A SUBPIXEL RESOLUTION HIERARCHICAL STEREO VISION SYSTEM

A Master's Thesis
Presented by
Uğur Murat Leloğlu

to
the Graduate School of Natural and Applied Sciences
of Middle East Technical University
in Partial Fulfillment for the Degree of

MASTER OF SCIENCE

in

ELECTRICAL AND ELECTRONICS ENGINEERING

MIDDLE EAST TECHNICAL UNIVERSITY
ANKARA
January, 1995

Aproval of the Graduate School of Natural and Applied Sciences.

Prof. Dr. İsmail TOSUN
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Kemal İNAN
Chairman of the Department

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science in Electrical and Electronics Engineering.

Prof. Dr. Mete SEVERCAN
Supervisor

Examining Committee in Charge:

Prof. Dr. Mübeccel Demirekler(Chairman)
Prof. Dr. Mete SEVERCAN
Prof. Dr. Ziya İder
Assoc. Prof. Dr. Uğur Halıcı
Dr. Gürhan Şaplakoğlu (TAEAGE)

ABSTRACT

A SUBPIXEL RESOLUTION HIERARCHICAL STEREO VISION SYSTEM

Uğur Murat Leloğlu
M. S. in Electrical and Electronics Engineering
Supervisor: Prof. Dr. Mete SEVERCAN
January, 1995, 73 pages.

A multi-resolution stereo vision system, which uses normalized cross-correlations and phases of band-pass filtered images as matching primitives, is described in this thesis. The coarse-to-fine strategy adopted leads to more accurate results in shorter time. At each level of the hierarchy, two modules, namely, a pixel matching module and a thin plate module, are employed. The former module determines the goodness of a possible match by using the correlation value and a neighborhood support function. Only unambiguous matches are accepted as true and these matches constrain neighbors through uniqueness and orderness constraints. As a result, ambiguity is resolved for some of the previously ambiguous matches. This spreading of constraints is iterated several times. Accepted matches are supplied to the thin plate module which attempts a subpixel resolution surface reconstruction while detecting depth discontinuities and occluded regions. The phases of band-pass filtered images are used as subpixel matching primitives and intensity edges are used for detection of depth discontinuities. Besides, small unmatched areas are interpolated by this module. This is not a blind interpolation since matching of phases guide interpolation. The stereo vision system presented is amenable to parallel implementation but is still reasonably fast in serial simulation. The robustness of the system is demonstrated on images from different domains.

Keywords: Stereo Correspondence, Surface Reconstruction, Phase Correlation, Multiresolution Image Analysis, Surface Interpolation, Relaxation, Band-Pass Filtering.

Science Code : 609.02.08

ÖZ

GÖZEARASI ÇÖZÜNÜRLÜKTE HİYERARŞİK BİR STEREO GÖRÜŞ DİZGESİ

Uğur Murat Leloğlu
Yüksek Lisans Tezi, Elektrik ve Elektronik Mühendisliği Anabilim Dalı
Tez Yöneticisi: Prof. Dr. Mete SEVERCAN
Ocak, 1995, 73 sayfa.

Bu tezde, eşlem öncülleri olarak düzgülenmiş çapraz ilintileri ve bant-geçiren süzülmüş görüntülerin fazlarını kullanan çok-çözünürlüklü bir stereo çözümleme dizgesi sunulmaktadır. Seçilen kabadan-inceye stratejisi daha kısa sürede daha doğru sonuç almayı sağlamaktadır. Hiyerarşinin her bir düzeyinde iki bölütten, göze-eşlem bölütü ve ince levha bölütünden yararlanılmaktadır. İlk bölüt olası bir eşlemin iyiliğini ilinti değeri ve bir yerel destek işlevi kullanarak saptar. Sadece kesin eşlemler doğru olarak kabul edilir ve bu eşlemler komşularını teklik ve sıralılık koşulları ile sınırlar. Sonuçta, daha önce muallakta olan bazı eşlemlerin belirsizliği çözülür. Bu şekilde koşulların yayılması birkaç kez tekrarlanır. Kabul edilen eşlemler, derinlik süreksizliklerini ve örtük alanları bulurken gözearası çözünürlükte yeniden yüzey kurmaya çalışan ince levha bölütüne verilir. Band-geçiren süzülmüş görüntülerin fazları gözearası eşlem öncülleri olarak kullanılırken, yoğunluk kenarları derinlik süreksizliklerini bulmak için kullanılır. Bunun yanı sıra, eşlemlenmemiş küçük alanlar da bu bölüt tarafından aradeğerlenir. Bu körlemesine bir aradeğerleme değildir çünkü fazların eşlemlenmesi aradeğerlemeye yol göstermektedir. Sunulan stereo aradeğerleme dizgesi koşut uygulamaya uygun olmakla birlikte seri benzeşimde de yeterince hızlıdır. Dizgenin dayanımlılığı çeşitli alanlardan görüntüler üzerinde gösterilmiştir.

Anahtar Sözcükler: Stereo Karşılıklılık, Yeniden Yüzey Kurma, Faz İlintisi, Çok-çözünürlüklü Görüntü Çözümleme, Yüzey Aradeğerleme, Gevşeme, Bant-Geçiren Süzme.

Bilim Dalı Sayısal Kodu : 609.02.08

ACKNOWLEDGEMENTS

First, I would like to express my gratitude to Prof. Dr. Mete Severcan for his patience, guidance and suggestions.

I thank my family for their encouragement and support. Special thanks are due to Zeynep Kurç for her care and support in my hard times. I thank Bora Nakiboğlu, Burak Tüzün and Atila Koç who contributed to this thesis in a paradoxical way. I also thank all members of laboratories at our storey, including those who are now at a distance either physically or socially.

Finally, I would like to acknowledge TÜBİTAK Ankara Electronics Research and Development Institute for continued support.

