論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Stereo Vision Based on Mathematical Modeling - Toward Adaptive and Precise 3-D Reconstruction from Projected Images
著者(和文)	奥富正敏
Author(English)	MASATOSHI OKUTOMI
出典(和文)	学位:工学博士, 学位授与機関:東京工業大学, 報告番号:乙第2507号, 授与年月日:1993年7月31日, 学位の種別:論文博士, 審査員:
Citation(English)	Degree:Doctor of Engineering, Conferring organization: Tokyo Institute of Technology, Report number:乙第2507号, Conferred date:1993/7/31, Degree Type:Thesis doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Stereo Vision Based on Mathematical Modeling

—— Toward Adaptive and Precise 3-D Reconstruction from Projected Images

Masatoshi Okutomi

February 1993.

Abstract

This thesis presents stereo vision based on physical and mathematical modeling. In our approach, physical phenomena which create 2-D images from the 3-D world are represented by simple models. Also, stereo methods which, to the contrary, extract 3-D information from the images are described mathematically. Both of them consequently establish a unified mathematical foundation. This foundation can explicitly involve many factors relating to both the imaging process and the stereo method, including intensity and disparity variations, noise, color, matching windows, and stereo baselines. The mathematical analysis based on the foundation enables us to understand various properties of stereo vision and gives us concrete algorithms which can overcome the problems of stereo matching. The further advantage is that since we know the characteristics of the algorithms, they are far more predictable and extensible for different situations than the algorithms based on heuristics.

One resultant algorithm is a locally adaptive window for matching. The goodness of a window depends on intensity change, disparity change, and noise involved in an image. What makes the problem more difficult is that these factors change from position to position in the same image and that the disparity is what we want to calculate and embedded in the intensity patterns. As a solution to the problem, we employ a statistical model that represents the uncertainty of the disparity of points over the window. This modeling enables us to compute both a disparity estimate *and* the uncertainty of the estimate obtained by using the particular window. So, the algorithm can search for a window that produces the estimate of disparity with the least uncertainty for each pixel of an image. The method controls not only the size but also the shape (rectangle) of the window.

Another challenging task in stereo mathcing is to overcome a trade-off problem between precision and accuracy in matching. That is, the estimated distance with a short baseline is less precise due to narrow triangulation. On the other hand, with a longer baseline, a larger disparity range must be searched and, as a result, matching is more difficult and there is a greater possibility of a false match. The stereo matching method presented in this thesis uses multiple stereo pairs with different baselines generated by a lateral displacement of a camera. A new evaluation function called SSSD-in-inverse-distance is defined to exploit the multiple pairs. We show that this new function exhibits a unique and clear minimum at the correct matching position even when the underlying intensity patterns of the scene include ambiguities or repetitive patterns. An advantage of this method is that we can eliminate false matches and increase precision without any search or sequential filtering.

Another aspect of stereo, the use of color information, is also presented. We analyze the effect of using color information in stereo matching and propose a color stereo algorithm for a medical application. In this application, the 3-D shape of optic nerve heads are measured using stereo fundus images for diagnosing and monitoring glaucoma.

Throughout this thesis, both theoretical and experimental results are presented to demonstrate the effectiveness of our mathematical analyses and the resultant algorithms.

Acknowledgements

First, I would like to thank Akira Kobayashi at the Tokyo Institute of Technology. As the head of my thesis committee, he gave me valuable advice and enabled me to complete my thesis. I would also like to thank the other members of the committee, Katsuhisa Furuta, Ato Kitagawa, Shigeo Hirose, and Makoto Sato, for making many useful comments on my thesis. I am also grateful to Masahiro Mori for enabling the committee to have Prof. Kobayashi as its head.

I would also like to thank people at Carnegie Mellon University. First of all, I have to express my deepest thanks to Takeo Kanade. Without him, this thesis could never exist. Larry Matthies introduced me to the world of stereo vision. His steady work in this field certainly influenced my research. John Krumm, an ex-officemate at CMU, read my manuscripts many times and gave me very useful comments both technically and grammatically. Rich Volpe, another ex-officemate, made the deserted research life enjoyable with his jokes. They made even our small and windowless office into a comfortable place. Carlo Tomasi answered any question very kindly. Tomoharu Nakahara performed the experiments with outdoor scenes and produced the result of figure 3.17. Steve Shafer created the Calibrated Imaging Laboratory where many images were taken for the experiments. Also thanks to many other people who made the days I spent in Pittsburgh wonderful and fruitful.

I would also like to thank people of Canon Inc. Hideyuki Tamura gave useful comments and prompted me to finish my thesis. Osamu Yoshizaki gladly supported me in writing this thesis. Masanori Nakamoto gave me the opportunity to accomplish this dissertation. Paul Otto made perceptive comments on my research. Haruo Shimizu helped me a lot in creating documents and figures. Yoshifumi Kitamura and Toshikazu Oshima wrote programs used in the experiments of this thesis.

Although I did not mention all the names, I am also grateful to my friends and many other people for their warm support.

Finally, but most of all, I would like to thank my wife Yuko and my daughter Lala who have kept bringing joy, motivation, and sometimes surprises to my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Reconstruction of the three-dimensional (3-D) structure of the world from its two-dimensional (2-D) projections is a major task in computer vision (figure 1.1). Stereo vision is one of the most promising methods and many algorithms have been proposed. However, most of these techniques have started from given images without explicitly considering the fact that they were created from the real 3-D world through physical phenomena. Consequently, they rely heavily on empirical rules or heuristics and tend to be ad hoc (figure 1.2). They may give good results for the images which happen to be used, but there is no guarantee for others. They don't have the ability to predict their performance for different situations or domains, nor do they give *understanding* about the vision problems.

On the other hand, work that takes account of how images were actually created in computer vision tasks has been recently done. Kanade and his research group at CMU are one of the most active forces. He called such an approach *physically based vision* [Kan91]. This thesis is on a similar track. Figure 1.3 illustrates our basic approach. In this approach, we model the physical phenomena that create images with simple equations. Also the methods which extract 3-D information from 2-D images are described mathematically. These two constitute a unified formulation which enables us to analyze the characteristics of the task and establishes a link between resultant distance estimates and various factors that may affect the performance. Algorithms are obtained as a result of the formulation through stochastic

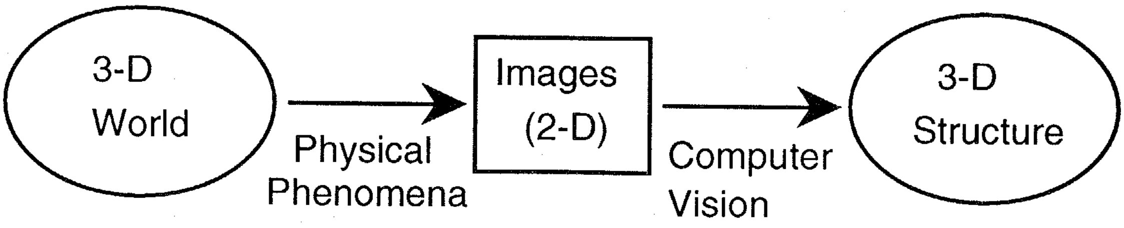


Figure 1.1: Reconstruction of 3-D structure from 2-D projections

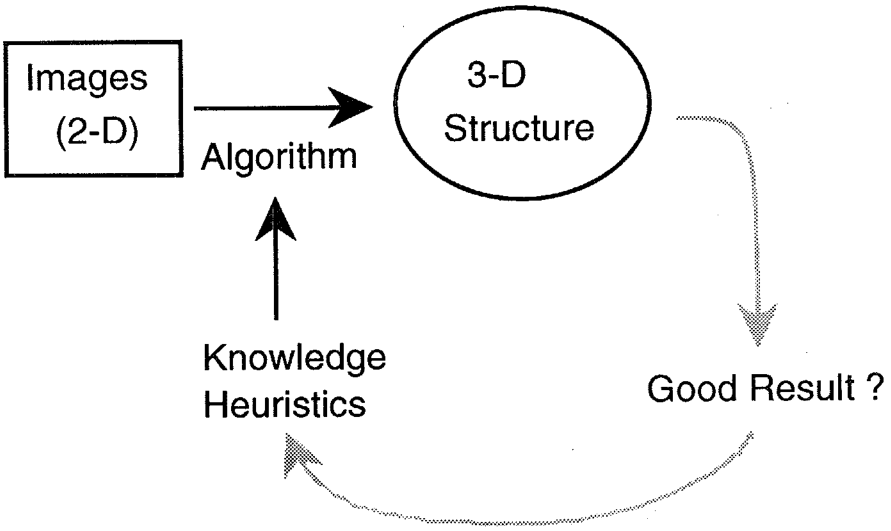


Figure 1.2: Heuristic-based approach

optimization. The important point is that the resultant algorithms are far more extensible, generalizable, and predictable for different situations than heuristic-based methods.

Now let us begin with a brief explanation of stereo vision and its related techniques. It is followed by a survey of previous work and main issues in stereo vision. They form the background for our work. We then summerize the contributions of the thesis. Finally, the organization of the reminder of this thesis is presented.

1.1 Stereo Vision

Figure 1.4 illustrates the principle of stereo vision. Assuming that the optical axes of the cameras are parallel to each other, the distance to the object z is computed from:

$$z = \frac{BF}{d}, \tag{1.1}$$

where B and F are the baseline (the distance between cameras) and focal length, respectively. d is the disparity, which is the difference in the x coordinate between left and right images for the same object point,

$$d = x_L - x_R. \tag{1.2}$$

That is, stereo's fundamental principle is *triangulation.* Once the corresponding points are found, which are the projections of the same physical point in space onto the left and the right images, the physical point's position in 3-D space is easily obtained using equation

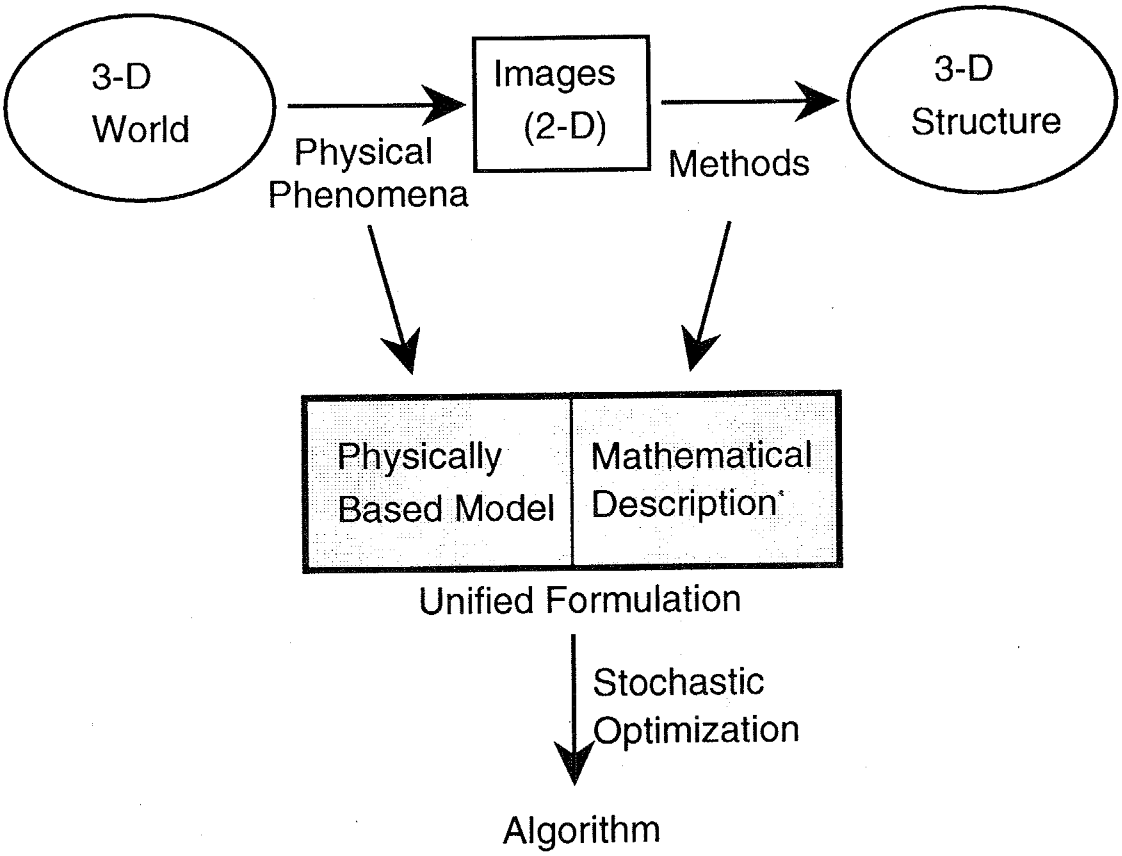


Figure 1.3: Our approach based on physical and mathematical modeling

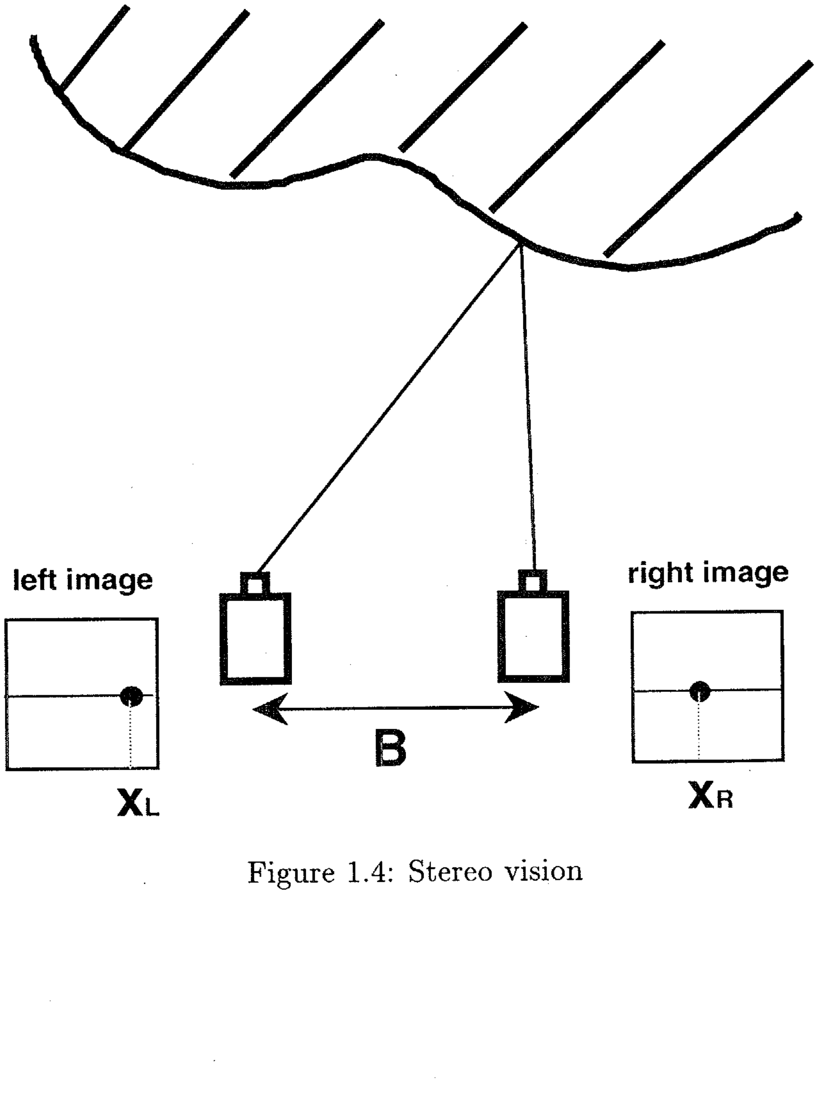


Figure 1.4: Stereo vision

active	time-of-flight	laser
		ultrasonic
	triangulation (structured-light projection)	spotlight
		light stripe
		2-D light pattern
	Ḿoire pattern	
passive (shape from . . .)	*stereo vision*	
	shading	Lambertian reflection
		specular reflection
		interreflection
	motion	
	geometrical constraint	parallel or perpendicular lines
		texture
	defocus	focus position
		aperture

Table 1.1: Noncontact 3-D sensing techniques.

(1.1) and simple geometry. So the most important and challenging task in this method is, for each point in one image, finding the corresponding point in another image.

On the other hand, we can regard this method as one of several 3-D sensing techniques that measure 3-D information of objects or scenes. Table 1.1 shows major noncontact 3-D sensing techniques. They can be divided into two types, *active* and *passive* methods. Active methods which cast energy onto objects are, in general, relatively reliable and practically used in industrial applications. However, there are some drawbacks due to *active*. They may affect the objects or the environment, interfere with each other when multiple sensors are used simultaneously, or consume relatively more power. On the other hand, passive methods are not as reliable as active methods so far. However, they are potentially more flexible and generally-applicable as the human visual system demonstrates. A lot of work is being done in the field of computer vision using various cues for estimating 3-D information, e.g., disparity caused by different view positions, shading on an object surface, motion, perspective, texture gradients, different degrees of defocussing, and so on as shown in table 1.1. Since disparity is one of the most powerful depth cues, a large amount of work on stereo vision has been done. We will survey them in the next section. More explanations about 3-D sensing techniques can be found in [Kan87][Shi87][San89][Wec90].

1.2 Previous Work and Main Issues in Stereo Vision

In this section, we survey previous work in stereo vision and depict generally what are the essential problems of this method.

Stereo vision algorithms are commonly divided into two types, area-based and feature-based. In area-based stereo matching, windows are set around candidate corresponding points in the left and right images to assess the likelihood of correspondence. Typically, a kind of correlation such as normalized cross correlation (the correlation coefficient) [Han74] [Gop77] [Mor81] [Qua84] [WTK87], covariance [SB76], and the sum of the squared differences (SSD) [Gen77] [Ana84] [MSK89] of the intensity values over the windows is computed. The algorithms are relatively simple and can produce dense depth maps. The major problems of area-based methods are:

- They tend to be affected by geometric distortions caused by the projection of non-frontoparallel surfaces onto the image planes, since if it is the case, the pixels over the window in one image are no longer just the shifted-in-parallel ones in another image. Quam [Qua84], Gruen [Gru85], and Otto et al. [OC89] warped (basically affine transformation) their windows to compensate for this distortion.

- An appropriate size of the correlation window has to be used for each application. Furthermore, one fixed-size window may not produce good measures for all portions of an image due to the locally different propeties in the image. Levine et al. [LOY73] and Dowman et al. [DH77] changed their window size using the variance of the intensities within the window.

- It is difficult to obtain depth measurements for the regions of constant intensity. Although this is often mentioned, it is not a problem peculiar to area-based matching, since, for example, there is essentially no way to find correspondense for a completely-constant-intensity region regardless whether area-based or feature-based matching is used. A possible claim of feature-based matching against area-based is that the *unfavorable* regions will be discarded in extracting features and never tried in matching while area-based methods tend to try to compute disparities for all pixels in the image. So what area-based methods need may be an additional mechanism to indicate unreliable measurements. Gennery [Gen80], Förstner et al. [FP86], and Matthies et al. [MSK89] gave estimates of the reliability of the match based on the statistical properties of random noise.

- Area-based methods consume relatively more time. Lucas et al. [LK81] and Nishihara [Nis87] proposed methods to reduce computational cost in calculating correlation. Also it should be noted that area-based algorithms are in general simple and amenable to parallel hardware implementations.

On the other hand, feature-based stereo matching consists of two main stages, i.e. (i) feature extraction from images and (ii) matching the extracted features between both images. There have been a wide variety of algorithms proposed depending on the methods adopted for each step. For example, feature points were extracted by the Hueckel operator [Arn78], an interest operator [BT80], zero-crossings [Gri85] [DP86], or the Canny edge detector [Pog88]. To reduce ambiguities, higher-order features were extracted such as line segments [HKK84] [MN85] [AL87], edge-delimited intervals [HMG79] [OK85], skeletons [BK88], rectangles [MN89], and curves [BB89]. Also many techniques were proposed in the matching stage including dynamic programming [BB81] [OK85], relaxation techniques [MP76] [BT80] [KA85], graph description matching [AL87], and simulated annealing [Bar89]. Though the feature-based algorithms are relatively fast and in general less sensitive to geometric distortions, they have the following drawbacks:

- The points where depths can be obtained are sparse, since only depths on extracted features can be directly computed. Furthermore, as Jenkin et al. [JJT91] pointed out: "The constraints on the density and on the ease of eliminating false targets are in direct opposition. In particular, the number of possible matches in a given region increases polynomially with the density of a given symbolic feature. Therefore the problem of finding the correct match can be expected to rapidly become more difficult as the density grows." To obtain a dense depth map from the sparse measurements, an interpolation (smoothing) process is often applied. However, it may degrade the already-computed depth measurements and add further complexity to the algorithms. Another direction may be obtaining 3-D structure directly from sparse measurements, though this needs some additional constraints or knowledge about the object and has its own difficult problems.

- All information contained in the images is not used. This is because feature extraction throws away a lot of information, e.g. a gradual change of intensity or shading. Moreover, as Otto et al. [OC89] mentioned, "small-scale 'texture' tends to confuse most feature detectors, and make their performance worse, whereas it improves that of most area-based detectors."

- The algorithms tend to be heuristic, e.g. a heuristic choice of information extracted, heuristic search for correspondence, and their heuristic combination. This is not necessarily a problem, but it is very difficult to understand the methods or know their characteristics in various situations.

Another important issue which is common to both area-based and feature-based methods is the elimination of ambiguities in matching. The major strategies for this are: (a) coarse-to-fine control strategy in which matching is done at a low resolution first to reduce ambiguities, and then the result is used to constrain the matching at a higher resolution [MP79] [Qua84]

[Gri85] [HA86] [Bar89] [Han89] [CM90], (b) best-first strategy in which matching for likely-to-be-good points is performed first which then constrains the matching at less promissing surrounding points [MKA73] [Han89] [OC89], and (c) using additional images taken from different view points [Mor79] [Tsa83] [MK85] [XTA85] [PH86] [YKK86] [II86] [AL87] [OWI86] [MO89] [YH92].

1.3 Our Approach and Main Contributions of the Thesis

In the previous sections, we have presented the background for our work. Now we revisit our approach and summerize the main contributions of the thesis.

The key point of our approach is the development of a mathematical framework in which both physical and methodological aspects in vision problems can be considered simultaneously as shown in figure 1.3. This approach can also be contrasted with that based on heuristics or that simulating the human vision system. The advantage of our approach is that we know the characteristics of the problems mathematically, and the resultant algorithm can be more extensible, generalizable, and predictable under different situations.

The main contributions of this thesis are to establish a mathematical foudation for our approach and to show that some major problems of stereo vision can be solved in a mathematically well-defined manner based on the same foundation. Throughout this thesis, both theoretical and experimental results are presented to show the effectiveness of our approach and the resultant algorithms. Additional details about the contributions are described below.

The mathematical foundation we establish consists of a physically-based model of stereo images[1] and a mathematical description of stereo matching[2]. Our mathematical formulation is based on statistical models, which consequently produces the most likely estimate of the disparity (distance) and the uncertainty of the estimate. The uncertainty estimation is very important, not only because it indicates unreliable measurements, but because it can relate the performance of stereo matching with various factors. They include those originating both in physical phenomena such as underling intensity patterns, disparity patterns (3-D structure of the scene), noise added in the images, and imaging geometry, and in methods such as matching windows, stereo baselines, the number of images, and colors to be used for matching. Analyzing these characteristics enables us to understand stereo problems and leads to concrete stereo algorithms as described below.

[1] The model is not necessarily a precise description of the real physical phenomena. It is enough as long as it is a resonable model or approximation of the real conditions which affect the performance of the resultant algorithm.

[2] In terms of the classification in the previous section, the matching method we employ is area-based, since it is much more suitable for our approach in which any ad hoc procedures and heuristics should be removed, and potentially more general as making use of all information contained in the images than feature-based.

One central problem in area-based stereo matching lies in selecting an appropriate window. The window size must be large enough to include enough intensity variation for reliable matching, but small enough to avoid the effects of projective distortion. If the window is too small and does not cover enough intensity variation, it gives a poor disparity estimate, because the signal (intensity variation) to noise ratio is low. If, on the other hand, the window is too large and covers a region in which the distance of scene points (i.e. disparity) varies, then the position of maximum correlation or minimum SSD may not represent correct matching due to different projective distortions in the left and right images. For this reason, a window size must be selected adaptively depending on local variations of intensity and disparity. The stereo algorithm we present selects a window adaptively by evaluating the local variation of the intensity *and* the disparity. We employ a statistical model that represents uncertainty of disparity of points over the window: the uncertainty is assumed to increase with the distance of the point from the center point. This modeling enables us to assess how disparity variation within a window affects the estimation of disparity. As a result, we can compute the uncertainty of the disparity estimate which takes into account both intensity and disparity variances. So, the algorithm can search for a window that produces the estimate of disparity with the least uncertainty for each pixel of an image. The method controls not only the size but also the shape (rectangle) of the window. The algorithm has been tested on both synthetic and real images, and the quality of the disparity maps obtained demonstrates the effectiveness of the algorithm.

Another challenging task in stereo vision is to obtain precise distance estimates without suffering from ambiguity. In stereo processing, a short baseline means that the estimated distance will be less precise due to narrow triangulation. For more precise distance estimation, a longer baseline is desired. With a longer baseline, however, a larger disparity range must be searched to find a match. As a result, matching is more difficult and there is a greater possibility of a false match. So there is a trade-off between precision and accuracy in matching. We introduce a stereo matching method that uses multiple stereo pairs with different baselines generated by a lateral displacement of a camera. Matching is performed simply by computing the sum of squared-difference (SSD) values. The SSD functions for individual stereo pairs are represented with respect to the inverse distance (rather than the disparity, as is usually done), and then are simply added to produce the sum of SSDs. This resulting function is called the SSSD-in-inverse-distance. We show that the SSSD-in-inverse-distance function exhibits a unique and clear minimum at the correct matching position even when the underlying intensity patterns of the scene include ambiguities or repetitive patterns. An advantage of this method is that we can eliminate false matches and increase precision without any search or sequential filtering. We show the advantage of the proposed method by analytical and experimental results.

Color is common information that has been ignored by most past stereo methods. We analyze the effect of using color information in stereo matching mathematically and experimentally, and propose a color stereo matching algorithm for a medical application. In this application, 3-D shapes of optic nerve heads are obtained using color stereo fundus images. The experimental results using real fundus images are encouraging, and they demonstrate that the method, together with various means of displaying the results, could give useful information for diagnosing and monitoring glaucoma, an eye disease which commonly causes blindness.

1.4 Thesis Overview

Figure 1.5 shows the organization of the following chapters in this thesis.

Chapter 2 presents a mathematical foundation for this thesis on which the following chapters are also based. We introduce a statistical formulation for stereo matching and analyze many characteristics mathematically using one-dimensional siganls to simplify the analysis. These analyses lead to a locally adaptive window which can control the size of the matching window adaptively for each position. Experimental results using synthesized signals are shown to demonstrate the effectiveness of the proposed method.

Chapter 3 deals with another important issue of stereo matching, i.e. how ambiguities in matching can be eliminated. We exploit multiple stereo image pairs with different baselines generated by a lateral displacement of a camera, and introduce a new evaluation function called SSSD-in-inverse-distance. It is shown both by analytical and experimental results that the resultant method can achieve two conflicting requirements in stereo matching, i.e. removing ambiguity and improving precision.

In **chapter 4**, the discussion in chaper 2 is extended to two dimension and a two-dimensional window control strategy is presented, in which the size and the shape (rectangle) of the window can be controlled. We show a detailed description of our stereo algorithm with an adaptive window using the multiple-baseline method introduced in chapter 3 as an initial estimation. The advantages of the method are shown by experimental results using both synthesized and real stereo images.

Chapter 5 presents color stereo matching. The effect of using color information in matching is analyzed mathematically and experimentally. Then, a medical application of the resultant color stereo algorithm is presented. With experiments using real stereo fundus images, it is shown that the stereo method could give useful information for diagnosing and monitoring glaucoma.

Chapter 6 summarizes the work of this thesis and presents directions of future research.

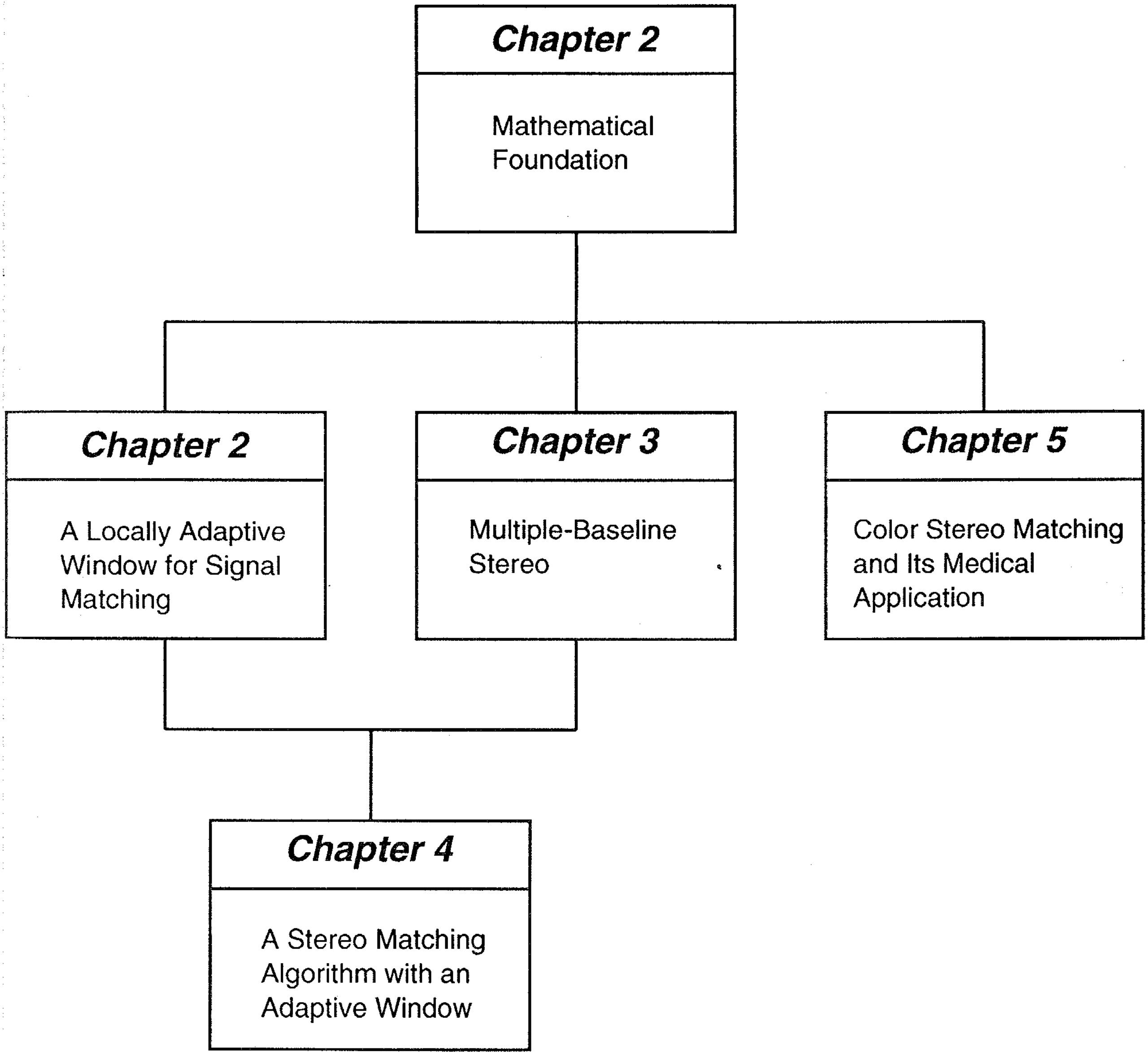

Chapter 2
Mathematical Foundation
Chapter 2
A Locally Adaptive Window for Signal Matching
Chapter 3
Multiple-Baseline Stereo
Chapter 5
Color Stereo Matching and Its Medical Application
Chapter 4
A Stereo Matching Algorithm with an Adaptive Window

Figure 1.5: Organization of this thesis

Chapter 2

Mathematical Foundation and a Locally Adaptive Window for Signal Matching

2.1 Introduction

In this chapter, we present mathematical foundations for our approach to stereo vision and analysis based on it. As a result, a locally adaptive window for matching is also introduced. Throughout this chapter, 1-D *signals* instead of 2-D *images* are used to facilitate the mathematical analysis and to discuss ideas in a more general manner. Now, let us begin by showing a typical problem which has not been systematically solved before.

One of the most basic methods for signal matching is calculating the sum of squared differences (SSD) between two signals over a certain window and locating the position of their minimum [MSK88] [FP86] [Woo83] [MKA73] [LOY73]. However, the precise localization of such minima, i.e., precise determination of disparity, is difficult and unreliable for two cases. The first is that when there is not enough signal variation, relative to noise, the SSD values are noisy and do not exhibit a clear and sharp minimum. The second case is that when there is too much disparity variation within the window, corresponding positions within the window are not equally shifted, and the differences are not calculated for the corresponding positions. As a result, the minimum SSD value may not occur at the correct match position. These properties pose conflicting requirements on the size of the window. Increasing the window size alleviates the first difficulty, since it increases the signal to noise ratio, while decreasing the window size alleviates the second difficulty, since it limits the SSD computation to only a local, likely relevant portion of the signals. Figure 2.1 illustrates this problem; it shows results of matching two intensity signals by calculating SSD's with different window sizes. The intensity signals are shown in figure 2.1 (a) and the true disparity is shown in figure 2.1

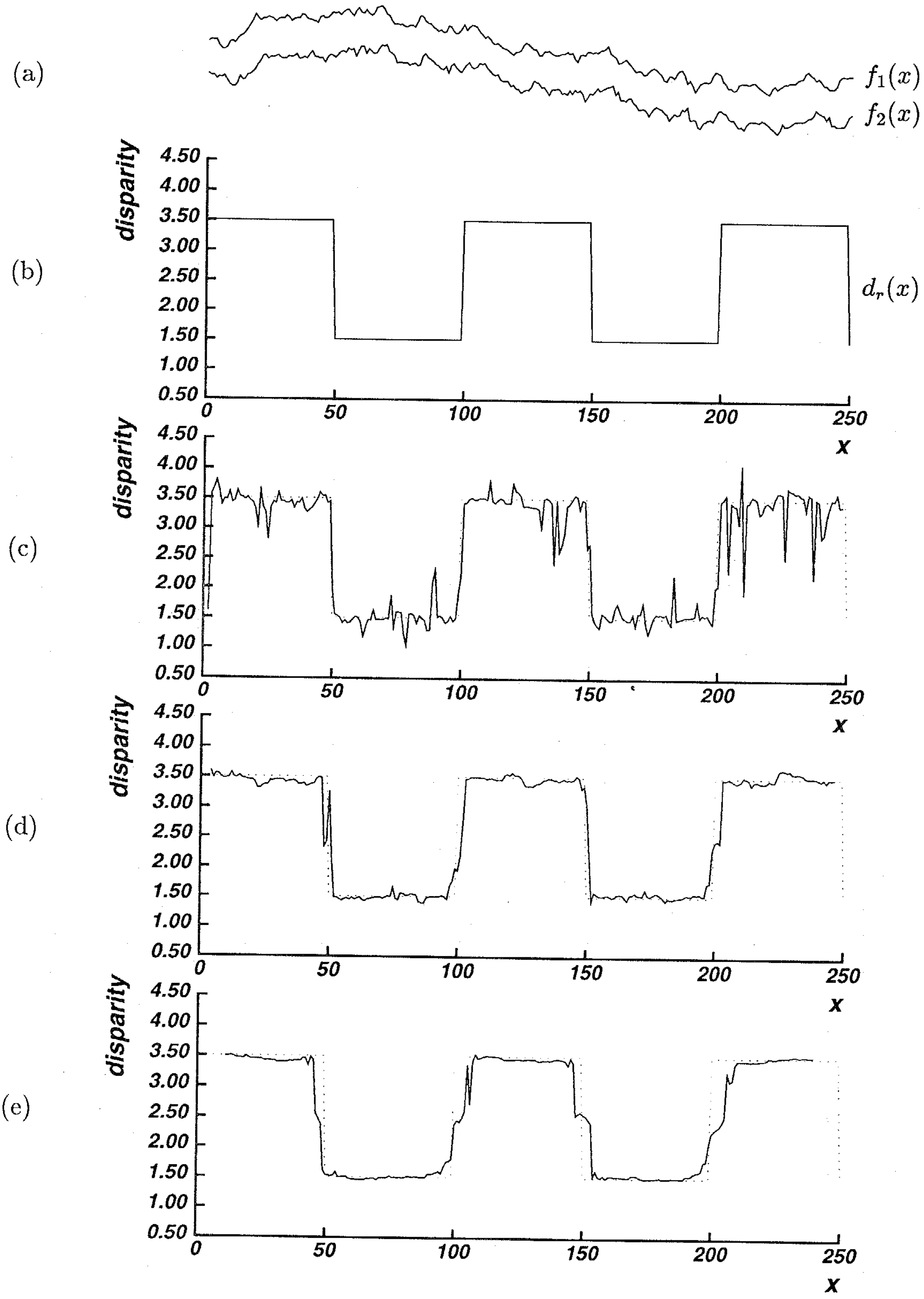


Figure 2.1: Matching by SSD with different window sizes. (a) Signals; (b) True disparity pattern; (c) Computed disparity, $w = 3$; (d) $w = 7$; (e) $w = 21$.

(b). We can observe that for the smaller window the computed disparity is noisy, but the disparity edges are sharp; while for the larger window the computed disparity is smoother, but the disparity edges are more blurred or misplaced. Clearly, we want a larger window for the flat regions (small disparity variation) and a smaller window near the disparity edges (large disparity variation).

There has been little work on a systematic method for automatically selecting a window size locally and adaptively for signal matching. In appearance, the problem seems to be very similar to that of smoothing a signal while detecting and preserving discontinuities, for which various powerful techniques have been developed including use of Markov Random Fields [Mar84], continuation methods [Ter86], weak continuity [BZ86], optimal amount of smoothing [Bou86], and robust M-estimation [BBW88]. These techniques, however, are for smoothing a signal itself by observing its properties, such as signal variation. The fundamental difference of our problem from smoothing, which makes it difficult, is that we have to deal with a disparity function which is embedded in the input signals. While the signal variation is measurable from the input, the disparity variation is not, since disparities are what we want to calculate.

As a solution to the problem, we propose to introduce a statistical model of the disparity pattern, which assumes that the disparity values within a window are generated by a random walk process starting from the value of the center point. Thus at each point within a window, its disparity is expected to be the same as that of the center point, but its variation is higher as the point is farther from the center. We employ this model to establish a link between the window size and the uncertainty of the computed disparity. This allows us to choose the window size that minimizes uncertainty in the disparity computed at each point.

In section 2.2 we present a mathematical framework necessary to discuss the statistical properties of disparity calculation based on SSD values in a window. Then in section 2.3, we generalize the previous formulation considering the disparity variation within a window and analyze how the disparity variation affects the estimation of disparity at the center point of the window. This analysis leads to an algorithm for signal matching with a locally adaptive window. In section 2.4, we analyze the uncertainty of the disparity for a few typical disparity patterns. Finally, we show experimental results in section 2.5 which demonstrate the advantages of the proposed method.

2.2 Modeling and Analysis of Signal Matching by Sum-of-Squared-Differences (SSD)

Signal matching algorithms that compute the sum of squared differences (SSD) of intensity patterns within a window implicitly assume that all points in the window have equal disparity; that is, all points within the window have been shifted in parallel. In this section, we will review the behavior of an iterative matching algorithm, similar to ones in [FP86] and [MO89], that makes this assumption. We will use this analysis to develop the new formulation in the next section.

2.2.1 Matching by SSD

Let $f_1(x)$ and $f_2(x)$ be signals that have a constant disparity d_r and that come from the same underlying intensity function with additive noise. Then, we can write;

$$f_1(x) - f_2(x + d_r) \;=\; n(x), \tag{2.1}$$

where $n(x)$ is Gaussian white noise such that

$$n(x) \sim N(0, 2\sigma_n^2). \tag{2.2}$$

Here, $N(a, b)$ denotes a Gaussian distribution with mean a and variance b and σ_n^2 is the power of the noise per image. The reason for $2\sigma_n^2$ is to account for noise added to two images.

