has developed into a new stage. Various wavelet functions have been constructed and applied to image decompositions [13]. Some matching constraints based on wavelets have also been put forward [15].

This paper generally reviews the techniques for conventional coarse to fine strategy and wavelet multiresolution analysis. At the end of the paper, a brief survey of current research on wavelet-based stereo matching is also presented.

## 2. Conventional Coarse to Fine Strategy

There are many problems to be solved in computer vision, for example vision modelling, the matching problem, motion analysis etc. Due to its fundamental importance, stereo matching is used as the main example for illustrating the topics covered in this paper.

Stereo matching is the problem of locating two corresponding points in two image planes which are the projections of the same physical point in 3D space. In stereo matching, the different matching primitives as well as the matching strategies determine the various approaches. Generally speaking, the matching solutions can be classified into three categories: area-based, feature-based and phase-based methods [9]. The area-based method compares directly the intensity values within small image patches of the left and right view, and tries to maximise the correlation between these patches. It is however difficult to determine an appropriate size of patch and to avoid expensive computational iterations. In the feature-based method, the features such as edge elements, corners or line segments need to be extracted first by feature detectors and then the matching process is applied to the attributes associated with the detected features. Its drawback is the output of a sparse depth map. Recently a third approach known as phase-based matching has been developed by matching the Fourier phase information of two images. In the process of the above algorithm development, a coarse to fine searching strategy has been widely used [11], [4], [9], [12], [14].

The conventional coarse to fine method works by applying a smoothing filter to the original images first. This is because vision problems like stereo matching or edge detection actually belong to the class of optimisation problems. It is much easier to locate the extrema of a smoothed version of the image than its original, which can then give a good starting point to locate the extrema of the original image.

In order to smooth images and detect intensity changes, a good filter should be used. Marr and Hildreth's work [9] shows that the most satisfactory operator is the filter $\nabla^2 G$ , where $\nabla^2$ is the Laplacian operator $(\partial^2/\partial x^2 + \partial^2/\partial y^2)$ and G denotes the two-dimensional Gaussian distribution $G(x,y) = e^{-\frac{x^2+y^2}{2\pi\sigma^2}}$ , $\sigma$ is the standard deviation. Then $\nabla^2 G$ is:

$$\nabla^2 G(x,y) = \frac{-1}{\pi\sigma^4}(1-\frac{x^2+y^2}{2\sigma^2})e^{\frac{-x^2-y^2}{2\sigma^2}} \qquad (1)$$

There are some advantages of choosing $\nabla^2 G$ [9]. Firstly, different $\sigma$ forming large or small filters enable the smoothing of an image at multiple scales. Secondly, the Gaussian part of the expression effectively removes all structures at scales much smaller than the space constant $\sigma$ of the Gaussian. Thirdly, the derivative part significantly reduces the amount of computation.

To detect the intensity extrema, filtering an image with the operator $\nabla^2 G$ at different scales, reflected by the space constant $\sigma$ of the Gaussian, needs to be done first. And then some algorithm is applied to the filtered images. Using this way, Mallat [6] detected zero-crossings and Canny [2] also put forward an effective edge detection approach.

In stereo matching, the search process starts at a coarse scale and the roughly matched results are used to guide searching at finer scales. It works as follows:

- generating a pair of image pyramids from the original image pairs so that only few and prominent features are present at the coarse levels. The original images are at the finest level of the image pyramids.
- starting the matching process at the coarsest level.
- using the matches obtained at the coarser level to guide the matching process gradually up to the finest level.

In 1988 Mallat proposed a complete theory for wavelet multiresolution analysis [5] by combining wavelets and the multiscale computer vision method. This theory provides a good hierarchical image decomposition method which facilitates improvements in coarse to fine strategy.

## 3. Wavelet Multiresolution Analysis

The multiresolution approach is used to decompose a signal at different resolutions. It produces a series of hierarchically organised decompositions. This section briefly introduces the Mallat's theory [5]. For simplicity, the basic principle of Multiresolution signal decomposition is described in terms of the one-dimensional case. The extension to two-dimensional case can be straightforwardly applied to images. The

Let φjk(x) be the orthogonal wavelet bases which are the dilations and translations of a function φ(x), then

$$\varphi_{jk}(x) = 2^{-j/2}\varphi(2^{-j}x - k) \qquad (2)$$

the space $V_j = span\{\varphi_{jk}(x) \mid (j,k \in \mathbf{Z})\}$ corresponds to the decompositions at different resolutions, then the wavelet transform of a function f(x) is defined as:

**Error! Objects cannot be created from editing field codes.** (3)

where $c_{jk} = \langle f(x) \bullet \varphi_{jk}(x) \rangle$ , $\bullet$ denotes the inner product. Equation (2) can also be written as the convolution form:

$$W_\varphi f = f(x) * \varphi_{jk}(x) \qquad (4)$$

The multiresolution analysis projects a signal into a set of subspaces, each of which includes a scaling space Vj and a wavelet space Wj at each level. Each Vj is contained in the next Vj+1, $V_1 \subset V_2 \subset \cdots \subset V_j \subset V_{j+1} \subset \cdots V_n$ . The projection on Vj gives the signal its identity. Whereas the wavelet space describes the difference between Vj+1 and Vj, $W_j = V_{j+1} - V_j$ , and Wj is orthogonal to Vj. For a given scale space Vj+1, the following expression holds:

$$V_{j+1} = V_0 + \sum_{i=0}^{j} W_i \qquad (5)$$

Let Aj denotes the approximation part of the projection on Vj and Dj denotes the detail of the projection on Wj. From the first level, a signal can be decomposed into:

S = A1+D1
= (A2+ D2)+D1
= (A3+D3)+D2+D1
= (A4+D4)+D3+D2+D1=.....

where $A_j = \sum_k c_{jk}\varphi_{jk}(x)$ , $c_{jk} = \langle f(x) \bullet \varphi_{jk}(x) \rangle$ , $\varphi_{jk}(x)$ is a scaling function and $D_j = \sum_k d_{jk}\phi_{jk}(x)$ , $d_{jk} = \langle f(x) \bullet \phi_{jk}(x) \rangle$ , $\phi_{jk}(x)$ is a wavelet function. Both $\varphi_{jk}(x)$ and $\phi_{jk}(x)$ possess the form of Equation (2).

Two-dimensional image decomposition can be constructed by the extension of one-dimensional case. The different resolution spaces are:

$V_j = span\{\varphi_{jk}(x)\varphi_{jm}(y) \mid (k,m \in \mathbf{Z})\}$. That is, the family of functions

$$\Phi_{jkm}(x,y) = \varphi_{jk}(x)\varphi_{jm}(y) = 2^{-j/2}\varphi(2^{-j}x-k)\varphi(2^{-j}y-m) \quad (6)$$

forms an orthonormal basis of Vj. Mallat [5] proved that the difference between Aj+1 and Aj is given by three detail parts represented by the following three wavelets:

$$\Psi^1_{jkm}(x,y) = \varphi_{jk}(x)\phi_{jm}(y),$$
$$\Psi^2_{jkm}(x,y) = \phi_{jk}(x)\varphi_{jm}(y),$$
$$\Psi^3_{jkm}(x,y) = \phi_{jk}(x)\phi_{jm}(y) \quad (7)$$

then

$$A_j = f(x,y) * \Phi_{jkm}(x,y)$$
$$D^1_j = f(x,y) * \Phi^1_{jkm}(x,y)$$
$$D^2_j = f(x,y) * \Phi^2_{jkm}(x,y)$$
$$D^3_j = f(x,y) * \Phi^3_{jkm}(x,y) \quad (8)$$

An example of Mallat image wavelet decomposition is given below. Figure 1 is an original image. Its decomposition results using wavelet bior1.5 [13] are displayed in Figure 2.

It can be seen that the wavelet multiresolution representation of the image is organised according to a coarse to fine hierarchy.