TABLE OF CONTENTS

ABSTRACT iii

ÖZ v

ACKNOWLEDGEMENTS vii

LIST OF TABLES xi

LIST OF FIGURES xii

CHAPTER I :INTRODUCTION 1

1.1 The Aim of the Thesis 2

1.2 The Organization of the Thesis 2

CHAPTER II :BACKGROUND 3

2.1 The Stereo Process 3

2.1.1 Image Acquisition 4

2.1.2 Determining the Geometry 6

2.1.3 Matching 6

2.1.4 Calculating the Actual Depth 6

2.2 Psychophysics and Neurophysiology of Stereopsis 7

2.2.1 The speed of human stereopsis 7

2.2.2 Stereopsis is a Low-Level Process 7

2.2.3 Limited Fusional Area 7

2.2.4 Effect of Contrast 8

2.2.5 Hyperacuity 8

2.2.6 Gradient Limit 8

2.2.7 Binocular Cells in Visual Cortex 9

2.2.8 Ocular Dominance 9

2.2.9 Orientation Selectivity 9

2.2.10 Frequency Selectivity 9

2.2.11 Receptive Fields Types 10

2.2.12 Binocular Sensitivity Types 10

2.2.13 Modelling Simple Cells 10

2.2.14 Coarse-to-Fine Structure 11

2.3 The Matching Primitives 11

2.3.1 Feature-Based Matching Primitives 11

2.3.2 Area-Based Matching Primitives 14

2.4 Problems in Matching 20

2.4.1 Photometric Variation 20

2.4.2 Lack of Texture 21

2.4.3 Repetitive Texture 21

2.4.4 Occlusion 21

2.5 Representing Disparity 21

2.6 Constraining the Solution 21

2.6.1 Smoothness 22

2.6.2 Opaqueness 23

2.6.3 Orderedness 24

2.6.4 Uniqueness 24

2.6.5 Compatibility 24

2.6.6 Epipolarity 25

2.6.7 Limited Disparity Range 25

2.7 Strategies 25

2.7.1 Multi-Channel Analysis 25

2.7.2 Relaxation 26

2.7.3 Dynamic Programming 30

2.7.4 Hybrid Approaches 31

2.7.5 Using Other Sources of Information 32

CHAPTER III :THE STEREO SYSTEM 34

3.1 An Overview of the System 34

3.2 Forming the Gaussian Image Pyramid 35

3.3 Pixel Matching 36

3.3.1 Calculation of Support 39

3.4 Edge Detection 40

3.5 Band-Pass Filtering 41

3.6 Thin Plate Module 44

3.6.1 The Phase Term 46

3.6.2 Interpolation 47

3.7 Detection of Occlusions 47

3.8 Increasing the Resolution of Disparity 47

3.9 Experimental Results 48

3.9.1 Fruits Stereo Pair 48

3.9.2 Pentagon Stereo Pair 48

3.9.3 Brutus Stereo Pair 49

CHAPTER IV :CONCLUSIONS 60

4.1 Conclusions 60

4.2 Directions For Further Research 61

REFERENCES 62

LIST OF TABLES

Table 3.1 The user time spent for each module at each level of the hierarchy for Fruits stereo pair. 50

Table 3.2 The user time spent for each module by one-level algorithm for Fruits stereo pair. 52

Table 3.3 The user time spent for each module at each level of the hierarchy for Pentagon stereo pair. 53

Table 3.4 The user time spent for each module at each level of the hierarchy for Brutus stereo pair. 59

LIST OF FIGURES

Figure 2.1 The Stereo Camera Geometry and the Epipolar Plane . . . 4

Figure 2.2 Disparity in Parallel and Non-parallel Axes Stereo Camera Geometries 5

Figure 2.3 %30 Random dot stereogram 8

Figure 2.4 Circularly symmetric LoG operator (inverted) 12

Figure 2.5 Real and imaginary parts of the complex Gabor filter with a bandwidth of 1 octave 16

Figure 2.6 Matches between rows R and L violating a) the uniqueness constraint, b) the compatibility constraint and c) the orderness constraint. d) A valid matching field with 2 occluded pixels in the row R. 25

Figure 2.7 Coarse-to-fine control strategy. 27

Figure 2.8 Cooperative integration of three stereo modules 32

Figure 3.1 Block Diagram of the Stereo System at one Level 35

Figure 3.2 Image Pyramid of Right Image of Fruits Stereo Pair 36

Figure 3.3 The matches forbidden by an accepted match due to uniqueness and orderness constraints 37

Figure 3.4 a) The matches forbidden by a match supplied by the coarser channel, b) the matches forbidden by a set of matches supplied by the coarser channel. 38

Figure 3.5 Support of a plane as a function of disparity gradient. . . . 40

Figure 3.6 Canny edges of fruits stereo pair 41

Figure 3.7 Right part of fruits stereo pair filtered with WFP filters of period 4, 8, 16, and 32 pixels. 42

Figure 3.8 Histograms of phase images for various filter sizes. 43

Figure 3.9 Results for Fruits stereo pair at 3 levels of the hierarchy. Left: Pixel matching module outputs. Right: Thin plate module outputs. 49

Figure 3.10 The final result of one-level algorithm on Fruits. 50

Figure 3.11 3-D rendered view of the result on Fruits at the finest level. 51

Figure 3.12 The discontinuities and occlusions for Fruits stereo pair. . 52

Figure 3.13 Pentagon Stereo Pair. 53

Figure 3.14 Results for Pentagon stereo pair at 3 levels of the hierarchy. Left: Pixel matching module outputs. Right: Thin plate module outputs. 54

Figure 3.15 3-D rendered view of the result on Pentagon at the finest level. 55

Figure 3.16 Brutus Stereo Pair. 56

Figure 3.17 Results for Brutus stereo pair at 2 levels of the hierarchy. Left: Pixel matching module outputs. Right: Thin plate sheet module outputs. 57

Figure 3.18 3-D rendered view of the result on Brutus at the finest level. 58

CHAPTER I

INTRODUCTION

Visual Perception which is an interpretation of 2-dimensional time-varying light information on the retina to form a spatio-temporal reconstruction of 3-dimensional world, has reached an astonishing complexity during the long course of evolution. Although “seeing” objects seem an effortless and automatic task to us, the underlying biological structure and processes are extremely complicated. Research on visual perception is an interdisciplinary area: philosophy, cognitive psychology, neurophysiology, psychophysics and computer science treat the subject from different points of view. Computer science which first involved in the subject for simulating some simple models of biological visual systems, later, proceeded in a rather independent way. By the rapid increase in computer hardware speeds, computer vision turned its eyes onto possible applications. Although machine vision applications use the results of visual perception research, they are not always concerned in being biologically plausible.

Among visual perception mechanisms like motion, color constancy etc, stereopsis, which is a passive way of determining the depth using the slight differences between the views of two eyes, draws considerable attention and much research is devoted to this area, because it has various military and civil applications such as determining the position of a target, robot navigation, aerial cartography, automatic surveillance, inspection of industrial parts, building models of objects for computer graphics, figure-ground separation for videophones, reconstruction of human retina for diagnosis and 3-D angiography. Besides, stereo systems are used in more comprehensive vision systems either to guide other low-level processes or to supply information for higher-level processes such as object recognition.

Existing stereo algorithms generally provide reliable and accurate data on a class of scenes but fail on others. The robustness of human stereopsis is not yet achieved by machine perception. This is partly due to deficiency of other visual tasks such as illusory contour detection, texture segmentation, featural grouping, etc. Another reason may be not using information from other sources as motion, accommodation of lenses, eye vergence, texture and others, so the trend in stereo research is towards integrating several vision modules. However, even in the absence of such cues, man-made systems perform worse than human, so there are more to do to improve stereo algorithms alone.

1.1 The Aim of the Thesis

The aim this thesis is to develop a robust stereo system that will work successfully on a large variety of images. The algorithm is expected to be amenable to parallel implementation but still fast in serial simulation. The system is not intended as a model of human stereopsis in any respect, but as much as possible from human stereopsis research is used, since human brain is the best working system. A dense disparity map with subpixel accuracy is attempted to be obtained with explicit localization of depth discontinuities and occlusions.

1.2 The Organization of the Thesis

Following this introductory chapter, previous work on stereo is reviewed in Chapter 2 titled “Background”. Then, in Chapter 3, the stereo algorithm developed in this thesis is presented with results on test images. The thesis ends in Chapter 4 with the conclusions drawn from this work.

CHAPTER II

BACKGROUND

Two valuable sources as reviews of computational stereo vision are, a survey by Barnard and Fischer published in 1982 [1] and, more recently, a review by Dhond and Aggarwal [2] in 1989. Here, after a short description of the stereo process, a more systematic review of the matching phase will be presented with special emphasis on more recent developments and on the issues closely related to the work presented in this thesis.

2.1 The Stereo Process

Stereopsis is the mechanism that fuses the images from both eyes to determine the depth. When we observe a point monocularly, we can say that it lies on a certain line in space. If we observe the same point from a different place, the actual position of that point is the intersection of the two lines determined by two eyes or two cameras. Consider two side-by-side cameras with almost parallel optical axes (Figure 2.1). The plane on which a point, its projections and focal points of the cameras lie is called the epipolar plane, and the intersection of this plane with the images is the epipolar line. The difference between the relative positions of the right and left projections of the same point is called disparity and this difference is a function of the distance of that point to the viewing system. Figure 2.2 illustrates this phenomenon for two different geometries with two points P_1 and P_2 on the same epipolar plane. If the axes of the cameras are parallel, then the objects at infinity have zero disparity while all other points have positive values. When the axes intersect at a point in front of the cameras, the object at that distance has zero disparity. Closer points and farther points have positive and negative disparities, respectively. This relation between disparity and depth is the basis of stereo vision.