If d_0 is an initial estimate of the disparity, by using the Taylor expansion,

$$f_2(x + d_r) \approx f_2(x + d_0) + \Delta d f_2'(x + d_0), \tag{2.3}$$

where $\Delta d = d_r - d_0$ is the correction. From (2.1) and (2.3),

$$f_1(x) - f_2(x + d_0) - \Delta d f_2'(x + d_0) = n(x). \tag{2.4}$$

Let

$$\begin{aligned} \psi_1(x) &= f_1(x) - f_2(x + d_0) \\ \psi_2(x) &= f_2'(x + d_0), \end{aligned} \tag{2.5}$$

then

$$\psi_1(x) - \Delta d \psi_2(x) = n(x). \tag{2.6}$$

For simplifying notation, suppose that the point whose disparity we would like to compute is at $x = 0$, and we set a window W around the candidate corresponding points of the two functions; that is at $x = 0$ for $f_1(x)$ and at $x = d_0$ for $f_2(x)$ as shown in figure 2.2.

Within the window, suppose that we select N sample points, $\xi_0, \xi_1, \ldots, \xi_{N-1}$, with equal intervals, and calculate the values of $\psi_1(\xi_i)$ and $\psi_2(\xi_i)$ from the sampled values of the intensity patterns. (Throughout this thesis, we will use ξ for a variable within a window and x for a general variable for functions.) Let us define η_i such that

$$\eta_i = \psi_1(\xi_i) - \Delta d \psi_2(\xi_i) \qquad i = 0, \ldots, N-1. \tag{2.7}$$

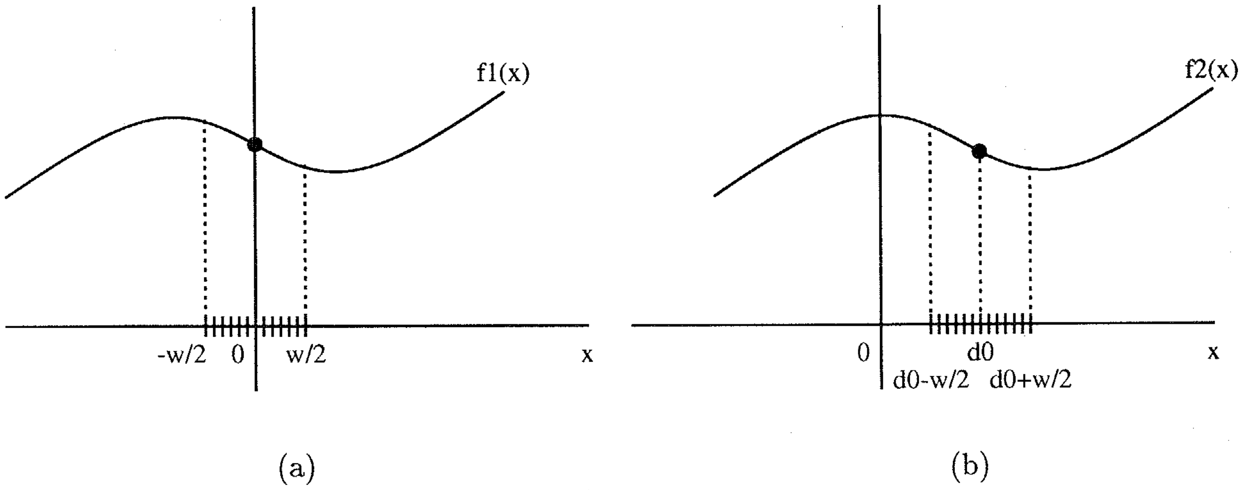


Figure 2.2: Signals and window settings

From equations (2.6) and (2.2), the conditional distribution function of η_i, given Δd, is

$$p(\eta_i|\Delta d) = \frac{1}{2\sqrt{\pi}\sigma_n} \exp\left(-\frac{(\psi_1(\xi_i) - \Delta d \psi_2(\xi_i))^2}{4\sigma_n^2}\right). \tag{2.8}$$

Since $n(x)$ is Gaussian white noise, the η_i's are independent of each other. So we get

$$p(\eta_0, \eta_1, \ldots, \eta_{N-1}|\Delta d) = \prod_{i=0}^{N-1} p(\eta_i|\Delta d). \tag{2.9}$$

From the continuous version of Bayes' theorem,

$$\begin{aligned} p(\Delta d|\eta_0, \eta_1, \ldots, \eta_{N-1}) &= \frac{p(\eta_0, \eta_1, \ldots, \eta_{N-1}|\Delta d)p(\Delta d)}{\int_{-\infty}^{\infty} p(\eta_0, \eta_1, \ldots, \eta_{N-1}|\Delta d)p(\Delta d)d(\Delta d)} \\ &= \frac{\prod_{i=0}^{N-1} p(\eta_i|\Delta d)p(\Delta d)}{\int_{-\infty}^{\infty} \prod_{i=0}^{N-1} p(\eta_i|\Delta d)p(\Delta d)d(\Delta d)}. \end{aligned} \tag{2.10}$$

Assuming no prior information, let $p(\Delta d) = constant$. Then we obtain

$$p(\Delta d|\eta_0, \eta_1, \ldots, \eta_{N-1}) = \frac{\prod_{i=0}^{N-1} p(\eta_i|\Delta d)}{\int_{-\infty}^{\infty} \prod_{i=0}^{N-1} p(\eta_i|\Delta d)d(\Delta d)}. \tag{2.11}$$

Substituting (2.8) into (2.11), we get

$$p(\Delta d|\eta_0, \eta_1, \ldots, \eta_{N-1}) = \frac{1}{\sqrt{2\pi}\sigma_{\Delta d}} \exp\left(-\frac{(\Delta d - \hat{\Delta d})^2}{2\sigma_{\Delta d}^2}\right), \tag{2.12}$$

where

$$
\begin{aligned}
\hat{\Delta d} &= \frac{\sum_{i=0}^{N-1}(\psi_1(\xi_i)\psi_2(\xi_i))}{\sum_{i=0}^{N-1}(\psi_2(\xi_i))^2} && (2.13)\\
\sigma^2_{\Delta d} &= \frac{2\sigma_n^2}{\sum_{i=0}^{N-1}(\psi_2(\xi_i))^2}. && (2.14)
\end{aligned}
$$

That is, the conditional probability density function of Δd, given the observed intensity pair, becomes a Gaussian distribution with mean $\hat{\Delta d}$ and variance $\sigma^2_{\Delta d}$. The $\hat{\Delta d}$ and $\sigma^2_{\Delta d}$ derived here are the same as those obtained from maximum likelihood estimation and standard error propagation techniques, which have been presented by multiple researchers including [MO89][FP86][RGH80].

2.2.2 Window Size and Uncertainty

Now, let

$$
\Delta\xi = \frac{w}{N} \qquad (2.15)
$$

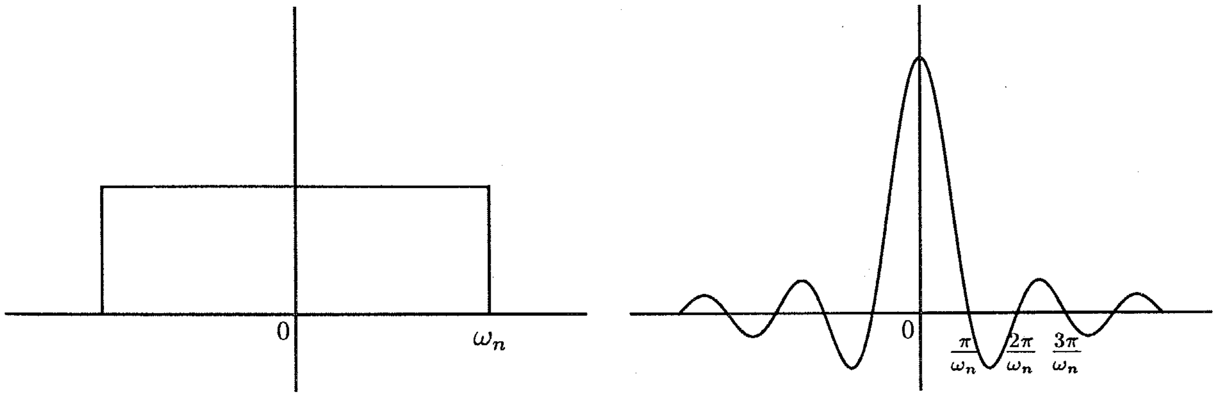
be the sampling interval, where w is the size of the window. Multiplying the numerators and the denominators of equations (2.13) and (2.14) by $\Delta\xi$, we obtain

$$
\begin{aligned}
\hat{\Delta d} &= \frac{\sum_{i=0}^{N-1}(\psi_1(\xi_i)\psi_2(\xi_i))\Delta\xi}{\sum_{i=0}^{N-1}(\psi_2(\xi_i))^2\Delta\xi} && (2.16)\\
\sigma^2_{\Delta d} &= \frac{2\sigma_n^2\Delta\xi}{\sum_{i=0}^{N-1}(\psi_2(\xi_i))^2\Delta\xi}. && (2.17)
\end{aligned}
$$

As $N \to \infty$,

$$
\begin{aligned}
\hat{\Delta d} &= \frac{\int_{-\frac{w}{2}}^{\frac{w}{2}}(\psi_1(\xi)\psi_2(\xi))d\xi}{\int_{-\frac{w}{2}}^{\frac{w}{2}}(\psi_2(\xi))^2 d\xi} && (2.18)\\
\sigma^2_{\Delta d} &= \frac{2\sigma_n^2\Delta\xi}{\int_{-\frac{w}{2}}^{\frac{w}{2}}(\psi_2(\xi))^2 d\xi} \quad \to 0. && (2.19)
\end{aligned}
$$

This result is somewhat counterintuitive. Here, it appears as if the variance of the estimated Δd could be made arbitrarily small by sampling as many points as we need. This would only be possible if $n(x)$ were actually *white*. However, in practice it is not, and we must model that $n(x)$ is *band-limited* Gaussian white noise. Figures 2.3 (a) and (b) show the power spectral density function and the autocorrelation function of band-limited white noise. We can see that points evenly spaced by $\frac{\pi}{\omega_n}$ are uncorrelated, where ω_n is the maximum frequency of $n(x)$. They are also independent of each other, since uncorrelated Gaussian variables



(a) Power spectral density function (b) Autocorrelation function

Figure 2.3: Band-limited white noise

are statistically independent [dC86]. Therefore, the sampling interval $\Delta\xi$ that keeps the samplings independent of each other and makes $\sigma^2_{\Delta d}$ the smallest is

$$\Delta\xi = \frac{\pi}{\omega_n}. \tag{2.20}$$

With this sampling period we can obtain

$$\sigma^2_{\Delta d} = \frac{2\pi\sigma_n^2}{\omega_n \int_{-\frac{w}{2}}^{\frac{w}{2}} (\psi_2(\xi))^2 d\xi}. \tag{2.21}$$

Substituting equation (2.5) into equations (2.18) and (2.21), we get

$$\hat{\Delta d} = \frac{\int_{-\frac{w}{2}}^{\frac{w}{2}} (f_1(\xi) - f_2(\xi + d_0)) f_2'(\xi + d_0) d\xi}{\int_{-\frac{w}{2}}^{\frac{w}{2}} (f_2'(\xi + d_0))^2 d\xi} \tag{2.22}$$

$$\sigma^2_{\Delta d} = \frac{2\pi\sigma_n^2}{\omega_n \int_{-\frac{w}{2}}^{\frac{w}{2}} (f_2'(\xi + d_0))^2 d\xi}. \tag{2.23}$$

These equations show that, given the signals $f_1(x)$ and $f_2(x)$ and an initial estimate d_0, we can obtain a disparity correction $\hat{\Delta d}$ and its uncertainty $\sigma^2_{\Delta d}$. We can revise the disparity estimate by replacing d_0 with $d_0 + \hat{\Delta d}$, and iterate the process. The results shown in figure 2.1 were obtained with this method.

Equation (2.23) indicates two characteristics. First, the larger the absolute value of the derivative of the signal is, the smaller the uncertainty. Second, the larger the window size is, the smaller the uncertainty. The first characteristic is intuitive. That is, the more fluctuation in the intensity pattern, the more reliable the matching. The second characteristic is also understandable because a large window can average out the effect of noise. However, this is true *only because* we have assumed that the two signals have equal disparity everywhere. If the disparity varies from position to position, as happens in practice especially near the disparity edges, a large window is likely to compare intensities which are not actually at corresponding positions, even when the center point of the window is at the corresponding position. As a result, a larger window has the effect that disparity edges are blurred and misplaced, while a smaller window yields sharp but noisy results. This is the phenomena that we have observed in figure 2.1. Clearly, there should be an appropriate window size for each position.

2.3 Generalized Formulation Considering Disparity Variation

In the previous section, we analyzed an iterative matching method with the assumption that the disparity within the window is constant. If this assumption were true, a larger window would be better. However, in practice the actual disparity varies from point to point, and too large a window is actually harmful. An appropriate window size must exist depending on local intensity *and* disparity patterns.

2.3.1 Assumption of Non-constant Disparity

In this section, we let the disparity be a function of position, i.e. $d_r(x)$. Therefore instead of equation (2.1), we have

$$f_1(x) - f_2(x + d_r(x)) = n(x). \tag{2.24}$$

As before, suppose that we want to compute the disparity at $x = 0$, i.e., the value $d_r(0)$. Using the Taylor expansion,

$$f_2(\xi + d_r(\xi)) \approx f_2(\xi + d_r(0)) + (d_r(\xi) - d_r(0))f_2'(\xi + d_r(0)). \tag{2.25}$$

Substituting equation (2.25) into equation (2.24), we get

$$f_1(\xi) - f_2(\xi + d_r(0)) = (d_r(\xi) - d_r(0))f_2'(\xi + d_r(0)) + n(\xi). \tag{2.26}$$

Comparing this with equation (2.1), we observe an additional term $e(\xi)$,

$$e(\xi) = (d_r(\xi) - d_r(0))f_2'(\xi + d_r(0)). \tag{2.27}$$

The interpretation of this term is the following: suppose that we place the center point of a window at the corresponding points of both signals and compare the intensity values at each position ξ within the window. Then $e(\xi)$ represents the *apparent* intensity difference between the two signals caused not by mismatch, but by the difference between the disparity $d_r(\xi)$ at that position and the disparity $d_r(0)$ at the center point of the window. Figure 2.4 illustrates this situation. We can see that although the centers of the windows are placed at the exactly corresponding points ($x = 0$ and $x = d_r(0)$), other points in the window do not correspond properly.

2.3.2 A Statistical Model of Disparity

In order to advance our analysis of matching, we must consider the effect of $e(\xi)$ on the disparity estimate. The difficulty is that $e(\xi)$ includes $d_r(\xi)$, which is what we wish to calculate. As a solution, we introduce the following statistical model for the disparity $d_r(\xi)$ within a window:

$$d_r(\xi) - d_r(0) \sim N(0, \alpha_d|\xi|), \tag{2.28}$$

where α_d is a constant that represents fluctuation of the disparity. This model assumes that the difference in disparity between a point ξ in the window and the center point of the window has a zero-mean Gaussian distribution with variance proportional to the distance between these points. In other words, the farther the point is from the center, the more likely it is that it will have a different disparity (though the expected value of the disparity is the same). This statistical model can be shown equivalent to assuming that $d_r(\xi)$ is generated by Brownian motion[1] (refer to [BN68][Vos87]). We also assume that the intensity derivatives $f_2'(\xi)$ within a window follow a zero-mean Gaussian white distribution[2]and are independent of $d_r(\xi)$.

From these assumptions and equation (2.27), we can show (see Appendix A) that $e(\xi)$ can be approximated by Gaussian white noise such that

$$e(\xi) \sim N(\mu_e, \sigma_e^2), \tag{2.29}$$

where

$$\begin{aligned}
\mu_e &= E[e(\xi)] \\
&= E[d_r(\xi) - d_r(0)]E[f_2'(\xi + d_r(0))] \\
&= 0 & (2.30) \\
\sigma_e^2 &= E[(e(\xi))^2] \\
&= E[(d_r(\xi) - d_r(0))^2]E[(f_2'(\xi + d_r(0)))^2] \\
&= \alpha_f \alpha_d |\xi|, & (2.31)
\end{aligned}$$

[1] More generally, we can assume $d_r(\xi)$ to be a fractal. This corresponds to choosing a different degree of $|\xi|$ in the variance in (2.28). The Brownian motion is the simplest case in which the degree is 1. However, our preliminary experiments have shown no noticeable advantage in using a general fractal assumption.

[2] This is also equivalent to assuming the pattern $f_2(x)$ to be the result of Brownian motion.

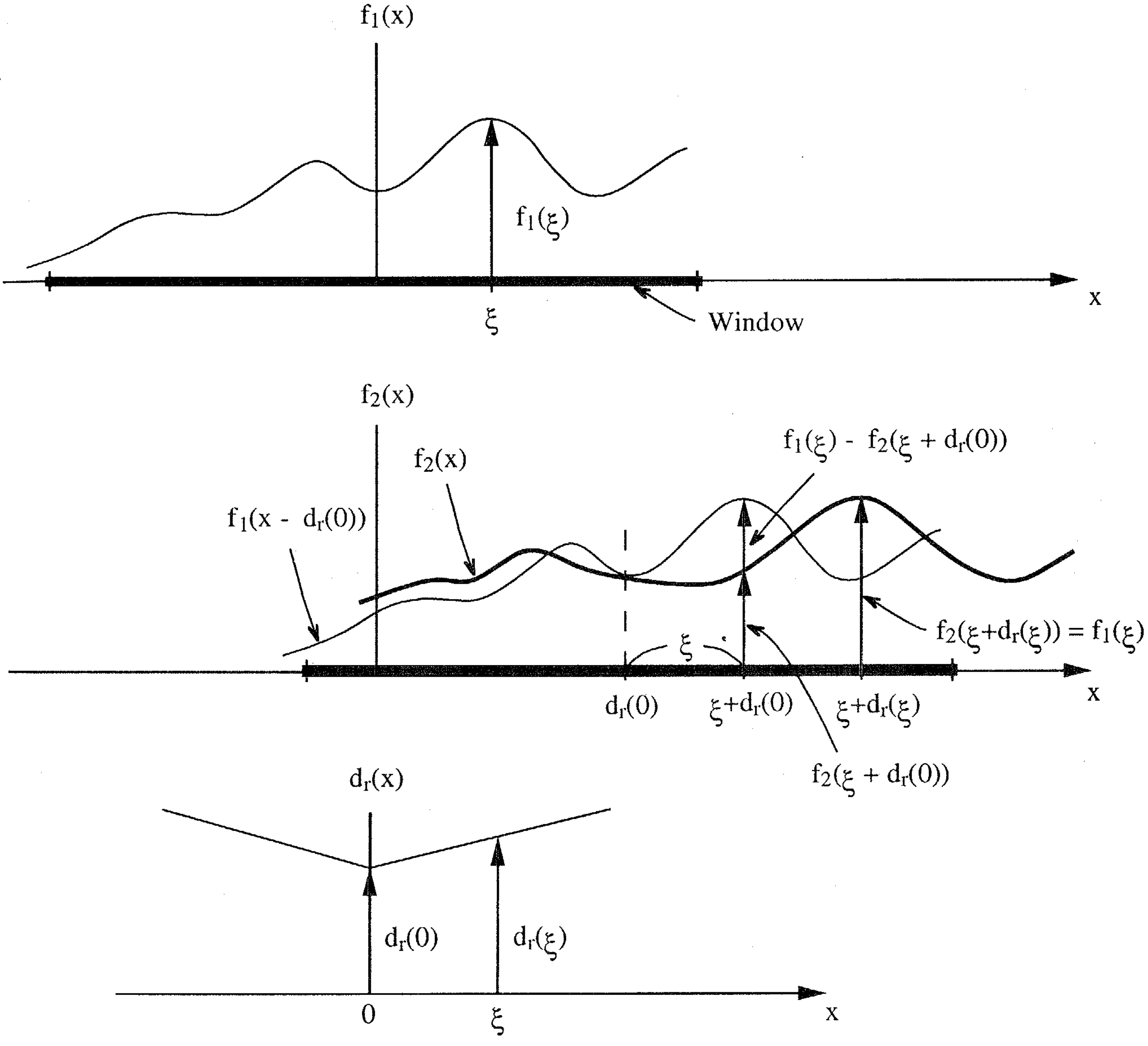


Figure 2.4: Illustration of $e(\xi)$. The graph at the top shows $f_1(x)$; the middle one, $f_2(x)$ (the thicker curve) with $f_1(x)$ shifted by $d_r(0)$ (the thinner curve); the bottom one, $d_r(x)$. The region indicated by the very thick lines on the axes indicate the region covered by the window.

and

$$\alpha_f = E[(f_2'(\xi + d_r(0)))^2]. \tag{2.32}$$

From equations (2.26), (2.27), (2.30), and (2.31), we obtain

$$\begin{aligned} f_1(\xi) - f_2(\xi + d_r(0)) &= e(\xi) + n(\xi) \\ &\equiv n_s(\xi), \end{aligned} \tag{2.33}$$

where $n_s(\xi)$ is Gaussian white noise such that

$$n_s(\xi) \sim N[0, \sigma_s^2(\xi)], \tag{2.34}$$

and

$$2\sigma_n^2 + \alpha_f \alpha_d |\xi| = \sigma_s^2(\xi). \tag{2.35}$$

Intuitively, $n_s(\xi)$ can be considered as the total noise added to signals, whose variance is the sum of a constant term and a term proportional to $|\xi|$. The first term comes from the noise added to the signals. The second term comes from an *uncertain local support*. That is, while the surrounding points of the center point in the window are used to support the matching for the center point, the information from these points adds some uncertainty, too, because of the disparity difference between the center point and the supporting points. This uncertainty is represented as if additional noise were added whose power is proportional to the distance from the center point in the window. If the disparity is constant over the window, i.e. $\alpha_d = 0$, this additional noise is zero. As the disparity varies more in the window (i.e., larger α_d), this additional noise becomes larger and the information from the supporting points becomes more uncertain. Also, note that the noise effect of the disparity variation is amplified by a factor of α_f, that is, by the amount of the intensity variation. This is because wrong correspondences due to disparity variation affect more severely when the intensity variation is higher.

More discussions about our statistical model of disparity (equation (2.28)) are presented in section 4.2.2 where we compare it with the assumptions about local support used in stereo algorithms.

2.3.3 Uncertainty of Estimation as a Function of Signal and Disparity Variations

Now, as we obtained equation (2.4) from equation (2.1) for iterative estimation, we can obtain from equation (2.33) the following:

$$f_1(\xi) - f_2(\xi + d_0) - \Delta d f_2'(\xi + d_0) = n_s(\xi). \tag{2.36}$$

Dividing both sides of this equation by $\sigma_s(\xi)$,

$$\frac{f_1(\xi) - f_2(\xi + d_0) - \Delta d f_2'(\xi + d_0)}{\sigma_s(\xi)} = \frac{n_s(\xi)}{\sigma_s(\xi)} \sim N(0, 1). \tag{2.37}$$

By letting

$$\phi_1(\xi) = \frac{f_1(\xi) - f_2(\xi + d_0)}{\sigma_s(\xi)} \tag{2.38}$$

$$\phi_2(\xi) = \frac{f_2'(\xi + d_0)}{\sigma_s(\xi)}, \tag{2.39}$$

we have

$$\phi_1(\xi) - \Delta d \phi_2(\xi) \sim N(0, 1), \tag{2.40}$$

which corresponds to equation (2.6) for the case of constant disparity within a window. So, instead of equation (2.18) and (2.21), we obtain

$$\hat{\Delta d} = \frac{\int_{-\frac{w}{2}}^{\frac{w}{2}} (\phi_1(\xi)\phi_2(\xi)) d\xi}{\int_{-\frac{w}{2}}^{\frac{w}{2}} (\phi_2(\xi))^2 d\xi} \tag{2.41}$$

$$\sigma_{\Delta d}^2 = \frac{\pi}{\omega_n \int_{-\frac{w}{2}}^{\frac{w}{2}} (\phi_2(\xi))^2 d\xi}. \tag{2.42}$$

Substituting equations (2.38), (2.39), and (2.35), we finally obtain

$$\hat{\Delta d} = \frac{\int_{-\frac{w}{2}}^{\frac{w}{2}} \frac{(f_1(\xi) - f_2(\xi + d_0)) f_2'(\xi + d_0)}{2\sigma_n^2 + \alpha_f \alpha_d |\xi|} d\xi}{\int_{-\frac{w}{2}}^{\frac{w}{2}} \frac{(f_2'(\xi + d_0))^2}{2\sigma_n^2 + \alpha_f \alpha_d |\xi|} d\xi} \tag{2.43}$$

$$\sigma_{\Delta d}^2 = \frac{\pi}{\omega_n \int_{-\frac{w}{2}}^{\frac{w}{2}} \frac{(f_2'(\xi + d_0))^2}{2\sigma_n^2 + \alpha_f \alpha_d |\xi|} d\xi}. \tag{2.44}$$

These are the equations for the correction of the disparity estimate and uncertainty of the correction under the assumption of non-constant disparity within a window. If $\alpha_d = 0$, that is, if all points have the same disparity, then these equations reduce to equations (2.22) and (2.23) in the previous section. The values of α_f and α_d depend on the intensity and disparity patterns respectively, and they change locally. Since they include $d_r(x)$, we cannot obtain their true values. Instead we calculate their estimates by using the current estimates of $\hat{d}_r(x)$. From equations (2.32) and (2.28).

$$\hat{\alpha_f} = \frac{1}{w} \int_{-\frac{w}{2}}^{\frac{w}{2}} (f_2'(\xi + \hat{d}_r(0)))^2 d\xi \tag{2.45}$$

$$\hat{\alpha_d} = \frac{1}{w} \int_{-\frac{w}{2}}^{\frac{w}{2}} \frac{(\hat{d}_r(\xi) - \hat{d}_r(0))^2}{|\xi|} d\xi. \tag{2.46}$$

These are the key values that furnish a link between the uncertainty of disparity estimate and the size of a window for matching.

2.3.4 Iterative Algorithm with a Locally Adaptive Window

We introduce the following algorithm that uses a locally adaptive window based on the preceding analysis:

1. Set $\hat{d}_r(x)$ to an initial estimate $d_0(x)$.
2. For a window size w, compute $\hat{\alpha_f}$ and $\hat{\alpha_d}$ from the signals $f_1(x)$, $f_2(x)$, and $\hat{d}_r(x)$ by using equations (2.45) and (2.46).
3. Compute a correction $\hat{\Delta d}$ and an uncertainty of the correction $\sigma^2_{\Delta d}$ by using equations (2.43) and (2.44).
4. Repeat steps 2 and 3 for various window sizes and use the one that provides the estimate with the smallest uncertainty.
5. Update $\hat{d}_r(x)$ by the amount $\hat{\Delta d}$.
6. Repeat steps 2 through 5 until convergence or up to a certain number of iterations.

2.4 Uncertainty Analysis for Typical Disparity Patterns

In the previous section, we introduced a statistical model for the disparity $d_r(x)$. We then derived equations that relate the disparity estimate and its uncertainty to intensity and disparity variations within the window. The value $\sigma^2_{\Delta d}$ in equation (2.44) is the variance of the estimated disparity or the uncertainty of the estimation. The adaptive window selection method proposed in the previous section chooses the window size that minimizes this value. In this section, we will analyze the behavior of $\sigma^2_{\Delta d}$ for a few typical disparity patterns: constant, linear, and step. By examining how $\sigma^2_{\Delta d}$ changes with the window size, we can tell how the proposed method will work for those typical cases.

Even when we fix the disparity pattern in the window, the value of $\sigma^2_{\Delta d}$ still depends on a particular intensity pattern $f_2(x)$. Therefore we will instead use the expected values of $\sigma^2_{\Delta d}$ over an ensemble of intensity patterns whose derivatives $f_2'(x)$ follow a zero-mean Gaussian white distribution such that

$$f_2'(x) \sim N(0, \alpha_f). \tag{2.47}$$

In the following analysis, we assume that α_f is constant, i.e. the intensity fluctuates equally over the whole signal. Then, we have

$$\begin{aligned} E\left[\int_{-\frac{w}{2}}^{\frac{w}{2}} \frac{(f_2'(x+d_0))^2}{2\sigma_n^2+\alpha_f\alpha_d|\xi|}dx\right] &= \alpha_f \int_{-\frac{w}{2}}^{\frac{w}{2}} \frac{1}{2\sigma_n^2+\alpha_f\alpha_d|\xi|}dx \\ &= \frac{2}{\alpha_d}\log\left(1+\frac{\alpha_d\alpha_f w}{4\sigma_n^2}\right). \end{aligned} \tag{2.48}$$

Substituting this into equation (2.44), the expected value of $\sigma_{\Delta d}^2$ is

$$E[\sigma_{\Delta d}^2] = \frac{\pi\alpha_d}{2\omega_n \log\left(1+\frac{\alpha_d r^2 w}{4}\right)}, \tag{2.49}$$

where

$$r = \frac{\sqrt{\alpha_f}}{\sigma_n}. \tag{2.50}$$

Since α_f is the variation in the first derivative of the intensity signal, and σ_n^2 is the power of additive noise, r represents *the ratio of the intensity fluctuation (i.e. signal) to the noise.* This is an important parameter in matching two intensity patterns.

2.4.1 Constant Disparity

Suppose all points have an equal disparity within a window. That is,

$$\alpha_d = 0. \tag{2.51}$$

Then, the measure of uncertainty is

$$\begin{aligned} \sqrt{E[\sigma_{\Delta d}^2]} &= \sqrt{\lim_{\alpha_d\to 0} \frac{\pi\alpha_d}{2\omega_n\log\left(1+\frac{\alpha_d r^2 w}{4}\right)}} \\ &= \frac{1}{r}\sqrt{\frac{2\pi}{\omega_n w}}. \end{aligned} \tag{2.52}$$

Here, we can see that the uncertainty is inversely proportional to the ratio r and to the square root of window size $\sqrt{w}$. Therefore our adaptive window method will choose the largest window size allowed. This is consistent with the observation made in section 2.2.

2.4.2 Linear Disparity

Suppose that the disparity changes linearly within a window as shown in figure 2.5; that is, the window includes a slanted surface:

$$d_r(\xi) = a\xi + b, \tag{2.53}$$

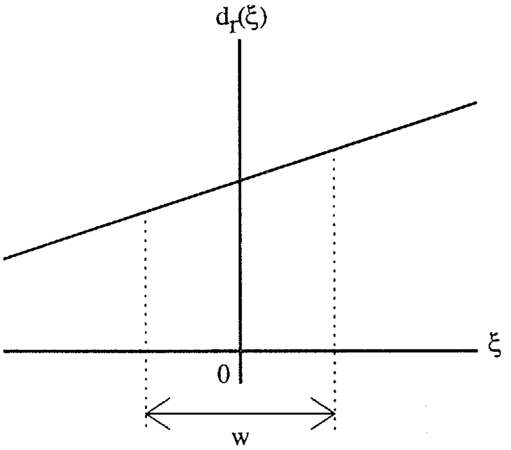


Figure 2.5: Linear disparity pattern

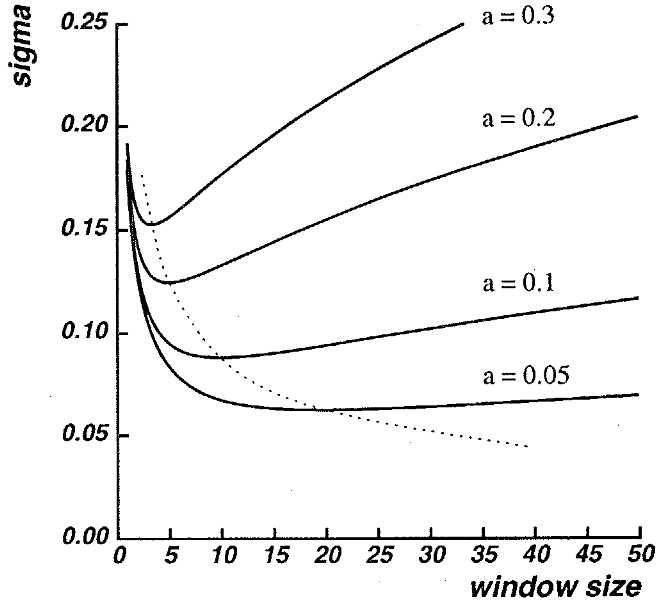


Figure 2.6: Uncertainty vs. window size w for a linear disparity pattern (r=8).

where a and b are constants. Then,

$$\alpha_d = \frac{a^2 w}{4}. \tag{2.54}$$

Substituting this into equation (2.49), the measure of uncertainty is

$$\sqrt{E[\sigma^2_{\Delta d}]} = \frac{|a|}{2}\sqrt{\frac{\pi w}{2\omega_n \log\left(1+\frac{a^2r^2w^2}{16}\right)}}. \tag{2.55}$$

Figure 2.6 shows how the uncertainty changes with the window size w for a few values of the slope a. We observe that for a fixed window size, the steeper the slope is, the larger the uncertainty. There is a minimum of uncertainty for each slope a. The dotted line connects those minima. If we denote by w_{opt} the window size that gives the minimum uncertainty, then

$$w_{opt} = \frac{4K}{|a|r}, \tag{2.56}$$

where K is a constant that satisfies

$$\log(1+K^2) - \frac{2K^2}{1+K^2} = 0. \tag{2.57}$$

We can see that given the ratio r, there is an optimal window size which is inversely proportional to the absolute value of the slope a.

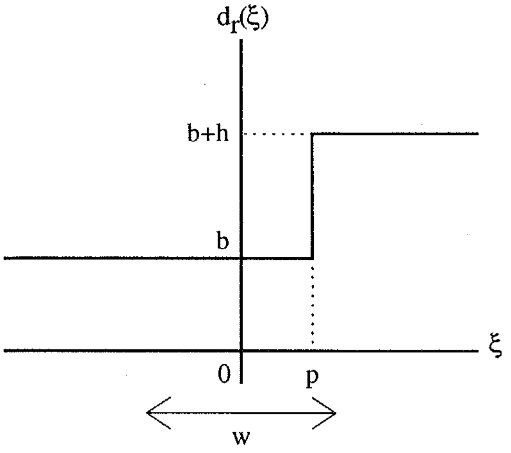


Figure 2.7: Step disparity pattern

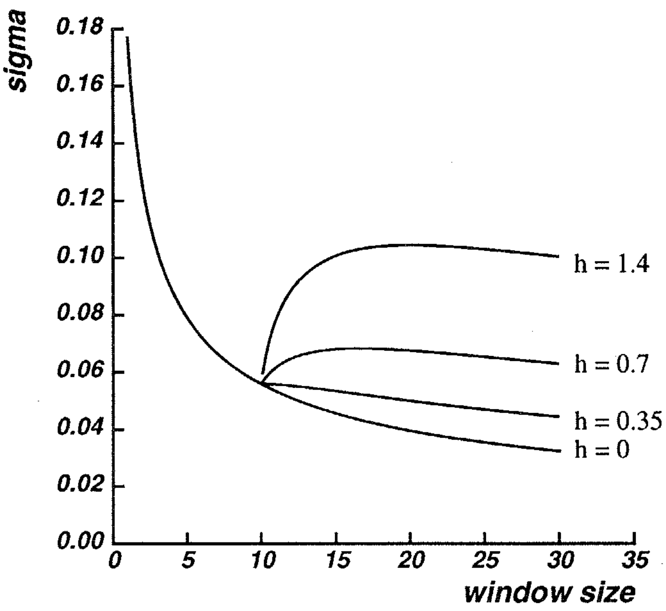


Figure 2.8: Uncertainty vs. window size w for a step disparity pattern ($p = 5$, $r = 8$).

2.4.3 Step Pattern of Disparity

Suppose that there is a disparity jump within the window, that is, the window includes an occluding edge. Figure 2.7 shows a step pattern of disparity, where the step of height h is positioned p away from the window center:

$$d_r(\xi) = \begin{cases} b & \xi < p \\ b+h & \xi \geq p. \end{cases} \tag{2.58}$$

Then,

$$\alpha_d = \begin{cases} 0 & \frac{w}{2} < p \\ \frac{h^2}{w} \log \frac{w}{2p} & \frac{w}{2} \geq p \end{cases} . \tag{2.59}$$

Substituting this into equation (2.49),

$$\sqrt{E[\sigma^2_{\Delta d}]} = \begin{cases} \frac{1}{r}\sqrt{\frac{2\pi}{\omega_n w}} & \frac{w}{2} < p \\ h\sqrt{\frac{\pi \log \frac{w}{2p}}{2\omega_n w \log(1+\frac{h^2 r^2}{4} \log \frac{w}{2p})}} & \frac{w}{2} \geq p \end{cases} . \tag{2.60}$$

Figure 2.8 shows plots of the uncertainty versus the window size w for different step sizes h. In general, the uncertainty has a local minimum at $w = 2p$, if $h^2 > 8/r^2$ (in the case of the figure, $h > 0.35$). That is, the uncertainty will reach a minimum when the window is

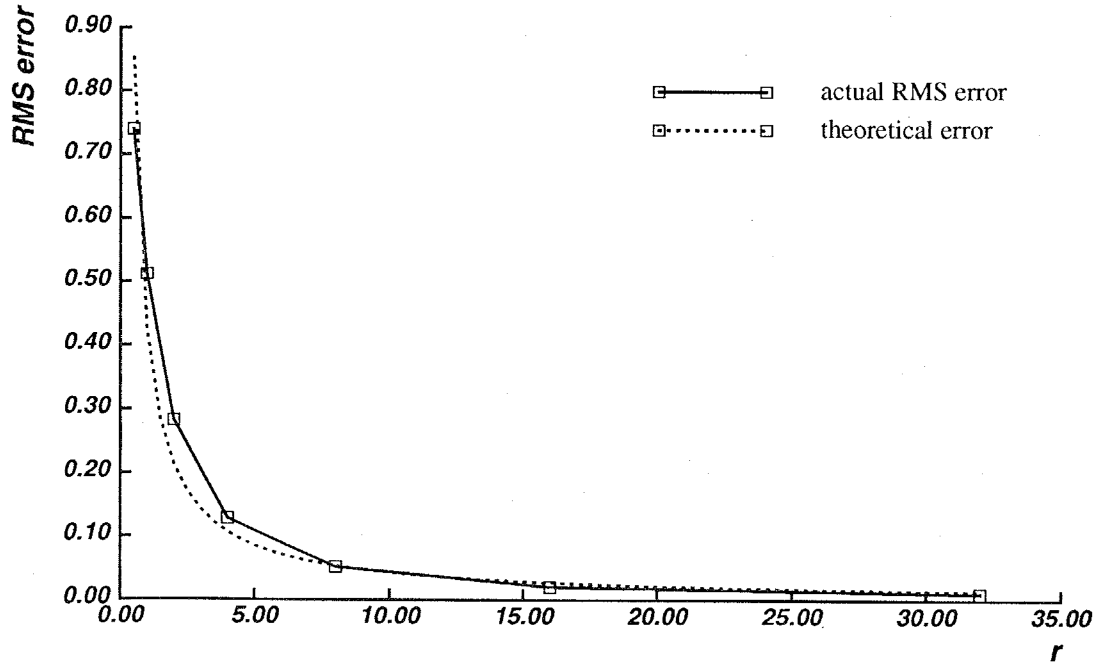


Figure 2.9: RMS error vs. r (constant disparity pattern).

just about to cover the step, unless the disparity step is very small relative to the intensity noise level. Therefore, the proposed adaptive window method will choose the largest window that doesn't cross the disparity edge.

2.5 Experimental Results

In this section, we show some experimental results of matching using synthesized signals and scanlines of real stereo images. The underlying intensity pattern of the synthesized signals is created by Brownian motion so that its derivatives are Gaussian white noise. The pattern is then transformed into the two intensity signals by adding Gaussian white noise and transforming according to an embedded disparity function $d_r(x)$. The signals which have been used in Figure 2.1(a) are examples where the disparity function is a square wave.

First we examine the case where the disparity pattern is constant ($d_r(x) = 2.5$). Figure 2.9 is a plot of the RMS error of the computed disparity vs. the ratio r. Figure 2.10 shows the RMS error as a function of the window size w. The dotted lines show the theoretical values from equation (2.52) in the previous section.