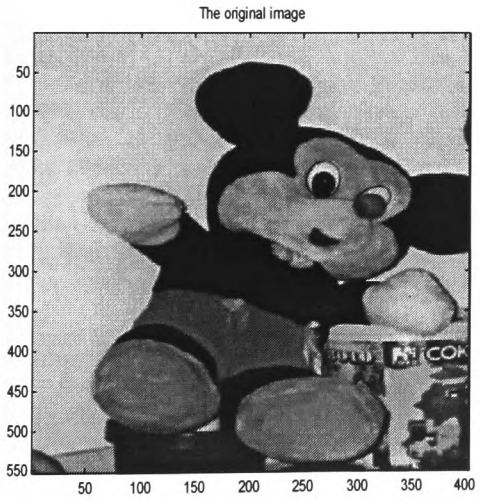



Figure 1. The Original Image

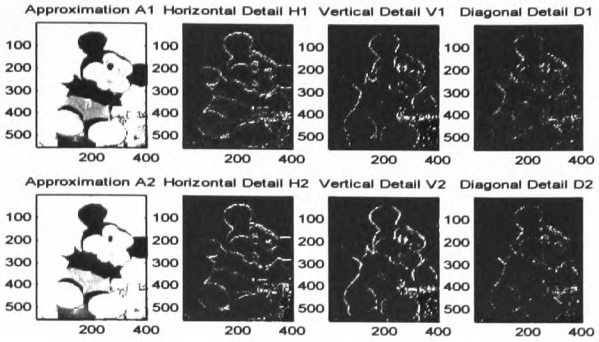



Figure 2. Wavelet (Bior1.5) Decompositions at Two Resolutions

## 4. Wavelets for Image Matching

Wavelet multiresolution Analysis is a good mathematical tool for image analysis, which can provide useful information at different resolutions. It has found many applications so far such as image compression, image noise reduction, feature detection etc. However, very little work has been reported in the field of wavelet-based stereo matching. After he created the theory for general signal wavelet decomposition, in [6], Mallat developed this theory by applying zero-crossings of wavelet transforms leading to a coarse to fine stereo matching algorithm based on the new representation. Some other work can be found on the application of orthogonal wavelet transforms to the stereo matching problem in [17], [10], [16].

There are many aspects to this field worthy of further study. The first issue is to choose appropriate wavelets with the view of extracting useful information for matching. Although Daubechies [3] created the important family of Daubechies wavelets, other wavelets have also emerged [8] and the procedure of how to establish wavelets is described in [5]. The second problem is that of selecting appropriate matching primitives in order to use coarse to fine matching strategy. Many research studies into general correlation- and feature-based matching have been reported. References on the phase-based matching using windowed Fourier analysis can also be found, but very few wavelet-based matching methods are developed. For example, Sanger [12] proposed a solution to stereo matching using the Gabor filter, in which disparities are measured directly as a function of image properties and therefore a dense depth map is given. Weng [14] introduced the Windowed Fourier Phase (WFP) as the primary matching primitive to stereo matching. He proved that the WFP is quasi-linear and dense and made the image zero-crossings and peaks correspond to phase values. This algorithm conveys dense disparity information.

## 5. Conclusion

As the end of this paper, the Mallat's remark in [7] is repeated here. That is, the wavelet mathematical theory is reaching a mature stage but how to make use of this multiscale information for information processing is not always clear.

Motivated by the previous work on phase- and wavelet-based techniques, work is currently being undertaken by authors with the aim of providing a wavelet-based stereo matching approach to produce dense disparity maps with few iterations.

## 6. References

[1] Barnard, S. T. Stochastic Stereo Matching over Scale. International Journal of Computer Vision, (3), pp. 17-32, 1986.

[2] Canny, J. A Computational Approach to Edge Detection. IEEE Trans. on Pattern Anal. and Machine Intell., 8 (6), pp. 679-698, 1986.

[3] Daubechies, I. Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics. 1992.

[4] Grimson, W. A Computer Implementation of a Theory of Human Stereo Vision. Phil. Trans. Royal Soc. London, B292, pp. 217-253, 1981.

[5] Mallat, S. A Theory for Multiresolution Signal Decomposition: the Wavelet Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (7), pp. 674-693, 1989.

[6] ---. Zeros-crossings of a Wavelet Transform. IEEE Transactions on Information Theory, 37 (4), pp. 1019-1033, 1991.

[7] ---. Wavelets for a Vision. Proceedings of the IEEE, 84 (4), pp. 604-614, 1996.

[8] ---. A Wavelet Tour of Signal Processing. Academic Press. 1998.

[9] Marr, D. Vision. New York: W. H. Freeman and Company. 1982.

[10] Pan, H. P. General Stereo Image Matching Using Symmetric Complex Wavelets. SPIE Proceedings. 1996.

[11] Rosenfeld, A., & Thurston, M. Coarse-fine Template Matching. IEEE Trans. Syst., Man, Cybern, 7, pp. 104-107, 1977.

[12] Sanger, T. D. Stereo Disparity Computations Using Gabor Filter. Biol. Cybern., 59, pp. 405-418, 1988.

[13] Strang, G., & Nguyen, T. Wavelets and Filter Banks. Wellesley-Cambridge Press. 1996.

[14] Weng, J. Image Matching Using the Windowed Fourier Phase. Int. J. of Computer Vision, 3, pp. 211-236, 1993.

[15] Zhong, S.,Shi, Q. Y., & Cheng, M. A Stereo Matching Method Based on Wavelet Transform. Pattern Recognition and Artificial Intelligence, 7 (1), pp. 27-33, 1994.

[16] Zhou, J.,Peng, J. X, & Ding, M. Y. Image Matching Based on Wavelet Features. Pattern Recognition and Artificial Intelligence, 9 (2), pp. 125-129, 1996.

[17] Zhou, X., & Dorrer, E. Automatic Image Matching Algorithm Based on Wavelet Decomposition. IAPRS. pp. 951-960, 1994.

# WAVELET TRANSFORMS FOR STEREO VISION

**Fangmin Shi, Neil Rothwell Hughes and Geoff Roberts**

Mechatronics Research Centre
University of Wales College, Newport
Allt-yr-yn Campus, P.O. Box 180
Newport NP20 5XR, United Kingdom

email: fangmin.shi@newport.ac.uk
neil.rothwell-hughes@newport.ac.uk
geoff.roberts@newport.ac.uk

## Abstract

*Wavelet Multi-resolution Analysis is an efficient tool for image processing. Due to its lack of shift-invariance, it is unstable with respect to translations of the original image. In practice, many vision tasks, especially stereo matching, require shift-invariant transforms. This paper reviews the different wavelet transforms focusing on their shiftability and information redundancy. Some strategies to achieve shift-invariance of the wavelet transform such as dyadic wavelet transform, zero crossings of wavelet transform and complex wavelet transform are presented. An application to image disparity computation is also given.*

## 1 Introduction

A wavelet is an oscillatory waveform that has finite duration. The wavelet transform decomposes a signal into shifted and scaled versions of the prototype (or mother) wavelet. It can be classified from many points of view for example, continuous or discrete, orthogonal or non-orthogonal, shift-invariant or shift-variant, and real or complex wavelet transforms. Shift-invariance and information redundancy are the main concern in this paper.

*Shift-invariance* (or *time-invariance*) means that if a signal is delayed in time, its transform result is delayed correspondingly. This property is essential for many vision tasks that are sensitive to the translation of spatial positions. Considering information redundancy, if a set of basis functions is orthogonal, then the signal representation with the orthogonal bases has no redundancy and perfect reconstruction is achievable. Such implementation is thus efficient. However, both shift-invariance and orthogonality cannot be achieved at the same time (Daubechies, 1992). A good transform should be approximately shift invariant with limited redundancy.

Section 2 discusses some existing wavelet transforms in terms of their shiftability and orthogonality. Section 3 presents the importance of shift-invariance to stereo matching since stereo vision deals with two stereo images which can be viewed as shifted versions

of each other. In section 4, three specially constructed wavelet approaches are discussed as good compromise solutions to stereo matching. Experimental results are given in section 5 and the section 6 is the conclusion and the future work.

# 2 Wavelet transforms

A family of wavelets $\varphi_{a,b}(x)$ is defined as dilations and translations of a mother wavelet $\varphi(x)$,

$$\varphi_{a,b}(x) = a^{-1/2}\varphi((x-b)/a) \tag{1}$$

where $a$ is the scale parameter, $b$ is the translation parameter, and $a, b \in R$. The wavelet Transform ($WT(b, a)$) represents a signal by such a family of wavelets:

$$WT(b,a) = \int f(x)a^{-1/2}\varphi((x-b)/a)dx \tag{2}$$

Some commonly used wavelet transforms and their shiftability and orthogonality are outlined as follows.

## 2.1 Continuous wavelet transform

Parameters $a$ and $b$ are continuous values in this case. $\{\varphi_{a,b}(x)\}$ constitutes a set of non-orthogonal over-complete bases (Teolls, 1998). The information is highly redundant with the continuous wavelet representation but it is shift invariant.

Let $f_s = f(x-s)$ be a shifted version of $f(x)$ by $s$, then its wavelet transform is:

$$WT_s(b,a) = \int f(x-S)a^{-1/2}\varphi((x-b)/a)dx = WT(b-s,a)$$

which shows that it is shift-invariant (Teolls, 1998).

## 2.2 Discrete wavelet transform

Parameters $a$ and $b$ are uniformly sampled in this case: $a = a_0^j$, $b = na_0^j b_0$, and n, $j \in Z$. Considering the assignment of $a_0$ and $b_0$, the shiftability and orthogonality of the wavelet transform depends on the choice of $a_0$ and $b_0$.