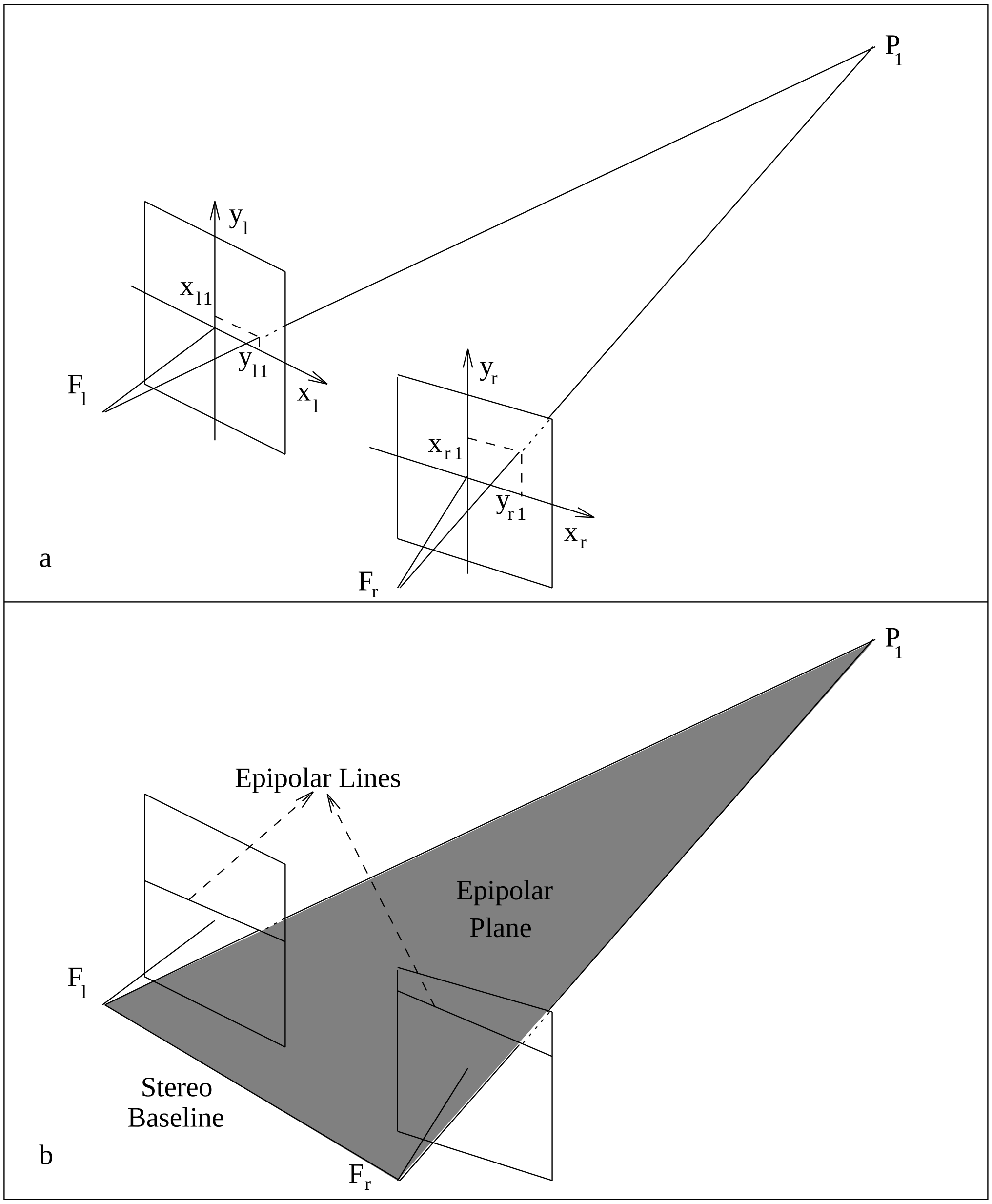

Figure 2.1. The Stereo Camera Geometry and the Epipolar Plane

The passive stereo process can be divided into following phases: 1) image acquisition, 2) determining the geometry, 3) matching and 4)calculating the actual depth. Below, each of these phases is introduced shortly.

2.1.1 Image Acquisition

The quality of image acquisition affects the results considerably. Low-noise cameras and good illumination conditions lead better results. When two cameras are used they should be identical, if special stereo cameras are not available.

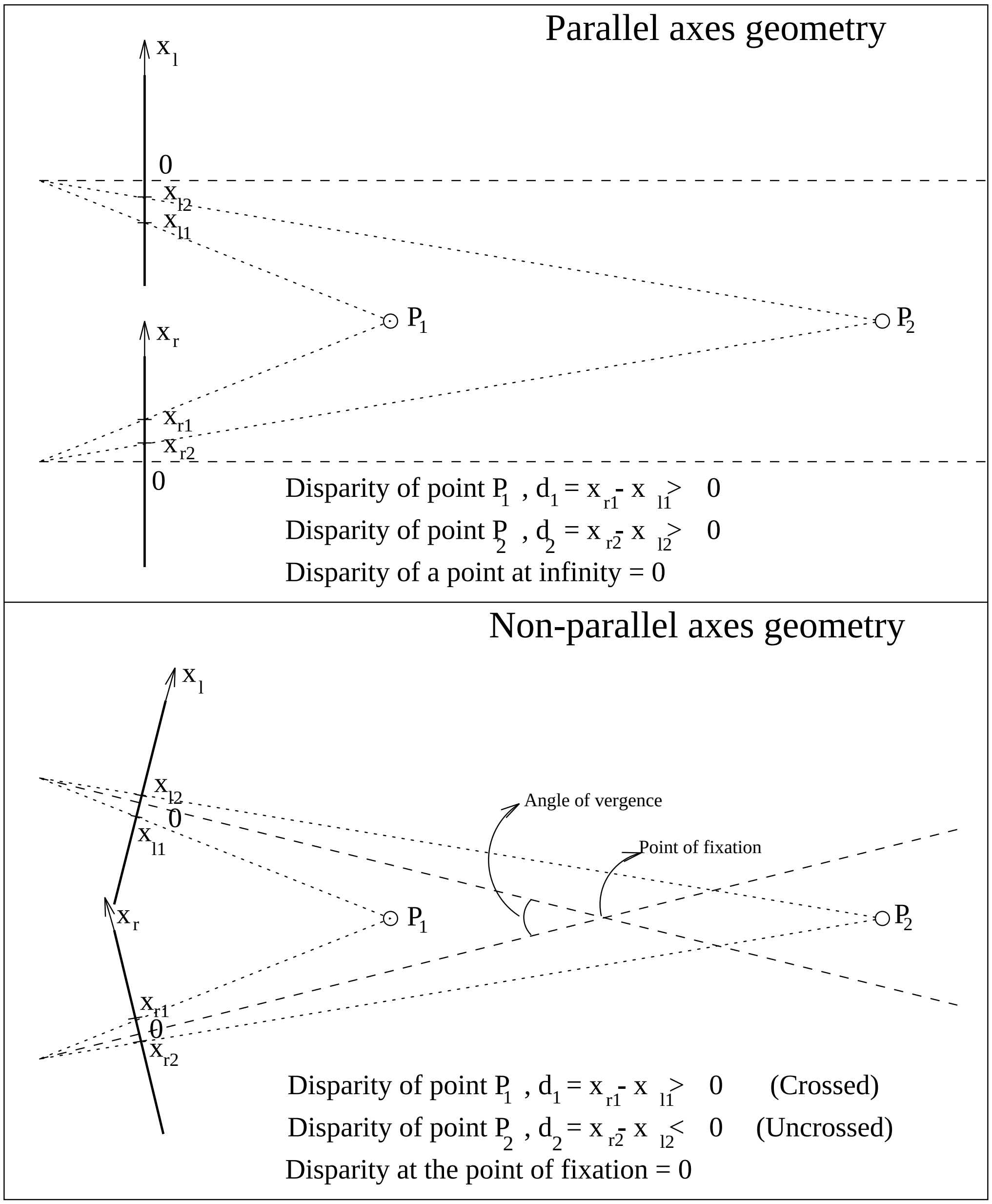
Parallel axes geometry
x_l
0
x_l2
x_l1
P_1
P_2
x_r
x_r1
x_r2
0
Disparity of point P_1 , d_1 = x_r1 - x_l1 > 0
Disparity of point P_2 , d_2 = x_r2 - x_l2 > 0
Disparity of a point at infinity = 0
Non-parallel axes geometry
x_l
x_l2
0
x_l1
Angle of vergence
Point of fixation
x_r
P_1
P_2
x_r1
0
x_r2
Disparity of point P_1 , d_1 = x_r1 - x_l1 > 0 (Crossed)
Disparity of point P_2 , d_2 = x_r2 - x_l2 < 0 (Uncrossed)
Disparity at the point of fixation = 0