Next we look at the case where the disparity changes linearly. In figure 2.11, the solid curve shows, for various window sizes, the actual RMS errors of the disparity values obtained by using equation (2.43) in the new formulation in section 2.3. The dashed line shows the

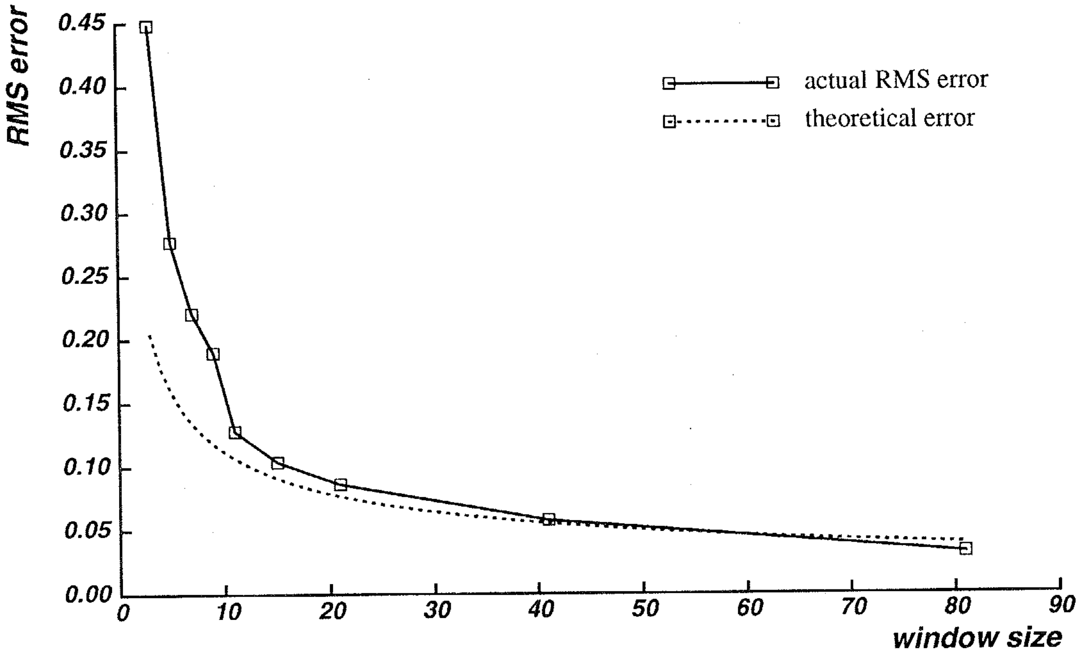


Figure 2.10: RMS error vs. window size (constant disparity pattern).

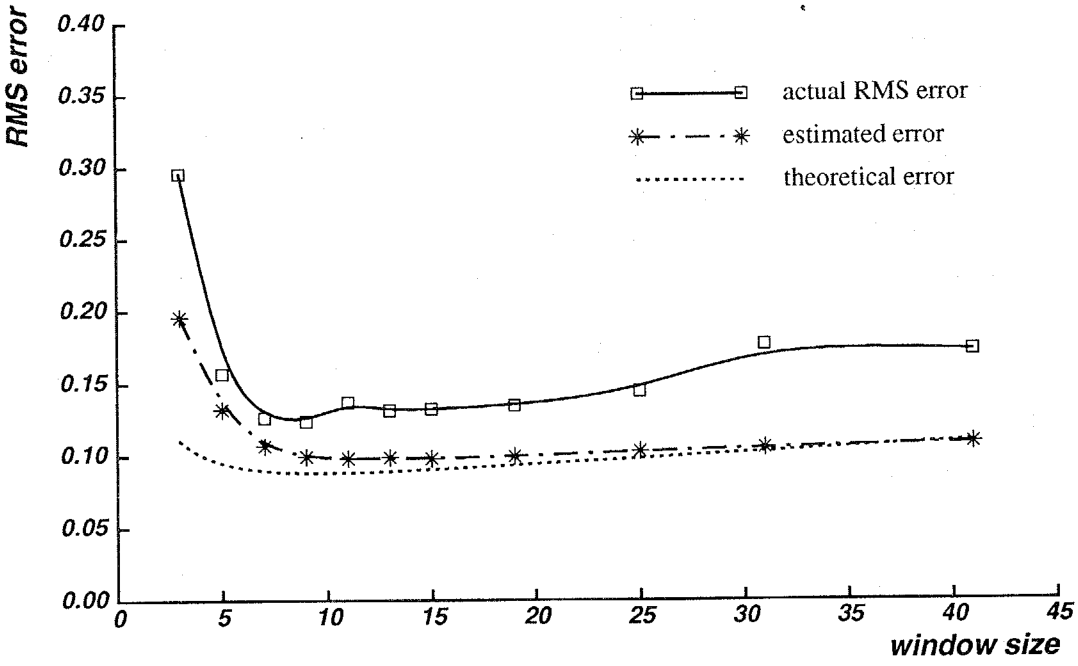


Figure 2.11: RMS error vs. window size (linear disparity pattern).

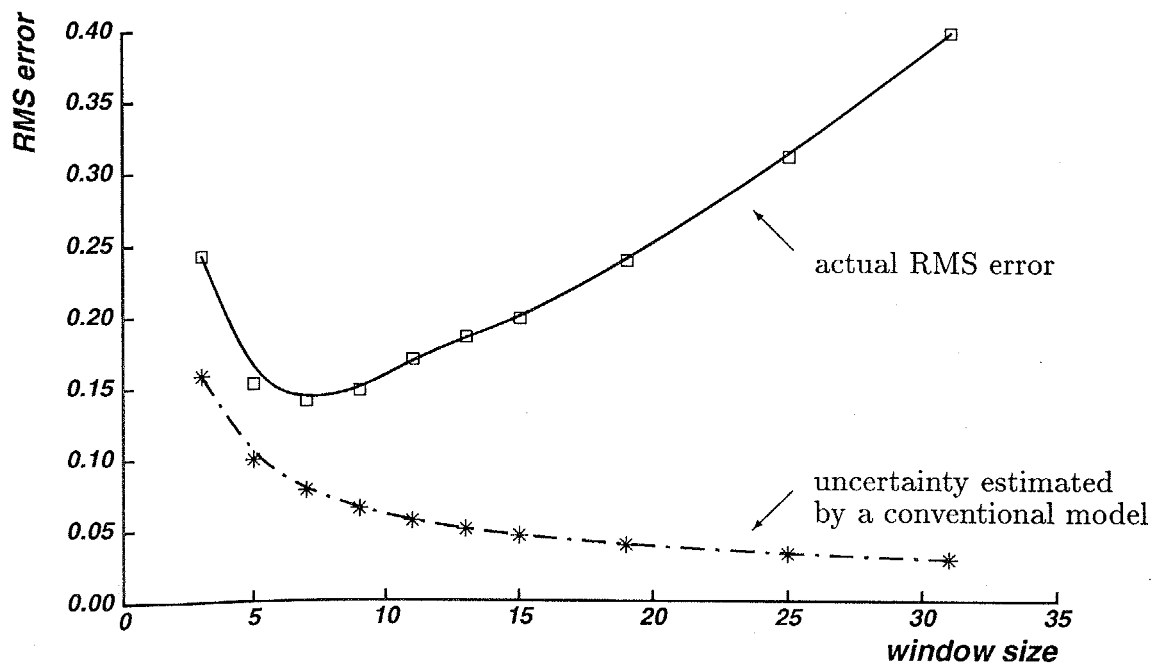


Figure 2.12: RMS error vs. window size (linear disparity pattern) by a conventional method.

estimated uncertainty $\sigma_{\Delta d}$ calculated by using equation (2.44). The dotted line shows the theoretically expected error from equation (2.55). For comparison, figure 2.12 shows the results from the conventional formulation with the (implicit) constant disparity assumption, i.e. equation (2.22) in section 2.2: the actual RMS errors are shown by the solid curve, together with the estimated uncertainty from equation (2.23). First of all, by comparing the solid curves in figures 2.11 and 2.12 we can see that the computed disparity from the new formulation has less RMS error than the conventional method over all of the window sizes. More importantly, the new formulation can give consistently better estimates for the uncertainty in computed disparity; that is, the estimated uncertainty is closer to the actual RMS error. This provides a solid foundation for automatic selection of an appropriate window size. In the conventional formulation, the estimated uncertainty is far from the actual RMS error as shown in figure 2.12. This is due to the inappropriate assumption of constant disparity as we mentioned before. Figure 2.13 shows the selected window sizes that give the minimum uncertainty estimation for each slope a. The dotted line shows the theoretically optimal window size from equation (2.56).

For the third experiment, figure 2.14 shows the computed disparities for a square-wave disparity pattern. In figure 2.1, we showed how the window size affects matching results. Comparing figure 2.14 with figure 2.1, clear improvement can be seen. Figure 2.14 (b) shows the window size that was actually selected by the algorithm (minimum and maximum sizes were limited to 3 and 21, respectively). We can observe that in general the method selects

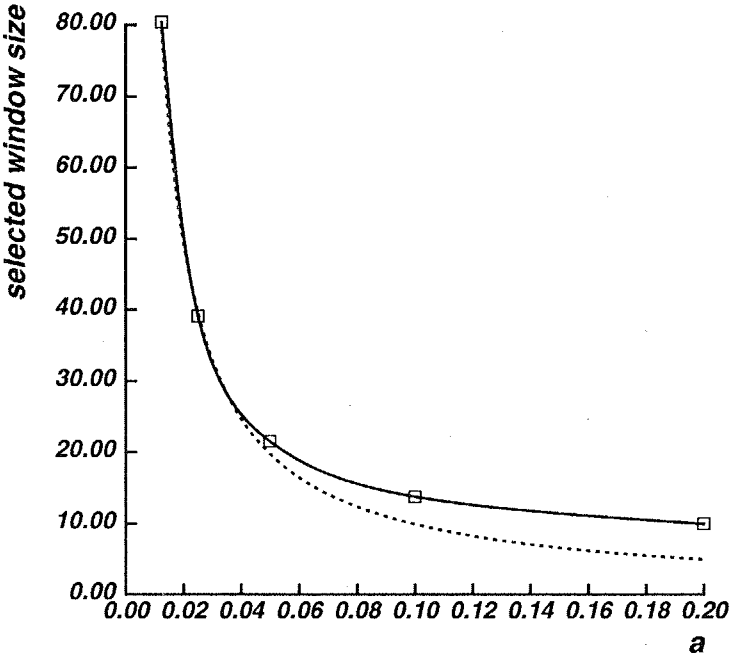


Figure 2.13: Selected window size vs. slope a.

a small window near disparity edges according to the distance to the edges and a larger window in the flat regions.

We also tested a more complicated disparity pattern, shown in figure 2.15 (b), which includes slopes, steps, and a curve. Ten experiments were performed using different but statistically identical intensity patterns and noise for the same disparity pattern to evaluate the robustness of the method. One of the ten signal pairs used is shown in figure 2.15 (a). Figure 2.15 (c) shows the computed disparities by the matching algorithm with a locally adaptive window: results for the ten experiments are overlaid. Again for comparison, the test signals were also subjected to the conventional matching method with a fixed size window. Figure 2.16 is a plot of the RMS errors over the whole pattern when various fixed window sizes are used. All of these RMS errors are larger than the RMS error (0.10 shown by the dotted line in the figure) achieved by the locally adaptive window method.

Figure 2.15 (d) shows the selected window sizes (averaged over ten experiments). Note that the window size is adaptively chosen near the step of disparity (e.g. $x = 50, 150, 190, 230, 260$), around the round shape of the disparity (i.e. $50 < x < 150$), and over the different disparity slope (i.e. $300 < x < 450$).

Figure 2.17 shows a scatter plot of the actual error vs. the estimated uncertainty $\sigma_{\Delta d}$ for all the points in figure 2.15; the □s indicate the RMS values of actual error. Despite the fact that they are the mixture of various disparity patterns and various window sizes, the plot demonstrates that the estimated uncertainty is a good measure of the actual error size, and

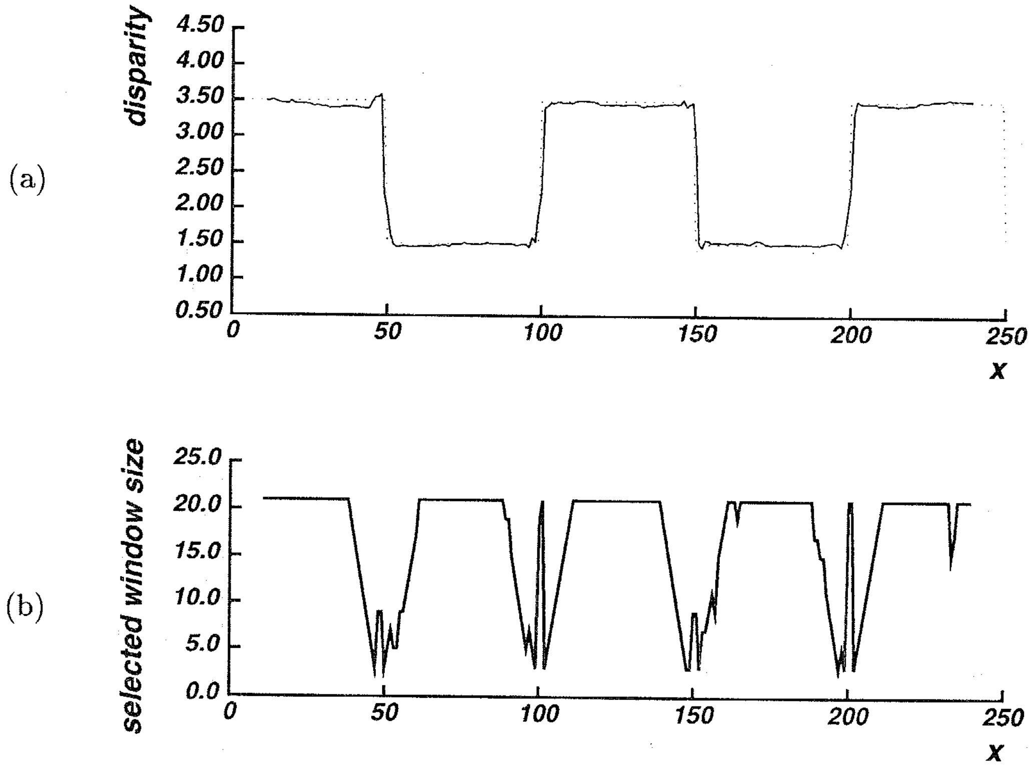


Figure 2.14: Step disparity (a) Computed disparity (b) Selected window size.

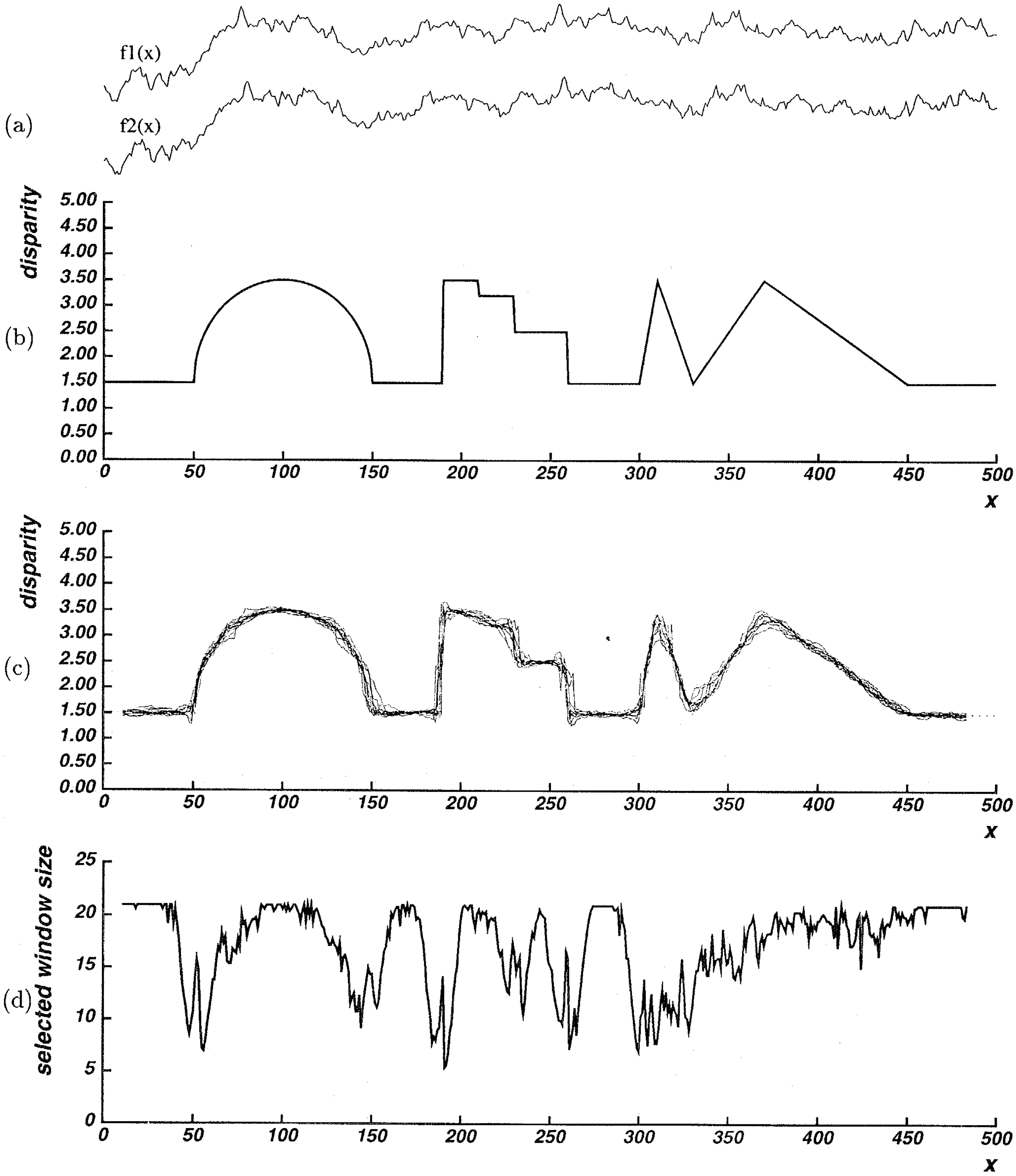


Figure 2.15: (a) Signals (b) True disparity (c) Computed disparity (d) Selected window size.

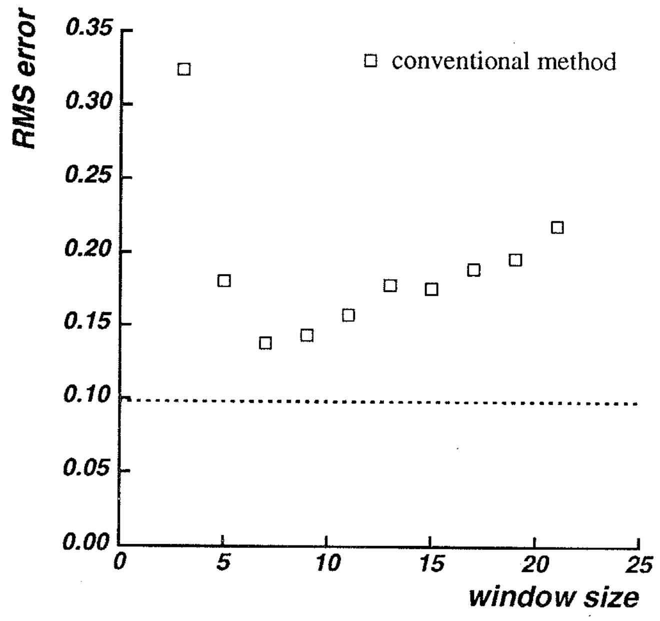


Figure 2.16: RMS errors of computed disparities for figure 2.15 by using various fixed size windows. The RMS error achieved by the locally adaptive window is 0.1, which is smaller than the RMS errors with any fixed size window.

that the actual RMS error is linearly related to the estimated uncertainty.

Lastly, we show disparity estimation by using signals from real stereo images. Figure 2.18 (a) shows the stereo images (top down views of a scale model of buildings). Scanlines marked by black lines are shown in figure 2.18 (b). Figure 2.18 (c) is the computed disparity. For comparison, the oblique view of the model is shown to the right. The selected window size is plotted in figure 2.18 (d).

2.6 Conclusions

We have presented a mathematical framework for our approach and mathematical analyses which enable us to understand many characteristics of stereo vision.

Then, a new signal matching method which can select appropriate window sizes adaptively has been proposed. This method is based on a statistical model of disparity distribution within the window. We assume that disparities have the same expected value, but their variation from that expected value increases with the distance from the center point of the window. This model has enabled us to correctly evaluate the influence of the disparity fluctuation within the window on the computation of disparity, so that the estimated un-

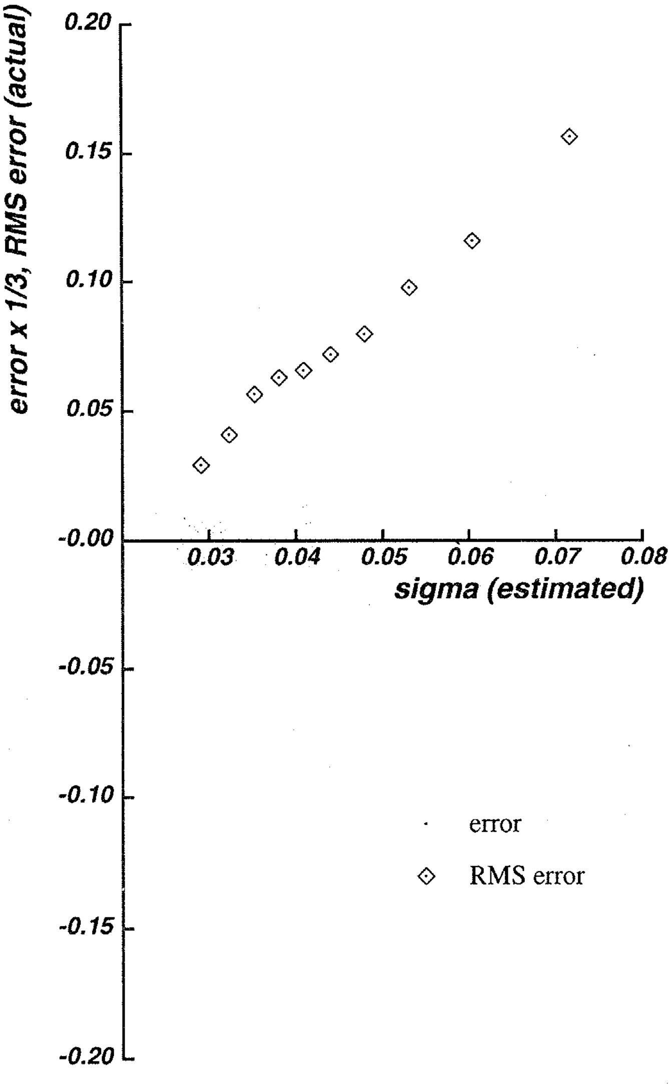


Figure 2.17: Actual error vs. estimated uncertainty. The vertical axis is scaled differently for actual error values and their RMS values so that they both can be fit in the same plot.

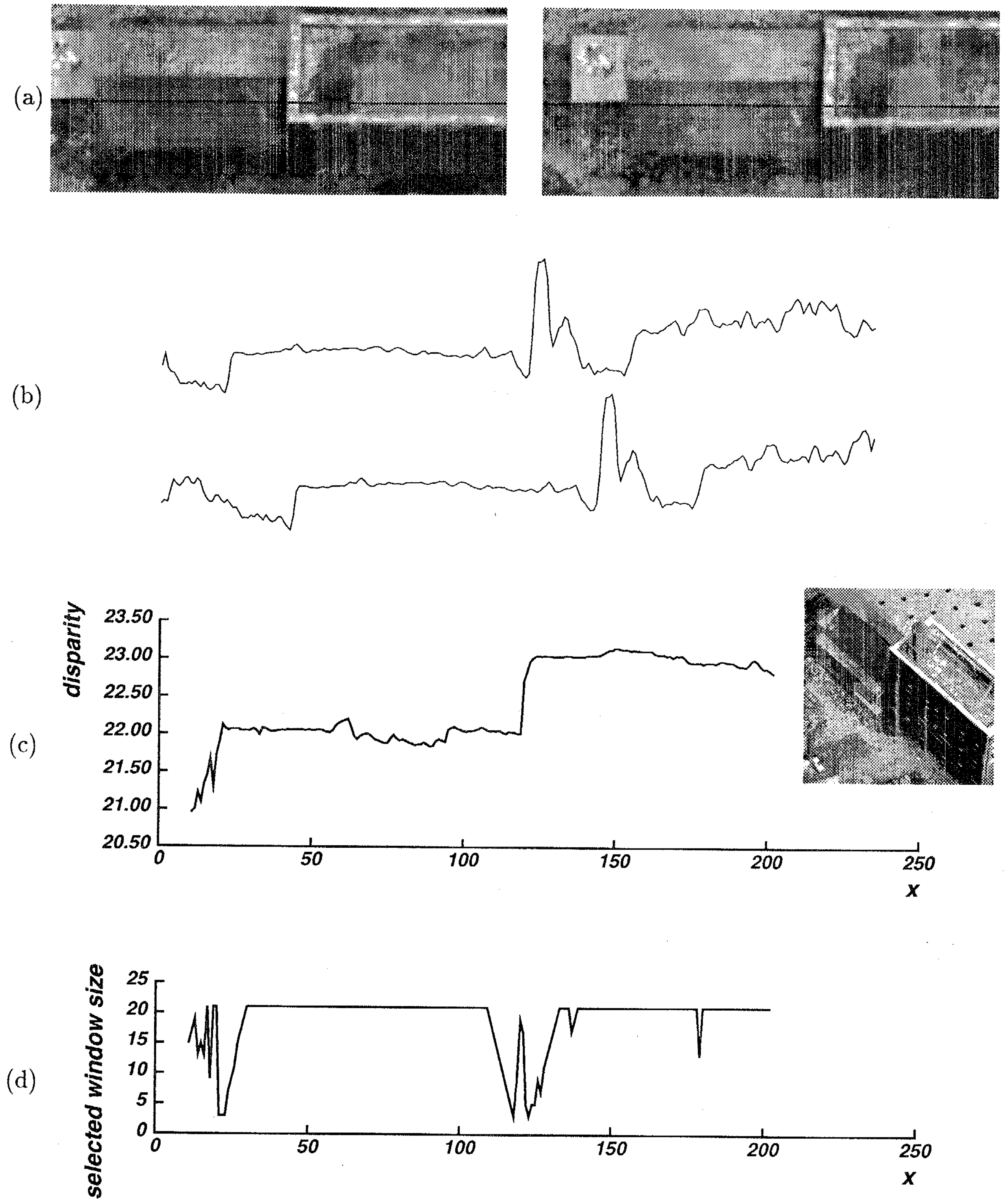


Figure 2.18: Matching signals from real images. (a) Stereo images; (b) Selected scan lines; (c) Computed disparity; (d) Selected window size.

certainty of the computed disparity is close to the real error of the computed disparity. As a result we can choose the window size that provides the disparity estimate with minimum uncertainty.

The analytical and experimental results have shown that the method is effective for various disparity and intensity patterns. The estimated disparities obtained here are better than those from a conventional method with any fixed window size, and the estimated uncertainties show good correlation with the actual RMS error of the estimated disparities. We will extend this idea to the two-dimensional case in chapter 4, where an appropriate size and shape of a window must be selected for stereo matching.

Chapter 3

A Multiple-Baseline Stereo

3.1 Introduction

This chapter deals with another important problem of stereo vision relating to the baseline, i.e. how we can obtain precise distance estimate without suffering from ambiguity?

In stereo matching, we measure the disparity d, which is the difference between the corresponding points of left and right images. The disparity d is related to the distance z by

$$d = BF\frac{1}{z}, \tag{3.1}$$

where B and F are baseline and focal length, respectively.

This equation indicates that for the same distance the disparity is proportional to the baseline, or that the baseline length B acts as a magnification factor in measuring d in order to obtain z. That is, the estimated distance is more precise if we set the two cameras farther apart from each other, which means a longer baseline. A longer baseline, however, poses its own problem. Because a longer disparity range must be searched, matching is more difficult and thus there is a greater possibility of a false match. So there is a trade-off between precision and accuracy (correctness) in matching.

One of the most common methods to deal with the problem is a coarse-to-fine control strategy [MP79] [Gri85] [Bar89] [Han89] [CM90]. Matching is done at a low resolution to reduce false matches, and then the result is used to limit the search range of matching at a higher resolution, where more precise disparity measurements are calculated. Using a coarse resolution, however, does not always remove false matches. This is especially true when there is inherent ambiguity in matching, such as a repeated pattern over a large part of the scene (e.g. a scene of a picket fence). Another approach to remove false matches and to increase precision is to use multiple images, especially a sequence of densely sampled images along a camera path [BBM87] [Yam88] [MSK89] [Hee89]. A short baseline between a pair of consecutive images makes the matching or tracking of features easy, while the structure imposed by the camera motion allows integration of the possibly noisy individual measurements into a precise estimate. The integration has been performed either by exploiting constraints on the EPI [BBM87][Yam88] or by a sequential Kalman filtering technique [MSK89][Hee89].

The stereo matching method presented in this chapter belongs to the second approach: use of multiple images with different baselines obtained by a lateral displacement of a camera. The matching technique, however, is based on the idea that global mismatches can be reduced by adding the sum of squared-difference (SSD) values from multiple stereo pairs. That is, the SSD values are computed first for each pair of stereo images. We represent the SSD values with respect to the inverse distance $1/z$ (rather than the disparity d, as is usually done). The resulting SSD functions from all stereo pairs are added together to produce the sum of SSDs, which we call SSSD-in-inverse-distance. We show that the SSSD-in-inverse-distance function exhibits a unique and clear minimum at the correct matching position even when the underlying intensity patterns of the scene include ambiguities or repetitive patterns.

There have been stereo techniques that use multiple image pairs taken by cameras which are arranged along a line [Mor79] [MO89] [YH92], in the form of a triangle [YKK86] [MK85] [AL87] (called trinocular stereo), or in other formations [Tsa83]. However, all of these techniques, except [YH92] and [Tsa83], decide candidate points for correspondence in each image pair and then search for the correct combinations of corręspondences among them using the geometrical consistencies that they must satisfy. Since the intermediate decisions on correspondences are inherently noisy, ambiguous and multiple, finding the correct combinations requires sophisticated consistency checks and search or filtering. In contrast, our method does not make any decisions about the correspondences in each stereo image pair; instead, it simply accumulates the measures of matching (SSDs) from all the stereo pairs into a single evaluation function, i.e. SSSD-in-inverse-distance, and then obtains one corresponding point from it. In other words, our method integrates *evidence* for a final decision, rather than filtering intermediate *decisions*. In this sense, Tsai [16] employed a strategy very similar to ours: he used multiple images to sharpen the peaks of his overall similarity measures, which he called JMM and WVM. However, the relationship between the improvement of the similarity measures and the camera baseline arrangement was not analyzed, nor was the method tested with real imagery. In this chapter we show both mathematical analysis and experimental results with real indoor and outdoor images. These demonstrate how the SSSD-in-inverse-distance function based on multiple image pairs from different baselines can greatly reduce false matches, while improving precision.

In the next section we present the method mathematically and show how ambiguity can be removed and precision increased by the method. Section 3.3 provides a few experimental results with real stereo images to demonstrate the effectiveness of the algorithm.

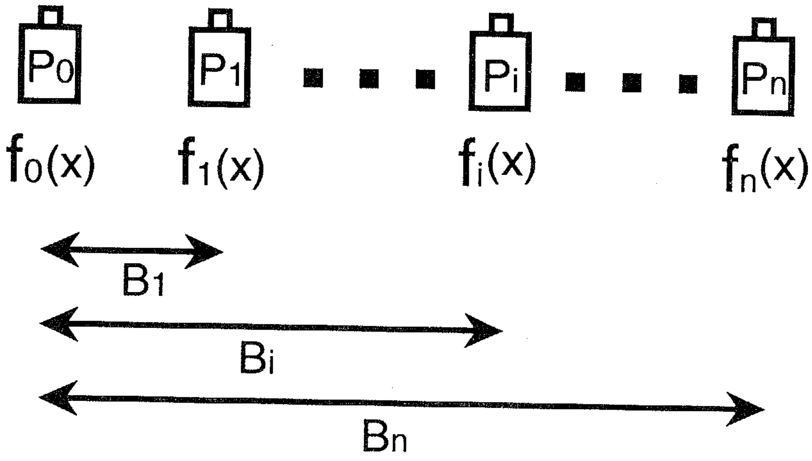


Figure 3.1: Camera arrangement for stereo

3.2 Mathematical Analysis

The essence of stereo matching is, given a point in one image, find in another image the corresponding point such that the two points are the projections of the same physical point in space. This task usually requires some criterion to measure similarity between images. The sum of squared differences (SSD) of the intensity values (or values of preprocessed images, such as bandpass filtered images) over a window is the simplest and most effective criterion. In this section, we define the sum of SSD with respect to the inverse distance (SSSD-in-inverse-distance) for multiple-baseline stereo, and mathematically show its advantages in removing ambiguity and increasing precision. For this analysis, we use 1-D stereo intensity signals, but the extension to two-dimensional images is straightforward. Also, we assume that the disparity within the window is constant in order to facilitate the analysis relating to the baseline in this section.

3.2.1 SSD Function

Suppose that we have cameras at positions $P_0, P_1, \ldots, P_n$ along a line with their optical axes perpendicular to the line and a resulting set of stereo pairs with baselines $B_1, B_2, \ldots, B_n$ as shown in figure 3.1. Let $f_0(x)$ and $f_i(x)$ be the image pair at the camera positions P_0 and P_i, respectively. Imagine a scene point Z whose distance is z. Its disparity $d_{r(i)}$ for the image pair taken from P_0 and P_i is

$$d_{r(i)} = \frac{B_i F}{z}. \qquad (3.2)$$

We model the image intensity functions $f_0(x)$ and $f_i(x)$ near the matching positions for Z as

$$\begin{aligned} f_0(x) &= f(x) + n_0(x) \\ f_i(x) &= f(x - d_{r(i)}) + n_i(x), \end{aligned} \tag{3.3}$$

assuming constant distance near Z and independent Gaussian white noise such that

$$n_0(x), n_i(x) \sim N(0, \sigma_n^2). \tag{3.4}$$

The SSD value $e_{d(i)}$ over a window W at a pixel position x of image $f_0(x)$ for the candidate disparity $d_{(i)}$ is defined as

$$e_{d(i)}(x, d_{(i)}) \equiv \sum_{j \in W} (f_0(x+j) - f_i(x + d_{(i)} + j))^2, \tag{3.5}$$

where the $\sum_{j \in W}$ means summation over the window. The $d_{(i)}$ that gives a minimum of $e_{d(i)}(x, d_{(i)})$ is determined as the estimate of the disparity at x. Since the SSD measurement $e_{d(i)}(x, d_{(i)})$ is a random variable, we will compute its expected value in order to analyze its behavior:

$$\begin{aligned} E[e_{d(i)}(x, d_{(i)})] &= E\left[\sum_{j \in W} (f(x+j) - f(x + d_{(i)} - d_{r(i)} + j) + n_0(x+j) - n_i(x + d_{(i)} + j))^2\right] \\ &= E\left[\sum_{j \in W} (f(x+j) - f(x + d_{(i)} - d_{r(i)} + j))^2\right] \\ &\quad + E\left[\sum_{j \in W} 2(f(x+j) - f(x + d_{(i)} - d_{r(i)} + j)(n_0(x+j) - n_i(x + d_{(i)} + j))\right] \\ &\quad + E\left[\sum_{j \in W} (n_0(x+j) - n_i(x + d_{(i)} + j))^2\right] \\ &= \sum_{j \in W} (f(x+j) - f(x + d_{(i)} - d_{r(i)} + j))^2 + 2N_w\sigma_n^2, \end{aligned} \tag{3.6}$$

where N_w is the number of the points within the window. For the rest of the chapter, $E[\,]$ denotes the expected value of a random variable. Equation (3.6) says that naturally the SSD function $e_{d(i)}(x, d_{(i)})$ is *expected* to take a minimum when $d_{(i)} = d_{r(i)}$, i.e., at the right disparity.

Let us examine how the SSD function $e_{d(i)}(x, d_{(i)})$ behaves when there is ambiguity in the underlying intensity function. Suppose that the intensity signal $f(x)$ has the same pattern around pixel positions x and $x + a$,

$$f(x+j) = f(x+a+j), \qquad j \in W \tag{3.7}$$

where $a \neq 0$ is a constant. Then, from equation (3.6)

$$E[e_{d(i)}(x, d_{r(i)})] = E[e_{d(i)}(x, d_{r(i)} + a)] = 2N_w\sigma_n^2. \tag{3.8}$$

This means that ambiguity is expected in matching in terms of positions of minimum SSD values. Moreover, the false match at $d_{r(i)} + a$ appears in exactly the same way for all i; it is separated from the correct match by a for all the stereo pairs. Using multiple baselines does not help to disambiguate.

3.2.2 Sum of SSDs with respect to Inverse Distance (SSSD-in-Inverse-Distance)

Now, let us introduce the *inverse distance* ζ such that

$$\zeta = \frac{1}{z}. \tag{3.9}$$

From equation and (3.2),

$$\begin{aligned} d_{r(i)} &= B_iF\zeta_r & (3.10) \\ d_{(i)} &= B_iF\zeta, & (3.11) \end{aligned}$$

where ζ_r and ζ are the real and the candidate inverse distance, respectively. Substituting equation (3.11) into (3.5), we have the SSD with respect to the inverse distance,

$$e_{\zeta(i)}(x, \zeta) \equiv \sum_{j \in W} (f_0(x+j) - f_i(x + B_iF\zeta + j))^2, \tag{3.12}$$

at position x for a candidate inverse distance ζ. Its expected value is

$$E[e_{\zeta(i)}(x, \zeta)] = \sum_{j \in W} (f(x+j) - f(x + B_iF(\zeta - \zeta_r) + j))^2 + 2N_w\sigma_n^2. \tag{3.13}$$

Finally, we define a new evaluation function $e_{\zeta(12\cdots n)}(x, \zeta)$, the sum of SSD functions with respect to the inverse distance (SSSD-in-inverse-distance) for multiple stereo pairs. It is obtained by adding the SSD functions $e_{\zeta(i)}(x, \zeta)$ for individual stereo pairs:

$$e_{\zeta(12\cdots n)}(x, \zeta) = \sum_{i=1}^{n} e_{\zeta(i)}(x, \zeta). \tag{3.14}$$

Its expected value is

$$\begin{aligned} E[e_{\zeta(12\cdots n)}(x, \zeta)] &= \sum_{i=1}^{n} E[e_{\zeta(i)}(x, \zeta)] \\ &= \sum_{i=1}^{n} \sum_{j \in W} (f(x+j) - f(x + B_iF(\zeta - \zeta_r) + j))^2 + 2nN_w\sigma_n^2. & (3.15) \end{aligned}$$

In the next three subsections, we will analyze the characteristics of these evaluation functions to see how ambiguity is removed and precision is improved.

3.2.3 Elimination of Ambiguity (1)

As before, suppose the underlying intensity pattern $f(x)$ has the same pattern around x and $x+a$ (equation (3.7)). Then, according to equation (3.13), we have

$$E[e_{\zeta(i)}(x,\zeta_r)] = E[e_{\zeta(i)}(x,\zeta_r + \frac{a}{B_iF})] = 2N_w\sigma_n^2. \tag{3.16}$$

We still have an ambiguity; a minimum is expected at a false inverse distance $\zeta_f = \zeta_r + \frac{a}{B_iF}$. However, an important point to be observed here is that this minimum for the false inverse distance ζ_f changes its position as the baseline B_i changes, while the minimum for the correct inverse distance ζ_r does not. This is the property that the new evaluation function, the SSSD-in-inverse-distance (3.14), exploits to eliminate the ambiguity. For example, suppose we use two baselines B_1 and B_2 ($B_1 \neq B_2$). From equation (3.15)

$$\begin{aligned} E[e_{\zeta(12)}(x,\zeta)] &= \sum_{j\in W}(f(x+j) - f(x+B_1F(\zeta-\zeta_r)+j))^2 \\ &+ \sum_{j\in W}(f(x+j) - f(x+B_2F(\zeta-\zeta_r)+j))^2 + 4N_w\sigma_n^2. \end{aligned} \tag{3.17}$$

We can prove that

$$E[e_{\zeta(12)}(x,\zeta)] > 4N_w\sigma_n^2 = E[e_{\zeta(12)}(x,\zeta_r)] \text{ for } \zeta \neq \zeta_r. \tag{3.18}$$

(refer to appendix B) In words, $e_{\zeta(12)}(x,\zeta)$ is *expected* to have the smallest value at the correct ζ_r. That is, the ambiguity is likely to be eliminated by use of the new evaluation function with two different baselines.