If the sampling interval $\tau = a_0^j b_0$ tends to be very small (minimum is zero), the discrete wavelet transform is close to the continuous wavelet transform above.

If the sampling interval $\tau$ is large relative to the rate of variation of the wavelet coefficients, then the transform may be not shift-invariant. This observation is particularly poignant for wavelet orthogonal bases where $a_0=2$ and $b_0=1$ (Mallat, 1998). This is a special discrete wavelet transform called Mallat's multiresolution analysis (MRA).

## 2.3 Wavelet multiresolution analysis

When $a_0$=2 and $b_0$=1, the wavelet family $\varphi_{nj}(x)=\frac{1}{\sqrt{2^j}}\varphi(\frac{x}{2^j}-n)$ constitute orthogonal bases. It can be hierarchically implemented by two channel filter banks (Mallat, 1989). The original signal is first decomposed into the first level approximation and detail parts by respectively passing through a low-pass and a high-pass filter. Downsampling (removing every other component of a sequence) by 2 follows each filter output. The approximation parts are hierarchically decomposed. One-level decomposition is illustrated in Figure 1. However, the downsampling is not shift-invariant (Strang and Nguyen, 1997).

MRA is such an efficient algorithm that it has led to the wide application of wavelets to many areas. Unfortunately, due to its shift-variance, it can not be directly used for tasks such as pattern recognition and computer vision that require shift-invariant transforms.

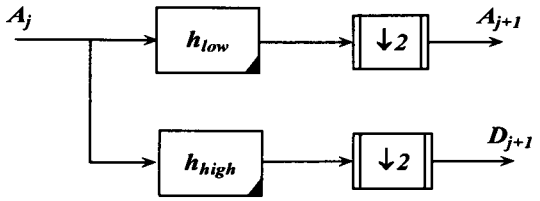



Figure 1 Illustration of one-level MRA

## 2.4 Dyadic wavelet transform

For the purpose of shift-invariance, the dyadic wavelet transform discretises the scale parameter $a$ along a dyadic sequence, e.g. $a=2^j$ ($j\in Z$), but the translation parameter $b$ remains continuous.

Mallat (Mallat, 1998) proved that if the Fourier transform, $\hat{\varphi}(2^j w)$, of the dyadic wavelets satisfies

$$A \le \sum \left| \hat{\varphi}(2^j w) \right|^2 \le B$$

for two constants $0 \le A \le B$, then the transform with such dyadic wavelet bases defines a complete and stable representation. However, due to the continuous translation parameter, the information is still redundant with the dyadic wavelet transform. However, the algorithm efficiency is greatly improved compared with the continuous wavelet transform.

## 2.5 Complex wavelet transform

All the above discussions are based on the real wavelets. If a mother wavelet is a complex function, then a complex wavelet transform is created. Its complex bases are definitely not orthogonal. Its shift-invariance depends on how $a$ and $b$ are discretised.

# 3 Importance of wavelet shift-invariance to vision

It is important to apply shift-invariant transforms to stereo matching. Stereo matching deals with the most difficult problem in stereo vision, which aims to find the corresponding point between two images taken at the same time with two identical cameras. Disparity is used to measure the pixel shift of one point in one image to its corresponding point in another image.

The projective geometry (Faugeras, 1993) of stereo vision is illustrated in Figure 2. ***C*** and ***C'*** are two optical centres of two identical cameras. ***M*** is a 3D point in the physical world and its projective points in two image planes are denoted by ***m*** and ***m'***. Then *disparity* = (***u-u'***), in which ***u*** and ***u'*** are 2D image co-ordinates in the two image planes, respectively.

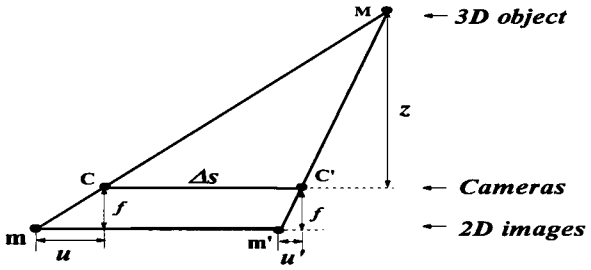



Figure 2 Geometry of stereo vision

To compute stereo disparity, intuitively, one image can be assumed to be the shifted version of the other. The shifted value with respect to each pixel is the disparity, which is dependent on the pixel position.

As a multiresolution analysis tool capable of highly efficient implementation, Mallat's MRA has found a lot of applications in the image processing domain such as image compression and noise reduction. However, as indicated in section 2, it lacks shift-invariance, which limits its application to a robust matching implementation.

As an example of the problem introduced by shift-variance, Figure 3 shows two epipolar lines (Faugeras, 1993) from two stereo images. Figure 4 displays the wavelet (bior3.7) decomposition results of 5 level approximations (*a1~a5*) and details (*d1~d5*) using MRA theory. Level 0 refers to the original signals. By comparing these decomposed plots, for instance, between left: d2 and right: d2, it can be seen that the details are not shifted according to the shifting of the original signals because the downsampled sequence depends on the parity of the shift value. This demonstrates that MRA is a shift-variant transform. Such transform does not therefore give robust results when applied to stereo matching. Approaches to shift-invariant wavelet transforms for matching are to be discussed in the next section.

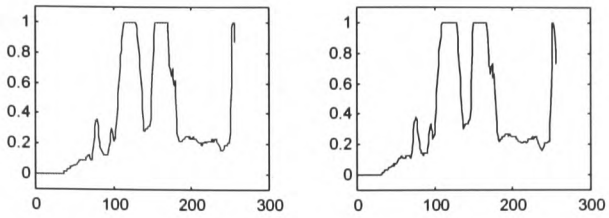


Figure 3. Scan lines from two stereo images

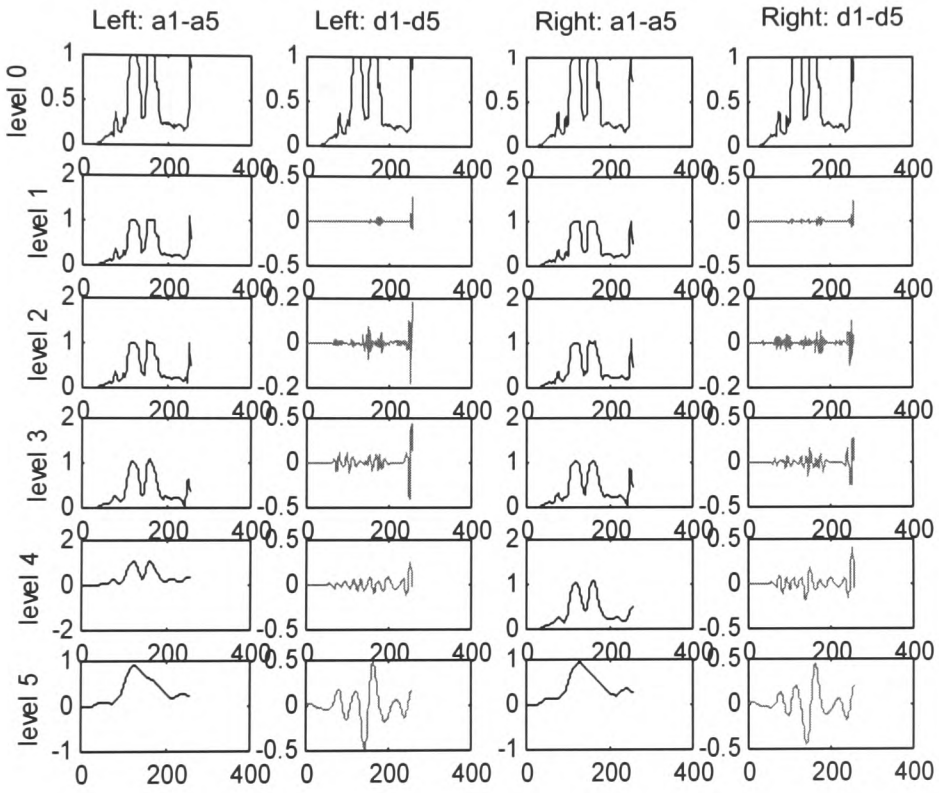



Figure 4. Wavelet 1D decompositions at 5 levels

# 4 Achieving shift-invariance for matching

There are some strategies to achieve shift-invariance of wavelet transforms. As discussed in section 2, the continuous and dyadic wavelet transform possess the property of shift-invariance. The discrete wavelet transform does not. In practice, any algorithm has to be implemented in the discrete form. Approximate shift-invariance with effective computation should be achieved. Mallat used the dyadic wavelet transform and zero-crossings of the dyadic wavelet transform (Mallat, 1991) to reduce the representation size. Simoncell (Simoncell *et al*, 1992) built steerable filters to achieve a shiftable transform that is jointly invariant in position, scale and orientation. More recently, a shift-invariant complex wavelet transform (Kingsbury, 1998) with perfect reconstruction has been constructed and applied to image processing and computer vision. This paper

will discuss the application of three wavelet approaches, dyadic wavelet transform, zero-crossings of a dyadic wavelet transform and complex wavelet transform, and present the result for the stereo matching using the dyadic wavelet transform.