Figure 2.2. Disparity in Parallel and Non-parallel Axes Stereo Camera Geometries

2.1.2 Determining the Geometry

The geometric parameters for stereo are camera parameters and the relative positions of the two cameras. The camera parameters including the focal length and distortion characteristics of the lens are known a priori and can be determined very precisely. The relative positions of the cameras must be determined a posteriori in some applications such as aerial cartography. When certain number of points are matched across images it is possible to find the rotation and translation matrices which relate the positions. Generally, affine transformations are applied to reproject one of the images or both so that image rows are aligned with the stereo baseline if it is not aligned already, since this facilitates the stereo task greatly in the next phase. Besides control points [3], correlations or Fourier transform [4] [5] can also be used to realign the images. Note that if the cameras are at different distances to the scene, the effect of this difference cannot be corrected by simple scaling because the projections are not orthographic. The correction for perspective projections requires depth information which is not known at this stage. However, if the depth variation is small compared to the viewing distance then the projection can be assumed to be orthographic. Refer to [6] [7] [8] for detailed information on this stage.

2.1.3 Matching

In this phase, the corresponding points in right and left images, that is, the image points which are the projections of the same physical point, are determined. Most of the work in this thesis deals with this problem. In matching stage the correspondences may not be obtained at every pixel of the images so an interpolation step is necessary if a dense depth map is required.

2.1.4 Calculating the Actual Depth

Once we know the geometry of the system and determine the disparity at all points, it is straightforward to calculate the actual depth. An analysis of the error in calculating the depth is found in [9] and [10].

Among the above phases the most difficult one is the matching phase. Although it is not known completely how biological systems solve this problem so quickly and accurately, a large body of knowledge is at hand through neurophysiological studies on higher vertebrates and psychophysical studies on human.

2.2 Psychophysics and Neurophysiology of Stereopsis

In this section, a short collection of some results from neurophysiology or psychophysics of stereopsis will be presented which seem to be closely related to and/or have influence on computer stereo vision. However, a detailed and exhaustive review is not intended. Interested reader is recommended to refer to [11] [12] and [13].

2.2.1 The speed of human stereopsis

A very remarkable feature of human stereopsis is its speed: it takes about 200 msecs from presentation of the stimulus to the occurence of depth perception [14] which is very close to the time needed for the information on the retina to reach to the visual cortex via the visual pathway.

2.2.2 Stereopsis is a Low-Level Process

Stereopsis is a low level process, that is, it does not require recognition or any abstract understanding of the image. It was first demonstrated by Julesz that [15] stereopsis survives in the absence of any monocular cue such as texture, a priori knowledge on the shapes and sizes of objects, shading etc. Figure 2.3 is an example of random dot stereograms which was invented by Julesz. One can see the floating square above the background when he fixates his eyes at a nearer point in such a way that the two images overlap in the center.

2.2.3 Limited Fusional Area

Only the surfaces within a specific disparity interval, so-called Panum's fusional area, can be fused. The extent of this range is measured between 10-40 minutes of arc depending on the data used. There is evidence that this range is larger for inputs with low frequency content compared to high frequency inputs [16] [17].

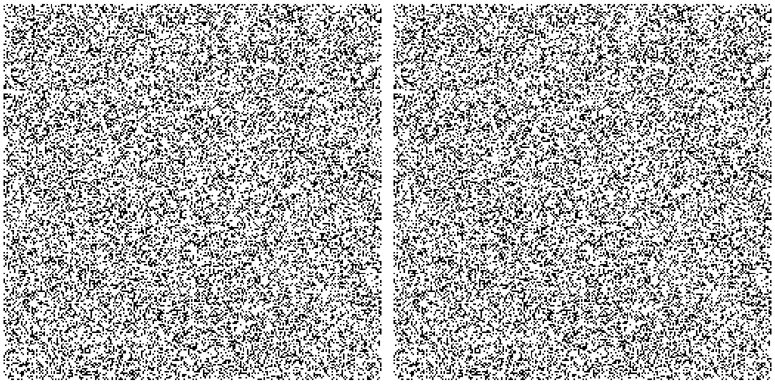

Figure 2.3. %30 Random dot stereogram

2.2.4 Effect of Contrast

It was shown by Julesz that [18] changes in the magnitude of the contrast across the images does not destroy stereopsis, but a change in the sign of contrast makes fusion of images impossible [15].

2.2.5 Hyperacuity

Even though the average distance among the light-sensitive cells of the retina (cones), is about 20-30 seconds of arc at the fovea where those cells are densest, the disparity differences down to 2 seconds of arc are detectable by the human visual system [19]. However, this hyperacuity drops drastically for non-zero disparities [20].

2.2.6 Gradient Limit

If the rate of change in disparity, that is, the disparity gradient, exceeds a certain limit the images cannot be fused and objects appear as double (diplopia) [21].

2.2.7 Binocular Cells in Visual Cortex

Although there is some interaction of information from both eyes on the way from retinae to cortex, the first place where cells differentially sensitive to binocular disparity are observed is the visual cortex in cats and monkeys. A considerable proportion of the cells at visual cortex are binocularly sensitive [22].

2.2.8 Ocular Dominance

Binocularly sensitive cells can be classified as balanced or unbalanced according to the type of their sensitivity [23]. Balanced cells respond equally to stimuli from each eye, but respond very strongly when stimulated binocularly. Unbalanced cells either respond stronger to one eye or exhibit a complex ocular dominance pattern.

A certain layer of the visual cortex (layer 4) is organized in ocular dominance columns. These vertical strips which are 1 mm thick in monkeys and 2 mm thick in humans respond alternatingly to left eye and right eye. Binocular cells are located above and below these monocular cells.

2.2.9 Orientation Selectivity

Almost all of the cells in visual cortex exhibit orientation selectivity at various angles. However, most of them respond best to bars oriented within ±20 degrees from the vertical [23].

2.2.10 Frequency Selectivity

Another important property of these cells is their frequency selectivity. The optimal spatial frequencies of these cells range from 0.3 to 3 cycles/degree in cats and 2 to 8 cycles/degrees in monkeys [12]. The bandwidth of the cells in the average is a little bit larger than one octave. The constancy of relative bandwidths over scales can be justified by the statistics of natural images [24]. There is almost constant energy in all channels, because the amplitude spectrum of natural images generally falls off with $1/f$.