We can illustrate this using synthesized data. Suppose the point whose distance we want to determine is at $x = 0$ and the underlying function $f(x)$ is given by

$$f(x) = \begin{cases} cos(\frac{\pi}{4}x) + 2 & \text{if } -4 < x < 12 \\ 1 & \text{if } x \leq -4 \text{ or } 12 \leq x. \end{cases} \tag{3.19}$$

Figure 3.2 (a) shows a plot of $f(x)$. Assuming that $d_{r(1)} = 5$, $\sigma_n^2 = 0.2$, and the window size is 5, the expected values of the SSD function $e_{d(1)}(x, d_{(1)})$ are as shown in figure 3.2 (b). We see that there is an ambiguity: the minima occur at the correct match $d_{(1)} = 5$ and at the false match $d_{(1)} = 13$. Which match will be selected will depend on the noise, search range, and search strategy. Now suppose we have a longer baseline B_2 such that $\frac{B_2}{B_1} = 1.5$. From equations (3.6) and (3.10), we obtain $E[e_{d(2)}]$ as shown in figure 3.2 (c). Again we encounter an ambiguity, and the separation of the two minima is the same.

Now let us evaluate the SSD values with respect to the inverse distance ζ rather than the disparity d by using equations (3.12) through (3.15). The expected values of the SSD measurements $E[e_{\zeta(1)}]$ and $E[e_{\zeta(2)}]$ with baselines B_1 and B_2 are shown in figures 3.2 (d) and (e), respectively (the plot is normalized such that $B_1F = 1$). Note that the minima at the correct inverse distance ($\zeta = 5$) does not move, while the minima for the false match changes its position as the baseline changes. When the two functions are added to produce the SSSD-in-inverse-distance, its expected values $E[e_{\zeta(12)}]$ are as shown in figure 3.2 (f). We can see that the ambiguity has been reduced because the SSSD-in-inverse-distance has a smaller value at the correct match position than at the false match.

3.2.4 Elimination of Ambiguity (2)

An extreme case of ambiguity occurs when the underlying function $f(x)$ is a periodic function, like a scene of a picket fence. We can show that this ambiguity can also be eliminated.

Let $f(x)$ be a periodic function with period T. Then, each $e_{\zeta(i)}(x, \zeta)$ is expected to be a periodic function of ζ with the period $\frac{T}{B_iF}$. This means that there will be multiple minima of $e_{\zeta(i)}(x, \zeta)$ (i.e., ambiguity in matching) at intervals of $\frac{T}{B_iF}$ in ζ. When we use two baselines and add their SSD values, the resulting $e_{\zeta(12)}(x, \zeta)$ will be still a periodic function of ζ, but its period T_{12} is increased to

$$T_{12} = LCM\left(\frac{T}{B_1F}, \frac{T}{B_2F}\right), \tag{3.20}$$

where $LCM()$ denotes Least Common Multiple. That is, the period of the expected value of the new evaluation function can be made longer than that of the individual stereo pairs. Furthermore, it can be controlled by choosing the baselines B_1 and B_2 appropriately so that the expected value of the evaluation function has only one minimum within the search range. This means that using multiple-baseline stereo pairs simultaneously can eliminate ambiguity, although each individual baseline stereo may suffer from ambiguity.

We illustrate this by using real stereo images. Figure 3.3(a) shows an image of a sample scene. At the top of the scene there is a grid board whose intensity function is nearly periodic. We took ten images of this scene by shifting the camera vertically as in figure 3.4. The actual distance between consecutive camera positions is 0.05 inches. Let this distance be b. Figure 3.3 shows the first and the last images of the sequence. We selected a point x within the repetitive grid board area in image9. The SSD values $e_{\zeta(i)}(x, \zeta)$ over 5-by-5-pixel windows are plotted for various baseline stereo pairs in figure 3.5. The horizontal axis of all the plots is the inverse distance, normalized such that $8bF = 1$. Figure 3.5 illustrates the trade-off between precision and ambiguity in terms of baselines. That is, for a shorter baseline, there are fewer minima (i.e. less ambiguity), but the SSD curve is flatter (i.e. less precise localization). On the other hand, for a longer baseline, there are more minima (i.e. more ambiguity), but the curve near the minimum is sharper; that is, the estimated distance is more precise if we can find the correct one.

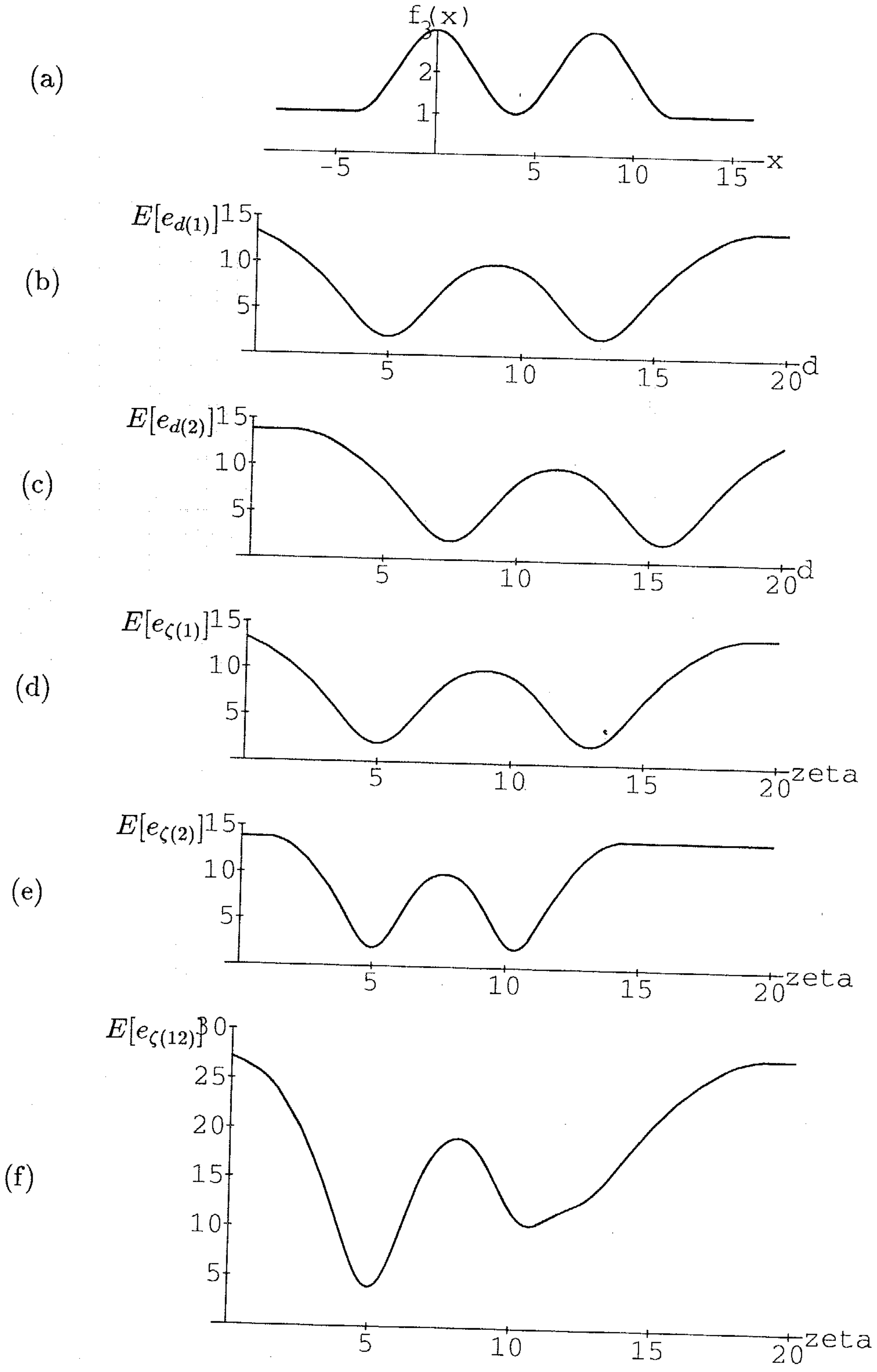


Figure 3.2: Expected values of evaluation functions: (a) Underlying function; (b) $E[e_{d(1)}]$; (c) $E[e_{d(2)}]$; (d) $E[e_{\zeta(1)}]$; (e) $E[e_{\zeta(2)}]$; (f) $E[e_{\zeta(12)}]$
.

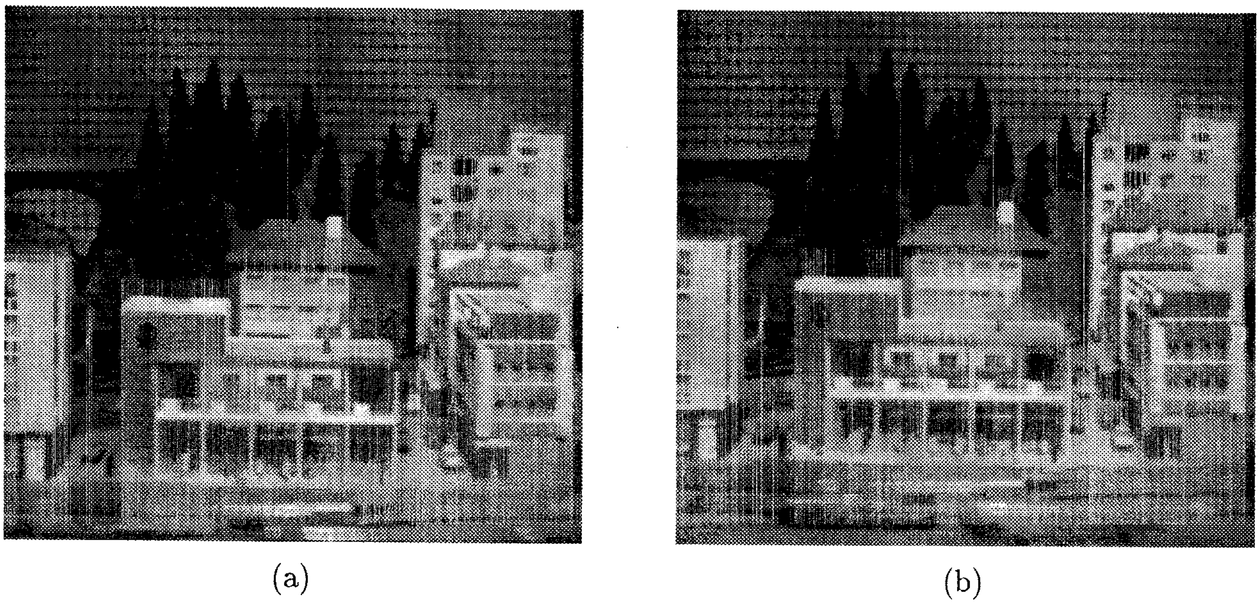

(a)
(b)

Figure 3.3: "Town" data set: (a) Image0; (b) Image9

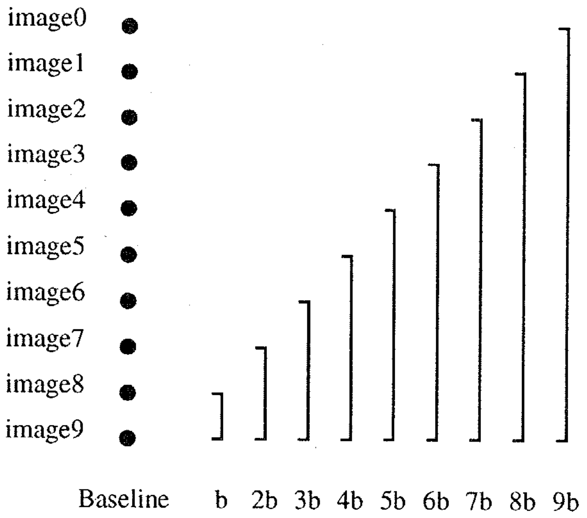


Figure 3.4: "Town" data set image sequence

Now, let us take two stereo image pairs: one with $B = 5b$ and the other with $B = 8b$. In figure 3.6, the dashed curve and the dotted curve show the SSD for $B = 5b$ and $B = 8b$, respectively. Let us suppose the search range goes from 0 to 20 in the horizontal axis, which in this case corresponds to 12 to ∞ inches in distance. Though the SSD values take a minimum at the correct answer near $\zeta = 5$, there are also other minima for both cases. The solid curve shows the evaluation function for the multiple-baseline stereo, which is the sum of the dashed curve and the dotted curve. The solid curve shows only one clear minimum; that is, the ambiguity is resolved.

So far, we have considered using only two stereo pairs. We can easily extend the idea to multiple-baseline stereo which uses more than two stereo pairs. Corresponding to equation (3.20), the period of $E[e_{\zeta(12\cdots n)}(x,\zeta)]$ becomes

$$T_{12,\ldots,n} = LCM\left(\frac{T}{B_1F}, \frac{T}{B_2F}, \ldots, \frac{T}{B_nF}\right) \tag{3.21}$$

where $B_1, B_2, \ldots, B_n$ are baselines for each stereo pair.

We will demonstrate how the ambiguity can be further reduced by increasing the number of stereo pairs. From the data of figure 3.4, we first choose image1 and image9 as a long baseline stereo pair, ie. (1) $B = 8b$. Then, we increase the number of stereo pairs by dividing the baseline between image1 and image9, i.e. (2) $B = 4b$ and $8b$, (3) $B = 2b$, $4b$, $6b$ and $8b$, (4) $B = b$, $2b$, $3b$, $4b$, $5b$, $6b$, $7b$ and $8b$. Figure 3.7 demonstrates that the SSSDs-in-inverse-distance shows the minimum at the correct position more clearly as more stereo pairs are used.

3.2.5 Increase of Precision

We have shown that ambiguities can be resolved by using the SSSD-in-inverse-distance computed from multiple baseline stereo pairs. The technique also increases precision in estimating the true inverse distance. We can show this by analyzing the statistical characteristics of the evaluation functions near the correct match.

From equations (3.3), (3.10), and (3.12), we have

$$e_{\zeta(i)}(x,\zeta) = \sum_{j\in W} (f(x+j) - f(x + B_iF(\zeta-\zeta_r)+j) + n_0(x+j) - n_i(x+B_iF\zeta+j))^2. \tag{3.22}$$

By taking the Taylor expansion about $\zeta = \zeta_r$ up to the linear terms, we obtain

$$f(x + B_iF(\zeta-\zeta_r)+j) \approx f(x+j) + B_iF(\zeta-\zeta_r)f'(x+j). \tag{3.23}$$

Substituting this into equation (3.22), we can approximate $e_{\zeta(i)}(x,\zeta)$ near ζ_r by a quadratic form of ζ:

$$\begin{aligned} e_{\zeta(i)}(x,\zeta) &\approx \sum_{j\in W} (-B_iF(\zeta-\zeta_r)f'(x+j) + n_0(x+j) - n_i(x+B_iF\zeta+j))^2 \\ &= B_i^2F^2a(x)(\zeta-\zeta_r)^2 + 2B_iFb_i(x)(\zeta-\zeta_r) + c_i(x), \end{aligned} \tag{3.24}$$

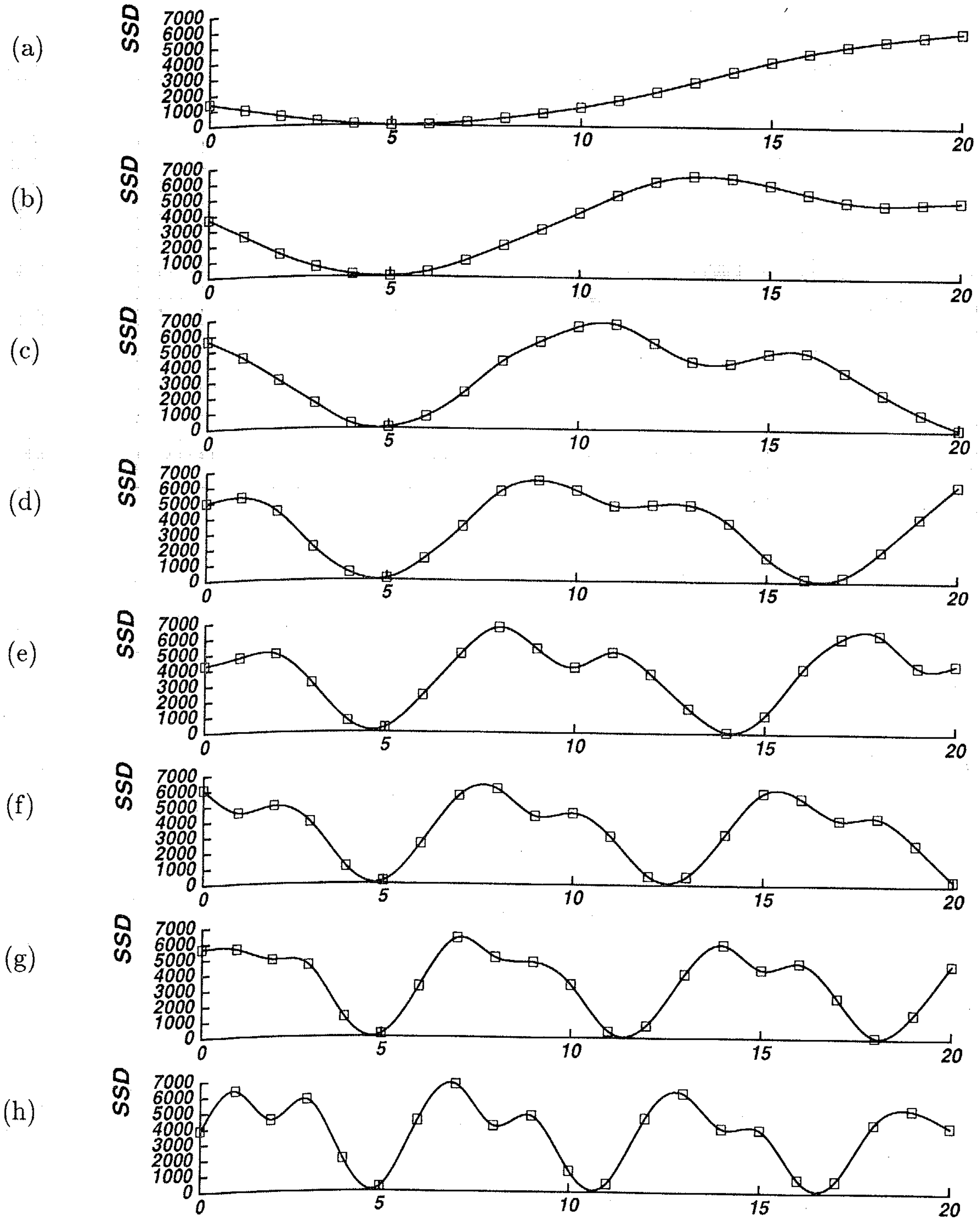


Figure 3.5: SSD values vs. inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

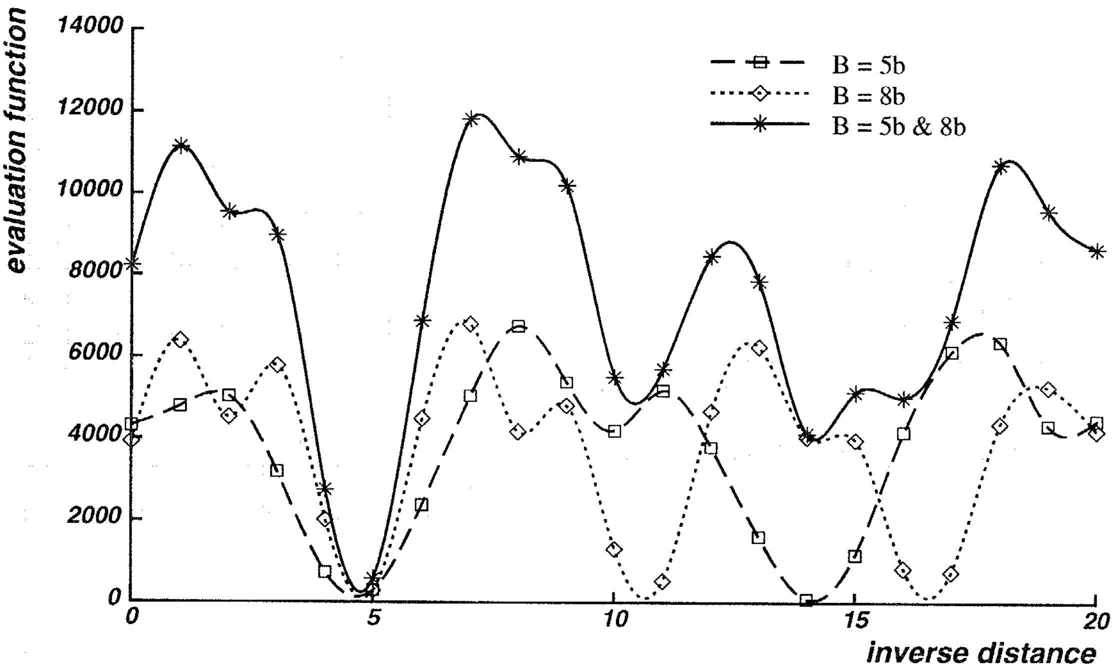


Figure 3.6: Combining two stereo pairs with different baselines

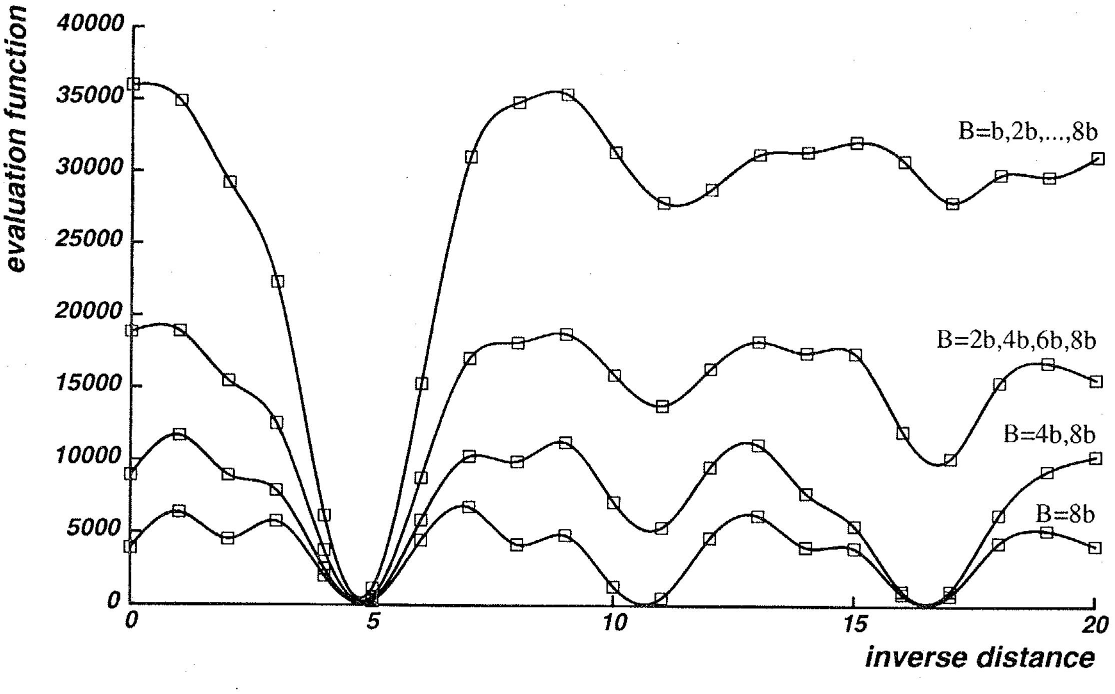


Figure 3.7: Combining multiple baseline stereo pairs

where

$$
\begin{aligned}
a(x) &= \sum_{j\in W}(f'(x+j))^2 && (3.25)\\
b_i(x) &= \sum_{j\in W} f'(x+j)(n_i(x+B_iF\zeta+j)-n_0(x+j)) && (3.26)\\
c_i(x) &= \sum_{j\in W}(n_i(x+B_iF\zeta+j)-n_0(x+j))^2. && (3.27)
\end{aligned}
$$

The estimated inverse distance $\hat{\zeta}_{r(i)}$ is the value ζ that makes equation (3.24) minimum;

$$\hat{\zeta}_{r(i)} = \zeta_r - \frac{b_i(x)}{B_iFa(x)}. \qquad (3.28)$$

Since $E[b_i(x)] = 0$, the expected value of the estimate $\hat{\zeta}_{r(i)}$ is the correct value ζ_r, but it varies due to the noise. The variance of this estimate is:

$$
\begin{aligned}
Var(\hat{\zeta}_{r(i)}) &= \frac{Var(b_i(x))}{B_i^2F^2(a(x))^2}\\
&= \frac{2\sigma_n^2}{B_i^2F^2a(x)}. && (3.29)
\end{aligned}
$$

Basically, this equation states that for the same amount of image noise σ_n^2, the variance is smaller (the estimate is more precise) as the baseline B_i is longer, or as the variation of intensity signal, $a(x)$, is larger.

We can follow the same analysis for $e_{\zeta(12\cdots n)}(x,\zeta)$ of (3.14), the new evaluation function with multiple baselines. Near ζ_r, it is

$$e_{\zeta(12\cdots n)}(x,\zeta) \approx \left(\sum_{i=1}^{n} B_i^2\right) F^2a(x)(\zeta-\zeta_r)^2 + 2F\left(\sum_{i=1}^{n} B_ib_i(x)\right)(\zeta-\zeta_r) + \sum_{i=1}^{n} c_i(x). \qquad (3.30)$$

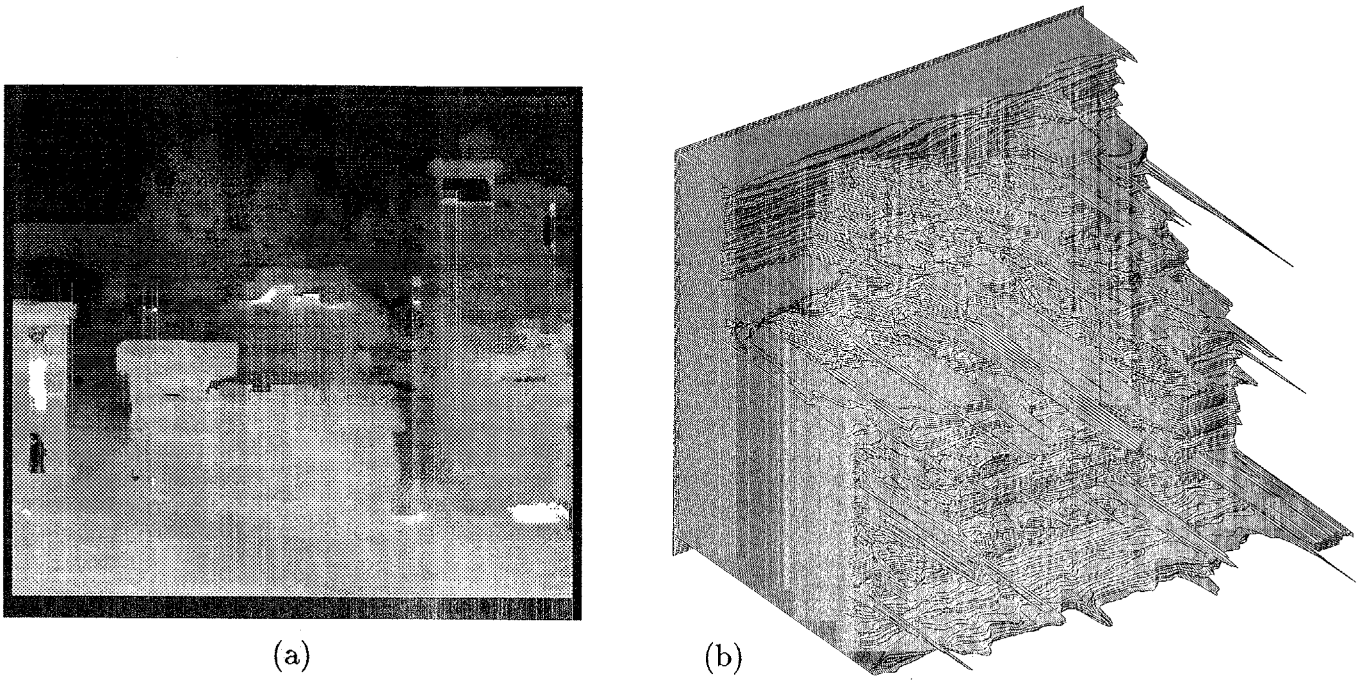
The variance of the estimated inverse distance $\hat{\zeta}_{r(12\cdots n)}$ that minimizes this function is

$$Var(\hat{\zeta}_{r(12\cdots n)}) = \frac{2\sigma_n^2}{\left(\sum_{i=1}^{n} B_i^2\right)F^2a(x)}. \qquad (3.31)$$

From equations (3.29) and (3.31), we see that

$$\frac{1}{Var(\hat{\zeta}_{r(12\cdots n)})} = \sum_{i=1}^{n}\frac{1}{Var(\hat{\zeta}_{r(i)})}. \qquad (3.32)$$

The inverse of the variance represents the precision of the estimate. Therefore, equation (3.32) means that by using the SSSD-in-inverse-distance with multiple baseline stereo pairs, the estimate becomes more precise. We can confirm this characteristic in figures 3.6 and 3.7 by observing that the curve around the correct inverse distance becomes sharper as more baselines are used.



(a) (b)

Figure 3.8: Result with a short baseline, $B = 3b$: (a) Distance map; (b) Isometric plot of the distance map from the upper left corner. The matching is mostly correct, but very noisy.

3.3 Experimental Results

This section presents experimental results of the multiple-baseline stereo based on SSSD-in-inverse-distance with real 2D images. A complete description of the algorithm is included in Appendix C.

The first result is for the "Town" data set that we showed in figure 3.3. Figures 3.8 (a) and (b) are the distance map and its isometric plot with a short baseline, $B = 3b$. The result with a single long baseline, $B = 9b$, is shown in figure 3.9. Comparing these two results, we observe that the distance map computed by using the long baseline is smoother on flat surfaces, i.e., more precise, but has gross errors in matching at the top of the scene because of the repeated pattern. These results illustrate the trade-off between ambiguity and precision. Figure 3.10, on the other hand, shows the distance map and its isometric plot obtained by the new algorithm using three different baselines, $3b$, $6b$, and $9b$. For comparison, the corresponding oblique view of the scene is shown in figure 3.11. We can note that the computed distance map is less ambiguous *and* more precise than those of the single-baseline stereo.

Figure 3.12 shows another data set used for our experiment. Figures 3.13 and 3.14 compare the distance maps computed from the short baseline stereo and the long baseline stereo: the longer baseline is five times longer than the short one. For comparison, the actual oblique view roughly corresponding to the isometric plot is shown in figure 3.15. Though

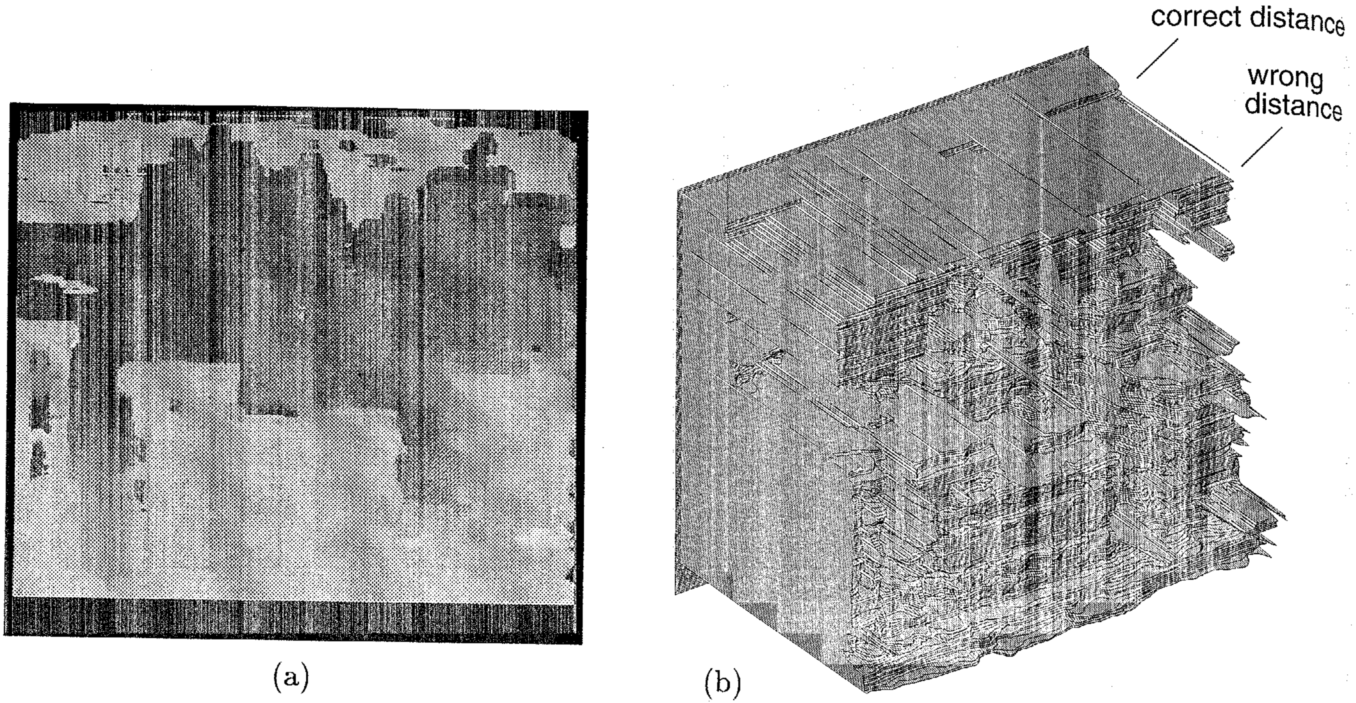


Figure 3.9: Result with a long baseline, $B = 9b$: (a) Distance map; (b) Isometric plot. The matching is less noisy when it is correct. However, there are many gross mistakes, especially in the top of the image where, due to a repetitive pattern, the matching is completely wrong.

no repetitive patterns are apparent in the images, we can still observe gross errors in the distance map obtained with the long baseline due to false matching. In contrast, the result from the multiple-baseline stereo shown in figure 3.16 demonstrates both the advantage of unambiguous matching with a short baseline and that of precise matching with a long baseline.

Figures 3.17 (a) and (b) show one of the real outdoor scenes to which the multiple-baseline stereo technique has been applied. The distance to the front object (curb) is roughly 20 m and it is another 8 m to the building wall. We used a Sony CCD camera with a 50 mm lens, and captured six images (five stereo image pairs) by moving the camera horizontally. The baseline between the neighboring camera positions is 1.9 cm, so that the disparity is of the order of a few pixels (thus less than 15 pixels for the image pair with the longest baseline). Figure 3.17(c) is the distance map obtained: we used a 9x9 window for SSD computation and used DOG-filtered images as input rather than the original intensity images in order to compensate for the change in sunlight during the data collection session. Pebbles on the road in front of the curb are detectable in the map, and the occlusion edges of the sign board are very sharp. Naturally, range measurements are noisy along the top edge of the curb, which is mostly horizontal. Note that the map is the direct output of the stereo algorithm with no smoothing or postprocessing applied.

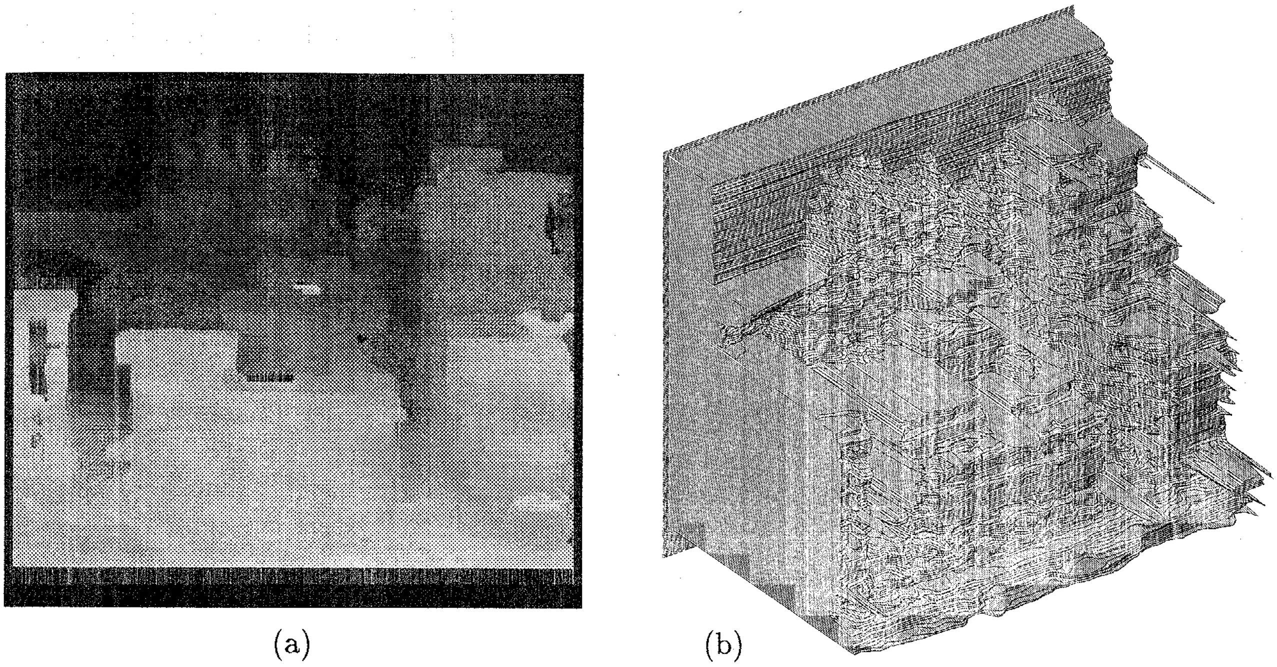


Figure 3.10: Result with multiple baselines, $B = 3b$, $6b$, and $9b$: (a) Distance map; (b) Isometric plot. Compared with figures 3.8(b) and 3.9(b), we see that the distance map is less noisy and that gross errors have been removed.