## 4.1 Dyadic wavelet transform for matching

As discussed in section 2.4, the dyadic wavelet transform decomposes the signal along the dyadic scales leaving the shift parameter continuous. It is suitable to apply the sum of squared difference (SSD) (Trucco and Verri, 1998) to the transformed signals to measure the similarity of the corresponding points. The smaller the SSD value, the more likely it is that the points correspond to each other. Unlike the conventional SSD directly applied to the image intensity value, the SSD here is applied to the wavelet coefficients. In addition, the windowing problem with the conventional SSD is naturally solved by using the wavelet transform because the time-frequency area of a wavelet transform is constant for all scales. Wavelet analysis uses large windows for low-frequency components and small windows for high-frequency components.

Let $f_1(x)$ and $f_2(x)$ be two scan lines of stereo images, their dyadic wavelet transforms are denoted by $DWT1(2^n, x)$ and $DWT2(2^n, x)$, respectively, where $n \in N$. The SSD measure ($ssd(n,x)$) is defined as:

$$ssd(n,x) = \sum_{\tau=x-2^n\sigma}^{x+2^n\sigma} \left| DWT1(2^n,\tau) - DWT2(2^n,\tau) \right|^2 \qquad (3)$$

where $\sigma$ is the size of an interval where the energy of the mother wavelet is mostly concentrated (Mallat, 1991).

Disparity is computed for each scale first. Then the results at all scales are combined.

## 4.2 Zero crossings of wavelet transform

Zero-crossings of the filtered images were used for stereo matching by Marr (Marr, 1982). The algorithm implemented by Grimson (Grimson, 1981) shows that only the positions of the zero-crossings are detected and used for matching. Mallat (Mallat, 1991) pointed out that such zero-crossing representation is not stable and only complete under some restrictive assumptions. He also proved that the zero-crossing representation of a wavelet transform is stable and complete and can be used to perfectly reconstruct the original signal (Mallat, 1998).

In Mallat's wavelet transform zero-crossing representaion, not only are the positions of the zero-crossings detected, but also the boundary values that are the integral between the neighbouring zero-crossings are recorded. Let $z(n,m)$ denote the horizontal co-ordinate of $mth$ zero-crossing at scale $n$, the boundary $b(n,m)$ is defined as:

$$b(n,m) = \int_{z(n,m)}^{z(n,m+1)} DWT(n,x)dx \qquad (4)$$

Matching using this method is therefore reduced to matching between the zero-crossings of the wavelet transform.

To determine the corresponding zero-crossings, the SSD measure is used. The algorithm efficiency is highly improved compared with the full dyadic wavelet transform because the number of zero-crossings is much less than the original pixel number used in section 2.3. However, the disparity map is not as dense as that obtained by the dyadic wavelet transform. Interpolation is therefore needed to produce the final depth map.

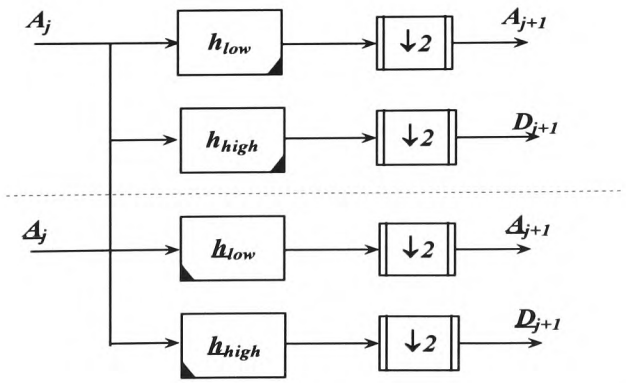
## 4.3 Dual-tree complex wavelet transform

The conventional solution to the problem of stereo matching is to apply constraints to determine which point in one image corresponds to a point in the second image. This is the so-called correspondence problem (Trucco and Verri, 1998). In 1988, Sanger (Sanger, 1988) first proposed a disparity computation method based on the Fourier shift theorem, which states the linear relationship between the signal's shift value in the time domain and its Fourier phase difference in the frequency domain. He employed the short time Fourier transform (STFT) to calculate the local shift value, which is assumed to be the disparity of two stereo signals, from the local phase difference. He called this the *correspondenceless* approach because the disparity results from direct calculation. Thus it has much potential as an efficient matching implementation with a dense depth map output.

The difficulty with phase-based disparity computation lies in the non-linearity of the local shift and local phase difference under STFT because the Fourier shift theorem does not hold for the STFT. It is proposed to introduce the complex wavelet transform (the normal wavelet transform is real) to compute disparity. It is believed that the multiresolutional decomposition structure of wavelet transform and carefully chosen complex filters can be applied to phase-based disparity determination.

To study motion estimation, Magarey and Kingsbury (Magarey and Kingsbury, 1995) developed a complex wavelet transform that provide approximate shift-invariance. In order to achieve perfect reconstruction and good frequency characteristics, the Dual–Tree Complex Wavelet Transform (DTCWT) (Kingsbury, 1998) has been proposed.

The implementation structure of DTCWT is similar to that of Mallat's MRA. But a dual-tree is added besides the conventional two channel filters. The delay of the two filters in the dual-tree is one sample offset respectively from those in the original tree.




At each level, the approximation parts are decomposed along two trees. Therefore, for a one dimensional signal, the DTCWT gives *2:1* redundancy ($2^m$*:1* for *m*-Dimensions). This makes the implementation quite efficient relative to the dyadic wavelet transform.

In (Kingsbury, 1998), it is claimed that the phases from the complex wavelet coefficients are approximately linearly to the shift of the signal. Therefore, the image disparity can be computed by measuring the phase difference between the corresponding wavelet coefficients.

# 5 Application to disparity map computation

This section gives an example of applying the dyadic wavelet transform to compute a dense disparity map. Initially, the commonly used random dot stereograms (Julesz, 1971) are constructed as test stereo pairs. Figure 5 shows a simple stereo pair, '*squares*', the first of which has three squares with different grey values and the central two squares right shifted by 4 and 8 pixels respectively forming the second image. Figure 6 shows the dyadic wavelet transforms with three scales of two scan lines (line 60). The image disparity map and the depth map are given in Figure 7.

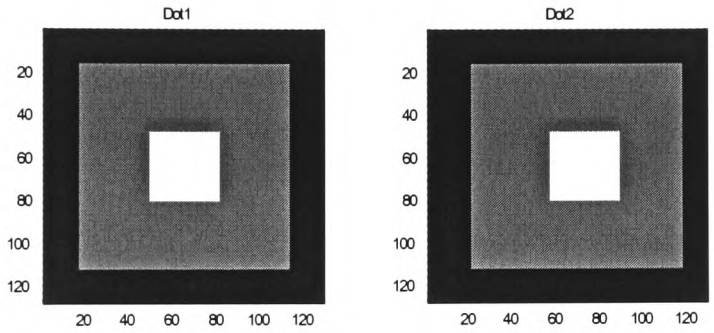



Figure 5 Stereo pair: squares

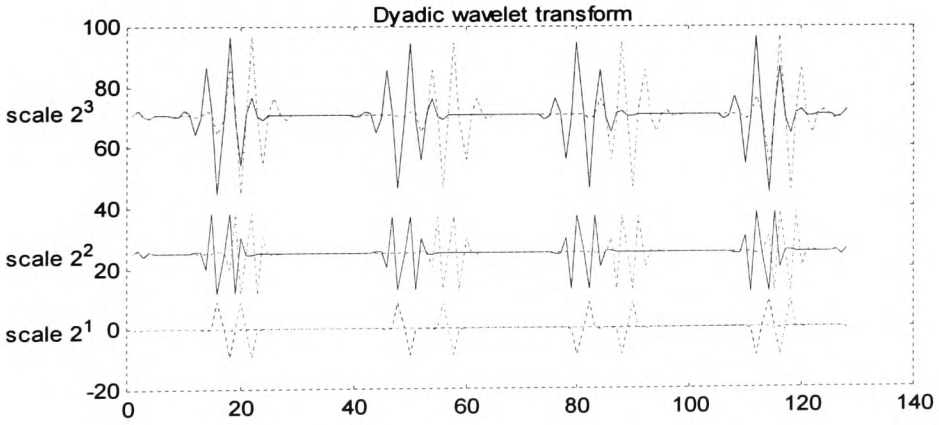



Figure 6 Dyadic wavelet decompositions, row 60 of the square images

Another random dot stereogram, '*Binary Squares*', is also constructed. In contrast with the *Squares* images, the pixel grey values in *Binary Squares* are either *0* or *1* and the density is *50%*. Figure 8 shows the synthesised stereo pair and the computational disparity result. A computation using real image pairs was also tested. This is shown in Figure 9.