2.2.11 Receptive Fields Types

Receptive field is the activation pattern of a cell as a function of stimulus position on the retina. According to the pattern of their receptive fields the cells in the visual cortex are classified as simple and complex cells [25] . Simple cells have smaller receptive fields and low spontaneous activity. Some parts of their receptive field respond the onset of the stimulus while some parts respond to the offset. On the contrary, complex cells respond both the onset and the offset. They have larger receptive fields and greater spontaneous activity.

2.2.12 Binocular Sensitivity Types

According to their binocular sensitivity, the cells in the visual cortex are classified into four groups by Poggio and Fischer [23] as tuned excitatory (TE), tuned inhibitory (TI), near and far. TE cells are excited by stimuli at the fixation distance. If the stimulus is disparate more than $\pm 0,1$ degrees then the cell activities are suppressed, that is, these cells are sharply tuned to zero disparity. The response pattern of TI cells as a function of disparity is the reverse of, but is not as sharp as, that of the TE cells. Near cells are sensitive to stimuli near than the fixation distance and far cells are visa versa. Among these cell groups only TE cells are ocularly balanced. Later, other kinds of cells are also identified and it is claimed that types according to binocular sensitivity belong to a continuum rather than discrete groups [26].

2.2.13 Modelling Simple Cells

The monocular receptive fields of simple cells are well described by Gabor functions [27] [28] which are filters limited in both space and frequency. See Subsection 2.3.2 for a detailed discussion on Gabor filters. There exists evidence that simple cells are found in pairs with an approximate phase difference of 90 degrees [29] which may compute real and imaginary parts of a complex Gabor filter. The integration of data from monocular receptive fields is modelled as linear summation by Ohzawa and Freeman [30] based on neurophysiological experiments. Nomura et al. [31] proposed a similar modelling where linear summation is followed by a non-linear smoothed thresholding function. This model predicts largely the binocular behavior of cells in the striate cortex. Freeman and Ohzawa observed that the phase difference sensitive responses of simple cells are not disturbed by large contrast differences across right and left eyes. Considering this observation, they proposed a monocular contrast gain mechanism that keeps the effect of contrast almost constant.

2.2.14 Coarse-to-Fine Structure

There is evidence that data from low-frequency channels constrain the matching at high frequencies. Wilson et al. [32] found that channels more than 2 octaves apart process independently, but closer channels interact. Low-frequency signals affect fusion in high-frequency channels but not vice versa. Watt [33] also concludes, after a series of experiments, that the human visual system uses a coarse-to-fine strategy.

2.3 The Matching Primitives

One of the major differences among the stereo algorithms in the literature is in the properties of the images they choose for matching. Considering this aspect, stereo vision algorithms are classified roughly into two main classes: feature-based and area-based. These two kinds of primitives will be discussed in detail in the following two subsections.

2.3.1 Feature-Based Matching Primitives

Image features which are chosen for matching are high interest points or point sets like edgels, edge segments or intervals between edges. The features can be localized very accurately (generally with sub-pixel resolution) which leads to accuracy in the computed disparity. They generally correspond to physical boundaries of objects, surface markings or other physical discontinuities, so provides valuable depth information. These features are typically sparse, that is, there are features at a very low percentage of pixels in an image. This speeds up processing since only features are tried to be matched which are small in number. On the other hand, disparities at non-feature points should be interpolated.

The fact that corners and vertices are highly distinguishable and generally sparse matching primitives makes the matching process less ambiguous. Corners or vertices can be detected by two methods: The first method involves analysis

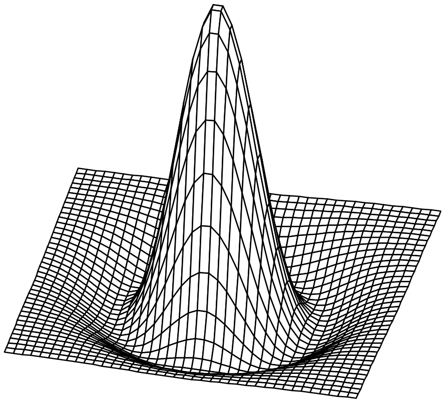

Figure 2.4. Circularly symmetric LoG operator (inverted)

of detected edges and the second works directly on grey-level images. For instance, Kim and Bovik [34] use high curvature points on edges. The Moravec operator which falls into the second category is frequently used [35] [36], though more recent and complicated operators [37] may be more powerful in resolving ambiguities.

Edgels seem to be the most common matching primitive in stereo research [38] [39] [40] [41] while zero-crossing of the Laplacian of Gaussian (LoG) seems to be the most common edge detection method [42] [43] [16] [44] [45] [34]. The LoG operator, so-called Mexican-hat operator (see Figure 2.4)

$$\bigtriangledown^2 G(r,\theta) = -\frac{1}{\sigma^3\sqrt{2\pi}}\left(1-\frac{r^2}{\sigma^2}\right)\ e^{(-r^2/2\sigma^2)} \tag{2.1}$$

which was first proposed by Marr and Hildreth [46] has several useful properties. The LoG is a Gaussian smoothing followed by a second derivative operation. The scale factor σ which is the standard deviation of the Gaussian, is inversely proportional to the average density of obtained edgels, s, with the equation [42]

$$s = 0.0945/\sigma \quad (edgels/pixel), \tag{2.2}$$

so we can control the ambiguity. Besides, Marr and Hildreth state that zero-crossings of LoG as an edge detector is biologically plausible. Even the large convolutions can be calculated quickly by either approximating the LoG by a difference of Gaussians or by decomposing the LoG [46] [47] [48] [49]. Although the zero-crossings of the LoG may detect spurious edges they can be eliminated with little extra computation [48] [50]. Another disadvantage of the LoG is the displacement of edges as σ grows. This behavior of edges is well-investigated in scale space [51] [52] [53] which is introduced by Witkin as [52]

$$f_{ss}(x,\sigma) \overset{\text{def}}{=} f(x) * g(x,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \frac{d^2}{dx^2} \int_{-\infty}^{\infty} F(\omega) e^{-(x-\omega)^2/2\sigma^2} \, d\omega \qquad (2.3)$$

where $g(x,\sigma)$ is the Gaussian function $e^{-x^2/2\sigma^2}$ and $F(\omega)$ is the Fourier transform of $f(x)$. Image of edges detected in scale space is called a scale map. Clark [51] showed that the error in zero-crossing based matching increases with sigma, disparity gradient and the angle of the scale map contour to vertical. Ulupınar and Medioni [50] proposed a technique to correct for this displacement. The direction of the edge is approximated as the direction of the gradient of the filtered image. In matching only edgels with the same sign and with roughly the same orientation are considered [54]. In recent stereo research Canny operator [55] and Deriche method [56] are also used as edge-detectors [57] [3] [58] [41]. Peaks of the LoG filtered image may also be used with zero-crossings [59].

A more abstract image feature is edge segments, either line segments [60] [58] [10] or curves [61] [62] [45]. Use of segments instead of edgels reduces the number of possible matches significantly. Besides one can define similarity measures between edge segments using the length, orientation, curvature, strength, coordinates of edge points, average intensity and intensity slope at each side or similar characteristics. Boyer and Kak [63] use structural primitives and a set of named relations over those primitives. The dissimilarities of primitives are obtained empirically using manually matched pairs.