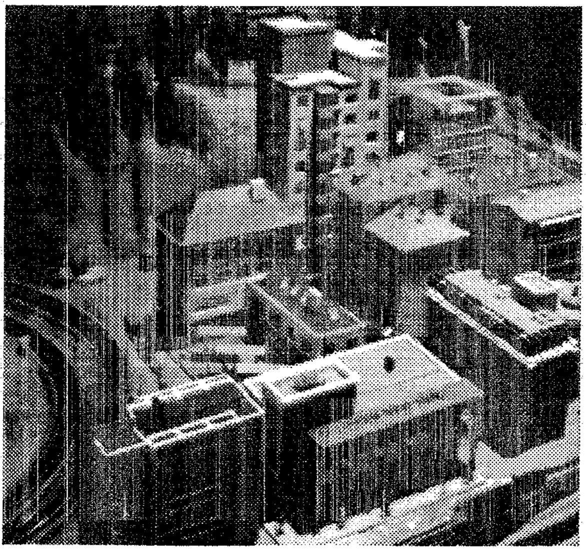
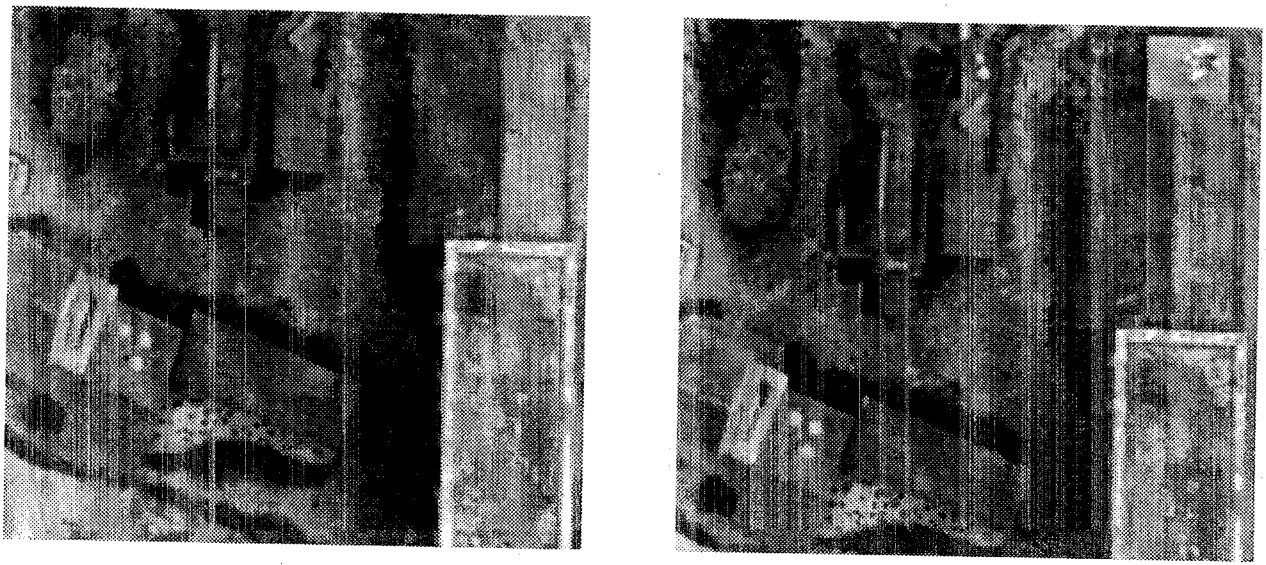


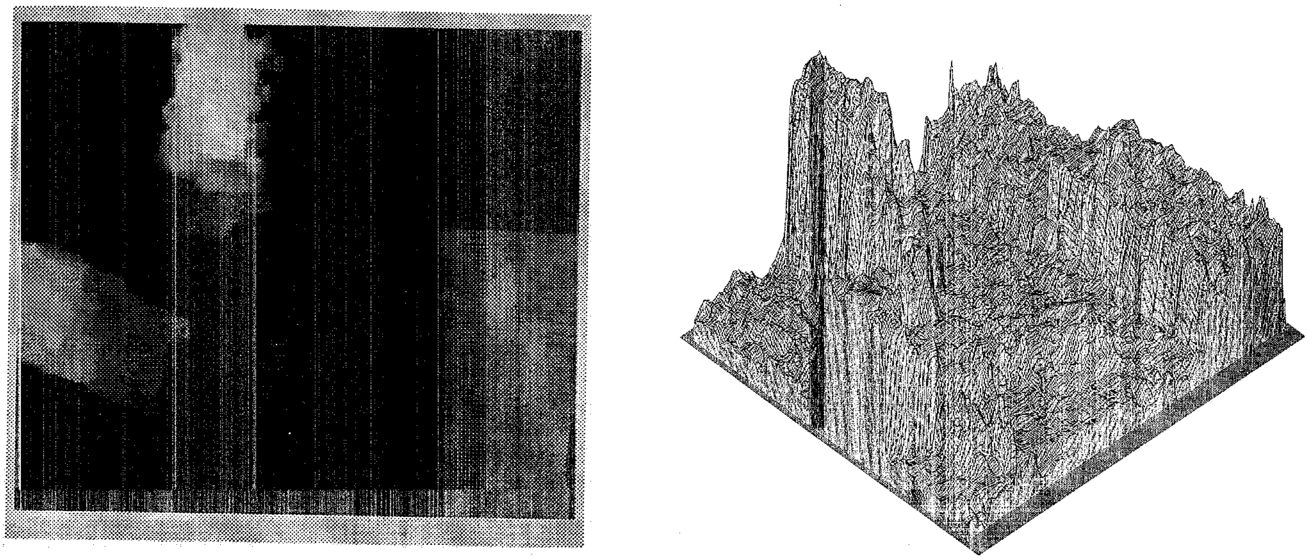
Figure 3.11: Oblique view



(a) (b)

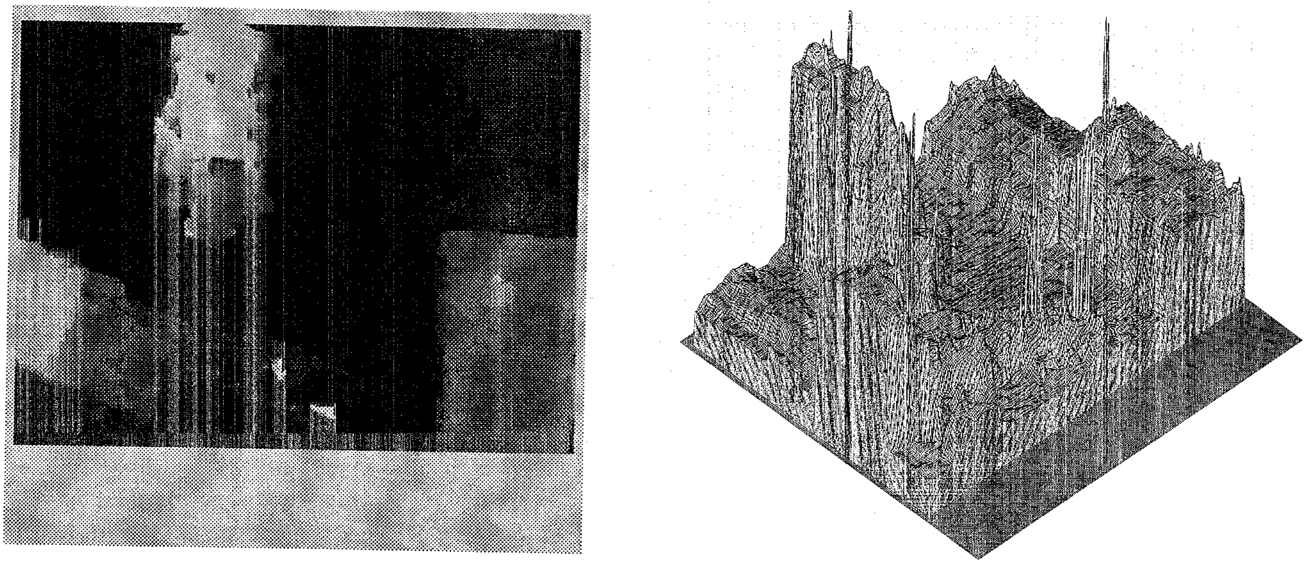
Figure 3.12: "Coal mine" data set, long-baseline pair

During the experiments with this and other scenes, we found that we invariably obtained better results by using relatively short baselines. As seen in figures 3.17 (a) and (b), the disparity is typically only 10 to 15 pixels even for the closest objects in the image pair with the largest baseline. This is somewhat surprising since for precision we anticipated that we would need much longer baselines, at least for one or two pairs. What is happening here seems to be the following. When the baselines become longer, the effect of photographic and geometric distortions, as well as occlusions, become severe. Use of the shorter baselines generally decreases precision, but alleviates these problems, making the SSD functions show more consistent behavior. Yet, since we accumulate multiple observations, sufficient precision is still achievable. This is, in fact, an advantage of the method, since it means fewer occluded parts in the final range map, and less computation as well, since the range of SSD computation is shorter. Moreover, after finding the unique minimum position of the SSSD function, we can compute the minimum positions of each individual pair's SSD functions near the overall minimum, their curvature at their minimums, and finally their minimum values. We have found some indication that these can be used to evaluate the uncertainty of the correctness of the matching, and further to classify the situation into occlusion, terminal edges, and specular reflections. We are investigating these issues further [KN91] [KON92].



(a) (b)

Figure 3.13: Result with a short baseline: (a) Distance map; (b) Isometric plot of the distance map viewed from the lower left corner



(a) (b)

Figure 3.14: Result with a long baseline: (a) Distance map; (b) Isometric plot

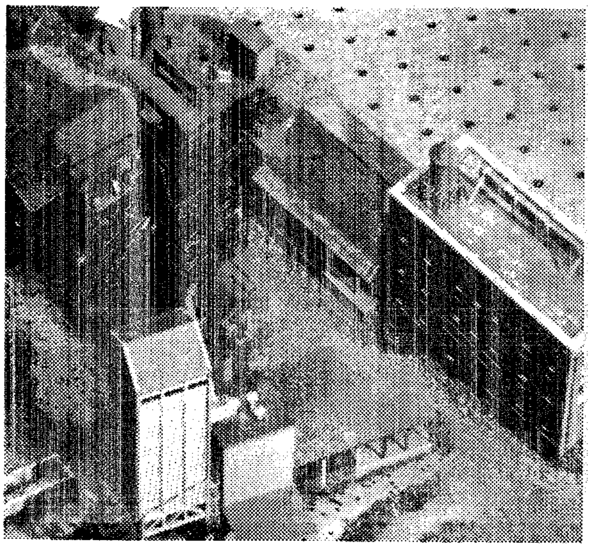
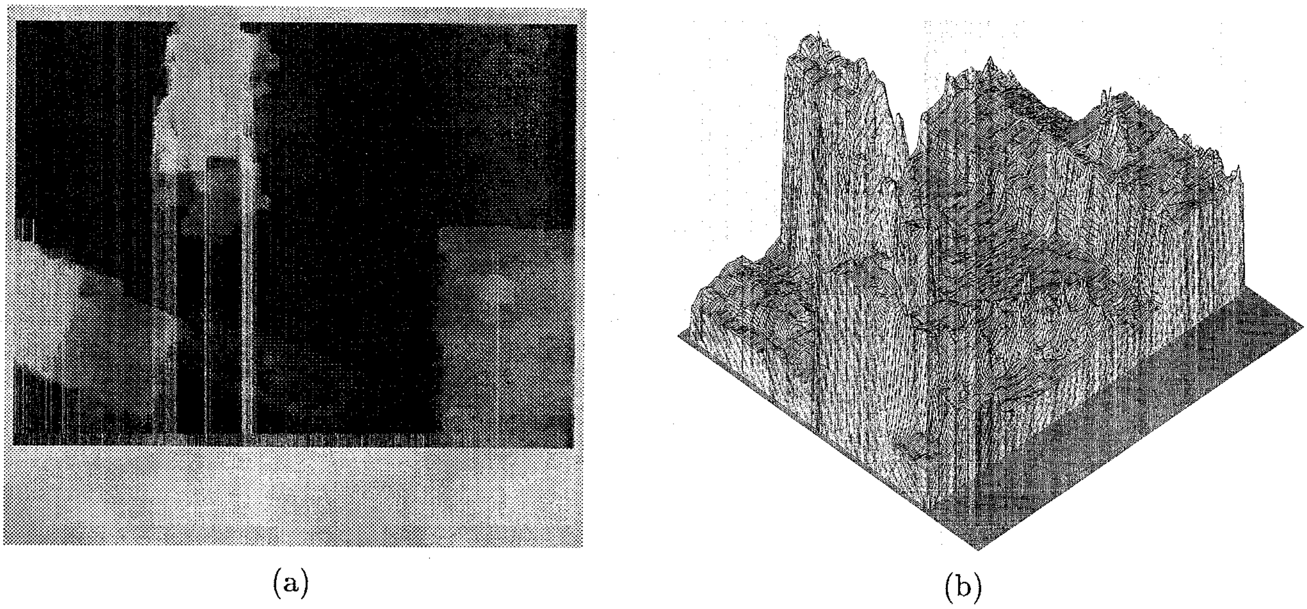


Figure 3.15: Oblique view



(a) (b)

Figure 3.16: Multiple baselines: (a) Distance map; (b) Isometric plot

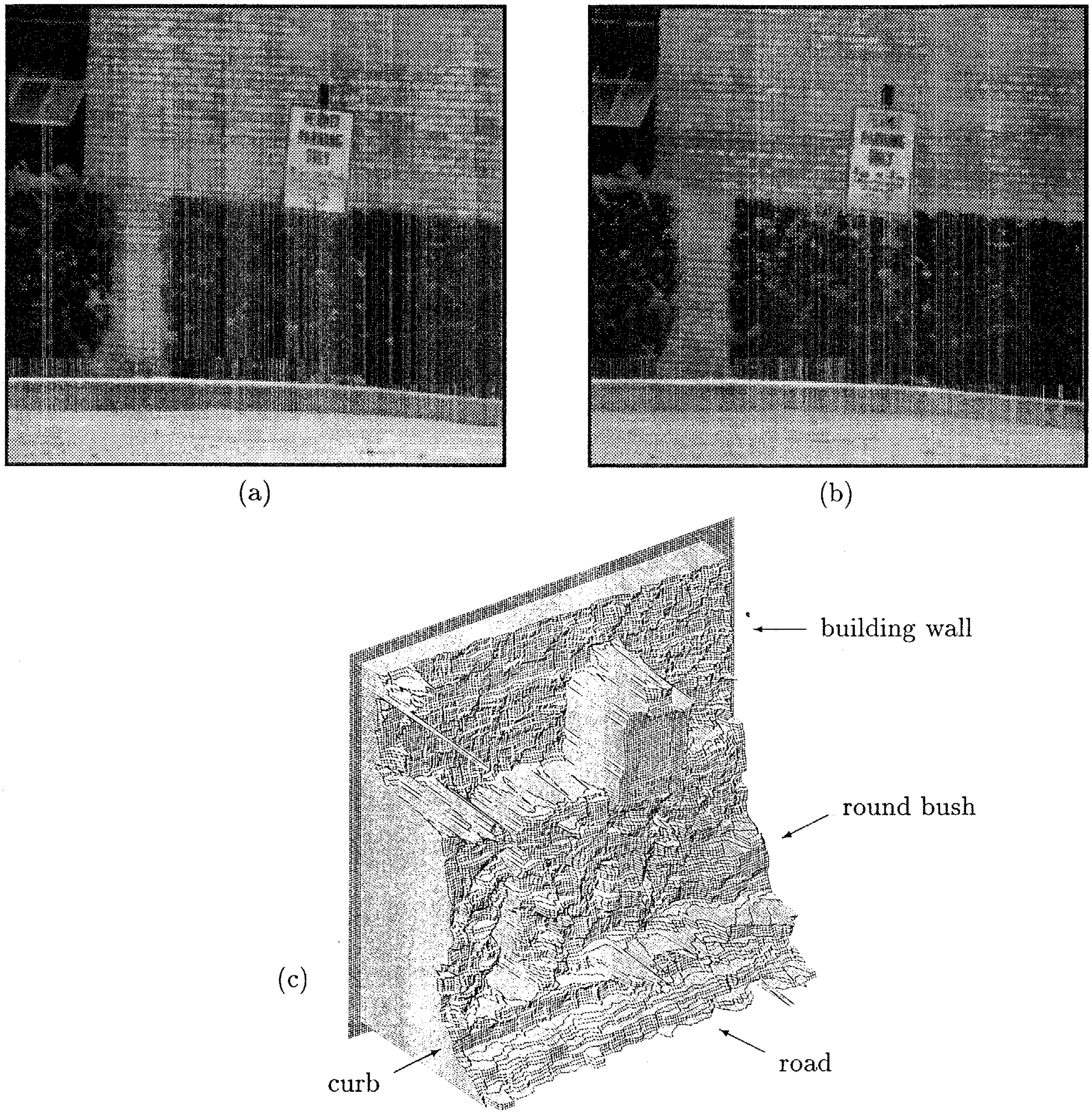


Figure 3.17: Result with a real outdoor scene: (a)(b) long baseline pair of images; (c) Isometric plot of the distance map.

3.4 Conclusions

In this chapter, we have presented a new stereo matching method which uses multiple baseline stereo pairs. This method can overcome the trade-off between precision and accuracy (avoidance of false matches) in stereo. The method is rather straightforward: we represent the SSD values for individual stereo pairs as a function of the inverse distance, and add those functions. The resulting function, the SSSD-in-inverse-distance, exhibits an unambiguous and sharper minimum at the correct matching position. As a result there is no need for search or sequential estimation procedures.

The key idea of the method is to relate SSD values to the inverse distance rather than the disparity. As an afterthought, this idea is natural. Whereas disparity is a function of the baseline, there is only one true (inverse) distance for each pixel position for all of the stereo pairs. Therefore there must be a single minimum for the SSD values when they are summed and plotted with respect to the inverse distance. We have shown the advantage of the proposed method in removing ambiguity and improving precision by analytical and experimental results.

Chapter 4

A Stereo Matching Algorithm with an Adaptive Window

4.1 Introduction

A locally adaptive window for signal matching has been introduced in chapter 2. Here, it is extended into the two-dimensional case, i.e. matching using stereo images. We extend and summarize the mathematical formulation with 2-D image functions. We also present a comparison of our statistical model of the disparity distribution with local support models (assumptions) which has been used by other stereo methods. Then, a detailed description of our stereo algorithm with 2-D adaptive window control and many experimental results using real images are presented. Now, we restate an important problem next.

Stereo matching by computing correlation or sum of squared differences (SSD) is a basic technique for obtaining a dense depth map from images [MSK89] [FP86] [Han89] [Woo83] [Pan78] [MKA73]. As Barnard and Fischler [BF87] point out, "a problem with correlation (or SSD) matching is that the patch (window) size must be large enough to include enough intensity variation for matching but small enough to avoid the effects of projective distortion." If the window is too small and does not cover enough intensity variation, it gives a poor disparity estimate, because the signal (intensity variation) to noise ratio is low. If, on the other hand, the window is too large and covers a region in which the depth of scene points (i.e. disparity) varies, then the position of maximum correlation or minimum SSD may not represent correct matching due to different projective distortions in the left and right images. For this reason, a window size must be selected adaptively depending on local variations of intensity and disparity.

However, most correlation- or SSD-based stereo methods in the past have used a window of a fixed size that is chosen empirically for each application. There has been little research for adaptive window selection. As a relevant technique, Panton [Pan78] warped the image to account for predicted terrain relief, but failed to consider the contribution due to intensity variation. In their coarse-to-fine stereo technique, Hoff and Ahuja [HA89] discuss the relationship between window shape and disparity, and argue how integrating the processes of matching, contour detection and surface interpolation can help reduce the problem. Levine et al. [LOY73] changed the window size locally depending on the intensity pattern, but uncertainty in matching due to the variation of unknown disparities was unaccounted for.

The difficulty of a locally adaptive window lies, in fact, in a difficulty in evaluating and using disparity variances. While the intensity variation is directly measurable from the image, evaluation of the disparity variation is not easy, since the disparity *is* what we intend to calculate as an end product of stereo. To resolve the dilemma, an appropriate model of disparity variation is required which enables us to assess how disparity variation within a window affects the estimation of disparity.

The stereo algorithm we propose in this chapter selects a window adaptively by evaluating the local variation of the intensity and the disparity. We employ a statistical model that represents the uncertainty of disparity of points over the window: the uncertainty is assumed to increase with the distance of the point from the center point. This modeling enables us to compute both a disparity estimate *and* the uncertainty of the estimate obtained by using the particular window. So, the algorithm can search for a window that produces the estimate of disparity with the least uncertainty for each pixel of an image. The method controls not only the size but also the shape (rectangle) of the window.

In this chapter, we first develop a model of stereo matching in section 4.2. Section 4.3 shows how to estimate the most likely disparity and the uncertainty of the estimate based on the modeling in section 4.2. These two sections provide theoretical grounds of our proposed algorithm. In section 4.4, we present a complete description of a stereo algorithm which selects the appropriate window size and shape adaptively for each pixel. Section 4.5 provides experimental results with real stereo images. The quality of the disparity maps obtained demonstrates the effectiveness of the algorithm.

4.2 Modeling Stereo Matching

We will first develop a statistical model of the distribution of the difference of intensities of two images within a window. This is an extension to two dimensions of the disparity distribution model presented in section 2.3. Then we will compare our model of the disparity distribution with local support models of other stereo methods.

4.2.1 Distributions of Intensity Differences and Disparities in a Window

Let the stereo intensity images (or results of some preprocessing) be $f_1(x,y)$ and $f_2(x,y)$. Without loss of generality, we can assume that the baseline is parallel to the x-axis. Further let us assume $f_1(x,y)$ and $f_2(x,y)$ come from the same underlying intensity function with a disparity function $d_r(x,y)$ and additive noise. Then f_1 and f_2 are related by

$$f_1(x,y) = f_2(x + d_r(x,y), y) + n(x,y), \tag{4.1}$$

where $n(x,y)$ is Gaussian white noise

$$n(x,y) \sim N(0, 2\sigma_n^2). \tag{4.2}$$

The value σ_n^2 is the power of the noise per image.[1]

To simplify the notation, suppose that we want to compute the disparity at $(x,y) = (0,0)$, i.e., the value $d_r(0,0)$. Also, suppose a window $W = \{(\xi,\eta)\}$ is placed at the correct corresponding positions in both images, that is, at $(0,0)$ in image $f_1(x,y)$ and at $(d_r(0,0),0)$ in image $f_2(x,y)$ (Refer to figure 2.4 for the situation). Then, the value of f_1 at (ξ,η) in the window is $f_1(\xi,\eta)$, and that of f_2 is $f_2(\xi + d_r(0,0),\eta)$. These values would be the same, except for the noise component, if the disparity $d_r(\xi,\eta)$ were constant and equal to $d_r(0,0)$, but in general they are not. By expanding $f_2(\xi + d_r(\xi,\eta),\eta)$ at $\xi + d_r(0,0)$ up to the linear term and using equation (4.1), we see that the difference of intensities between f_1 and f_2 at (ξ,η) in the window can be approximated as

$$\begin{aligned} & f_1(\xi,\eta) - f_2(\xi + d_r(0,0),\eta) \\ \approx \; & (d_r(\xi,\eta) - d_r(0,0))\frac{\partial}{\partial\xi} f_2(\xi + d_r(0,0),\eta) + n(\xi,\eta). \end{aligned} \tag{4.3}$$

At this point, let us introduce the following statistical model for the disparity $d_r(\xi,\eta)$ within a window:

$$d_r(\xi,\eta) - d_r(0,0) \sim N\left(0, \alpha_d\sqrt{\xi^2 + \eta^2}\right), \tag{4.4}$$

where α_d is a constant that represents the amount of fluctuation of the disparity. That is, this model assumes that the difference in disparity at a point (ξ,η) in the window from that of the center point $(0,0)$ has a zero-mean Gaussian distribution with variance proportional to the distance between these points. In other words, the expected value of the disparity at (ξ,η) is the same as the center point, but it is expected to fluctuate more as the point is farther from the center.[2] Or, in terms of the scene, the small surface corresponding to the window in the image is statistically expected to be locally flat and parallel to the baseline,

[1] We use $2\sigma_n^2$ in equation (4.2) as the variance of $n(x,y)$ to indicate that it accounts for noise added to both f_1 and f_2.

[2] The statistical model of (4.4) can be shown equivalent to assuming that $d_r(\xi,\eta)$ is generated by Brownian motion (refer to [BN68][Vos87]). More generally, we can assume $d_r(\xi,\eta)$ to be a fractal. This corresponds to choosing a different degree of $\xi^2 + \eta^2$ between $(0,1)$ in the variance in (4.4). Brownian motion is the simplest case in which the degree is $\frac{1}{2}$ (see Appendix D). However, our preliminary experiments have shown no noticeable advantage of using a general fractal assumption.

but the expectation becomes less certain as the window becomes larger. More discussi on the implication of this assumption are presented in 4.2.2.

In order to facilitate the mathematical derivations to follow, we make an additio assumption that the image intensity $f_2(\xi, \eta)$ within the window is also generated by anot Brownian process which is independent of the one that has generated $d_r(\xi, \eta)$. This mea that the image intensity at a point within a window is expected to be the same as t center point, but that expectation is less certain the farther the point is from the cent In terms of the distribution of image intensity derivatives $\frac{\partial}{\partial \xi} f_2(\xi, \eta)$ within a window, t assumption is mathematically equivalent to assuming that they follow a zero-mean Gaussi white distribution which is independent of the distribution of disparities $d_r(\xi, \eta)$.[3]

Now we are ready to develop a statistical distribution of the intensity difference (4 between a pair of stereo images. Let us denote the right hand side of equation (4.3) $n_s(\xi, \eta)$. First, we compute the mean and variance of $n_s(\xi, \eta)$:

$$\begin{aligned} E\left[n_s(\xi, \eta)\right] &= E[d_r(\xi, \eta) - d_r(0,0)] E\left[\frac{\partial}{\partial \xi} f_2(\xi + d_r(0,0), \eta)\right] + E[n(\xi, \eta)] \\ &= 0 \qquad (4 \end{aligned}$$

$$\begin{aligned} E\left[(n_s(\xi, \eta))^2\right] &= E\left[\left((d_r(\xi, \eta) - d_r(0,0)) \frac{\partial}{\partial \xi} f_2(\xi + d_r(0,0), \eta)\right)^2\right] \\ &\quad + E\left[2(d_r(\xi, \eta) - d_r(0,0)) \left(\frac{\partial}{\partial \xi} f_2(\xi + d_r(0,0), \eta)\right) n(\xi, \eta)\right] \\ &\quad + E\left[(n(\xi, \eta))^2\right] \\ &= E\left[(d_r(\xi, \eta) - d_r(0,0))^2\right] E\left[\left(\frac{\partial}{\partial \xi} f_2(\xi + d_r(0,0), \eta)\right)^2\right] + E\left[(n(\xi, \eta))^2\right. \\ &= 2\sigma_n^2 + \alpha_f \alpha_d \sqrt{\xi^2 + \eta^2}, \qquad (4. \end{aligned}$$

where

$$\alpha_f = E\left[\left(\frac{\partial}{\partial \xi} f_2(\xi + d_r(0,0), \eta)\right)^2\right]. \qquad (4.$$

[3] Given no prior knowledge of a particular class of images or scenes, this assumption is justifiable on so grounds. Brownian motion is the simplest form of fractals which are often used to create natural text patterns. In television transmission technologies, it has been known that the image intensity difference sig z follows approximately an exponential distribution of the form $e^{-\alpha |z|^\beta}$ where α and β depend on the ty of the image. Also, except along occluding edges where intensity change and disparity change tend to occ simultaneously, intensity patterns can in general be independent of surface shapes.

$n_s(\xi, \eta)$ can be approximated by a Gaussian white distribution with the above mean and variance (refer to Appendix A). That is,

$$n_s(\xi, \eta) \;\approx\; f_1(\xi, \eta) - f_2(\xi + d_r(0,0), \eta) \;\sim\; N\left(0, 2\sigma_n^2 + \alpha_f \alpha_d \sqrt{\xi^2 + \eta^2}\right). \qquad (4.8)$$

The intuitive interpretation of (4.8) is as follows. Referring to figure 2.4, $n_s(\xi, \eta)$ is the difference between f_1 and f_2 at (ξ, η) within a window when the window is placed at the corresponding positions for obtaining the disparity at $(0, 0)$. If there is no additive noise $n(x, y)$ in the image (i.e., $\sigma_n = 0$) and the disparity is constant within the window (i.e., $\alpha_d = 0$), then the two images match exactly, and $n_s(\xi, \eta)$ must be null. Otherwise, however, the difference has a value which shows a combined noise characteristic which comes from both intensity and disparity variations. As derived in (4.8), we can model it by zero-mean Gaussian noise whose variance (power) is a summation of a constant term and a term proportional to $\sqrt{\xi^2 + \eta^2}$. The constant term is from the noise added to the image intensities. The second term is from *uncertain local support*. That is, while the points surrounding the center point in the window are used to support the matching for the center point, it should be noted that these points may actually increase, rather than decrease, the error in computing the disparity of the center point. This is because, in general, the disparity of the surrounding points deviates from that of the center point. This uncertainty is represented as if the intensity signals have additional noise whose power is proportional to the distance from the center point in the window. If the disparity is constant over the window (i.e., the surface is frontoparallel and $\alpha_d = 0$), the additional noise is zero. If the disparity changes more in the window (i.e., the larger α_d is), its effect becomes larger and the information contributed by the surrounding points becomes more uncertain. Also, note that the noise effect of the disparity variation is amplified by a factor of α_f, that is, by the amount of the intensity variation. This is because wrong correspondences due to disparity variation affect more severely when the intensity variation is higher.

4.2.2 Models of Disparity Distribution and Local Support in Stereo

Binocular stereo matching is in general ambiguous: there are often multiple equally good matches if the matching quality is evaluated independently at each point purely by using image properties, such as area correlation, edge orientation, and slope of Laplacian zero-crossing. In order to increase the reliability of matching, all the stereo matching algorithms developed so far examine the candidate matches by calculating how much support they receive from their local neighborhood. The manner in which this support from the local neighborhood is calculated varies between algorithms and is related to fundamental assumptions the algorithms make about the scene and its surfaces. Some algorithms state such assumptions very explicitly and others rather implicitly. It is interesting and revealing to compare our statistical model of the disparity distribution (equation (4.4)) with the assumptions about local support used in other stereo algorithms.

Hakkarainen, Little, Lee and Wyatt [HLLW91] present an excellent comparison of the three most representative local support assumptions: the surface continuity assumption in Marr-Poggio stereo [MP76][DP86], the disparity gradient limit by Pollard, Mayhew and Frisby [PMF85], and the disparity similarity function of Prazdny [Pra85]. The original cooperative algorithm by Marr and Poggio (MP) [MP76] uses a surface continuity assumption about the scene, and a match at a point looks for support from the matches in its local neighborhood which have the same disparity. Following Hakkarainen, Little, Lee and Wyatt [HLLW91], the diagram shown in figure 4.1 (a) provides a graphical representation of this local support assumption. A one-dimensional case is shown for simplicity where only a neighborhood along an epipolar line is considered. The horizontal and vertical axes represent the pixel position ξ and the disparity d, respectively, relative to those of the match of concern indicated by □. The thick segment on the horizontal axis on both sides of the origin indicates the region (ie., the combinations of ξ and d) that can contribute to support the match at $\xi = 0$ for $d = 0$ (relatively speaking): that is, the neighborhood ($|\xi| \leq \xi_{max}$) whose matches have the same disparity ($d = 0$) provides support. Basically, the MP stereo assumes frontoparallel surfaces, and disparity changes are discouraged. Grimson [Gri85] relaxes this assumption and allows neighboring points with disparities within a certain range to provide local support. Thus, the support assumption of Grimson's stereo can be represented as a rectangular region as shown in figure 4.1 (b).

Pollard, Mayhew and Frisby (PMF) [PMF85] place a limit on the disparity gradient for acceptable matches, where a disparity gradient is defined as the ratio of the disparity difference between two points to their distance apart; the disparity gradient limit assumption means $|d/\xi| \leq g_m$. This assumption is based on the observation that the disparity gradient for correct matches is small in most cases of binocular stereo. The gradient limit constrains the relative "jaggedness" of surfaces. In its implementation, the PMF stereo computes a local support in such a way that a match at a point receives support from neighboring matches that satisfy the disparity gradient limit; the support is weighted so that a closer neighborhood with a better match gives more support. Figure 4.1 (c) shows the region in the d - ξ plane which can provide support to the match at the origin. We see that the MP assumption corresponds to the case where $g_m = 0$.

Prazdny [Pra85] argues that the major mechanism in disambiguating disparity assignments is the "coherence principle", which states that neighboring disparities, if corresponding to the same 3D object, should be similar. Two neighboring pixels with similar disparities should support (or facilitate) each other, while pixels with dissimilar disparities should not inhibit (or interact) with each other. To incorporate this idea into a stereo algorithm, a function is needed which specifies the amount of support between neighboring points based on their disparities. Prazdny set three requirements for the function: it should be inversely proportional to the difference of disparities; more distant points should exert less influence;

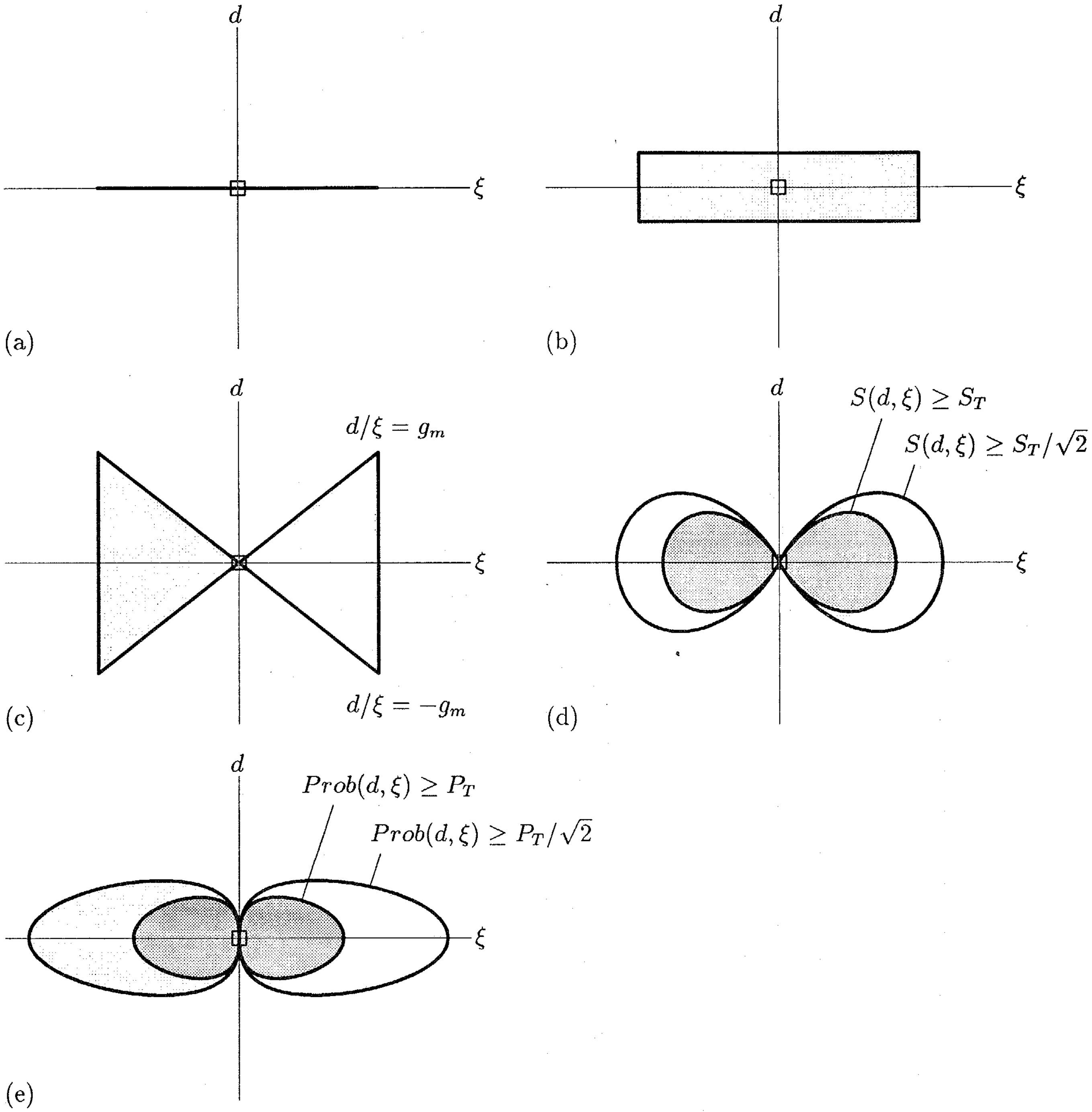


Figure 4.1: Graphical representation of various local-support assumptions. Each diagram shows the local support region that provides support to the match of concern □: (a) Marr-Poggio continuity assumption; (b) Grimson's neighboring points with disparities within a certain range; (c) Pollard-Mayhew-Frisby disparity gradient limit assumption; (d) Prazdny's support in his similarity function; (e) the disparity distribution model of (4) in this paper.

and the more distant the two points are, the less seriously should their disparity difference be considered. As a function which satisfies these requirements, Prazdny chose

$$s(i,j) = \frac{1}{c|i-j|\sqrt{2\pi}} e^{-\frac{|d_j - d_i|^2}{2c^2|i-j|^2}} \tag{4.9}$$

where $s(i,j)$ represents the amount of support that disparity d_i at pixel i receives from disparity d_j at j, $|i-j|$ is the distance between the two pixels, and c is a scaling constant. Graphically, figure 4.1 (d) shows the region which exerts a support more than a certain threshold, ie, $s(i,j) \geq s_T$. Note that $d_j - d_i$ corresponds to d and $i - j$ to ξ in our diagram.

Prazdny's similarity function (4.9) is exactly the same as our model of disparity distribution (4.4) in that the disparity difference between two pixels $d_j - d_i$ follows a zero-mean Gaussian distribution whose variance increases with their distance apart. The only difference is that in (4.9) the variance is proportional to the square of the distance between the pixels, instead of the distance itself as in (4.4). In fact, equation (4.9) is the limiting case of $H \to 1$ in equation (D.1) in Appendix D in which a general fractal surface assumption is discussed. Figure 4.1 (e) shows the region for which $Prob(d_r(\xi) - d_r(0)) \geq P_T$. Note that in both figures 4.1 (d) and 4.1 (e) the support becomes stronger as the disparities become similar ($d \to 0$) and the pixels become closer ($\xi \to 0$).

Prazdny presented several computational and psychophysical arguments to justify his choice of the support function, including the relationship of the term of the exponent $|d_j - d_i|/|j-i|$ to the disparity gradient. The function represents the bias toward frontoparallel planes, but as all the diagrams in figure 4.1 show, it is a graceful mix of the distance and the disparity difference into a support score. If we view the likelihood of disparity relative to the neighboring disparity as the support score, our model provides probabilistic justification for Prazdny's selection of the support function, and Prazdny's arguments provide psychophysical justification for our model.

All of the above assumptions on local support and disparity distribution emphasize frontoparallel planes. As pointed out by many researchers [HLLW91] [MP76] [DP86] [Gri85] [PMF85] [Pra85], however, this is not necessarily a problem. Acceptable results have been obtained for a scene which contains slanted surfaces, curved surfaces and even disparity jumps by stereo algorithms which use these and similar methods for computing local support. The diagrams in figure 4.1 indicates in fact that, relatively speaking, our model emphasizes (allows) more local variations of disparity than the support function of MP, Grimson, PMF or Prazdny stereo. Also, it should be noted that any stereo algorithm which involves some kind of smoothing or averaging over an area does indeed assume or have a bias towards frontoparallel planes, whether or not the assumption is stated explicitly.

Matching by SSD calculation requires in theory (assumes implicitly) the surface to be covered by a window to have the same disparity (i.e. a frontoparallel plane) in order for it to generate an exact estimate of disparity. Otherwise, the estimate becomes uncertain. The disparity distribution model of (4.4) has been introduced to allow us to evaluate that uncertainty, so that we can choose an appropriate window size which will generate the most certain estimate of disparity. The model does not necessarily limit the surface types to which the resultant stereo method will be applicable.