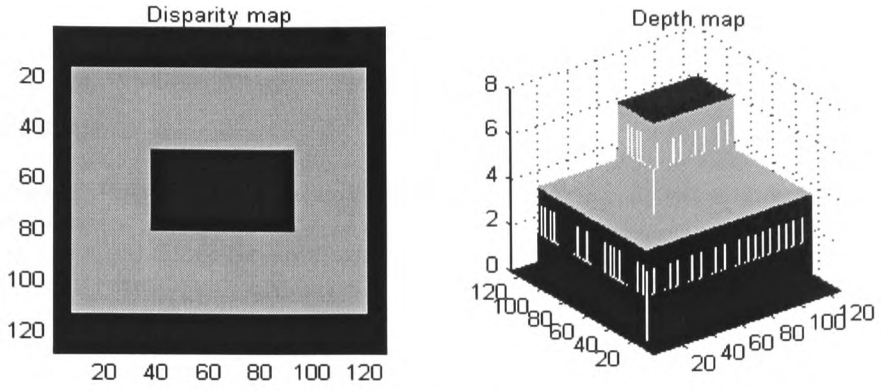



Figure 7 Disparity map and depth map: *Squares*

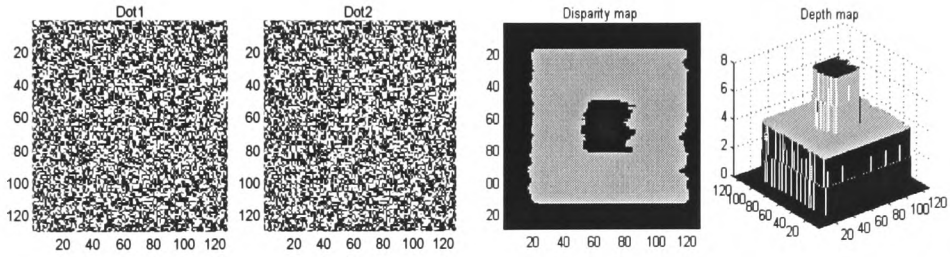



Figure 8 Stereo pair: *Binary Squares,* disparity map and depth map

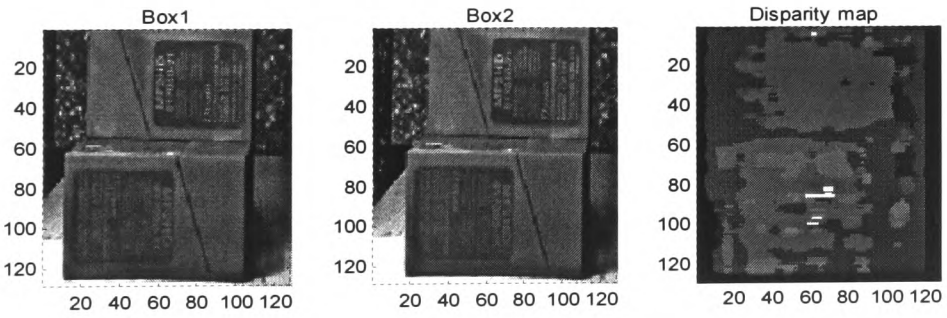



Figure 9 Real image stereo pair and the disparity map

# 6 Conclusion and future work

Based on the discussion above, it can be concluded that:

- Shift-invariance is of vital importance when dealing with the stereo correspondence problem.
- The disparity computation result using the dyadic wavelet transform demonstrates the viability of applying the wavelet transform approach to stereo matching. This encourages further investigation of other wavelet transform techniques such as wavelet zero-crossings and dual-tree complex wavelet transform applied to the matching problem.

These further methods are being investigated. A comparison of all the three strategies will be undertaken in future work.

## References

Daubechies, I. (1992), *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics.

Faugeras, O. (1993), *Three Dimensional Computer Vision: a Geometric Viewpoint.* UK: The MIT Press.

Grimson, W. (1981), A Computer Implementation of a Theory of Human Stereo Vision. *Phil. Trans. Royal Soc. London*, **292**: 217-253.

Julesz, B. (1971), *Foundations of Cyclopean Perception.* Chicago: University of Chicago Press.

Kingsbury, N. G. (1998), *The Dual-tree Complex Wavelet Transform: A New Technique for Shift Invariance and Directional Filters.* IEEE Digital Signal Processing Workshop, DSP 98. (86) Bryce Canyon.

Magarey, J. & Kingsbury, N. G. (1995), Report, Department of Engineering, Cambridge University.

Mallat, S. (1989), A Theory for Multiresolution Signal Decomposition: the Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11** (7): 674-693.

Mallat, S. (1991), Zero-crossings of a Wavelet Transform . *IEEE Transactions on Information Theory*, **37** (4): 1019-1033.

Mallat, S. (1998), *A Wavelet Tour of Signal Processing.* Academic Press.

Marr, D. (1982), *Vision.* New York: W. H. Freeman and Company.

Sanger, T. D. (1988), Stereo Disparity Computations Using Gabor Filter. *Biol. Cybern.*, **59**: 405-418.

Simoncell, E. P., Freeman, W. T., Adelson, E. H. & Heeger, D. J. (1992), Shiftable Multiscale Transforms. *IEEE Trans. Information Theory*, **38** (2): P587-607.

Strang, G. & Nguyen, T. (1997), *Wavelets and Filter Banks.* Revised end. Wellesley-Cambridge Press.

Teolls, A. (1998), *Computational Signal Processing with Wavelets.* Birkhauser Boston.

Trucco, E. & Verri, A. (1998), *Introductory Techniques for 3-D Computer Vision.* New Jersey: Prentice Hall.

# SSD Matching Using Shift-Invariant Wavelet Transform

Fangmin Shi, Neil Rothwell Hughes and Geoff Roberts
Mechatronics Research Centre
University of Wales College, Newport
Allt-Yr-Yn Campus
PO Box 180
Newport NP20 5XR, UK
fangmin.shi@newport.ac.uk
neil.rothwell-hughes@newport.ac.uk
geoff.roberts@newport.ac.uk

**Abstract**

The conventional area-based stereo matching algorithm suffers from two problems, the windowing problem and computational cost. Multiple scale analysis has long been adopted in vision research. Investigation of the wavelet transform suggests that -- dilated wavelet basis functions provide changeable window areas associated with the signal frequency components and hierarchically represent signals with multiresolution structure. This paper discusses the advantages of applying wavelet transforms to stereo matching and the weakness of Mallat's multiresolution analysis. The shift-invariant dyadic wavelet transform is exploited to compute an image disparity map. Experimental results with synthesised and real images are presented.

# 1 Introduction

Finding correspondence is an ill-posed problem in stereo vision. Area-based stereo matching is one of the conventional solutions. It compares the intensity similarity between windowed areas of two stereo images. The sum of squared difference (SSD) [1] is commonly used as the similarity measure:

$$ssd(x) = \sum_{\tau=x-\sigma}^{x+\sigma} \left|L(\tau) - R(\tau)\right|^2 \qquad (1)$$

where $x$ is the pixel index over the stereo images $L$ and $R$, $\tau$ indexes over the local area around $x$ within $\pm\sigma$. It is well known that this method suffers from the windowing problem and computational cost [1].

In order to alleviate these problems, multistage strategies were developed by vision

researchers such as multistage matching by dividing images into small blocks of equal size [2], multiscale matching based on Gaussian filtered images [3], [4], [5], and the pyramid structure that generates sets of low-pass and band-pass filtered images [6]. A more general hierarchical architecture and fast implementation was created by Mallat in 1989 following a study of wavelet concepts. This is known as wavelet multiresolution analysis (MRA) [7].

The advantage of the wavelet transform is that it uses wide windows for low-frequency components and narrow windows for high-frequency components [8]. These windows are formed by dilations and translations of a prototype (or mother) wavelet. Thus, if SSD matching is performed on the wavelet transforms of signals, the windowing problem with the conventional SSD approach is naturally solved. However, Mallat's multiresolution analysis lacks shift-invariance, which will be discussed in the following section. Stereo matching requires a shift-invariant transform because stereo image pairs can be considered as the shifted versions of each other (with distortions). This makes MRA unsuitable for matching.