The interval between two edges along a scanline is also used as a matching primitive. Lloyd [43] defined the dissimilarity of two intervals I_1 and I_2 as

$$\delta(I_1, I_2) = \frac{1}{f_l}|l_1 - l_2| + \frac{1}{f_a}(D(\theta_{r1}, \theta_{r2}) + D(\theta_{l1}, \theta_{l2})) + \frac{1}{f_g}(|g_{r1} - g_{r2}| + |g_{l1} - g_{l2}|) \quad (2.4)$$

where $l_i, \theta_{ri}, \theta_{li}, g_{ri}$ and g_{li} are the length, the angles of edges bounding the interval from right and left of the interval I_i and average gray levels at each end of the interval, respectively. f_l, f_a and f_g are fixed weights to adjust the contribution of each feature and $D(\theta_1, \theta_2) = 1 - cos(\theta_1 - \theta_2)$. The cost function of Ohta and Kanade [64] for intervals is based on intensities: they assume that the pixels from two matching intervals are from the same homogeneous surface so they must have similar intensities. The cost of matching two intervals with lengths l and k, respectively, is $C_p = \sigma\sqrt{k^2 + l^2}$ where the variance and mean of all pixels in the two intervals are computed as

$$m = \frac{1}{2}\left(\frac{1}{k}\sum_{i=1}^{k} a_i + \frac{1}{l}\sum_{j=1}^{k} b_j\right) \tag{2.5}$$

$$\sigma^2 = \frac{1}{2}\left(\frac{1}{k}\sum_{i=1}^{k} (a_i - m)^2 + \frac{1}{l}\sum_{j=1}^{k} (b_j - m)^2\right) \tag{2.6}$$

A similar cost function used by Ito and Ishii [65] for one side of edges is

$$E = \alpha|m_r - m_l|/(m_r + m_l + 1) + (1-\alpha)|\sigma_r^2 - \sigma_l^2|/(\sigma_r^2 + \sigma_l^2 + 1) \tag{2.7}$$

where m and σ^2 denote average and variance of pixels, respectively.

2.3.2 Area-Based Matching Primitives

Area properties are those which are available at almost every point in an image. The simplest area property is the image intensity. The smallest is the difference between the pixels to be matched, the better is the match. This measure is very sensitive to noise as well as to brightness differences across images. On the other hand, it is very simple and easy to compute, so it was used by several researchers [66] [67] [68]. Also Jordan and Bovik [69] have developed a difference metric for color images based on intensities. Another area property is the derivative of intensity. This measure is not much sensitive to illumination differences but it is still sensitive to noise. Zhou and Chellappa [70] have fit polynomials to the image intensity to reduce the effect of noise. Kass [71] uses a vector formed by first and second derivatives of the image at two orthogonal directions as a matching primitive that is robust to moderate noise.

A common way to match areas directly is to find correlations between areas from left and right images [40] [72] [73] [3] [74]. The cross-correlation and normalized cross-correlation at position (i, j) of the right image with disparity d are

$$C(i,j,d) = \sum_{x=-N}^{N}\sum_{y=-M}^{M} I_r(i+d+x, j+y)\, I_l(i+x, j+y) \tag{2.8}$$

and

$$C(i,j,d) = \frac{\left(\sum_{x=-N}^{N}\sum_{y=-M}^{M} I_r(i+d+x,j+y)\, I_l(i+x,j+y)\right)^2}{\sum_{x=-N}^{N}\sum_{y=-M}^{M} I_r^2(i+d+x,j+y)\sum_{x=-N}^{N}\sum_{y=-M}^{M} I_l^2(i+d+x,j+y)} \tag{2.9}$$

respectively. There are several other correlation-like measures of which the most frequently used one is the sum of squared differences:

$$C(i,j,d) = \sum_{x=-N}^{N}\sum_{y=-M}^{M}\left(I_r(i+d+x,j+y) - I_l(i+x,j+y)\right)^2 \tag{2.10}$$

A comparison of several correlation-like functions can be found in [75] and [73]. Although correlation techniques are successful at textured areas, they fail around depth discontinuities, since the area inside the correlation window belongs to at least two different surfaces at different depths, so the window does not match totally at any disparity value. They also suffer from disparity gradients because one of the signals is scaled compared to other. Besides, the accuracy obtained is lesser when compared to feature-based matches. Another drawback of the correlation technique is its computational complexity. As the size of the correlation window gets larger, the computational complexity and the uncertainty in disparity increase as well as problematic regions near discontinuities get larger, however, the match becomes more robust to noise. To overcome this problem adaptive window sizes can be used [76]. Nishihara who used correlations of the signs of LoG filtered images [77] showed that correlation of sign representation has good localization properties, that is, it has a sharp autocorrelation function, and it is not sensitive to contrast differences. Besides, the multiplication in correlation is replaced with an exclusive or function.

Another dense property to match is local frequency components [51] [78] [79] [80] [81] [82] [83] [84]. The Fourier theorem states that when a function $f(x)$ with Fourier transform $F(u)$ is shifted by an amount of Δx then the Fourier transform of the shifted function $f_s(x-\Delta x)$ is $e^{-ju\Delta x}F(u)$, so a shift in the spatial domain corresponds to a phase shift in the frequency domain. If the left view had been a shifted version of the right view it would have been possible to determine the amount of shift from the phase of the Fourier transforms of both images. However, since the shift, i.e. the disparity, is different in various regions of the images, one needs a local frequency filter to determine the phase differences. A natural choice for such a function is the Gabor filter [85] which is a bandpass

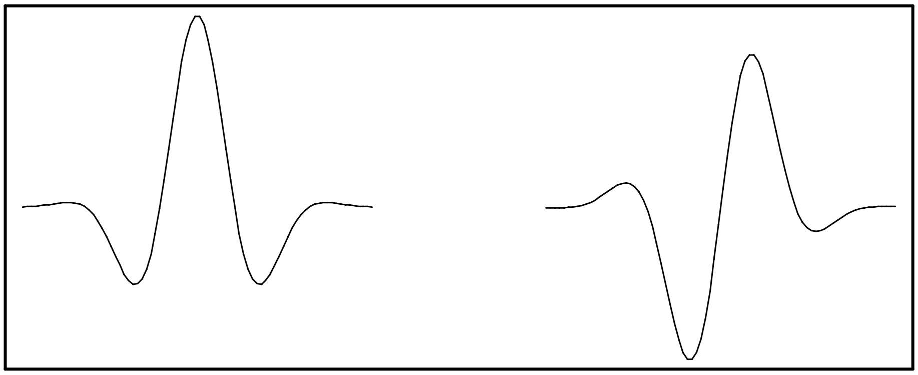

Figure 2.5. Real and imaginary parts of the complex Gabor filter with a bandwidth of 1 octave

filter with limited spatial width:

$$g_{\omega_0}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} e^{j\omega_0 x} \tag{2.11}$$

whose Fourier transform is

$$G_{\omega_0}(\omega) = e^{-(\omega-\omega_0)^2/2\tau^2} \tag{2.12}$$

where the product $\tau\sigma$ is 1 which is the theoretical minimum of any linear complex filter [85]. This choice is also biologically plausible since the receptive fields of simple cells are not statistically distinguishable from Gabor filters [27]. Besides, simple cells are found in pairs with an approximate phase difference of 90 degrees [29] and this justifies the use of complex filters. If the ratio of the spatial width, σ, to the period of the filter, $\omega_0/2\pi$, is held constant, then the shape of the filter and the relative bandwidth given by

$$\lambda = log_2 \left(\frac{\omega_0 + \tau}{\omega_0 - \tau} \right) \tag{2.13}$$

in octaves remain unchanged. Figure 2.5 shows the real and imaginary parts of a Gabor filter with a bandwidth of 1 octave. The 2-dimensional extension of the filter is

$$g_{uv}(x,y) = e^{-\left(\frac{x}{\sigma_x}+\frac{y}{\sigma_y}\right)^2} e^{j(ux+vy)} \tag{2.14}$$

Note that the filter is separable, so computational complexity is reduced from $O(N^2)$ to $O(N)$. The filtered versions of right and left images $I_r(x,y)$ and $I_l(x,y)$

are

$$R^r_{\omega_0}(x,y) = I_r(x,y) * g_{\omega_0}(x,y) \tag{2.15}$$

$$R^l_{\omega_0}(x,y) = I_l(x,y) * g_{\omega_0}(x,y). \tag{2.16}$$

So that the Gabor filtered image is a band-pass signal, it can be modelled (in 1-D for simplicity) as [79]

$$R_{\omega_0}(x) = \rho(x)e^{j(\omega_0 x+\psi(x))} \tag{2.17}$$

where ω_0 is the center frequency equal to the frequency of the filter. The local frequency is defined as [86] $\omega_l = \phi\prime(x)$ where $\phi(x) = \omega_0 x + \psi(x)$. If we assume perfect sinusoids that is $\psi(x) = kx$ then we can estimate the disparity as [78]

$$d_r(x) = \frac{\phi_l(x) - \phi_r(x)}{\omega_0}. \tag{2.18}$$

Since the bandwidth of the filter is non-zero, $\psi(x)$ may vary around zero and disturb the linearity. However, in real images with sufficient texture the phase is almost linear over the image except some regions. Fleet et al. [79] showed that the bandpass phase is not sensitive to typical distortions that exist between right and left images. Later, Fleet and Jepson [87] did a more in-depth treatment of the phase stability for several complex band-pass filters.