4.3 Estimating Disparity and Its Uncertainty

Now we will show how the disparity and its uncertainty can be estimated based on the modeling presented in the previous section. Let $d_0(x,y)$ be an initial estimate of the disparity $d_r(x,y)$. By using the Taylor expansion, the variable $n_s(\xi,\eta)$ in equation (4.8) is equal to

$$f_1(\xi,\eta) - f_2(\xi + d_0(0,0),\eta) - \Delta d \frac{\partial}{\partial \xi} f_2(\xi + d_0(0,0),\eta), \tag{4.10}$$

where $\Delta d = d_r(0,0) - d_0(0,0)$. Note that Δd is an incremental correction of the estimate to be made. Let us denote

$$\begin{aligned} \phi_1(\xi,\eta) &= f_1(\xi,\eta) - f_2(\xi + d_0(0,0),\eta) & (4.11) \\ \phi_2(\xi,\eta) &= \frac{\partial}{\partial \xi} f_2(\xi + d_0(0,0),\eta). & (4.12) \end{aligned}$$

Functions ϕ_1 and ϕ_2 are the image differences and image derivatives, respectively, within a window which is placed according to the initial estimate d_0. 'Using these notations we can rewrite equation (4.8) as

$$n_s(\xi,\eta) = \phi_1(\xi,\eta) - \Delta d \phi_2(\xi,\eta) \sim N(0,\sigma_s^2(\xi,\eta)), \tag{4.13}$$

where

$$\sigma_s^2(\xi,\eta) = 2\sigma_n^2 + \alpha_f \alpha_d \sqrt{\xi^2 + \eta^2}. \tag{4.14}$$

Now, by sampling image values f_1 and f_2 at (ξ_i,η_j) in the window W we obtain a sample φ_{ij} of $n_s(\xi,\eta)$

$$\varphi_{ij} = n_s(\xi_i,\eta_j) = \phi_1(\xi_i,\eta_j) - \Delta d \phi_2(\xi_i,\eta_j). \tag{4.15}$$

From (4.13), the conditional probability density function of φ_{ij} given Δd is

$$p(\varphi_{ij}|\Delta d) = \frac{1}{\sqrt{2\pi}\sigma_s(\xi,\eta)} \exp\left(-\frac{(\phi_1(\xi_i,\eta_j) - \Delta d \phi_2(\xi_i,\eta_j))^2}{2\sigma_s^2(\xi,\eta)}\right). \tag{4.16}$$

Since $n_s(\xi,\eta)$ is white noise, the φ_{ij}'s are mutually independent. So the joint distribution of φ_{ij}'s for the points in the window is

$$p(\varphi_{ij}(i,j \in W)|\Delta d) = \prod_{i,j \in W} p(\varphi_{ij}|\Delta d), \tag{4.17}$$

where $\prod_{i,j \in W}$ denotes the product over the window.

The task is to estimate Δd given measurements φ_{ij}'s. Therefore, using the continuo version of Bayes' theorem we compute

$$p(\Delta d | \varphi_{ij}(i,j \in W)) = \frac{p(\varphi_{ij}(i,j \in W)|\Delta d)p(\Delta d)}{\int_{-\infty}^{\infty} p(\varphi_{ij}(i,j \in W)|\Delta d)p(\Delta d)d(\Delta d)}. \qquad (4.1$$

Assuming no prior information of Δd (i.e., $p(\Delta d)$ = constant), substitution of (4.16) a (4.17) into (4.18) yields (see Appendix E for derivation):

$$p(\Delta d | \varphi_{ij}(i,j \in W)) = \frac{1}{\sqrt{2\pi}\sigma_{\Delta d}} \exp\left(-\frac{(\Delta d - \hat{\Delta d})^2}{2\sigma_{\Delta d}^2}\right), \qquad (4.1$$

where

$$\hat{\Delta d} = \frac{\sum_{i,j \in W}(\phi_1(\xi_i,\eta_j)\phi_2(\xi_i,\eta_j)/\sigma_s^2(\xi_i,\eta_j))}{\sum_{i,j \in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2} \qquad (4.2$$

$$\sigma_{\Delta d}^2 = \frac{1}{\sum_{i,j \in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2}, \qquad (4.2$$

where $\sum_{i,j \in W}$ denotes the summation over the window. Or, by substituting equations (4.11 (4.12), and (4.14) into equations (4.20) and (4.21), we obtain

$$\hat{\Delta d} = \frac{\sum_{i,j \in W} \frac{(f_1(\xi_i,\eta_j) - f_2(\xi_i + d_0(0,0),\eta_j))\frac{\partial}{\partial \xi} f_2(\xi_i + d_0(0,0),\eta_j)}{2\sigma_n^2 + \alpha_f \alpha_d \sqrt{\xi_i^2 + \eta_j^2}}}{\sum_{i,j \in W} \frac{(\frac{\partial}{\partial \xi} f_2(\xi_i + d_0(0,0),\eta_j))^2}{2\sigma_n^2 + \alpha_f \alpha_d \sqrt{\xi_i^2 + \eta_j^2}}} \qquad (4.2$$

$$\sigma_{\Delta d}^2 = \frac{1}{\sum_{i,j \in W} \frac{(\frac{\partial}{\partial \xi} f_2(\xi_i + d_0(0,0),\eta_j))^2}{2\sigma_n^2 + \alpha_f \alpha_d \sqrt{\xi_i^2 + \eta_j^2}}}. \qquad (4.2$$

Equation (4.19) says that the conditional probability density function of Δd given the ol served stereo image intensities over the window becomes a Gaussian probability densit function. The mean value and the variance of the Gaussian probability are $\hat{\Delta d}$ and σ_{Δ}^2 computed with equations (4.22) and (4.23). That is, $\hat{\Delta d}$ and $\sigma_{\Delta d}^2$ provide the maximur likelihood estimate of the disparity (increment) and the uncertainty of the estimation for th given window W, respectively.

The parameters α_d and α_f represent the disparity fluctuation and the intensity fluctuation, respectively. We estimate them locally within the window from equations (4.4) an (4.7),

$$\hat{\alpha_d} = \frac{1}{N_w} \sum_{i,j \in W} \frac{(d_0(\xi_i,\eta_j) - d_0(0,0))^2}{\sqrt{\xi_i^2 + \eta_j^2}} \qquad (4.24$$

$$\hat{\alpha_f} = \frac{1}{N_w} \sum_{i,j \in W} \left(\frac{\partial}{\partial \xi} f_2(\xi_i + d_0(0,0), \eta_j)\right)^2, \qquad (4.25$$

where N_w is the number of the samples within the window. These parameters change as the shape and size of a window changes.

In summary, given images f_1 and f_2, a window W, and the current estimate of disparities $d_0(\xi, \eta)$ within the window, use of equations (4.22) - (4.25) will enable us to calculate a better estimate of disparity $d_0(0,0) + \Delta d$ at the center of the window, as well as the uncertainty of this estimation. The goal of our stereo algorithm will now become finding the disparity estimate with the lowest uncertainty.

4.4 Iterative Stereo Algorithm with an Adaptive Window

In the previous sections we have developed a theory for computing the estimates of the disparity increment and its uncertainty, which take into account the fact that not only the intensity but also the disparity varies within a window. We now present a complete description of our stereo algorithm with an adaptive window:

1. Start with an initial disparity estimate $d_0(x,y)$. We obtained this at pixel resolution by using the multiple-baseline stereo matching method as presented in section C.1 for better initial estimates without suffering from ambiguity.[4]

2. For each point (x,y), we want to choose a window that provides the estimate of disparity increment having the lowest uncertainty. For the chosen window, calculate the disparity increment by (4.22) and update the disparity estimate by $d_{i+1}(x,y) = d_i(x,y) + \Delta d(x,y)$.

   Here we need a strategy to select a window that results in the disparity estimate having the lowest uncertainty. In the discussions so far, the shape of the window can be arbitrary. In practice we limit ourselves to a rectangular window, as illustrated in figure 4.2, whose width and height can be independently controlled in all four directions. Our strategy is as follows:

   (a) Place a small 3×3 window centered at the pixel, and compute the uncertainty by using (4.24), (4.25), and (4.23).

   (b) Expand the window by one pixel in one direction, e.g., to the right $x+$, for trial, and compute the uncertainty for the expanded window. If the expansion increases the uncertainty, the direction is prohibited from further expansions. Repeat the

[4] In the context of the discussion of this chapter, this initial estimate can be obtained by any existing stereo algorithm. One alternative is a simple SSD-based method, i.e. finding the corresponding point which gives the minimum SSD over a window at the pixel resolution. We used this method for the following examination with synthesized images shown in this section.

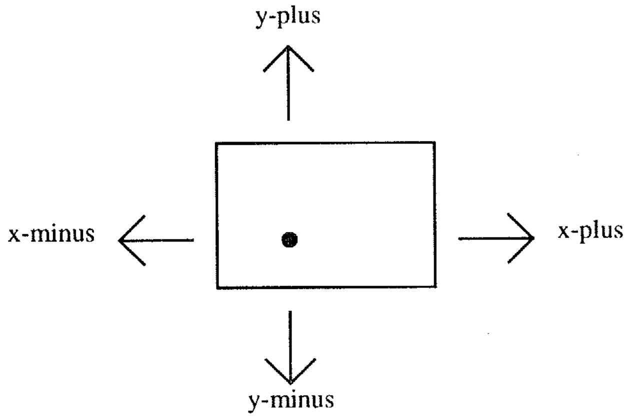


Figure 4.2: Window expansion

same process for each of the four directions $x+, x-, y+$, and $y-$ (excludi: already prohibited ones).

(c) Compare the uncertainties for all the directions tried and choose the di: which produces the minimum uncertainty.

(d) Expand the window by one pixel in the chosen direction.

(e) Iterate steps (b) to (d) until all directions become prohibited from expans until the window size reaches a limit that is previously set.

Thus, our strategy is basically a sequential search for the best window by ma: descent starting with the smallest window

3. Iterate the above process until the disparity estimate $d_i(x, y)$ converges, or u certain maximum number of iterations.

Now, by using synthesized data, we will examine how the window is adaptively the stereo algorithm for each position in an image and demonstrate its advantage. I 4.3 (a) and (b) show the left and the right images of the test data. In generating th set, a linear ramp in the direction of the baseline is used as the underlying intensity p. It is deformed according to the disparity pattern in figures 4.3 (c) and (d), and Ga noise is added independently to both images. We apply the iterative stereo algorithm resultant data.

First, we will examine the result of window selection. The four images in figure 4. the length (increasing brightness corresponds to increasing length) by which the wind been extended in each of the four directions.[5] For example, the vertical dark stripes ir

[5] Actually these are the average of ten runs with different noises to obtain the general tendenc) than an accidental set up.

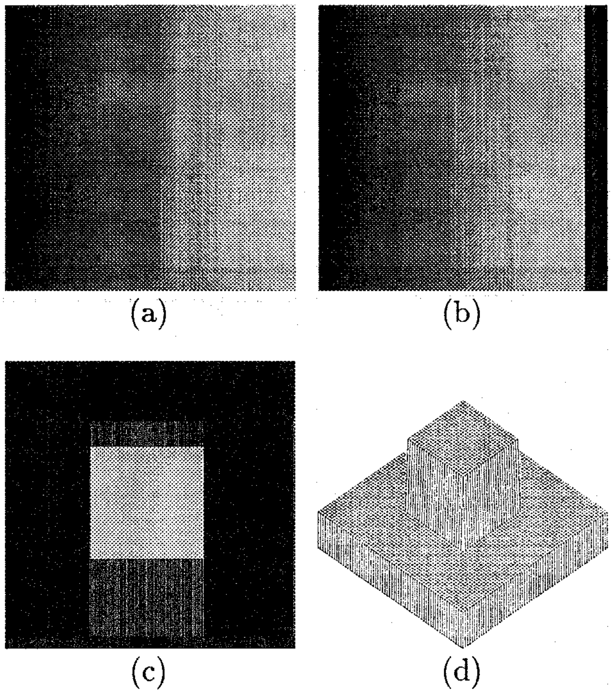


Figure 4.3: Synthesized stereo images, with a ramp intensity pattern with Gaussian noise: (a) Left image; (b) Right image; (c) Disparity pattern; (d) An isometric plot of the disparity pattern.

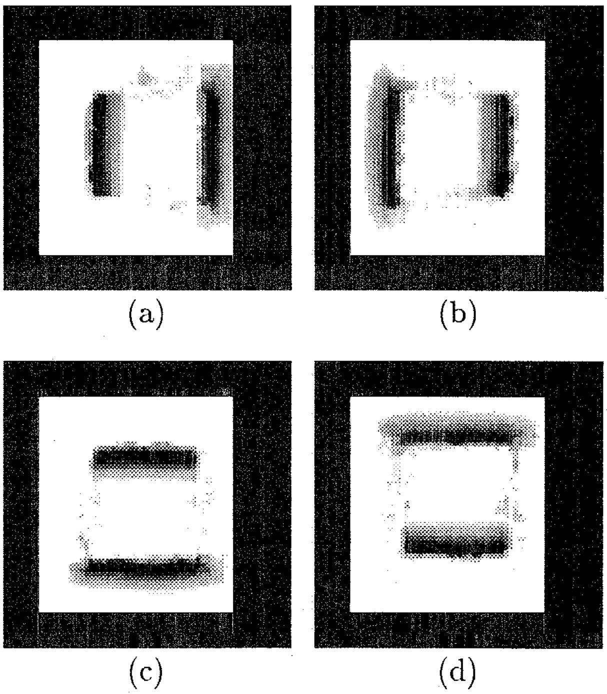


Figure 4.4: Extent of window-size expansion for each direction: (a) Left (X-minus) direction; (b) Right (X-plus) direction; (c) Down (Y-minus) direction (d) Up (Y-plus) direction.

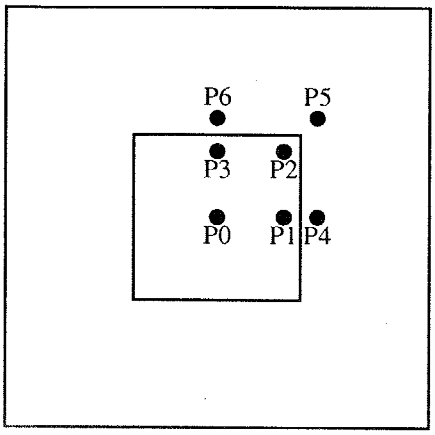


Figure 4.5: Positions for which size and shape of selected windows are examined.

4.4 (a) on the right hand side of the vertical disparity edge show that the windows for points are not extended to the left so that the windows do not cross the disparity edg region of different disparity. We observe the same phenomena in the other directions. W examine the size and shape of the selected windows at several representative positions s in figure 4.5. The windows selected at those positions are drawn by dashed lines in 4.6 relative to the disparity edges drawn by solid lines. For example, at $P0$ a windo been expanded to the limit for all directions, whereas at $P1$ expansion to the right has stopped at the disparity edge. At $P5$, a window is elongated either vertically or horizon depending on the image noise, but consistently avoids the corner of the disparity jum

Next, let us examine the computed disparities. For comparison, we also have com disparities by an iterative fixed-window-size SSD-based stereo method, that is, by ru the same iterative algorithm except that in Step 2 of the stereo algorithm a wind predetermined size is used assuming a constant disparity over the window. We run three window sizes, 3 × 3, 7 × 7, and 15 × 15. Figures 4.7 (a), (b) and (c) show the produced by fixed window sizes, and (d) by the adaptive-window algorithm. We can c see the problem with using a predetermined fixed window size. A larger window is go flat surfaces, but it blurs the disparity edges. In contrast, a smaller window gives sh disparity edges at the expense of noisy surfaces. The computed disparity by the ada window algorithm shown in figure 4.7 (d) shows both smooth flat surfaces and sharp disp edges. The improvements are further visible by plotting the absolute difference betwee computed and true disparities as shown in figure 4.8, with a table that lists their mean values. The adaptive-window algorithm has the smallest mean error, but more import we should observe that the algorithm has reduced two types of errors. A small fixed wi results in large random error everywhere. A large fixed window removes the random but introduces systematic errors along the disparity edges. The adaptive-window

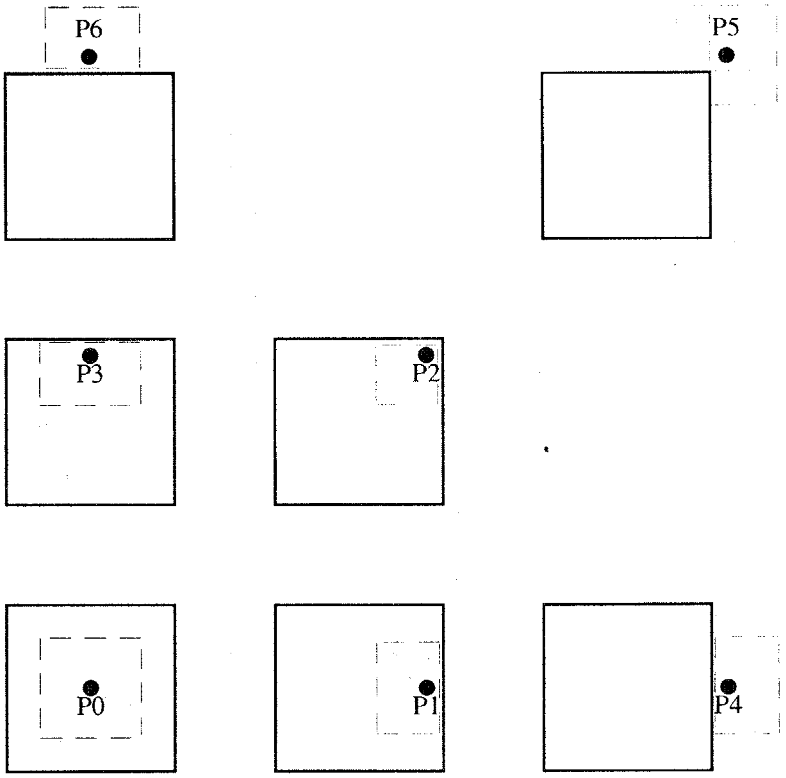


Figure 4.6: Selected windows for each position

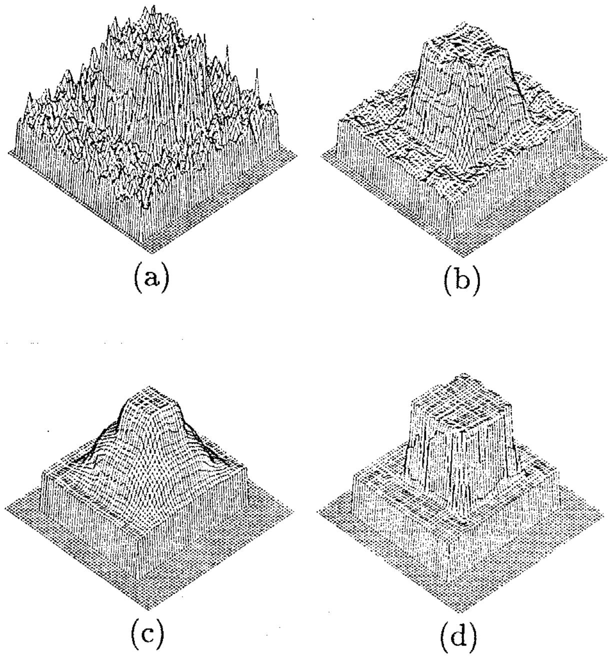


Figure 4.7: Isometric plots of the computed disparity by: (a) a 3×3 window; (b) a 7×7 window; (c) a 15×15 window; (d) the adaptive window algorithm.

method generates small errors of both types simultaneously.

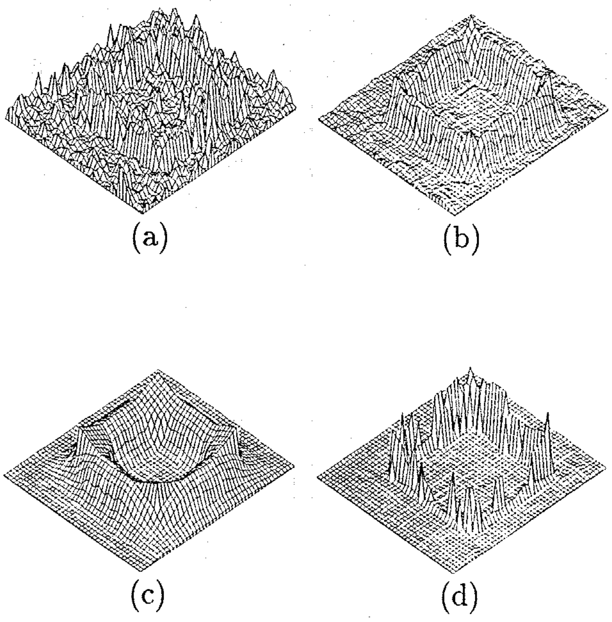
Figures 4.9 (a) and (b) show another example of synthesized test data. Figure 4.10 presents the computed disparity by the new method in (d), together with the results produced by fixed window sizes in (a) to (c) for comparison. As with the previous example, we clearly see better performance with the new method. The behavior of the window-size adaptation has been analyzed theoretically and tested with synthesized signals for various cases of disparity patterns including step, linear, and quadratic functions in chapter 2.

4.5 Experimental Results with Real Images

We have applied the adaptive-window based stereo matching algorithm presented in this paper to real stereo images.

Figure 4.11 shows images of a town model that were taken by vertically-displaced cameras. The disparity, therefore, is in the vertical direction. To give an idea of the arrangement of objects in the scene, a picture taken from an oblique angle is given in figure 4.11 (c).

For initial disparity estimates, we have used multiple-baseline stereo matching presented



Window	Mean Error Value (pixel)
3 × 3	0.22
7 × 7	0.20
15 × 15	0.34
Adaptive Window	0.08

Figure 4.8: Difference between the true disparity and the computed disparity: (a) by a 3 × 3 window; (b) by a 7 × 7 window; (c) by a 15 × 15 window; (d) by the adaptive window.

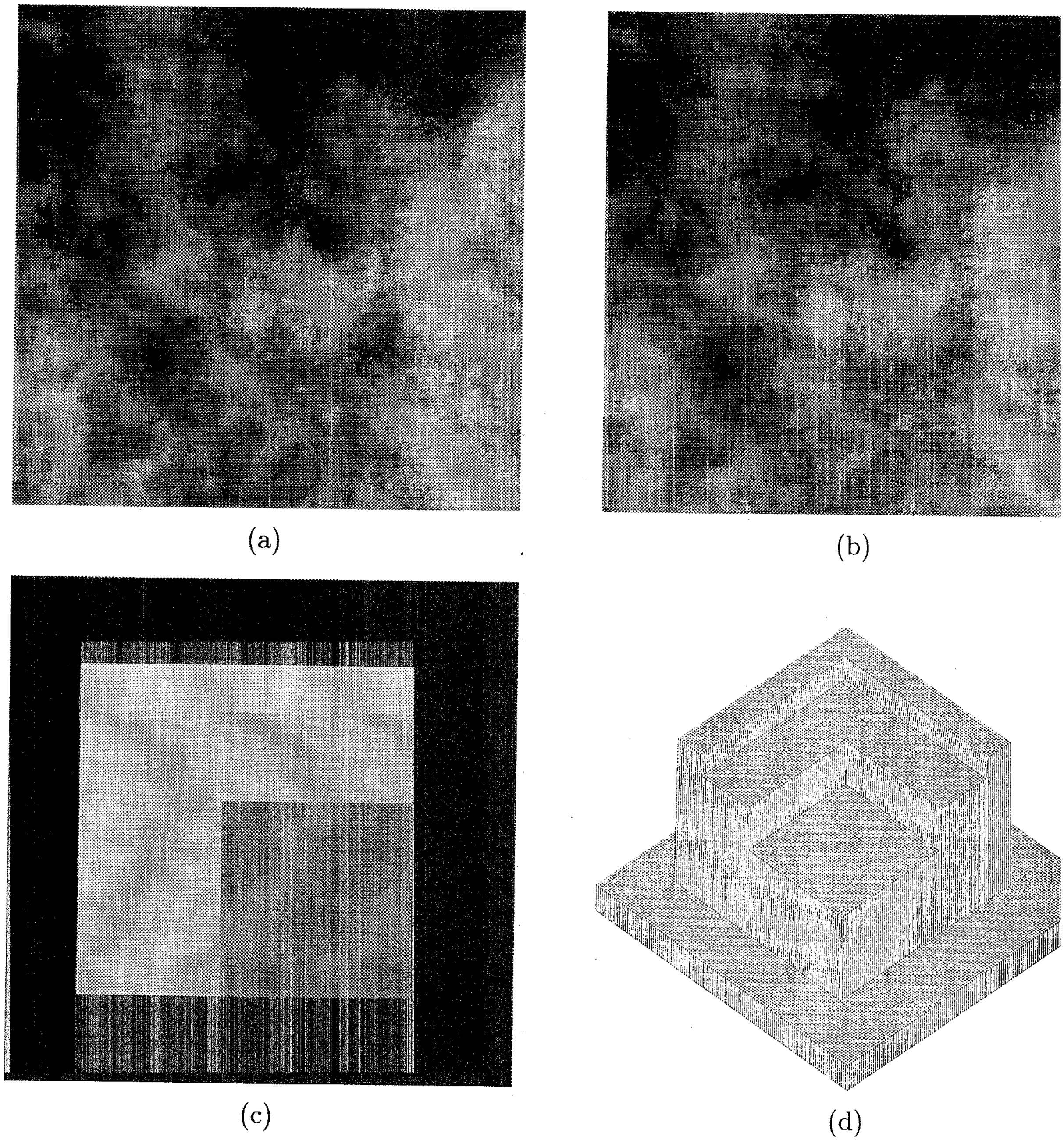


Figure 4.9: Synthesized stereo images no. 2: (a) Left image; (b) Right image; (c) Dispa pattern; (d) Isometric plot of the disparity pattern shown in (c).

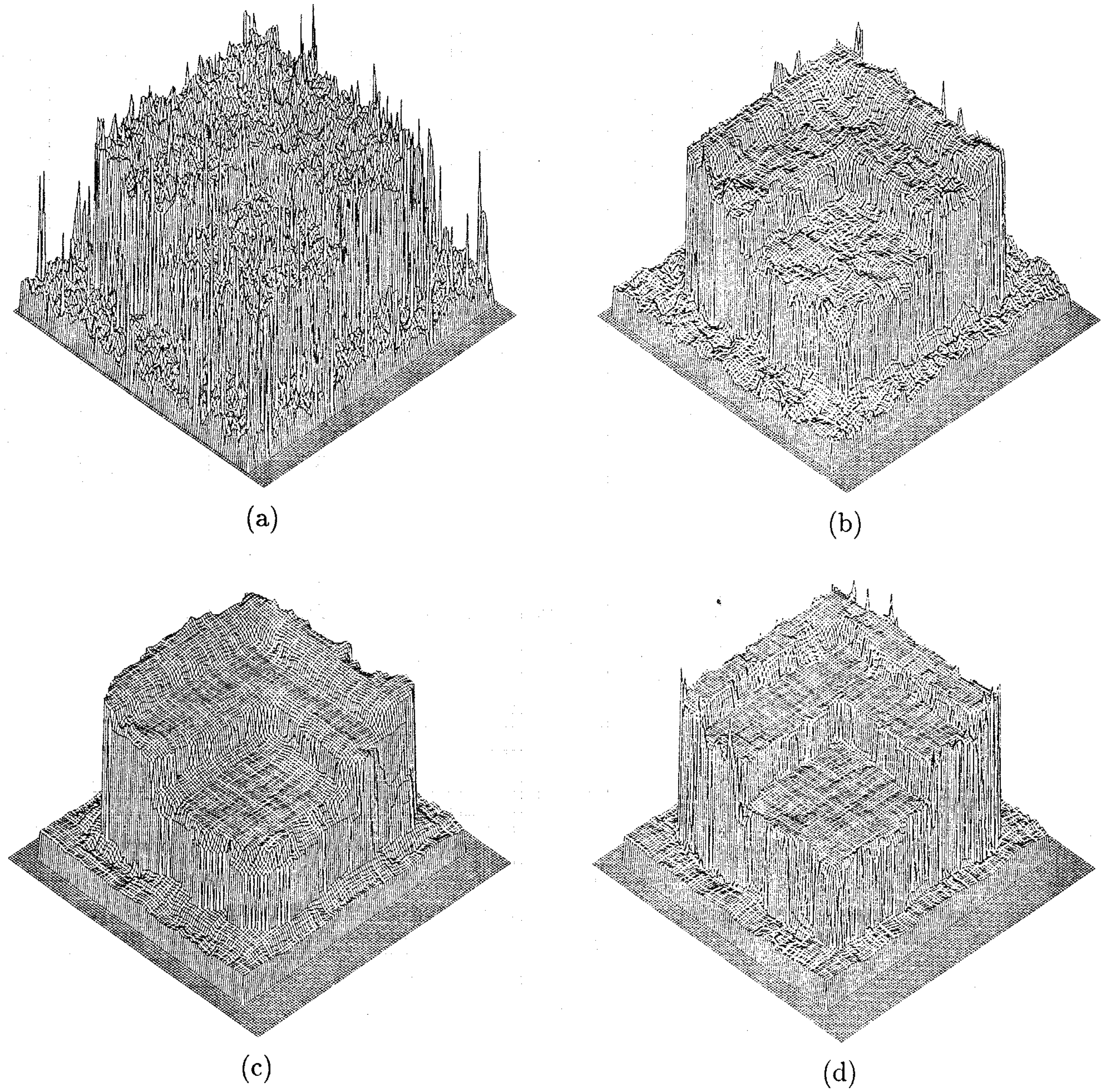
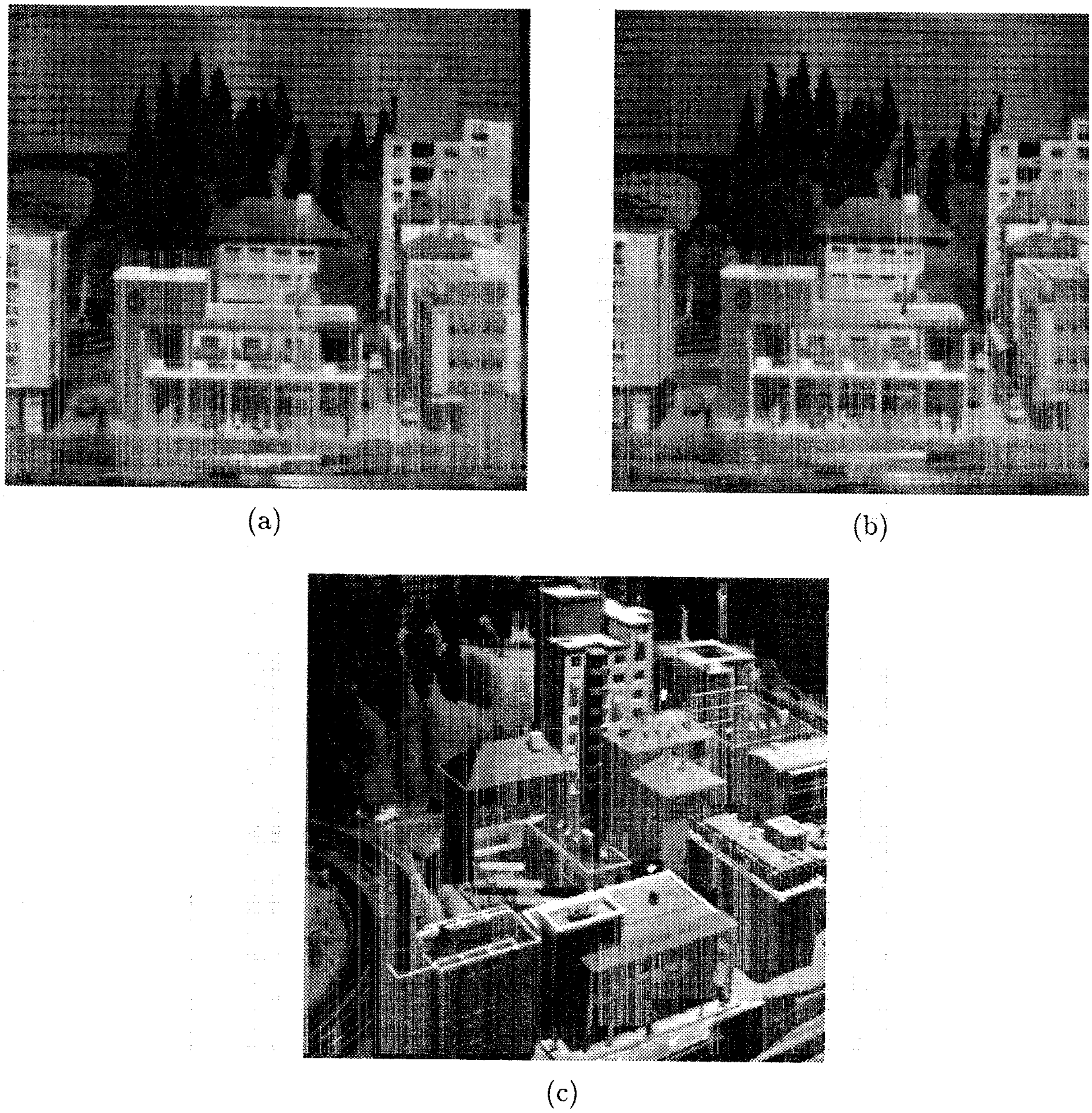


Figure 4.10: Computed disparities by: (a) a fixed 3 × 3 window; (b) a fixed 7 × 7 window; (c) a fixed 15 × 15 window; (d) the adaptive window.



(a)

(b)

(c)

Figure 4.11: "Town" stereo data set: (a) Upper image of stereo;(b) Lower image of ster (c) Oblique view.

in chapter 3 which can remove matching ambiguities due to repetitive patterns, especially in the top portion of the image. The number of iterations in step 3 of the algorithm description was set to 5. Figure 4.12 (a) shows the disparity map computed by the adaptive window algorithm. In addition, the uncertainty estimate computed by the algorithm is shown in figure 4.12 (b): increasing brightness corresponds to higher uncertainty. With this uncertainty estimate we can locate the regions whose computed disparity is not very reliable (very white regions in figure 4.12 (b)). In this example, they are either due to aliasing caused by the fine texture of roof tiles of a building (in the middle part of the image) or due to occlusion (the others). The isometric plot of the disparity map is shown in figure 4.12 (c), which roughly corresponds to the viewing angle of figure 4.11 (c). We can see that each building wall has a smooth surface and yet is clearly separated from others, and the shape of the distant bridge (on the left) is recovered. For comparison, the resultant isometric plots of the disparity maps with fixed window sizes are shown in figures 4.13 (a) 3×3, (b) 7×7, and (c) 15×15. We observe noisy surface reconstraction by a small window and over-smoothing of disparity edges by a large window.

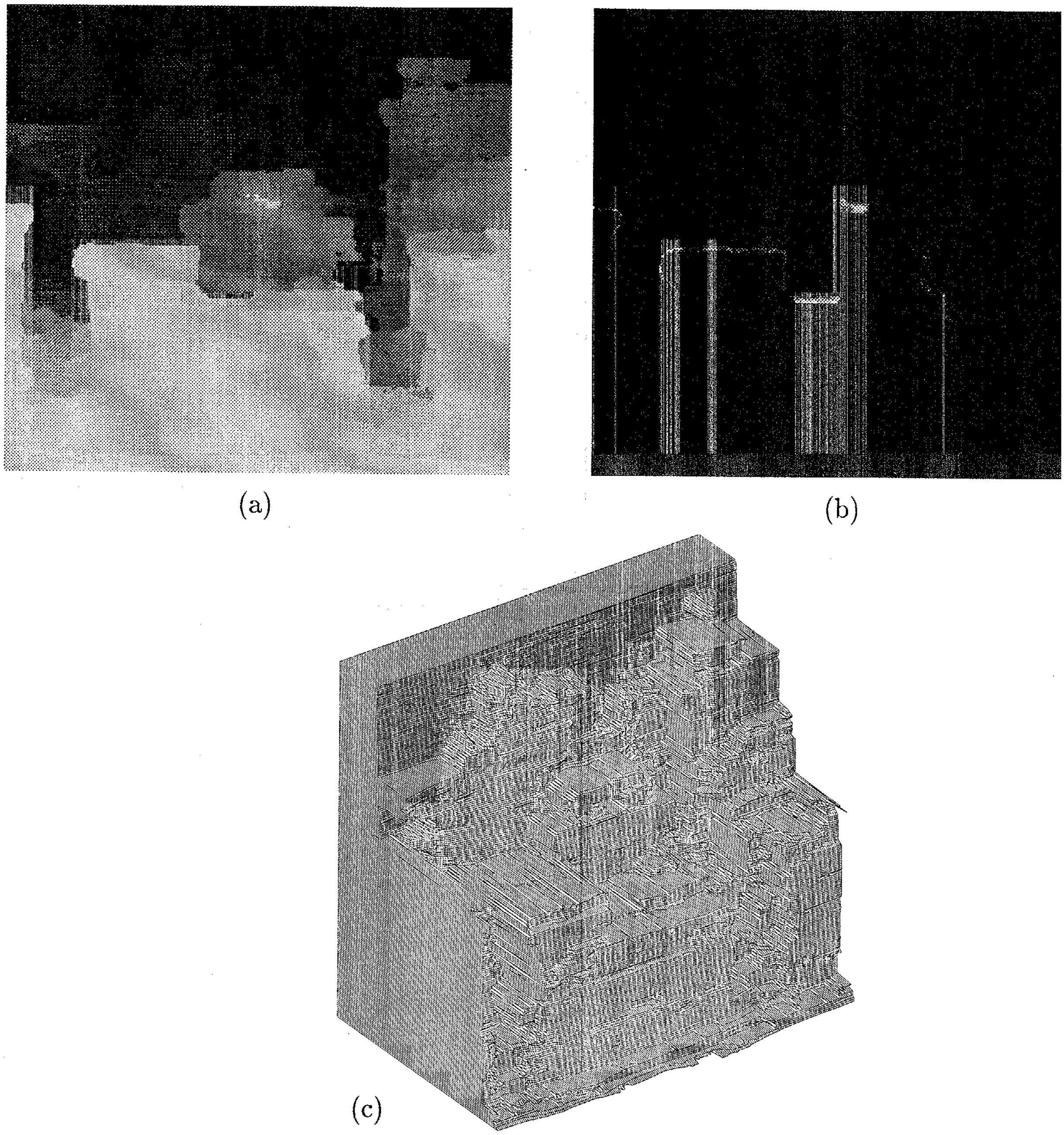
Figure 4.14 presents perspective views of the recovered scene by texture mapping the original intensity image on the constructed disparity map shown in figure 4.12 and generating views from new positions which are outside of the original stereo views. They can give an idea of the quality of reconstruction. This stereo data set is the same one used in [MSK89]. We can observe a noticeable improvement of the result over the previous result. Also it should be noted that this is extremely narrow baseline stereo: the baseline is only 1.2 cm long and the scene is about 1m away from the camera, thus the depth to the baseline ratio is approximately 80.

Figures 4.15 (a) and (b) show another set of real stereo images which are top views of a coal mine model. Figures 4.16 (a) and (c) show the isometric plots of the computed disparity. For comparison, actual pictures of the model taken from roughly the same angles are given in figures 4.16 (b) and (d). The shapes of buildings, a Λ-shaped roof, a water tank on the roof, and flat ground have been recovered without blurring the edges.

4.6 Conclusions

In this chapter, we have presented an iterative stereo matching algorithm using an adaptive window. The algorithm selects a window adaptively for each pixel so that it produces the disparity estimate having the least uncertainty. By evaluating both the intensity and the disparity variations within a window, we can compute both the disparity estimate and its uncertainty which can then be used for selecting the locally adaptive window.

The key idea for the algorithm is that it employs a statistical model that represents uncertainty of disparity of points over the window: the uncertainty is assumed to increase with the distance of the point from the center point. This model has enabled us to assess



(a)

(b)

(c)

Figure 4.12: Disparity and uncertainty computed by the adaptive window algorithm for t "town" stereo data: (a) Disparity map; (b) Uncertainty; (c) Isometric plot of the dispar map.