This paper discusses the strengths of wavelet transforms when applied to stereo matching and some alternative wavelet methods to Mallat's MRA. The Dyadic wavelet transform is exploited to compute a dense disparity map. Experimental results using synthesised and real images are presented.

# 2 Properties of the Wavelet Transform for Stereo Matching

Modern wavelet theory was motivated initially for the sake of a better time-frequency signal representation than the short time Fourier transform (STFT) and to overcome its drawbacks. In contrast with the STFT that uses a constant window for the whole signal, the wavelet transform uses wide windows for low-frequency components and narrow windows for high-frequency components [8]. It achieves this by decomposing a signal into the dilations and translations of a mother wavelet.

Let $\varphi(t)$ denote a mother wavelet, which is a small oscillatory function with finite support. A family of wavelets $\varphi_{a,b}(t)$ is then represented by

$$\varphi_{a,b}(t) = a^{-1/2}\varphi((t-b)/a) \tag{2}$$

where $a$ is the scale parameter, $b$ is the translation parameter, and $a, b \in R$. The wavelet transform represents a signal $x(t)$ by an infinite set of such basis functions:

$$WT(b,a) = \int x(t)a^{-1/2}\varphi((t-b)/a)dt \tag{3}$$

## 2.1 Automatic Windowing Analysis

In order to illustrate the time-frequency resolution of a wavelet transform, Figure 1 shows the coverage of a wavelet in the time-frequency plane. It is evident that when the frequency interval goes up by a scale factor, the time interval goes down by the same factor. Let $\Delta t$ and $\Delta f$ denote the window width of the mother wavelet in time and in the spectral domain, and $\Delta t_{ab}$ and $\Delta f_{ab}$ are the corresponding denotations of the scaled and shifted wavelet. That is:

$$\Delta t \cdot \Delta f = \Delta t_{\tau s} \cdot \Delta f_{\tau s} \quad (4)$$

This reveals that the product of the window width of time and frequency is constant at all scales [8]. This property is one of the most important advantages that wavelet transforms provide. In contrast with the STFT applying either narrow or wide window (but not both) to the whole signal, wavelet transforms are able to analyse high-frequency components using small windows and low-frequency components using big windows. This property is ideal when dealing with non-stationary signals that contain both short high-frequency components and long low-frequency components.

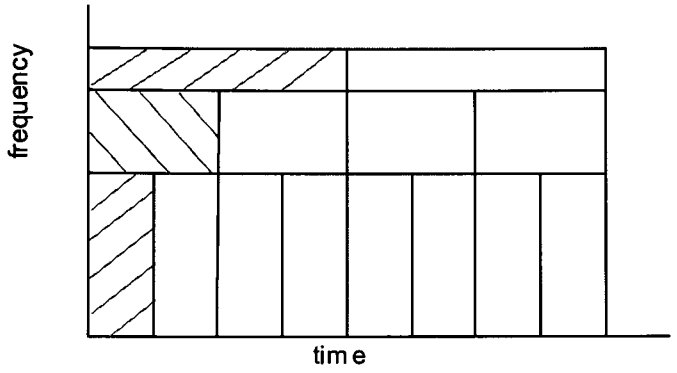



**Figure 1 Time-frequency plane of wavelet transform: window area is constant at all scales**

An image is a typical non-stationary signal, which consists of a slowly changing background and rapidly changing details. After decomposing an image, a set of images at different resolutions is obtained. At coarser resolutions, the matching is performed by comparing wider areas leading to larger uncertainty in disparity localisation. At finer scales, the compared areas tend to be more localised and smaller uncertainty in disparity localisation is expected. The windowing problem that occurs with SSD matching is then naturally solved by choosing the right wavelet.

## 2.2 Why MRA Is not Suitable for Matching

In equation (3), if parameter $a$ and $b$ are continuous real values, then the transform is called a continuous wavelet transform. The wavelet basis functions constitute an overcomplete representation in which information is highly redundant. This redundancy can be reduced by discretising $a$ and $b$. Daubechies [9] found that when $a=2^n$, $b=k2^n$, $n$, $k \in \boldsymbol{Z}$, the basis functions $\{ \varphi_{kn}(x) = 2^{-n}\varphi(2^{-n}x - k) \}$ are orthogonal for certain choices of wavelet. Stimulated by the pyramidal approach in vision, Mallat, as a former vision researcher, proposed a fast implementation for the wavelet orthogonal decomposition [7]. This is the well known multiresolution analysis (MRA).

MRA decomposes a signal into the same size subimages at dyadic scales. At each scale an approximation part and a detail part are formed by passing the signal through a half-band low-pass filter and a half-band high-pass filter, and subsequently downsampling them by two. The approximation part is then hierarchically decomposed. Downsampling a signal simply discards every other sampling point. This

operation reduces the number of signal samples by a half when the scale is doubled.

*Shift-invariance* (or *time-invariance*) means that if a signal is delayed in time, its transform result is delayed as well. Downsampling is not shift-invariant. Neither, therefore, is MRA. This issue was discussed by Strang [10] and the shift-invariance problem was considered to be the main drawback of MRA.

For stereo matching, intuitively, one image can be assumed to be the shifted version of the other. The shifted value with respect to each pixel is the disparity, which is dependent on the pixel position. Only shift-invariant wavelet transforms can be used for matching.

## 2.3 Shift-Invariant Wavelet Transforms

The continuous wavelet transform possesses the property of shift-invariance. However, its high redundancy gives rise to high computational cost. Approximate shift-invariance with effective computation needs to be achieved. Mallat used the dyadic wavelet transform and the zero-crossings of the dyadic wavelet transform to reduce the representation size [11]. Simoncell [12] built steerable filters to achieve a shiftable transform that is jointly invariant in position, scale and orientation. More recently, a shift-invariant complex wavelet transform [13] with perfect reconstruction has been constructed and applied to image processing and computer vision.

The wavelets used in these papers could be applied to the matching problem. This paper will discuss the application of the dyadic wavelet transform to disparity computation. The motivation for this is discussed in the next section.

# 3 Correspondence Matching Using the Dyadic Wavelet Transform

To simplify the numerical computations and maintain shift-invariance, the scale parameter $a$ of equation (3) is discretised along a dyadic sequence $\{2^n, n \in Z\}$ while leaving the shift parameter $b$ continuous. The dyadic wavelet transform ($DWT(b, j)$)has the following form:

$$DWT(b,n) = \int x(t)2^{-n/2}\varphi((t-b)/2^{-n})dt \qquad (5)$$

Mallat [14] proved that under certain condition dyadic wavelet transform defines a complete and stable representation. The algorithmic efficiency if gretly improved compared with the continuous wavelet transform. The information is still redundant due to the continuous translation parameter. However, it is good for matching task aiming at dense disparity map output.

Figure 2 gives two signals (epipolar lines from two synthesised stereo images). The decomposition results at three scales, $2^1$, $2^2$ and $2^3$, are shown in Figure 3.

The sum of squared difference (SSD) [15] is applied to the transformed signals to measure the similarity of the corresponding points. The smaller the SSD value, the more likely it is that the points correspond to each other. Unlike the conventional SSD, directly applied to the image intensity value, the SSD here is applied to the wavelet coefficients.

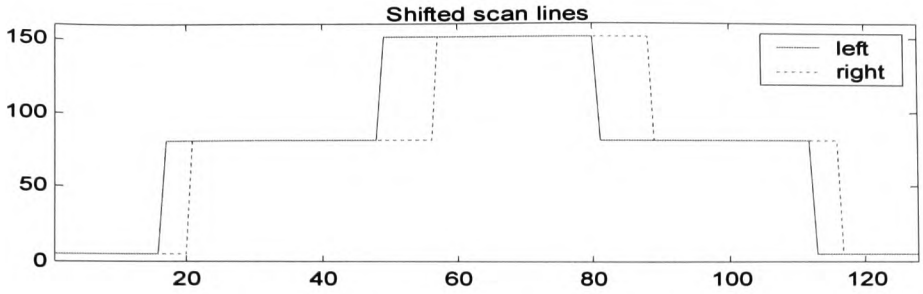



**Figure 2 Scan lines from stereo images**

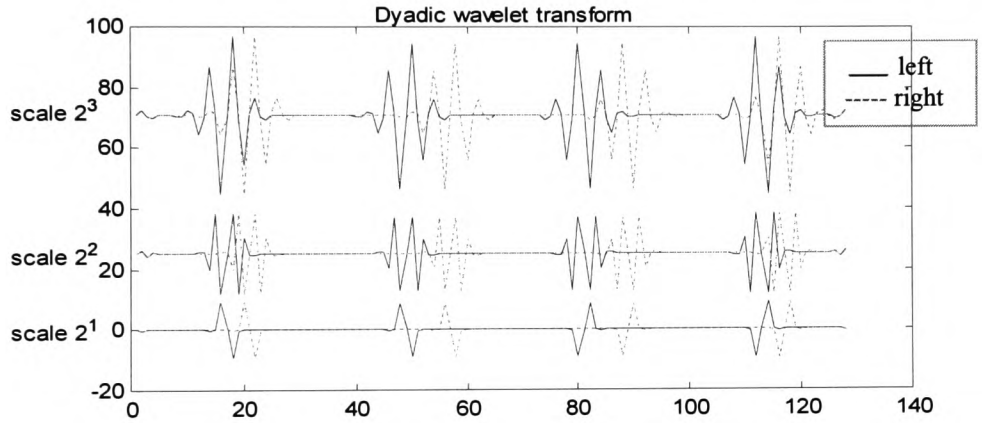



**Figure 3 1D dyadic wavelet transform at three scales**

Let $x_1(t)$ and $x_2(t)$ be two scan lines of stereo images, their dyadic wavelet transforms are denoted by $DWT1(2^n,t)$ and $DWT2(2^n,t)$, respectively, where $n \in N$. The SSD measure ($ssd(n,x)$) is defined as:

$$ssd(n,x) = \sum_{\tau=x-2^n\sigma}^{x+2^n\sigma} \left| DWT1(2^n,\tau) - DWT2(2^n,\tau) \right|^2 \qquad (6)$$

where $\sigma$ is the size of an interval where the energy of the mother wavelet is mostly concentrated [11].

From equation (6), it can be seen that at each scale the searching area is ($-2^n\sigma$, $2^n\sigma$). Matching should be taken at as much scales as possible. The maximum scale ($nmax$) should be determined by: $2^{n\,max}\,\sigma \leq signal\ \ length$ .

Besides the *epipolar constraint* [16], other constraints e.g. *similarity*, *uniqueness*, *ordering* and *continuity* [17] are also applied along with equation (6). Figure 4 gives the computed disparity result at three scales. The corresponding SSD values are also recorded as a measure of the matching confidence. The smaller the SSD value is, the higher is the confidence of the matching. For comparison, the parameter at three scales is ploted in one figure, see Figure 5.

After the computation from the above steps, each pixel corresponds to two

parameters, its disparity value, *d(n,x)*, and its SSD value, SSD *(n,x)*. The most intuitive way is to choose the right scale $N$ for each pixel disparity so that at that scale its SSD value is a minimum of all the scales. That is, if $N = \min_{n}\{ssd(n,x)\}$, then *d(x)=d(N,x)*.

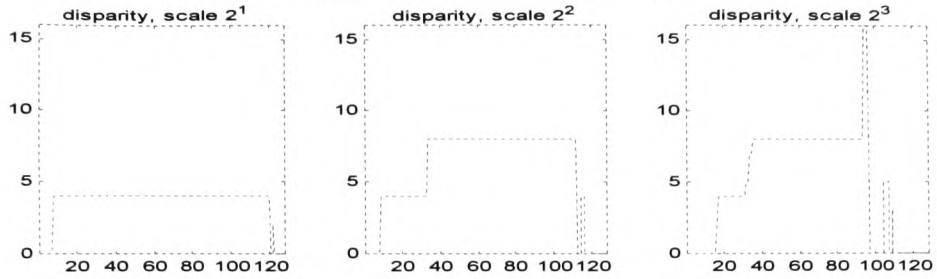



**Figure 4 Disparity at three scales**

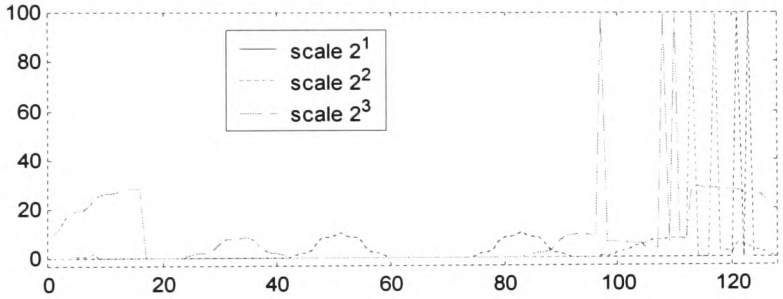



**Figure 5 SSD at three scales**

One of the big problems of SSD matching is its noise characteristics. Figure 6(a) shows the computational result using above method. It can be seen that the disparity value at the right end, i.e. around pixel 120 is not very good. This is the general case when the same program is tested with some other more complicated image pairs, which show worse results at some points. In some cases, for example, the points at the border of the image may not have matches, or images corrupted with noise give rise to unstable fluctuating results. Thus noise reduction is needed.

The noise problem can be dealt with using one of two possible methods, both of which use an additional matching constraint to remove the points with higher SSD value than a threshold. Hard thresholding and soft thresholding [18] are employed in the two methods, respectively. The standard deviation is adopted as a soft threshold in this paper.

Figure 6(b) gives the computational results using soft thresholding, which shows sharp edges and stable disparity values.

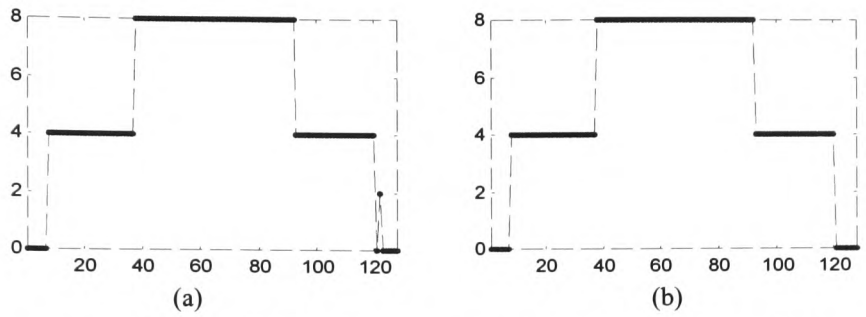



**Figure 6 Computed disparity, (a) without threshold, *(b)* soft threshold**

# 4 Experimental Results with Images

For the initial test, the commonly used random dot stereograms [19] are constructed. Figure 7 shows the synthesised '*Squares*' images of size 128*128. The central two squares are right shifted by 4 and 8 pixels, respectively, between the two images.

The image matching is performed along the theoretical epipolar lines of the images. The disparity map and the depth map for *Squares* are given in Figure 8.

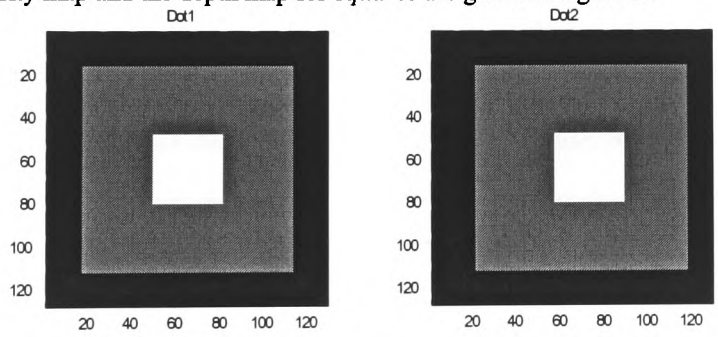



**Figure 7 Stereo pair: *Squares***

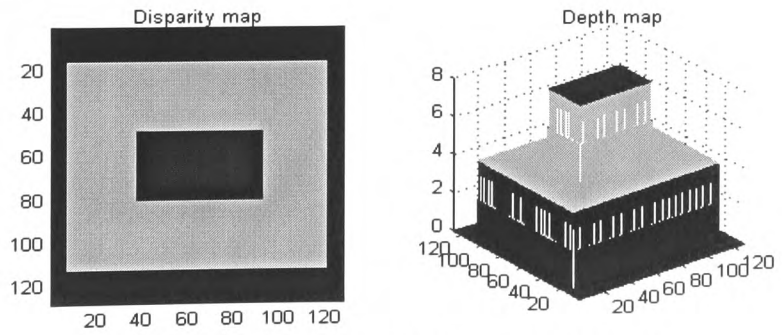



**Figure 8 Disparity map and depth map: *Squares***

To increase the complexity of the image features, another pair of random dot stereograms, *Random Square* showed in Figure 9 , is tested. In contrast with the *Squares* images, the pixels in *Random Squares* are random values between *0* and *1*. In Figure 10, the left figure gives the ground truth disparity map and right shows the computational results using the above method.