Note that the phase measurements give the disparity directly, so a search is not performed for the best fit, because of this phase-based techniques are sometimes called “correspondenceless”. It is worth mentioning that matching phases is a general case of matching zero-crossings because the zero-crossings of band-pass filters such as LoG correspond roughly to level curves at $\pi/2$ and $-\pi/2$ of the phase signal. Another advantage of the phase-measurements is that they provide sub-pixel measurements without explicitly reconstructing the signal between pixels. This is also in accordance with biological findings.

Phase measurements are valid within a limited range of disparity because of the wrap-around problem: we measure only the principal component of the phase in the range $[-\pi, \pi]$, so a filter of fundamental frequency ω_0 signals only disparities of $-\pi/\omega_0$ to π/ω_0. Sanger [78] uses only the phase of the signal and the magnitudes are used independently to generate confidence values.

Fleet et al. [79] iterate the basic measurements to obtain more accurate results. They investigate the Gabor filter in Gabor scale space which is defined as

$$S(x,\omega) = (e^{-x^2/2(\sigma(\omega))^2} e^{j\omega x}) * I_x \tag{2.19}$$

and

$$\sigma(\omega) = \frac{1}{\omega}\left(\frac{2^\beta + 1}{2^\beta - 1}\right) \tag{2.20}$$

where β is the relative bandwidth in octaves. In Gabor scale space there are some points where the magnitude is zero, so the phase is undefined. At a neighborhood of these singular points the local frequency varies very rapidly and the disparity measurements are not reliable. Fleet et al. [79] use thresholds on local frequency deviation and local amplitude variation to detect singularity neighborhoods. They also show on real images how disparity measurements improve when we detect and correct for singularities. Weng [81] considers a drawback of the Gabor filter: the DC residual of the even component which disturbs the linearity of the phase. Instead he proposes another local frequency filter, namely, Windowed Fourier Phase (WFP) where the Gaussian envelope of the Gabor filter is replaced with a rectangular one:

$$h(x,y) = w_M(x,y)e^{j(ux+vy)} \tag{2.21}$$

where

$$w_M(x,y) = \begin{cases} 1 & \text{if } |x| \leq M/2 \text{ and } |y| \leq M/2 \\ 0 & \text{otherwise.} \end{cases}$$

He also convolves this filter with a Gaussian filter to remove the high frequency noise in the signal. The resultant filter is very similar to the Gabor filter but has no DC when M is a multiple of the period. However, Fleet and Jepson [87] state that this filter has a bias towards phase values of $\pm\pi/2$ because of the difference of amplitude spectra of real and imaginary parts. Westelius [80] compares several types of complex filters, namely, Gabor, lognorm, non-ringing and difference of Gaussians and concludes that the choice of filters depends on the intended application.

Nomura [82] introduced a fundamental equation for binocular disparity,

$$\frac{dI}{do} = \nabla I \, d + \frac{\partial I}{\partial o} = 0 \tag{2.22}$$

where o is the eye position, I is the intensity and d is the disparity. This equation is a variation of the gradient model of optical flow field. Substituting Gabor filtered image in place of I, he obtained

$$\frac{\partial R_l(x,y)}{\partial x} d + (R_l(x,y) - R_r(x,y)) = 0. \tag{2.23}$$

Besides he showed that the terms other than d can be approximated as linear combinations of far, near and tuned inhibitory type simple cells.

Another correspondenceless method is the phase correlation that was introduced by Kuglin and Hines [88]. If an image patch $l(x,y)$ from the left image and the corresponding image patch $r(x,y)$ in the right image have a disparity d then we can write $l(x,y) = r(x,y) * \delta(x+d,y)$. Solving for the disparity term in the Fourier domain gives

$$e^{-jud} = \frac{L(u,v)}{R(u,v)} = \frac{L(u,v)R^*(u,v)}{|R(u,v)|^2}. \tag{2.24}$$

Since the magnitudes of $L(u,v)$ and $R(u,v)$ are identical

$$e^{-jud} = \frac{L(u,v)R^*(u,v)}{|L(u,v)R^*(u,v)|}. \tag{2.25}$$

The phase correlation is defined as

$$P_{lr}(x,y) \overset{\text{def}}{=} \mathcal{F}^{-1}\left\{\frac{L(u,v)R^*(u,v)}{|L(u,v)R^*(u,v)|}\right\} \tag{2.26}$$

which equals to

$$\mathcal{F}^{-1}\left\{e^{-jud}\right\} = \delta(x-d,y). \tag{2.27}$$

So, we can obtain the disparity directly by locating the impulse. This is similar to cross-correlation because

$$C_{lr}(x,y) \overset{\text{def}}{=} E\{l(x-d,y)r^*(x,y)\} = \mathcal{F}^{-1}\left\{L(u,v)R^*(u,v)\right\}. \tag{2.28}$$

Note that, in phase correlation, the magnitude of the signals in the Fourier domain are forced to unity. This technique is better than cross-correlation when there exists band-limited noise like the effect of illumination changes that are concentrated at low frequencies, because all frequencies contribute to the result equally [6].

A related method is the cepstral filtering approach of Yeshurun and Schwartz [14]. Cepstral filtering is a Fourier transformation followed by a logarithm and an inverse Fourier transform. Yeshurun and Schwartz append $l(x,y)$ to the left of $r(x,y)$. Assume that the width of the patches is D and $r(x,y)$ is equal to

$l(x-d,y)$ where d is the disparity to be computed. Then the compound image $f(x,y)$ can be written as

$$f(x,y) = l(x,y) * \{\delta(x,y) + \delta(x-D-d,y)\} \tag{2.29}$$

with the Fourier transform

$$F(u,v) = L(u,v) \cdot \{1 + e^{-j(D+d)u}\}. \tag{2.30}$$

When we take the logarithm of $F(u,v)$, the product becomes a sum:

$$log(F(u,v)) = log(L(u,v)) + log(1 + e^{-j(D+d)u}). \tag{2.31}$$

Taking the Inverse Fourier Transform, we obtain

$$\mathcal{F}^{-1}\{log(F(u,v))\} = \mathcal{F}^{-1}\{log(L(u,v))\} + \mathcal{F}^{-1}\left\{log(1 + e^{-j(D+d)u})\right\}. \tag{2.32}$$

The second term equals

$$\mathcal{F}^{-1}\left\{log(1 + e^{-j(D+d)u})\right\} = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{\delta(x-n(D+d))}{n}. \tag{2.33}$$

Thus, we can find the disparity of the patch by locating the largest delta function.

2.4 Problems in Matching

The matching is not a trivial task: one has to choose the best (according to some criteria) match among a number of possible matches, and this match is not necessarily the true one. The sources of ambiguity and error in stereo matching can be classified as photometric variation, lack of texture, repetitive texture and occlusions [3]. Following subsections briefly explain these sources.