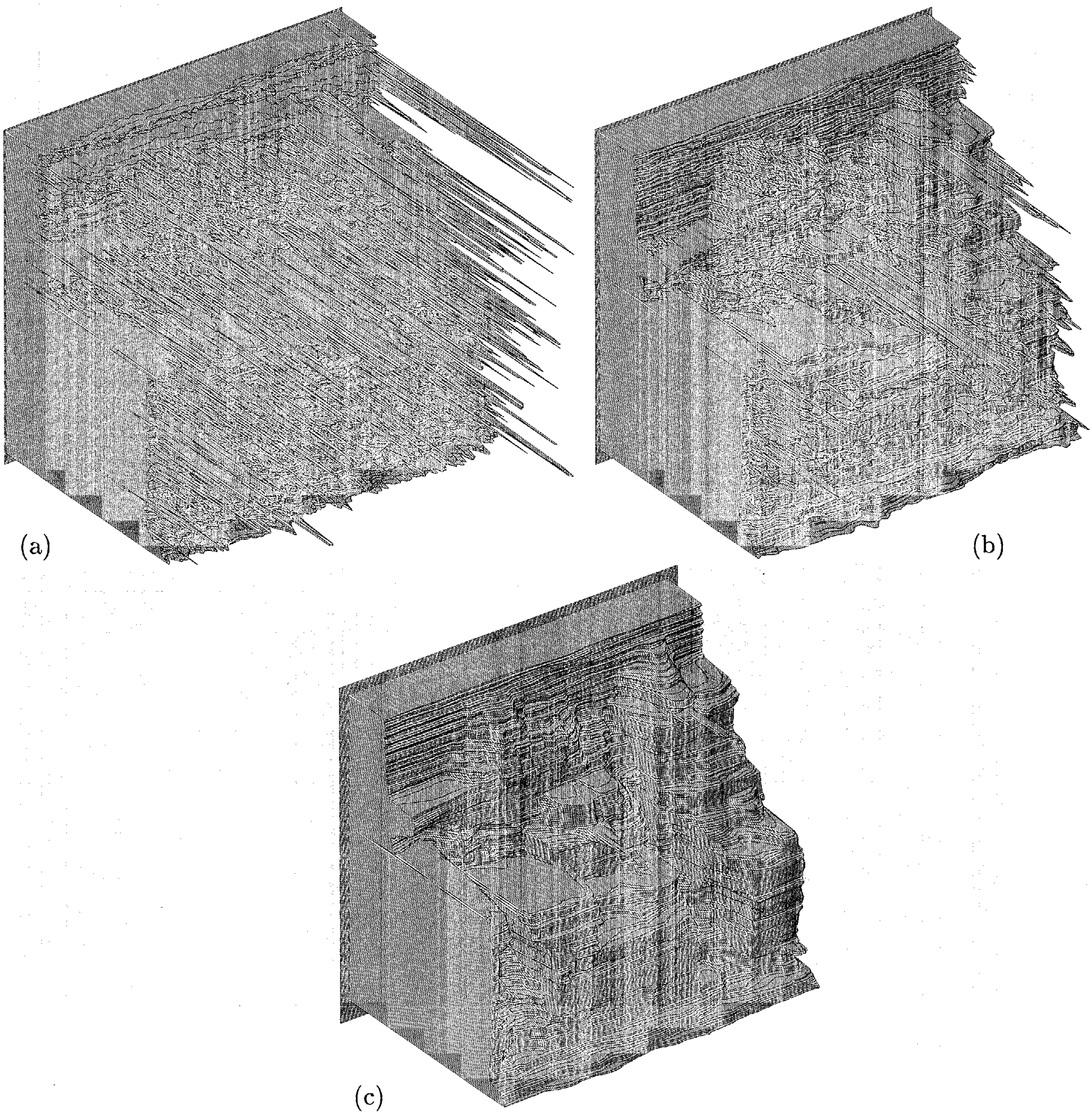


Figure 4.13: Isometric plots of the disparity maps computed by fixed-size windows: (a) 3 × 3; (b) 7 × 7; (c) 15 × 15.

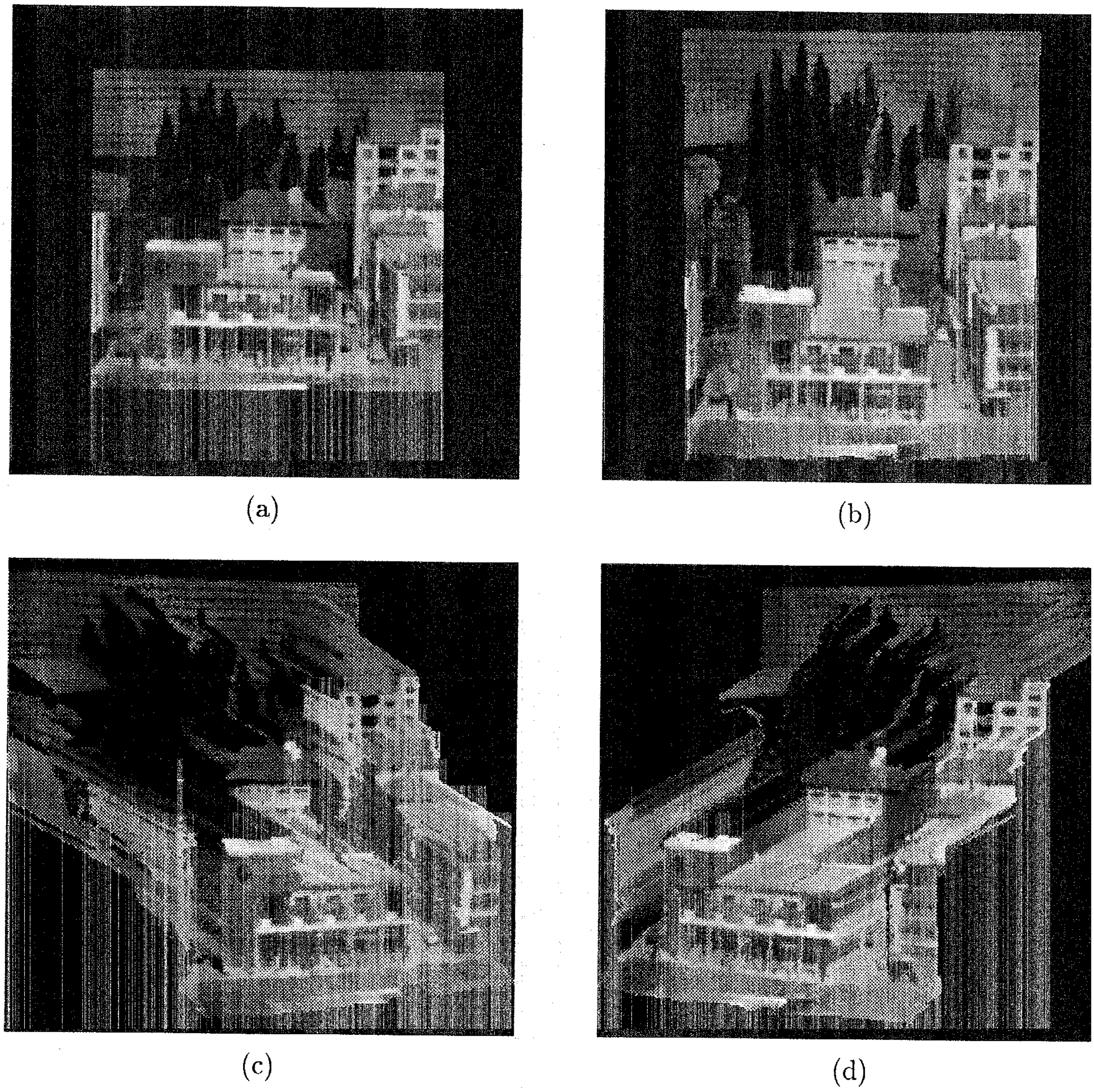
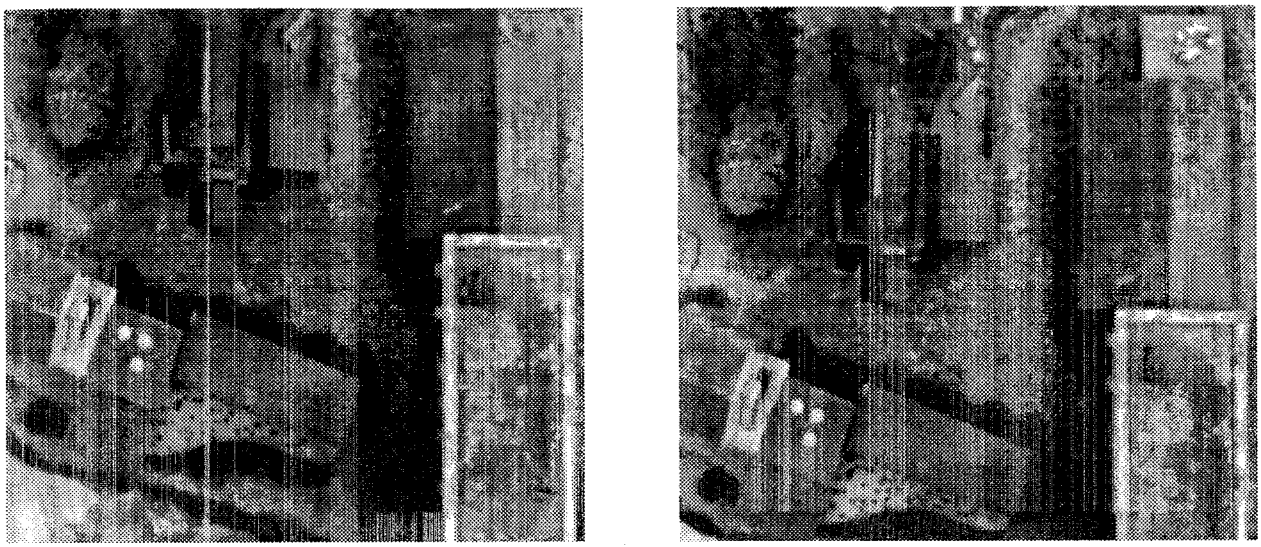


Figure 4.14: Perspective views of the recovered scene: (a) from the original camera positioı (b) from an upper position; (c) from an upper left position; (d) from an upper right positioı



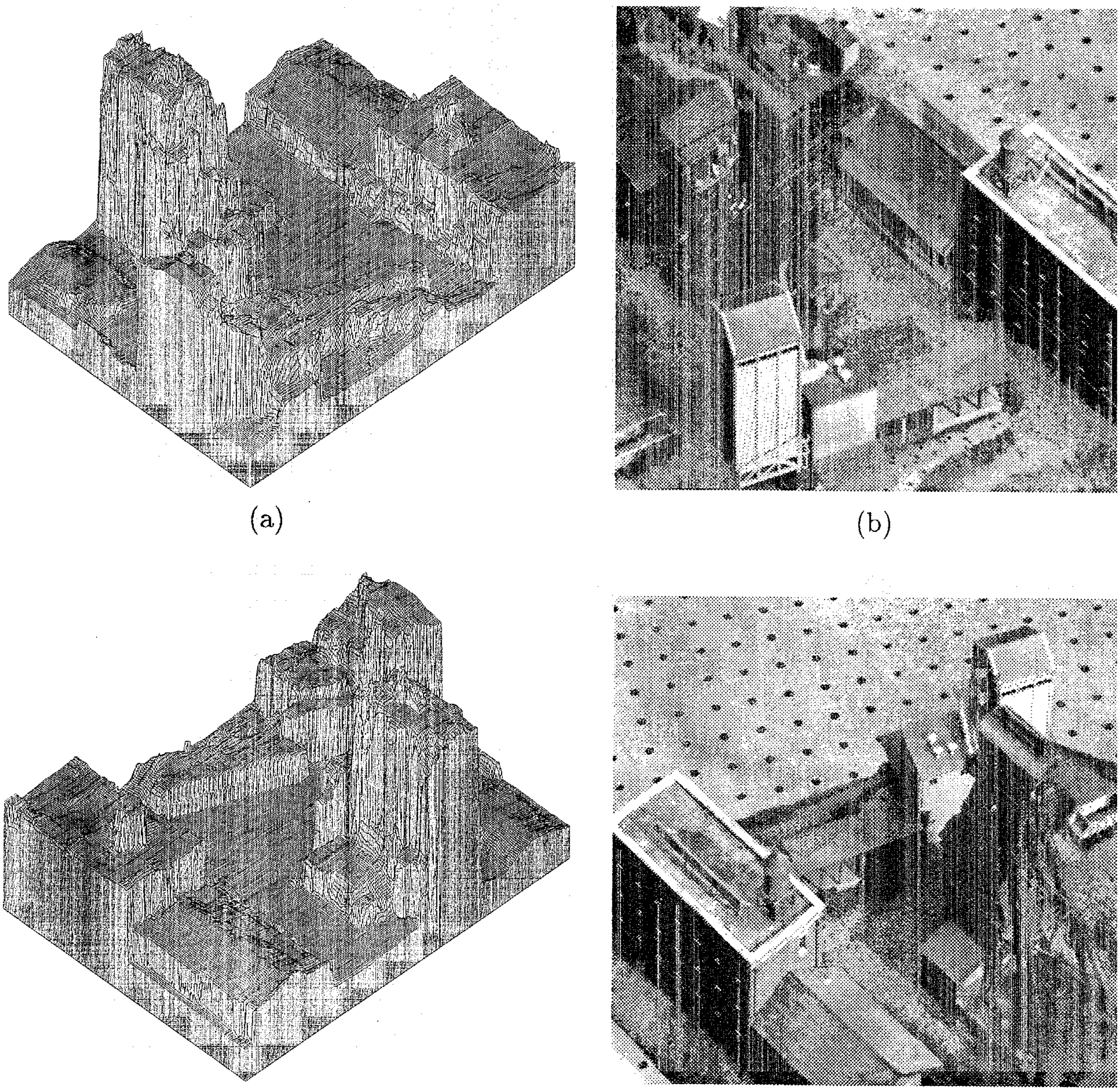
(a) (b)

Figure 4.15: "Coal mine" stereo data set: (a) Lower image; (b) Upper image.

how disparity variation within a window affects the estimation of disparity.

An important feature of the algorithm is that it is completely local and does not include any global optimization. Also, the algorithm does not use any post-processing smoothing, but smooth surfaces are recovered as smooth while sharp disparity edges are retained.

The experimental results have demonstrated a clear advantage of this algorithm over algorithms with a fixed-size window both on synthetic and on real images.



(a)

(b)

(c)

(d)

Figure 4.16: Isometric plots of the computed disparity map and their corresponding actu view: (a) (b) Isometric plot and corresponding view from the lower left corner; (c) (c Isometric plot and corresponding view from the upper right corner.

Chapter 5

Color Stereo Matching and Its Medical Application

5.1 Introduction

Almost all of the stereo algorithms proposed previously have used only gray-level intensity information, although the possibility of using color information to improve stereo matching has been sometimes mentioned, e.g. [DP86]. Recently Jordan et. al [IB91] investigated edge-based stereo correspondence which uses chromatic information.

In this chapter, we discuss this aspect, i.e. use of color information, in stereo vision. First we show the effect of using color information mathematically in area-based stereo matching which is potentially more general than feature-based matching since the former usually uses all information involved in the images and does not select or reject any information like the latter. We then describe a stereo algorithm which uses color stereo images.

Next, we show experimental results with synthesized images to illustrate the effect of using color information in our method. Also, experimental results by applying the method to real stereo images of ocular funduses are presented. Glaucoma is an eye disease which is a common cause of blindness. The 3-D shape of the optic nerve head is one of the most important information in diagnosing the disease at an early stage and for monitoring its development. As relevant work for this purpose, Lee et. al [LB91] introduced a technique which integrates stereo and photometric stereo. We have made a system which measures the 3-D shape of the optic nerve head by using the proposed color stereo algorithm and displays the result in various ways. The experimental results demonstrate the effectiveness of our method and the possibility of giving useful information about the disease.

5.2 Color Stereo Matching

5.2.1 Effect of Using Color Information

We have described in chapter 3 that "precision" and "accuracy" in matching should be considered separately in evaluating stereo matching. Roughly speaking, precision is mainly related to random noise added to the images and accuracy to the ambiguity inherent in the underling intensity pattern itself, though they interact with each other. In this section, we mathematically show that precision will be increased by using color images for stereo matching. For simplicity, we use one dimensional intensity signals for analysis, but the extension to two dimensional images is straightforward.

We model the color stereo images $f_{R1}(x)$, $f_{G1}(x)$, $f_{B1}(x)$ and $f_{R2}(x)$, $f_{G2}(x)$, $f_{B2}(x)$ as:

$$f_{R1}(x) = f_R(x) + n_{R1}(x) \tag{5.1}$$

$$f_{R2}(x) = f_R(x - d_r) + n_{R2}(x) \tag{5.2}$$

(same for $f_{G1}(x)$, $f_{G2}(x)$, $f_{B1}(x)$, and $f_{B2}(x)$),
assuming disparity d_r is constant near x and $n_{R1}(x)$ and $n_{R2}(x)$ are independent Gaussian white noise such that:

$$n_{R1}(x), n_{R2}(x) \sim N(0, \sigma_n^2) \tag{5.3}$$

(same for $n_{G1}(x)$, $n_{G2}(x)$, $n_{B1}(x)$, and $n_{B2}(x)$).

Then, we use the sum of squared differences (SSD) of the intensity values over a window as a criterion for correspondence between left and right images. For color stereo images, it is computed such that

$$e_{RGB}(x,d) = \sum_{Q=R,G,B} \sum_{j \in W} (f_{Q1}(x+j) - f_{Q2}(x+j+d))^2, \tag{5.4}$$

where $\sum_{j \in W}$ means summation over the window. This criterion represents the squared norm of the difference of two vectors which consist of the intensity values over the windows in the left and right images. The d in (5.4) which produces the minimum SSD gives the estimate of the disparity.

Substituting equations (5.1) to (5.3) into equation (5.4),

$$e_{RGB}(x,d) = \sum_{Q=R,G,B} \sum_{j \in W} (f_Q(x+j) - f_Q(x+j+d-d_r) + n_Q(x+j))^2, \tag{5.5}$$

where $n_Q(x)$ is Gaussian white noise such that

$$n_Q(x) \sim N(0, 2\sigma_n^2). \tag{5.6}$$

By taking the Taylor expansion about $d = d_r$ up to the linear term, we obtain

$$f_Q(x+j+d-d_r) \approx f_Q(x+j) + (d-d_r) f'_Q(x+j). \tag{5.7}$$

Substituting this into equation (5.5), we can approximate $e_{RGB}(x,d)$ near d_r by a quadratic form of d:

$$e_{RGB}(x,d) = a(x)(d-d_r)^2 - 2b(x)(d-d_r) + c(x), \tag{5.8}$$

where

$$\begin{aligned} a(x) &= \sum_{Q=R,G,B} \sum_{j \in W} (f'_Q(x+j))^2 & (5.9) \\ b(x) &= \sum_{Q=R,G,B} \sum_{j \in W} f'_Q(x+j) n_Q(x+j) & (5.10) \\ c(x) &= \sum_{Q=R,G,B} \sum_{j \in W} (n_Q(x+j))^2. & (5.11) \end{aligned}$$

The estimated disparity $\hat{d}_r$ is the value d which makes equation (5.8) minimum;

$$\hat{d}_{r[RGB]} = d_r + \frac{b(x)}{a(x)}. \tag{5.12}$$

We can see that the expected value of the estimate $\hat{d}_{r[RGB]}$ is the correct value d_r, but each individual estimate varies due to the noise. The variance of the estimate by using color images is given by

$$\begin{aligned} Var(\hat{d}_{r[RGB]}) &= Var\left(\frac{b(x)}{a(x)}\right) \\ &= \frac{2\sigma_n^2}{\left(\sum_{Q=R,G,B}\sum_{j\in W}(f'_Q(x+j))^2\right)^2} \sum_{j \in W} \left[\sum_{Q=R,G,B} (f'_Q(x+j))^2 \right. \\ &\quad +2\{\rho_{RG} f'_R(x+j) f'_G(x+j) + \rho_{GB} f'_G(x+j) f'_B(x+j) \\ &\quad \left. +\rho_{BR} f'_B(x+j) f'_R(x+j)\}\right], & (5.13) \end{aligned}$$

where ρ_{RG} is the correlation coefficient of $n_R(x)$ and $n_G(x)$ (same for ρ_{GB} and ρ_{BR}). Assuming no correlation among noise added to red, green, and blue images, i.e. $\rho_{RG} = \rho_{GB} = \rho_{BR} = 0$, we obtain

$$Var(\hat{d}_{r[RGB]}) = \frac{2\sigma_n^2}{\sum_{Q=R,G,B}\sum_{j\in W}(f'_Q(x+j))^2}. \tag{5.14}$$

On the other hand, we can easily find that the variance of the disparity estimate by using a single gray image is given by

$$Var(\hat{d}_{r[R]}) = \frac{2\sigma_n^2}{\sum_{j\in W}(f'_R(x+j))^2} \tag{5.15}$$

(same for $Var(\hat{d}_{r[G]})$ and $Var(\hat{d}_{r[B]})$).
From equations (5.14) and (5.15), the ratio of the variance of the estimate by using color images to that by using a single gray image, e.g. red, is

$$\frac{Var(\hat{d}_{r[RGB]})}{Var(\hat{d}_{r[R]})} = \frac{\sum_{j \in W}(f'_R(x+j))^2}{\sum_{Q=R,G,B}\sum_{j \in W}(f'_Q(x+j))^2}. \tag{5.16}$$

We can easily find that this ratio ranges between 0 and 1. Also from equations (5.14) and (5.15), we obtain the following relation:

$$\frac{1}{Var(\hat{d}_{r[RGB]})} = \frac{1}{Var(\hat{d}_{r[R]})} + \frac{1}{Var(\hat{d}_{r[G]})} + \frac{1}{Var(\hat{d}_{r[B]})}. \tag{5.17}$$

That is, the variance by using color images is smaller, i.e. the precision is higher, than that by using any single gray image. This is especially important when the original images contain many colors. In this case, which gives the minimum variance of the estimate among red, green, and blue varies depending on the position in the image. The estimate by using color images has always smaller variance than the smallest one in three colors at any position in the image.

Furthermore, when we use the intensity average of the red, green, and blue intensity values such that $f_{I1}(x) = (f_{R1}(x) + f_{G1}(x) + f_{B1}(x))/3$ and $f_{I2}(x) = (f_{R2}(x) + f_{G2}(x) + f_{B2}(x))/3$, then the variance of the estimate is given by

$$Var(\hat{d}_{r[I]}) = \frac{6\sigma_n^2}{\sum_{j \in W}(\sum_{Q=R,G,B} f'_Q(x+j))^2}. \tag{5.18}$$

From equations (5.14) and (5.18), the ratio of the variance is

$$\frac{Var(\hat{d}_{r[RGB]})}{Var(\hat{d}_{r[I]})} = \frac{\sum_{j \in W}(\sum_{Q=R,G,B} f'_Q(x+j))^2}{3\sum_{Q=R,G,B}\sum_{j \in W}(f'_Q(x+j))^2}. \tag{5.19}$$

This also ranges between 0 and 1, where 1 is the case if and only if $f_R(x) = f_G(x) = f_B(x)$. That is, the variance of the estimate by using color images is always smaller than or at worst equal to that by using the average of the three images.

In section 5.3.1, it is shown that accuracy in addition to precision can be improved experimentally by using synthesized color stereo images.

5.2.2 Color Stereo Matching Algorithm

In this section, we describe a color stereo matching algorithm which we propose for a medical application presented in the next section. This method is based on what we have presented in the previous chapters. That is, this method takes into account both the intensity and the disparity variations within the matching window and estimates the disparity at subpixel resolution using an iterative procedure. In addition, the method is extended so as to take advantage of color stereo images.

Let $f_{Q1}(\xi,\eta)$ and $f_{Q2}(\xi,\eta)$ $(Q = R,G,B)$ be the stereo images and suppose that the point whose disparity we would like to compute is placed at the origin $(0,0)$. The algorithm is as follows:

1. Compute initial disparity estimates d_0 at pixel resolution by finding the minimum of the summation of the SSD values for each color
$$e_{RGB}(d) = \sum_{Q=R,G,B} \sum_{i,j\in W} \left(f_{Q1}(\xi_i,\eta_j) - f_{Q2}(d+\xi_i,\eta_j)\right)^2, \tag{5.20}$$
where $\sum_{i,j\in W}$ means summation over the window.

2. Compute the parameters α_f and α_d which represent the local intensity and disparity variations within the window respectively.
$$\begin{aligned}\hat{\alpha_f} &= \frac{1}{3N_w}\sum_{Q=R,G,B}\sum_{i,j\in W}\left(\frac{\partial}{\partial\xi}f_{Q2}(\xi_i+d_0(0,0),\eta_j)\right)^2 & (5.21)\\ \hat{\alpha_d} &= \frac{1}{N_w}\sum_{i,j\in W}\frac{(d_0(\xi_i,\eta_j)-d_0(0,0))^2}{\sqrt{\xi_i^2+\eta_j^2}}, & (5.22)\end{aligned}$$
where N_w is the number of the points within the window.

3. Compute the correction of the disparity $\hat{\Delta d}$ and its uncertainty (variance) $\sigma^2_{\Delta d}$ such that
$$\begin{aligned}\hat{\Delta d} &= \frac{\sum_{Q=R,G,B}\sum_{i,j\in W}(\phi_{Q1}(\xi_i,\eta_j)\phi_{Q2}(\xi_i,\eta_j))}{\sum_{Q=R,G,B}\sum_{i,j\in W}(\phi_{Q2}(\xi_i,\eta_j))^2} & (5.23)\\ \sigma^2_{\Delta d} &= \frac{1}{\sum_{Q=R,G,B}\sum_{i,j\in W}(\phi_{Q2}(\xi_i,\eta_j))^2}, & (5.24)\end{aligned}$$
where
$$\begin{aligned}\phi_{Q1}(\xi,\eta) &= \frac{f_{Q1}(\xi,\eta)-f_{Q2}(\xi+d_0(0,0),\eta)}{\sqrt{2\sigma_n^2+\alpha_f\alpha_d\sqrt{\xi^2+\eta^2}}} & (5.25)\\ \phi_{Q2}(\xi,\eta) &= \frac{\frac{\partial}{\partial\xi}f_{Q2}(\xi+d_0(0,0),\eta)}{\sqrt{2\sigma_n^2+\alpha_f\alpha_d\sqrt{\xi^2+\eta^2}}}. & (5.26)\end{aligned}$$

4. Compute steps 2 and 3 for all pixels in the image.

5. Update the disparity estimates by $d_0 \leftarrow d_0 + \hat{\Delta d}$ and iterate steps 2 to 4 until the disparity estimates converge or up to a certain maximum number of iterations.

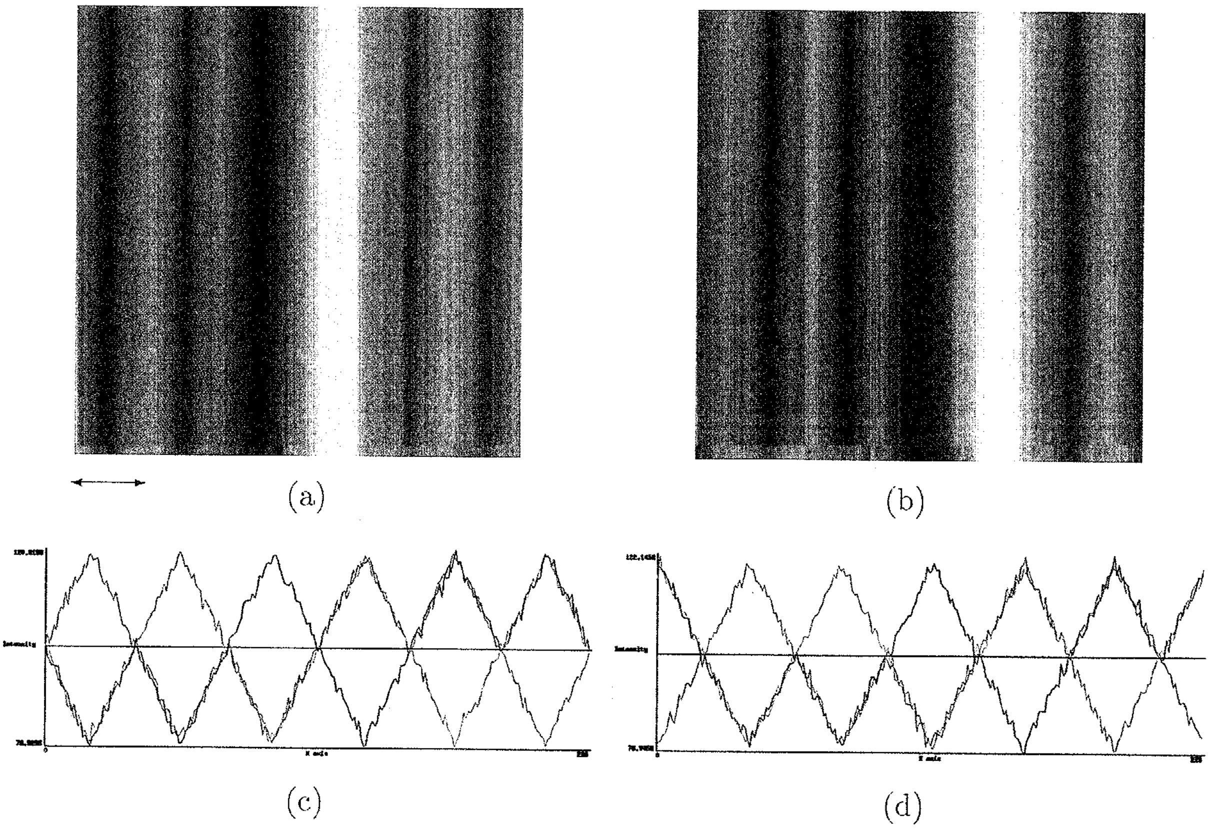


Figure 5.1: (a),(b) Synthesized color stereo images. ”↔” indicates where disparities are computed. (c),(d) Cross sections of scanlines in (a) and (b), respectively.

5.3 Experimental Results

5.3.1 Experiments Using Synthesized Images

In section 5.2.1, we have presented mathematically that ”precision” can be increased by using color images for stereo matching. In this section, we show that color stereo matching can improve not only ”precision” but also ”accuracy” in matching by using synthesized images. Figures 5.1 (a) and (b) show the synthesized color stereo images with Gaussian white noise added. The cross sections of scanlines in (a) and (b) are shown in (c) and (d), respectively. The actual disparity is constant (20.5 pixels) over the image. Disparities for the red part (indicated by ↔) in figure 5.1 (a) were computed. Figure 5.2 plots the histogram of the disparities computed by using (a) only the red images, (b) only the green images, (c) only the blue images, (d) the intensity images which are the averages of the three color images, and (e) the color images, i.e. all of the red, green, and blue images. First, we can see that

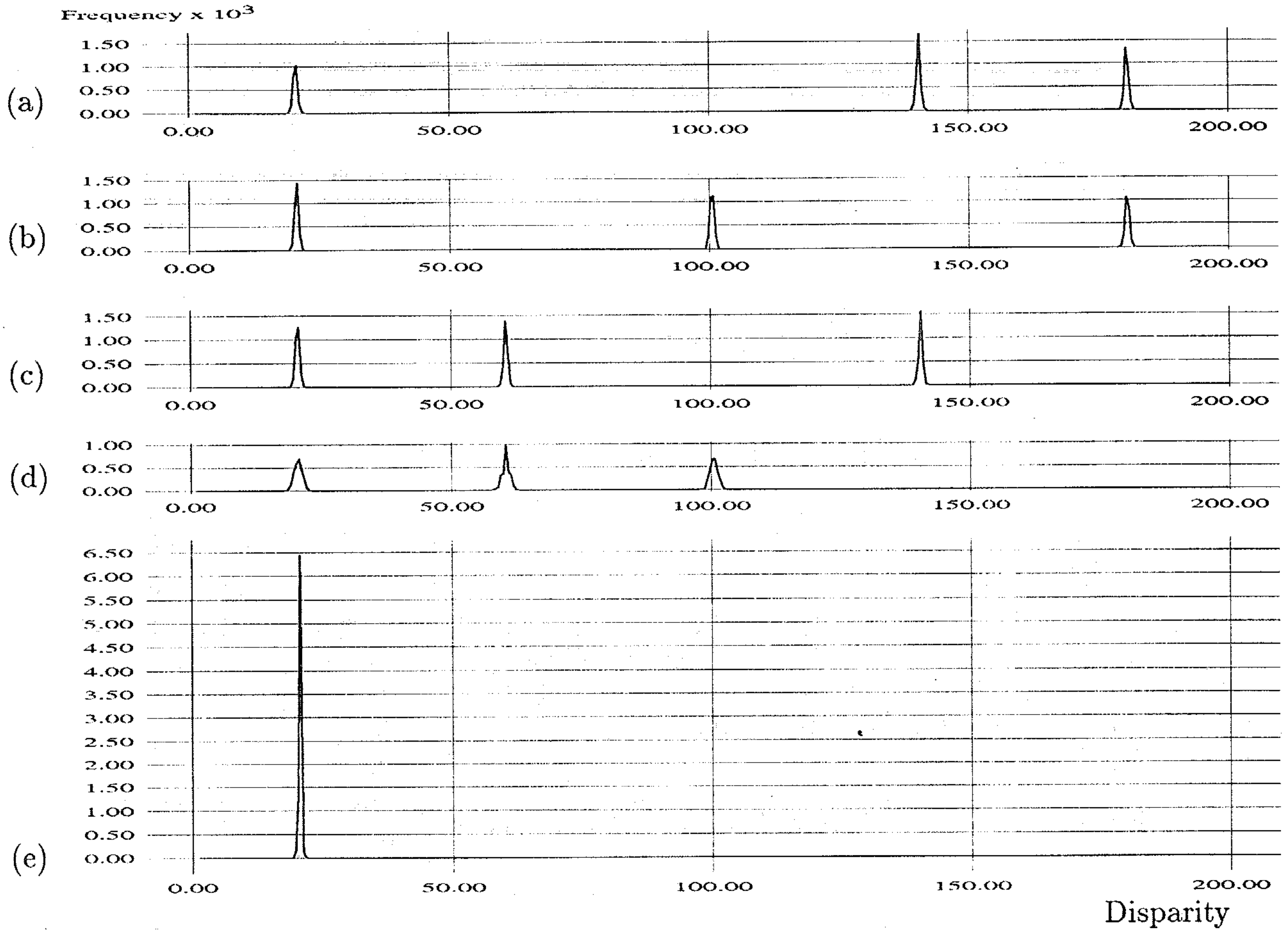


Figure 5.2: Histogram of the disparities computed by using (a) the red images; (b) the green images; (c) the blue images; (d) the intensity images which are the averages of the red, green, and blue images; (e) the color images.

	(a)	(b)	(c)	(d)	(e)
mean	20.48	20.54	20.48	20.52	20.50
variance	0.15	0.12	0.13	0.56	0.029
theoretical variance	0.080	0.080	0.080	0.24	0.027

Table 5.1: Variance at the correct disparity.

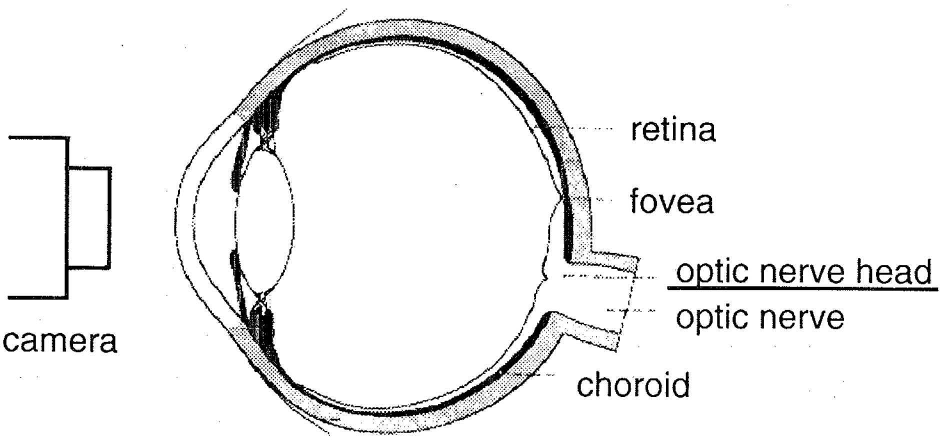


Figure 5.3: Section of a human eye

there are false matches, i.e. the matching is not accurate, in (a) to (d), while no false match in (e). Second, the peak at the correct disparity (20.5) in (e) is sharper, i.e. the matching is more precise, than any other ones in (a) to (d). Table 5.1 shows the mean and the variance of the peaks at the correct disparity and the theoretical variance obtained by equations (5.14), (5.15), and (5.18). These results may look obvious for such peculiar images as shown in figure 5.1. Still it shows the potential ability of the color stereo matching.

5.3.2 Experiments Using Ocular Fundus Images

This section presents experimental results with real stereo images of ocular funduses. Figure 5.3 shows a section of a human eye. An optic nerve head is where optic nerves and retinal vessels gather and go out from the eyeball. The 3-D shape of the optic nerve head gives very important information for diagnosing and monitoring glaucoma, an eye disease which commonly causes blindness to many people.

Figure 5.4 (a) shows the original stereo fundus images of a patient who suffers from glaucoma. The round bright parts are optic nerve heads. The images are preprocessed with a Laplacian of Gaussian (LOG) filter to reduce photometric distortion and noise. Figure 5.4 (b) shows the LOG-filtered pair of green images. The resultant disparity map computed by our color stereo matching algorithm is presented in figure 5.5 (a). Figure 5.5 (b) is the isometric plot of the disparity map. We can observe a deep hole at the optic nerve head, which is a typical phenomenon of glaucoma. Figure 5.5 (c) shows the isometric plot computed by using only green images for comparison. We can see large errors due to false matching in (c). Other regions look quite similar between (b) and (c). Since the stereo images were smoothed by a LOG filter and a relatively large window was used for matching in this case, the small variation of the computed disparity around the correct disparity is not noticeable,

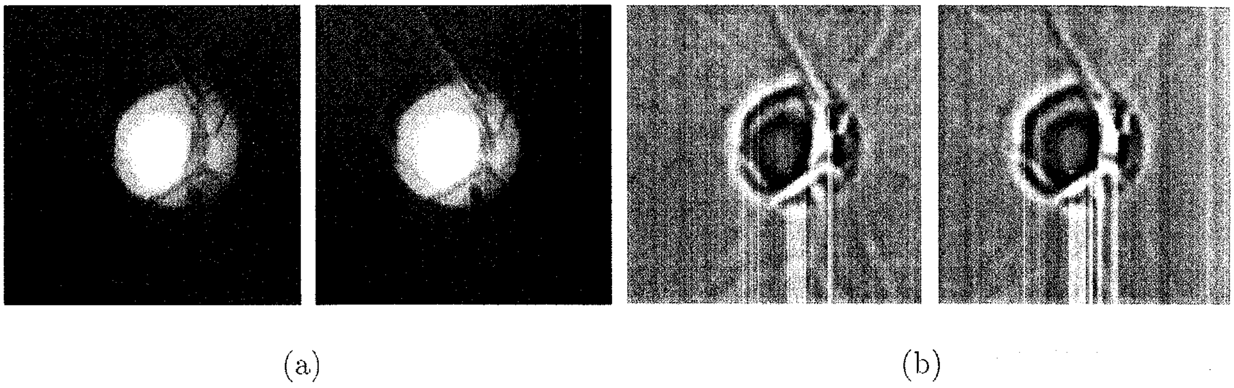


Figure 5.4: Stereo image pair of an optic nerve head suffering from glaucoma: (a) original images; (b) LOG-filtered images.

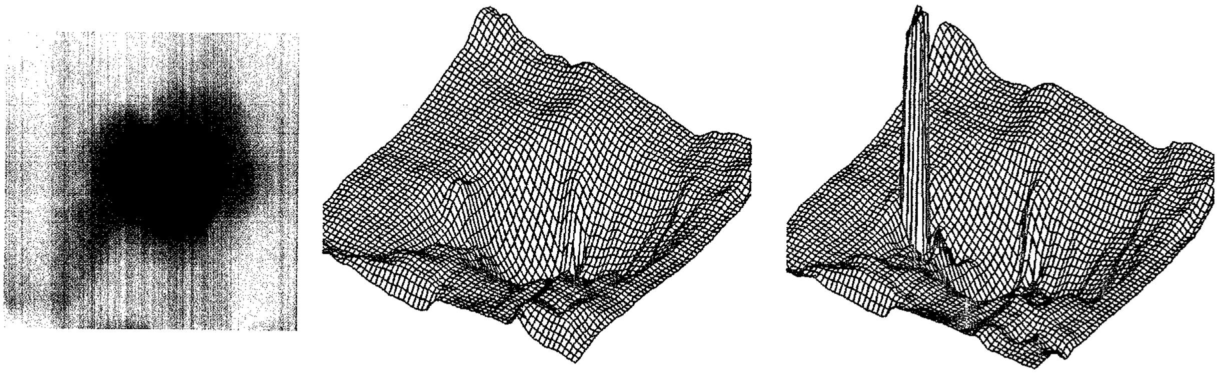
i.e. no obvious difference of "precision" is apparent between (b) and (c).