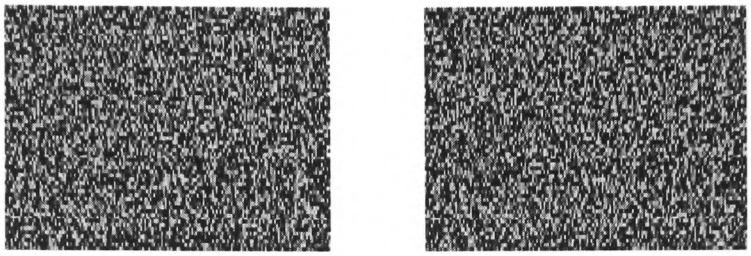


**Figure 9 Stereo pair 2: *Random Squares***

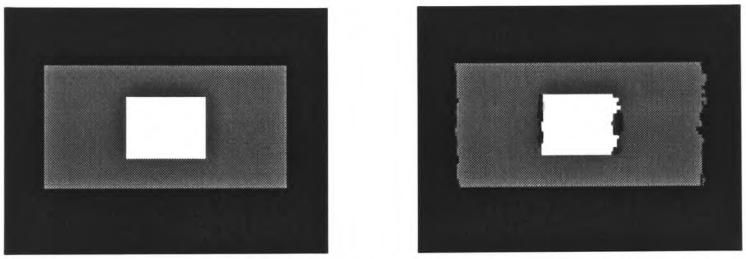


**Figure 10 A comparison of the ground truth data and computed data (I)**
**Left: ground truth disparity map, Right: disparity map using wavelets**

Computation with real image pairs is also tested. One imagery popularly used is shown in Figure 11, which can be downloaded from the web site, http://www.research.microsoft.com/~szeleski/stereo. The ground truth and estimated disparity maps using dyadic wavelet transform are displayed in Figure 12.

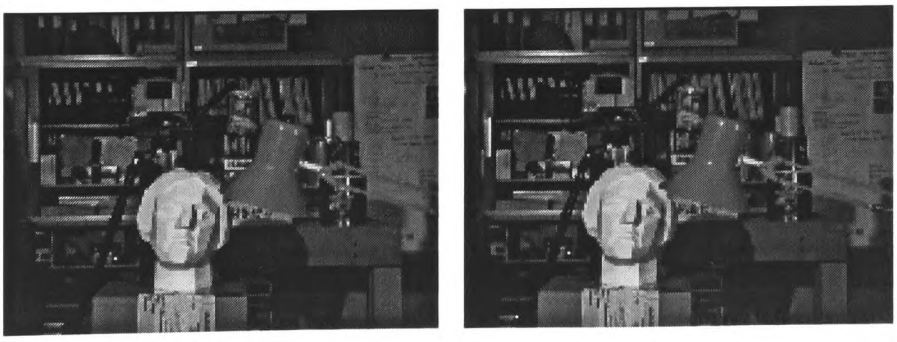


**Figure 11 Stereo pair 3: Tsukuba images**

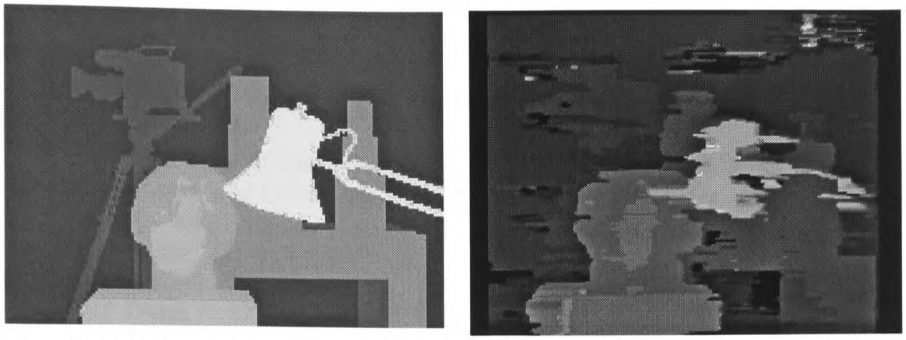


**Figure 12 A comparison of the ground truth data and computed data (II) Left: ground truth disparity, right: estimated disparity result using wavelets**

Another pair of real images was taken in the authors' laboratory and used in [20] is shown in Figure 13, the right figure of which gives the computed disparity map.

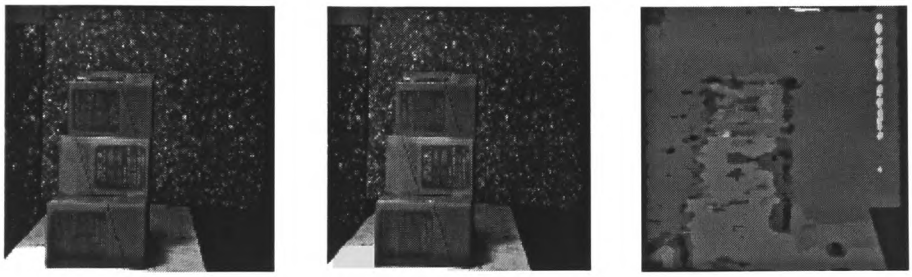


**Figure 13 Real stereo pairs 4 and the disparity map**

# 5 Conclusion and Future Work

This paper has presented a wavelet approach to computing disparity maps. It is of vital importance to apply shift-invariant wavelet transforms when using wavelet techniques for stereo matching. As an initial approach to the application of wavelet transforms to stereo matching, the dyadic wavelet transform was used to develop a matching algorithm. The sum of squared difference is defined based on the values of dyadic wavelet transform coefficients. The experimental results give rise to promising disparity maps. This demonstrates the viability of applying the wavelet transform approach to stereo matching.

However, a better compromise between the algorithmic efficiency and information redundancy could be made because the translation parameter of the dyadic wavelet transform remains continuous. Further investigation of other wavelet transform techniques such as wavelet zero-crossings and dual-tree complex wavelet transforms applied to the matching problem is therefore being carried out. And the comparison of wavelet-based algorithms with standard matching algorithms will be made in the future work.

# 6 References


[1] Trucco, E. and Verri, A., *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.

[2] Rosenfeld, A. and Thurston, M., Coarse-fine Template Matching, *IEEE Trans. System, Man, and Cybernetics*, vol. 7, pp. 104-107, 1977.

[3] Marr, D. and Poggio, T., A Computational Theory of Human Stereo Vision, *Proc. R. Soc. Lond.*, vol. 204, pp. 301-328, 1979.

[4] Grimson, W., A Computer Implementation of a Theory of Human Stereo Vision, *Phil. Trans. Royal Soc. London*, vol. V292, pp. 217-253, 1981.

[5] Grimson, W. E. L., Computational Experiments with a Feature-Based Stereo Algorithm, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 7, pp. 17-34, 1985.

[6] Burt, P. and Adelson, E. H., The Laplacian Pyramid as a Compact Image Code, *IEEE Trans. Communications*, vol. 31 , pp. 532-540, 1983.

[7] Mallat, S., A Theory for Multiresolution Signal Decomposition: the Wavelet Representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674-693, 1989.

[8] Chui, C. K., *Wavelets: A Tutorial in Theory and Applications*, Academic Press, 1992.

[9] Daubechies, I., Orthonormal Bases of Compactly Supported Wavelets, *Communications on Pure Applied Mathematics*, vol. 41, pp. 906-996, 1988.

[10] Strang, G. and Nguyen, T., *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Second Ed., 1997.

[11] Mallat, S., Zero-crossings of a Wavelet Transform, *IEEE Transactions on Information Theory*, vol. 37, pp. 1019-1033, 1991.

[12] Simoncell, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J., Shiftable Multiscale Transforms, *IEEE Trans. Information Theory*, vol. 38, pp. P587-607, 1992.

[13] Kingsbury, N. G., The Dual-tree Complex Wavelet Transform: A New Technique for Shift Invariance and Directional Filters, *IEEE Digital Signal Processing Workshop, DSP 98,* Bryce Canyon, pp. Paper no 86, 1998.

[14] Mallat, S., *A Wavelet Tour of Signal Processing*, Academic Press, 1998

[15] Anandan, P., Computing Dense Displacement Fields with Confidence Measures in Scenes Containing Occlusion, *Proceedings DARPA Image Understanding Workshop,* pp. 236-246, 1984.

[16] Faugeras, O., *Three Dimensional Computer Vision: a Geometric Viewpoint*, The MIT Press, 1993.

[17] Marr, D., *Vision*, W. H. Freeman and Company, 1982.

[18] Chambolle, A., Devore, R. A., Lee, N. Y., and Lucier, B. J., Nonlinear Wavelet Image Processing: Variational Problems, Compression, and Noise Removal Through Wavelet Shrinkage, *IEEE Transactions on Image Procesing*, vol. 7, pp. 319-334, 1998.

[19] Julesz, B., *Foundations of Cyclopean Perception*, University of Chicago Press, 1971.

[20] Rothwell Hughes, N., *Fuzzy Filters for Depth Map Smoothing*, PhD Thesis, University of Wales, 1999.