Next, we show the change of the shape of an optic nerve head. Figure 5.6 shows the stereo fundus images of a monkey, where (a) is the stereo image pair of the normal fundus and (b) of the same fundus but suffering from glaucoma. The isometric plots of the computed disparities for both stereo images are shown in figure 5.7. We can see a clear difference in shape between normal (a) and glaucoma (b). Figure 5.8 shows perspective views of the optic nerve heads produced by using the computed disparity maps and the original color images.

5.4 Conclusions

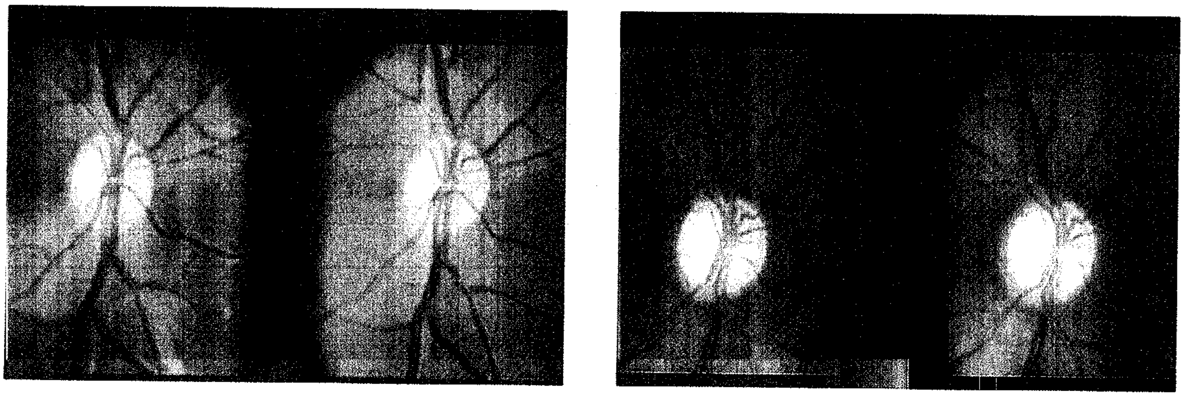
In this chapter, we have presented the effect of using color information in stereo matching and introduced a color stereo algorithm. In short, the degree of the effectiveness of using color depends entirely on the images used, as shown by equations (5.16) and (5.19). The improvement may not be as much as expected in most cases, and it is still an open question how relatively good in real scenes. However, our mathematical and experimental results showed that the results obtained by using color information are always better or at least equal to those by using gray images. So, if robustness is a more important requirement than the computational cost, color stereo matching should be worth applying.

Our algorithm was applied to the 3-D measurement of optic nerve heads, and the experimental results have demonstrated the effectiveness of the method for diagnosing and monitoring glaucoma. We are currently making more investigation so that slight changes which may be due to the progress of the disease or due to a treatment can be observed.



(a) (b) (c)

Figure 5.5: Computed disparity: (a) disparity map; (b) isometric plot of the disparity map; (c) isometric plot obtained by using only green images.



(a) (b)

Figure 5.6: Stereo fundus images of the same monkey: (a) normal fundus; (b) glaucomatous fundus.

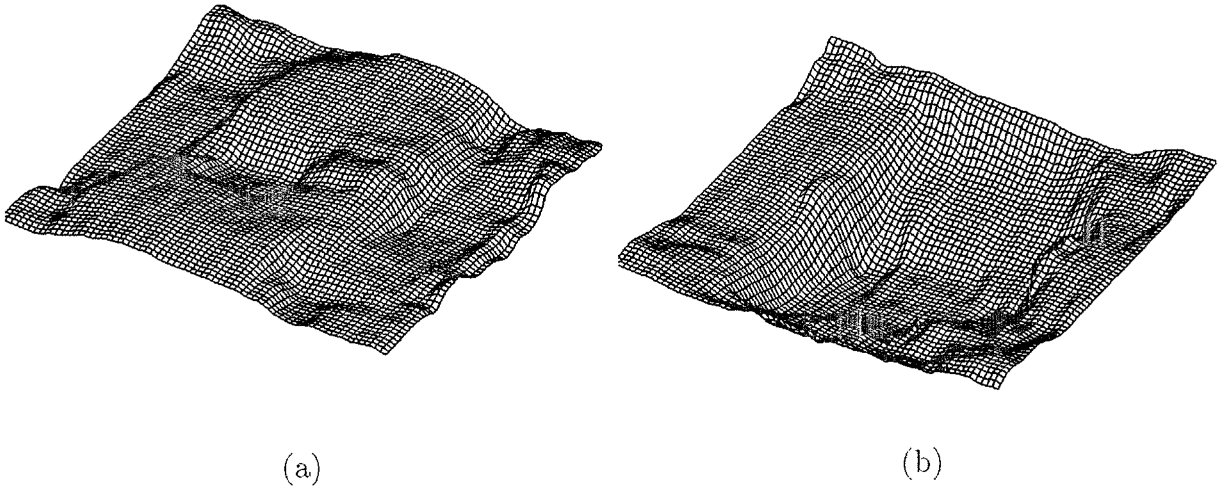


Figure 5.7: Isometric plots of the computed disparities at the optic nerve heads: (a) normal; (b) glaucoma.

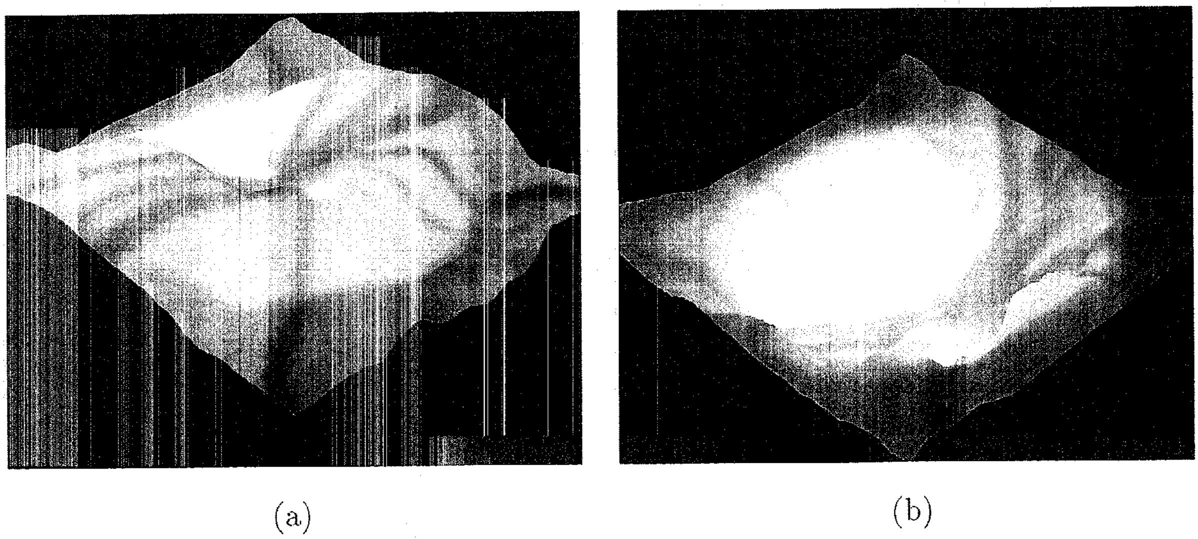


Figure 5.8: Perspective views of the optic nerve heads produced by the computed disparity maps and the original color images: (a) normal; (b) glaucoma.

Chapter 6

Conclusions

6.1 Thesis Summary

In this thesis, we have presented stereo vision based on physical and mathematical modeling. Our approach can be contrasted with that based on heuristics or that simulating the human vision system phenomenologically. In this approach, physical phenomena which create 2-D images from the 3-D world are modeled and stereo methods which, to the contrary, extract 3-D information from the images are described in the same mathematical formulation as shown in figure 1.3. This enables us to understand various aspects of stereo matching and leads to stereo algorithms which can adapt beforehand or automatically to the situation. The important point is that these algorithms are far more predictable and extensible for different situations than the algorithms based on heuristics.

The statistical formulation we have developed produces both an estimate of disparity and an uncertainty of the estimation. The uncertainty estimation enables us to analyze many properties of stereo matching relating to various factors such as intensity variation, disparity variation, noise, color, window size in matching, and stereo baseline. This anylysis then results in algorithms which solve problems in stereo vision.

We have developed a locally adaptive window. In general, an appropriate size of the matching window depends on intensity change, disparity change, and noise involved in an image. The adaptive window we have proposed can select the appropriate window size and shape (rectangle) automatically depending on those local properties at each position in an image in stereo matching. The key idea for the method is that it employs a statistical model of disparity distribution within the window. We assume that disparities have the same expected value, but their variation from that expected value increases with the distance from the center point of the window. This model has enabled us to correctly evaluate the influence of the disparity fluctuation within the window on the computation of disparity, so that the estimated uncertainty of the computed disparity is close to the real error of the computed disparity. As a result we can choose the window that provides the disparity estimate with minimum uncertainty. Both analytical and experimental results have demonstrated a clear advantage of the adaptive window over a fixed-size window.

We have also presented a new stereo matching method which uses multiple baseline stereo pairs. We have shown the trade-off problem relating to the baseline in stereo matching. That is, with a short baseline, the estimated distance is less precise due to narrow triangulation, while with a longer baseline, a larger disparity range must be searched to find a match and consequently matching is more difficult and there is a greater possibility of a false match. The proposed method can overcome the trade-off between precision and accuracy (avoidance of false matches) in stereo. The method is rather straightforward: we represent the SSD values for individual stereo pairs as a function of the inverse distance, and add those functions. The resulting function, the SSSD-in-inverse-distance, exhibits an unambiguous and sharper minimum at the correct matching position. Again, both analytical and experimental results have been provided to show the effectiveness of the proposed method in removing ambiguity and improving precision.

The effect of using color information in stereo matching has been also analyzed, and a color stereo algorithm has been proposed in the thesis. Although the degree of the effectiveness of using color depends on the images used, the mathematical and experimental results have shown that the disparity estimated by using color information is better or at least equal to that by using gray images. The color stereo algorithm has been used for a medical application: 3-D measurement of optic nerve heads using color stereo fundus images. The experimental results have demonstrated that the proposed stereo algorithm, together with various means of displaying the results, could give useful information for diagnosing and monitoring glaucoma.

6.2 Future Research

As described above, many aspects concerning stereo vision have been analyzed in this thesis. We also believe that the mathematical framework established in this thesis can be extended so that other issues in stereo vision can be dealt with in the same way. In the following, we briefly summerize several aspects untaken in the thesis and directions of future research.

Occlusion In the experiments with this thesis, we used relatively short baselines, meaning that occlusion does not cause a severe problem. In fact, these regions are so small that we can hardly recognize their existance. However, occluding edges often give important information in interpreting a scene. So detecting explicitly and making active use of these edges could be fruitful.

Specular reflection This also was not explicitly taken account of in this thesis, since, if it happens, it generally occurs in a very small portion in an image. Still, there is an ongoing investigation concerning these issues using the multiple-baseline stereo approach, in which we expect several causes of mismatches including occlusions and specular reflections can be classified by analyzing SSD curves with different baselines [KN91] [KON92].

Calibration Calibrating cameras will be important for practical use of a stereo vision technique. The necessity and method of calibration need to be discussed relating to the equipment to be used in aquiring images and their usage, e.g. using multiple cameras or a single camera shifting laterally.

Parallel implementation The methods presented in this thesis are relatively simple and local in the sense that the same computations are iterated independently for each pixel. So they are amenable to parallel hardware implementaion. For example, the implementation of the multiple-baseline algorithm has recently been done on MasPar, a Single Instruction Multiple Data (SIMD) machine with 4096 processors [KON92]. It should be noted that enormous speedup of the computational time not only gives just quantitative improvement, but also could lead to qualitative improvement, e.g. a more sophisticated window control strategy in the adaptive-window stereo method.

Other depth cues The human vision system can reach one consistent and stable interpretation of a scene using many depth (or shape) cues simultaneously such as disparity, shading, texture, motion, perspective, and defocus. Cooperation with the other various cues, or sometimes with other active sensors, is an attractive and challenging task. Uncertainty estimation could be a key in this work.

Bibliography

[AL87] N. Ayache and F. Lustman. Fast and reliable passive trinocular stereo vision. In *Proc. ICCV'87*, pages 422–426, 1987.

[Ana84] P. Anandan. Computing dense displacement fields with confidence measures in scenes containing occlusion. In *Proc. DARPA Image Understanding Workshop*, pages 236–246, 1984.

[Arn78] R. D. Arnold. Local context in matching edges for stereo vision. In *Proc. DARPA Image Understanding Workshop*, pages 65–72, 1978.

[Bar89] Stephen T. Barnard. Stochastic stereo matching over scale. *International Journal of Computer Vision*, pages 17–32, 1989.

[BB81] H. H. Baker and T. O. Binford. Depth from edge and intensity based stereo. In *Proc. IJCAI'81*, 1981.

[BB89] A. T. Brint and M. Brady. Stereo matching of curves by least deformation. In *Proc. IROS*, 1989.

[BBM87] R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to deterniming structure from motion. *International Journal of Computer Vision*, 1(1), 1987.

[BBW88] Paul J. Besl, Jeffrey B. Birch, and Layne T. Watson. Robust window operators. In *Proc. Int'l Conf. on Computer Vision*, pages 591–600, 1988.

[BF87] S.T. Barnard and M.A. Fischler. Stereo vision. In *Encyclopedia of Artificial Intelligence*, pages 1083–1090. Wiley, New York, 1987.

[BK88] K. L. Boyer and A. C. Kak. Structural stereopsis for 3-d vision. *IEEE Trans. PAMI*, 10(2), 1988.

[BN68] B.B.Mandelbrot and B.J.Van Ness. Fractional brownian motion, fractional noises and applications. *SIAM*, 10(4):422–438, 1968.

[Bou86] T. E. Boult. *Information Based Complexity in Non-Linear Equations and Compiuter Vision.* PhD thesis, Dept. of Computer Science, Columbia University, 1986.

[BT80] S. T. Barnard and W. B. Thompson. Disparity analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(4):333–340, July 1980.

[BZ86] A. Blake and A. Zisserman. Invariant surface reconstruction using weak continuity constraint. In *International Conference on Computer Vision and Pattern Recognition*, pages 62–68. IEEE, 1986.

[CM90] Jer-Sen Chen and Gerard Medioni. Parallel multiscale stereo matching using adaptive smoothing. In *ECCV90*, pages 99–103, 1990.

[dC86] Frederic de Coulon. *Signal Theory and Processing.* Artech House, Inc., 1986.

[DH77] I. J. Dowman and A. Haggag. Digital image correlation along epipolar lines. In *Proc. International Symposium on Image Processing, Interactions with Photogrammetry and Remote Sensing*, pages 47–49, 1977.

[DP86] M. Drumheller and T. Poggio. On parallel stereo. In *Proc. Int'l. Conf. Robotics and Automation*, pages 1439–1448, 1986.

[FP86] Wolfgang Forstner and Alfred Pertl. *Photogrammetric Standard Methods and Digital Image Matching Techniques for High Precision Surface Measurements*, pages 57–72. Elsevier Science Publishers B.V., 1986.

[Gen77] Donald B. Gennery. A stereo vision system for an autonomous vehicle. In *Proc. IJCAI*, 1977.

[Gen80] Donald B. Gennery. Object detection and measurement using stereo vision. In *Proc. ARPA Image Understanding Workshop*, pages 161–167, April 1980.

[Gop77] Wolfgang M. Gopfert. Digital cross-correlation of complex exponentiated inputs. In *Proc. International Symposium on Image Processing, Interactions with Photogrammetry and Remote Sensing*, pages 63–66, 1977.

[Gri85] W. E. L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(1):17–34, January 1985.

[Gru85] A. W. Gruen. Adaptive least squares correlation: A powerful image matching technique. *S. Afr. J. of Photogrammetry Remote Sensing and Cartography*, 14(3), 1985.

[HA86] William Hoff and Narendra Ahuja. Surfaces from stereo. In *Proc. ICPR'86*, 1986.

[HA89] William Hoff and Narendra Ahuja. Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection. *Trans. on PAMI*, 11(2), 1989.

[Han74] M. J. Hannah. *Computer Matching of Areas in Stereo Images.* PhD thesis, Computer Science Department, Stanford University, 1974. STAN-CS-80-805.

[Han89] M.J. Hannah. A system for digital stereo image matching. *Photogrammetric Engineering and Remote Sensing*, 55(12):1765–1770, Dec 1989.

[Hee89] Joachim Heel. Dynamic motion vision. In *Proceedings of the DARPA Image Understanding Workshop*, pages 702–713, Palo Alto, Ca, May 23-26 1989.

[HKK84] M. Herman, T. Kanade, and S. Kuroe. Incremental aquisition of a three-dimensional scene model from images. *IEEE Trans. PAMI*, PAMI-6(3):331–340, 1984.

[HLLW91] J. M. Hakkarainen, J. J. Little, H. Lee, and J. L. Wyatt. Interaction of algorithm and implementation for analog vlsi stereo vision. In *SPIE Visual Information Processing: From Neurons to Chips*, 1991.

[HMG79] R. L. Henderson, W. J. Miller, and C. B. Grosch. Automatic stereo reconstruction of man-made targets. *SPIE*, 186, 1979.

[IB91] J. R. Jordan III and A. C. Bovik. Using chromatic information in edge-based stereo correspondence. *CVGIP: Image Understanding*, 54(1):98–118, 1991.

[II86] M. Ito and A. Ishii. Three view stereo analysis. *IEEE Tran. on PAMI*, PAMI-8(4):534–531, 1986.

[JJT91] Michael R.M. Jenkin, Allan D. Jepson, and John K. Tsotsos. Techniques for disparity measurement. *CVGIP: Image Understanding*, 53(1), January 1991.

[KA85] Y. C. Kim and J. K. Aggarwal. Finding range from stereo images. In *CVPR*, 1985.

[Kan87] Takeo Kanade, editor. *Three-Dimensional Machine Vision.* Kluwer Academic Publishers, 1987.

[Kan91] Takeo Kanade. Computer vision as a physical science. In *CMU Computer Science: a 25th Anniversary Commemorative*, chapter 14, pages 345–369. ACM Press, 1991.

[KN91] T. Kanade and T. Nakahara. Experimental results of multibaseline stereo. In *IEEE Special Workshop on Passive Ranging*, October 1991. Princeton, NJ.

[KO91] Takeo Kanade and Masatoshi Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. In *Proc. of Int'l Conf. on Robotics and Automation*, April 1991. Also appeared in Proc. of DARPA Image Understanding Workshop, 1990.

[KON92] Takeo Kanade, Masatoshi Okutomi, and Tomoharu Nakahara. A multi-baseline stereo method. In *DARPA Image Understanding Workshop*, January 1992.

[LB91] Simon Lee and Michael Brady. Integrating stereo and photometric stereo to monitor the development of glaucoma. *Image and Vision Computing*, 9(1):39–44, 1991.

[LK81] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. IJCAI*, 1981.

[LOY73] Martin D. Levine, Douglas A. O'Handley, and Gary M. Yagi. Computer determination of depth maps. *Computer Graphics and Image Processing*, 2(4):131–150, 1973.

[Mar84] J. L. Marroquin. Surface reconstruction preserving discontinuities. Technical Report A. I. Memo 792, MIT, 1984.

[MK85] Victor J. Milenkovic and Takeo Kanade. Trinocular vision using photometric and edge orientation constraints. In *Proceedings of the Image Understanding Workshop*, pages 163–175, Miami Beach, Florida, December 1985.

[MKA73] Kenichi Mori, Masatsugu Kidode, and Haruo Asada. An iterative prediction and correction method for automatic sterocomparison. *Computer Graphics and Image Processing*, 2:393–401, 1973.

[MN85] G. Medioni and R. Nevatia. Segment-based stereo matching. *Computer Vision, Graphics, and Image Processing*, 31(1):2–18, 1985.

[MN89] R. Mohan and R. Nevatia. Using perceptual organization to extract 3-d structure. *Tran. on PAMI*, 11(11), 1989.

[MO89] Larry Matthies and Masatoshi Okutomi. A bayesian foundation for active stereo vision. In *SPIE, Sensor Fusion II: Human and Machine Strategies*, pages 62–74, November 1989.

[Mor79] Hans P. Moravec. Visual mapping by a robot rover. In *Proc. IJCAI'79*, pages 598–600, 1979.

[Mor81] Hans P. Moravec. Rover visual obstacle avoidance. In *Proc. IJCAI'81*, 1981.

[MP76] D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, October 1976.

[MP79] D. Marr and T. Poggio. A theory of human stereo vision. In *Proc. Roy. Soc. London*, volume vol. B 204, pages 301–328, 1979.

[MSK88] Larry Matthies, Richard Szeliski, and Takeo Kanade. Incremental estimation of dense depth maps from image sequenses. In *Proc. CVPR88*, pages 366–374, June 1988.

[MSK89] Larry Matthies, Richard Szeliski, and Takeo Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3:209–236, 1989.

[Nis87] H. K. Nishihara. Practical real-time imaging stereo matcher. In *Redings in Computer Vision*. Morgan Kaufmann Publishers, Inc., 1987.

[OC89] G. P. Otto and T. K. W. Chau. 'region-growing' algorithm for matching of terrain images. *Image and Vision Computing*, 7(2), 1989.

[OK85] Yuichi Ohta and Takeo Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(2):139–154, March 1985.

[OK91] Masatoshi Okutomi and Takeo Kanade. A multiple-baseline stereo. In *Proc. of Computer Vision and Pattern Recognition*, June 1991.

[OK92] Masatoshi Okutomi and Takeo Kanade. A locally adaptive window for signal matching. *International Journal of Computer Vision*, 7(2):143–162, 1992. Also appeared in Proc. of Int'l Conf. on Computer Vision, 1990.

[OWI86] Yuichi Ohta, Masaki Watanabe, and Katsuo Ikeda. Improving depth map by right-angled trinocular stereo. In *Proc. ICPR*, pages 519–521, 1986.

[OYT92] Masatoshi Okutomi, Osamu Yoshizaki, and Goji Tomita. Color stereo matching and its application to 3-d measurement of optic nerve head. In *Proc. of Int'l Conf. on Pattern Recognition*, August 1992.

[Pan78] D. J. Panton. A flexible approach to digital stereo mapping. *Photogram. Eng. Remote Sensing*, 44(12):1499–1512, Dec 1978.

[PH86] Matti Pietikainen and David Harwood. Depth from three camera stereo. In *Proc. CVPR*, 1986.

[PMF85] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby. Pmf: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449–470, 1985.

[Pog88] T. Poggio. The mit vision machine. In *Proc. Image Understanding Workshop*, 1988.

[Pra85] K. Prazdny. Detection of binocular disparities. *Biol. Cybern.*, 52:93–99, 1985.

[Qua84] L. H. Quam. Hierarchical warp stereo. In *Proc. Image Understanding Workshop*, 1984.

[RGH80] T. W. Ryan, R. T. Gray, and B. R. Hunt. Prediction of correlation errors in stereo-pair images. *Optical Engineering*, 19(3):312–322, May 1980.

[San89] Jorge L. C. Sanz, editor. *Advances in Machine Vision.* Springer-Verlag, 1989.

[SB76] Frank A. Scarano and Gerald A. Brumm. A digital elevation data collection system. *Photogrammetric Engineering and Remote Sensing*, 42(4):489–496, 1976.

[Shi87] Yoshiaki Shirai. *Three-Dimensional Computer Vision.* Springer-Verlag, 1987.

[Ter86] Demetri Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE transaction on pattern analysis and machine inteligence*, 8(4):413–424, July 1986.

[Tsa83] Roger Y. Tsai. Multiframe image point matching and 3-d surface reconstraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(No.2), March 1983.

[Vos87] Richard F. Voss. Fractals in nature. In *Course note on FRACTALS: Introduction, Basics, and Perspectives*, 1987.

[Wec90] Harry Wechsler. *Computational Vision.* Academic Press, 1990.

[Woo83] G.A. Wood. Realities of automatic correlation problem. *Photogrammetric Engineering and Remote Sensing*, 49:537–538, April 1983.

[WTK87] Andrew Witkin, Demetri Terzopoulos, and Michael Kass. Signal matching through scale space. *International Journal of Computer Vision*, pages 133–144, 1987.

[XTA85] G. Xu, S. Tsuji, and M. Asada. Coarse-to-fine control strategy for matching motion stereo pairs. In *Proc. IJCAI*, pages 892–894, 1985.

[Yam88] Masanobu Yamamoto. The image sequence analysis of three-dimensional dynamic scenes. Technical Report 893, Electrotechnical Laboratory - Agency of Industrial Science and Technology, Tsukuba, Ibaraki, Japan, May 1988.

[YH92] Kazuhiro Yoshida and Shigeo Hirose. Real-time stereo vision with multiple arrayed camera. In *Proc. Int'l Conf. on Robotics and Automation*, pages 1765–1770, 1992.

[YKK86] M. Yachida, Y. Kitamura, and M. Kimachi. Trinocular vision: New approach for correspondence problem. In *Proc. ICPR*, pages 1041–1044, 1986.

Appendix A

Approximating the Distribution of $e(\xi)$

We will examine the statistical properties of $e(\xi)$, i.e. equation (2.27)

$$e(\xi) = (d_r(\xi) - d_r(0)) f_2'(\xi + d_r(0)) \tag{A.1}$$

We see that $e(\xi)$ is product of u and v where

$$\begin{aligned} u &= d_r(\xi) - d_r(0), \\ v &= f_2'(\xi + d_r(0)), \end{aligned} \tag{A.2}$$

Our assumptions are: u is zero-mean Gaussian noise; v is zero-mean Gaussian white noise; and u and v are statistically independent.

Let $p_u(u)$, σ_u^2, and $R_u(\tau)$ denote the density function, variance, and autocorrelation function of u, respectively.

$$p_u(u) = \frac{1}{\sqrt{2\pi}\sigma_u} e^{-\frac{u^2}{2\sigma_u^2}}$$

We define notations for v in the same manner. Also since v is white, we have

$$R_v(\tau) = a\delta(\tau),$$

where $\delta(\tau)$ is the delta function and a is a constant.

Since u and v are independent, the autocorrelation function of z is given by (see [dC86]):

$$R_z(\tau) = R_u(\tau) R_v(\tau) = b\delta(\tau)$$

where b is a constant. Therefore, z is also white.

The density function $p_z(z)$ can be calculated as

$$\begin{aligned} p_z(z) &= \int_{-\infty}^{\infty} \frac{1}{|u|} p_u(u) p_v\left(\frac{z}{u}\right) du \\ &= \frac{1}{\pi \sigma_u \sigma_v} \int_0^{\infty} \frac{1}{u} \exp\left(-\frac{u^2}{2\sigma_u^2} - \frac{z^2}{2\sigma_v^2 u^2}\right) du \\ &= \frac{1}{\pi \sigma_u \sigma_v} K_0\left(\frac{|z|}{\sigma_u \sigma_v}\right), \end{aligned}$$

where $K_0(z)$ is the modified Bessel function of order 0

$$K_0(z) = \frac{1}{2} \int_0^{\infty} u^{-1} \exp\left(-u - \frac{z^2}{4u}\right) du.$$

The thick curve in Figure A.1 shows this density function. $p_z(z)$ is a monomodal distribution which is symmetrical about the mode at $z = 0$. For simplicity, it is reasonable to approximate the distribution by a Gaussian distribution that has the same mean and variance as $p_z(z)$, which are

$$\begin{aligned} E[z] &= E[u]E[v] \\ &= 0 \\ E[(z - E[z])^2] &= E[(uv)^2] = E[u^2]E[v^2] \\ &= \sigma_u^2 \sigma_v^2 \end{aligned}$$

The faint curve in figure A.1 shows the zero-mean Gaussian distribution $N(0, \sigma_u^2 \sigma_v^2)$. Hence, equation (2.29).

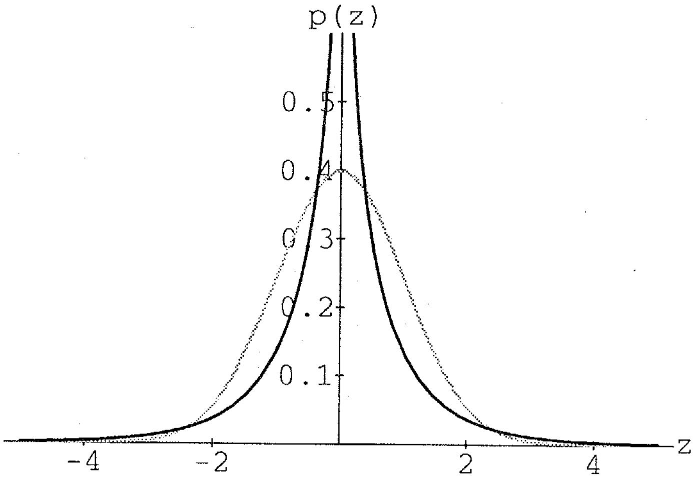


Figure A.1: Probability density functions, $\frac{1}{\pi\sigma_u\sigma_v}K_0\left(\frac{|z|}{\sigma_u\sigma_v}\right)$ and $N(0,\sigma_u^2\sigma_v^2)$. The horizontal axis is normalized; i.e., $z' = \frac{z}{\sigma_u\sigma_v}$

Appendix B

SSSD-in-inverse-distance for Ambiguous Pattern

Proposition: Suppose that there are two and only two repetitions of the same pattern around positions x and $x+a$ where $a \neq 0$ is a constant. That is, for $j \in W$

$$f(x+j) = f(\xi + j), \qquad \text{if and only if } \xi = x \text{ or } \xi = x + a. \tag{B.1}$$

Then, if $B_1 \neq B_2$, for $\forall\zeta$, $\zeta \neq \zeta_r$,

$$\begin{aligned} E[e_{\zeta(12)}(x,\zeta)] &= \sum_{j \in W} (f(x+j) - f(x + B_1 F(\zeta - \zeta_r) + j))^2 \\ &+ \sum_{j \in W} (f(x+j) - f(x + B_2 F(\zeta - \zeta_r) + j))^2 + 4N_w \sigma_n^2 \\ &> 4N_w \sigma_n^2 \ = \ E[e_{\zeta(12)}(x, \zeta_r)]. \end{aligned} \tag{B.2}$$

Proof: Tentatively suppose that for $\exists\zeta_f$, $\zeta_f \neq \zeta_r$,

$$\sum_{j \in W} (f(x+j) - f(x+B_1 F(\zeta_f - \zeta_r)+j))^2 + \sum_{j \in W} (f(x+j) - f(x+B_2 F(\zeta_f - \zeta_r)+j))^2 = 0. \tag{B.3}$$

Then, it must be the case that

$$\begin{aligned} f(x+j) &= f(x+a_1+j) \\ \text{and} \quad f(x+j) &= f(x+a_2+j), \end{aligned} \tag{B.4}$$

for $j \in W$, where

$$\begin{aligned} a_1 &= B_1 F(\zeta_f - \zeta_r) \\ a_2 &= B_2 F(\zeta_f - \zeta_r). \end{aligned}$$

Since $B_1 \neq B_2$ and $\zeta_r \neq \zeta_f$,

$$a_1 \neq a_2. \tag{B.5}$$

So, we have

$$f(x+j) = f(\xi + j), \qquad \text{for } \xi = x,\ x + a_1, \text{ or } x + a_2. \tag{B.6}$$

Since this contradicts assumption (B.1), equation (B.3) does not hold. Its left hand side must be positive. Hence (B.2) holds.

Appendix C

Multiple-Baseline Stereo Algorithm

We present a complete description of the stereo algorithm using multiple-baseline stereo pairs. The task is, given n stereo pairs, find the ζ that minimizes the SSSD-in-inverse-distance function,

$$SSSD(x, \zeta) = \sum_{i=1}^{n} \sum_{j \in W} (f_0(x + j) - f_i(x + B_i F \zeta + j))^2. \tag{C.1}$$

We will perform this task in two steps: one at pixel resolution by minimum detection and the other at sub-pixel resolution by iterative estimation.

C.1 Minimum of SSSD at Pixel Resolution

For convenience, instead of using the inverse distance, we normalize the disparity values of individual stereo pairs with different baselines to the corresponding values for the largest baseline. Suppose $B_1 < B_2 < \cdots < B_n$. We define the baseline ratio R_i such that

$$R_i = \frac{B_i}{B_n}. \tag{C.2}$$

Then,

$$B_i F \zeta = R_i B_n F \zeta = R_i d_{(n)}, \tag{C.3}$$

where $d_{(n)}$ is the disparity for the stereo pair with baseline B_n. Substituting this into equation (C.1),

$$SSSD(x, d_{(n)}) = \sum_{i=1}^{n} \sum_{j \in W} (f_0(x + j) - f_i(x + R_i d_{(n)} + j))^2. \tag{C.4}$$

We compute the SSSD function for a range of disparity values at the pixel resolution and identify the disparity that gives the minimum. Note that pixel resolution for the image pair with the longest baseline (B_n) requires calculation of SSD values at sub-pixel resolution for other shorter baseline stereo pairs.

C.2 Iterative Estimation at Sub-pixel Resolution

Once we obtain the disparity at pixel resolution for the longest baseline stereo, we improve the disparity estimate to sub-pixel resolution by an iterative algorithm presented in section 2.2. For this iterative estimation, we use only the image pair $f_0(x)$ and $f_n(x)$ with the longest baseline. This is because of a few reasons. First, since the pixel-level estimate was obtained by using the SSSD-in-inverse-distance, the ambiguity has been eliminated and only improvement of precision is intended at this stage. Second, using only the longest-baseline image pair reduces the computational requirement for the SSD calculation by a factor of n, and yet does not degrade precision too significantly.

In the experiments shown in section 3.3, we used the following algorithm for sub-pixel estimation: Let $d_{0(n)}$ be the initial disparity estimate obtained at pixel resolution. Then, a more precise estimate is computed by calculating the following two quantities:

$$\Delta\hat{d}_{(n)} = \frac{\sum_{j\in W}(f_0(x+j) - f_n(x+d_{0(n)}+j))f'_n(x+d_{0(n)}+j)}{\sum_{j\in W}(f'_n(x+d_{0(n)}+j))^2} \tag{C.5}$$

$$\sigma^2_{\Delta d_{(n)}} = \frac{2\sigma_n^2}{\sum_{j\in W}(f'_n(x+d_{0(n)}+j))^2}, \tag{C.6}$$

The value $\Delta\hat{d}_{(n)}$ is the estimate of the correction of the disparity to further minimize the SSD, and $\sigma^2_{\Delta d_{(n)}}$ is its variance. We iterate this procedure by replacing $d_{0(n)}$ by

$$d_{0(n)} \leftarrow d_{0(n)} + \Delta\hat{d}_{(n)} \tag{C.7}$$

until the estimate converges or up to a certain maximum number of iterations.

Appendix D

Assuming $d_r(\xi, \eta)$ to Be Fractal

Here we assume $d_r(\xi, \eta)$ to be fractal, then instead of equation (4.4), we have

$$d_r(\xi, \eta) - d_r(0,0) \quad \sim \quad N\left(0, \alpha_d(\xi^2 + \eta^2)^H\right), \tag{D.1}$$

where the parameter H has a value $0 < H < 1$. When $H = \frac{1}{2}$, this equation represents the case that $d_r(\xi, \eta)$ is Brownian motion.
Then, instead of the final equations (4.22) and (4.23), we get

$$\begin{aligned} \hat{\Delta d} &= \frac{\sum_{i,j \in W} \frac{(f_1(\xi_i,\eta_j) - f_2(\xi_i + d_0(0,0),\eta_j))\frac{\partial}{\partial \xi} f_2(\xi_i + d_0(0,0),\eta_j)}{2\sigma_n^2 + \alpha_f \alpha_d(\xi_i^2 + \eta_j^2)^H}}{\sum_{i,j \in W} \frac{(\frac{\partial}{\partial \xi} f_2(\xi_i + d_0(0,0),\eta_j))^2}{2\sigma_n^2 + \alpha_f \alpha_d(\xi_i^2 + \eta_j^2)^H}} \\ \sigma_{\Delta d}^2 &= \frac{1}{\sum_{i,j \in W} \frac{(\frac{\partial}{\partial \xi} f_2(\xi_i + d_0(0,0),\eta_j))^2}{2\sigma_n^2 + \alpha_f \alpha_d(\xi_i^2 + \eta_j^2)^H}}. \end{aligned}$$

Furthermore, instead of equation (4.24), we obtain:

$$\hat{\alpha_d} = \frac{1}{N_w} \sum_{i,j \in W} \frac{(d_0(\xi_i, \eta_j) - d_0(0,0))^2}{(\xi_i^2 + \eta_j^2)^H}$$

Appendix E

Derivation of Equations in Section 4.3

We will show the derivation of equations (4.19) to (4.21). Substituting equations (4.16) and (4.17) into equation (4.18),

$$
\begin{aligned}
& p(\Delta d|\varphi_{ij}(i,j \in W)) \\
= & \frac{\prod_{i,j\in W} \frac{1}{\sqrt{2\pi}\sigma_s(\xi_i,\eta_j)} \exp\left(-\frac{(\phi_1(\xi_i,\eta_j)-\Delta d\phi_2(\xi_i,\eta_j))^2}{2\sigma_s^2(\xi_i,\eta_j)}\right)}{\int_{-\infty}^{\infty} \prod_{i,j\in W} \frac{1}{\sqrt{2\pi}\sigma_s(\xi_i,\eta_j)} \exp\left(-\frac{(\phi_1(\xi_i,\eta_j)-\Delta d\phi_2(\xi_i,\eta_j))^2}{2\sigma_s^2(\xi_i,\eta_j)}\right) d(\Delta d)} \\
= & \frac{\exp\left(\sum_{i,j\in W}\left(-\frac{(\phi_1(\xi_i,\eta_j)-\Delta d\phi_2(\xi_i,\eta_j))^2}{2\sigma_s^2(\xi_i,\eta_j)}\right)\right)}{\int_{-\infty}^{\infty} \exp\left(\sum_{i,j\in W}\left(-\frac{(\phi_1(\xi_i,\eta_j)-\Delta d\phi_2(\xi_i,\eta_j))^2}{2\sigma_s^2(\xi_i,\eta_j)}\right)\right) d(\Delta d)} \\
= & \frac{\exp\left(-\frac{\sum_{i,j\in W}\phi_2(\xi_i,\eta_j)^2}{2\sigma_s^2(\xi_i,\eta_j)}\left(\Delta d-\frac{\sum_{i,j\in W}(\phi_1(\xi_i,\eta_j)\phi_2(\xi_i,\eta_j)/\sigma_s^2(\xi_i,\eta_j))}{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2}\right)^2\right)}{\int_{-\infty}^{\infty} \exp\left(-\frac{\sum_{i,j\in W}\phi_2(\xi_i,\eta_j)^2}{2\sigma_s^2(\xi_i,\eta_j)}\left(\Delta d-\frac{\sum_{i,j\in W}(\phi_1(\xi_i,\eta_j)\phi_2(\xi_i,\eta_j)/\sigma_s^2(\xi_i,\eta_j))}{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2}\right)^2\right) d(\Delta d)} \\
= & \sqrt{\frac{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2}{2\pi}} \exp\left(-\frac{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2}{2}\right. \\
& \left.\left(\Delta d-\frac{\sum_{i,j\in W}(\phi_1(\xi_i,\eta_j)\phi_2(\xi_i,\eta_j)/\sigma_s^2(\xi_i,\eta_j))}{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2}\right)^2\right),
\end{aligned}
$$

where $\sum_{i,j\in W}$ denotes the summation over the window. From this equation, we can see that $p(\Delta d|\varphi_{ij}(i,j \in W))$ becomes a Gaussian probability density function. Its mean value $\hat{\Delta d}$ and variance $\sigma_{\Delta d}^2$ are

$$
\hat{\Delta d} = \frac{\sum_{i,j\in W}(\phi_1(\xi_i,\eta_j)\phi_2(\xi_i,\eta_j)/\sigma_s^2(\xi_i,\eta_j))}{\sum_{i,j\in W}(\phi_2(\xi_i,\eta_j)/\sigma_s(\xi_i,\eta_j))^2}
$$

$$\sigma^2_{\Delta d} = \frac{1}{\sum_{i,j \in W} (\phi_2(\xi_i, \eta_j)/\sigma_s(\xi_i, \eta_j))^2}.$$

Hence, equation (4.19) to (4.21).