2.4.1 Photometric Variation

Stereo images are not simply warped versions of each other. The sensor and quantization add noise to the images. Besides, the non-linearities in the lens-camera system distort them. If the images are taken simultaneously, using two cameras, little variations between the gains or optical characteristics of the cameras may affect the result. On the other hand, if the images are taken successively, using the same camera, the lighting conditions may change meanwhile. Sometimes the view changes physically. Non-Lambertian surfaces are another source of error, since the corresponding light intensities of a point change across the images with the viewing angle.

2.4.2 Lack of Texture

In some areas of images there is no significant texture, so nothing can be matched. The disparity can only be interpolated in these areas. Some examples of such areas are clear sky, painted flat surfaces like walls and too dark or too bright areas where all the pixels have the minimum or maximum possible value, respectively.

2.4.3 Repetitive Texture

If a texture is repeated in horizontal direction (e.g. a chess table), there will be more than one good matches. In this case, one cannot decide which one is the true match using only local information.

2.4.4 Occlusion

When there is a discontinuity in depth, some parts of the scene are seen only by one camera, so these areas do not have a counterpart to match in the other image.

2.5 Representing Disparity

When a very sparse set of matching primitives is used, the disparities can be stored internally as a list of coordinates of features and associated disparities. In other cases, an array is used to store the disparity field. Generally the maximum disparity range is smaller than 256 pixels so the arrays are of type byte. If subpixel accuracy is aimed floating point type can be used. Some researchers keep two disparity arrays for right and left images respectively. Other alternatives are either to use centralized coordinates or to define a dominant eye. Besides a special marker can be used in the disparity field to model occluded regions.

2.6 Constraining the Solution

Since the correspondence problem is ill-posed in nature, that is, the existence, uniqueness and stability of the problem is not guaranteed, some a priori data need to be used about the disparity field. The assumptions made are imposed on the algorithms as constraints. Every stereo algorithm uses some of these constraints implicitly or explicitly.

2.6.1 Smoothness

Marr and Poggio [89] stated that matter is cohesive, that is, “it is separated into objects, and the surfaces of objects are generally smooth in the sense that the surface variation due to roughness cracks , or other sharp differences that can be attributed to changes in distance from the viewer, are small compared with the overall distance from the viewer”[16]. The disparity field produced by such surfaces is smooth (varies continuously) everywhere except at object boundaries, which occupy only a small portion of an image. Some researchers [68] [69] [70] [90] [91] use this constraint implicitly as an additional term, to an energy function to be minimized, similar to the variance of disparity or magnitude of the gradient field. Note that such a term tries to make neighboring disparities equal, so it favors frontoparallel surfaces. A true surface smoothness constraint must force neighboring surface normals to be parallel. An example to this kind of smoothness constraint is minimization of the squared magnitude of the second derivative.

Blind use of the smoothness constraint can cause problems at depth discontinuities. Several methods are proposed to avoid smoothing of the disparity field at and near these areas. Weng [81] used the term $\lambda|d(i,j) - \bar{d}(i,j)|^2$ where $\bar{d}(i,j)$ is the average disparity in the neighborhood of the point (i,j), but when calculating $\bar{d}(i,j)$, points on the other side of intensity discontinuities, which correspond to possible discontinuities of disparity, were not used. Another method of handling disparity discontinuities while using the smoothness constraint is using line processes [39] [91] where the smoothness constraint is broken. Sometimes the smoothness constraint is used explicitly where an accepted match is used to guide matches in the neighborhood [61].

A weaker form of the smoothness constraint is the figural continuity constraint [42][62] which was first exploited by Mayhew and Frisby [59]. This constraint implies smooth variation of disparity along edges, because the edgels on the same edge segment are assumed to belong to the same object and this assumption is almost always valid. Mohan et al. [92] use figural continuity for correction of disparity after obtaining matches by any algorithm. Note that the figural continuity constraint is automatically satisfied when contours are used as matching primitives, so the above correction cannot be applied.

Another form of smoothness constraint is fitting models where planar and quadratic patches are fit to the estimated disparity field [93] [94]. Smoothness constraint can also be expressed as a gradient limit on disparity that is known to be used in human stereopsis. Generally, the support from a neighboring match to a potential match is inversely scaled by the disparity gradient between the two matches [95] [96] [61]. Cox et al. [67] do not use the smoothness constraint, but they choose, among several solutions for each scanline, that one which has minimum number of discontinuities. Poggio et al. [97] showed that smoothness assumptions in early visual processing including stereo are related to regularization theory which is a branch of mathematics dealing with ill-posed problems.

2.6.2 Opaqueness

This assumption is violated if there are semi-transparent surfaces in the image, but this is very rare in natural images except objects like fence or bush which occludes background partially. In case of transparency, continuity constraint is not applicable, since the disparity field switches frequently between background and foreground. To handle transparency as well as discontinuities at object boundaries, Prazny introduced the coherence principle which states that the world is made of (either opaque or transparent) objects each occupying a well defined 3D volume. So “a discontinuous disparity may be a superposition of a number of several interlaced continuous disparity fields each corresponding to a piecewise smooth surface” as a result “Two disparities are either similar, in which case they facilitate each other because they possibly contain information about the same surface, or dissimilar in which case they are informationally orthogonal, and should not interact at all because they potentially carry information about different surfaces”[95]. He proposed the support function

$$s(\boldsymbol{i},\boldsymbol{j}) = \frac{1}{c|\boldsymbol{i}-\boldsymbol{j}|\sqrt{2\pi}} e^{-\frac{|d_i-d_j|^2}{2c^2|\boldsymbol{i}-\boldsymbol{j}|^2}} \tag{2.34}$$

where $s(\boldsymbol{i},\boldsymbol{j})$ is the support from the neighboring point $\boldsymbol{j}$ to point $\boldsymbol{i}$. Among possible matches at point $\boldsymbol{j}$ only the one with minimum disparity difference $|d_j - d_i|$ is used in calculation of support. The term $\frac{|d_i-d_j|}{|\boldsymbol{i}-\boldsymbol{j}|}$ on the exponent is the disparity gradient so the support function imposes a disparity gradient limit implicitly. Pollard et al. [98] developed independently a very similar support function which used a disparity gradient limit explicitly. Both algorithms performed well on images involving transparencies and depth discontinuities. Szeliski and Hinton [99] proposed a local function, the difference of heat equations, which leads to a function similar to that of Prazdny when applied iteratively:

$$s(\boldsymbol{i},\boldsymbol{j}) = \frac{1}{\sqrt{(\kappa_1|\boldsymbol{i}-\boldsymbol{j}|)^2 + |d_i - d_j|^2}} - \frac{1}{\sqrt{(\kappa_2|\boldsymbol{i}-\boldsymbol{j}|)^2 + |d_i - d_j|^2}} \tag{2.35}$$

This function can be implemented as a relaxation process, but it lacks the ability to choose the match with minimum disparity gradient.

2.6.3 Orderedness

Assume a point A, and a point B which is right to A match points A' and B' in the other image. Then, this constraint states that B' cannot be at the left side of A'. Resulting disparity constraint violates this assumption if the disparity difference between a figure and its background is larger than the width of the figure in the image. Such objects, like columns, ropes etc. are rare in natural images, so this constraint is frequently used to reduce ambiguity [43] [64] [96] [61] [62] [74]. Human visual system also prefers order-preserving solutions [100].

2.6.4 Uniqueness

This constraint states that a point in one image matches only one point in the other image, that is, the disparity field is a single valued function. In stereo pairs involving only opaque surfaces, this constraint greatly reduces the number of possible solutions. If human visual system uses this constraint or not is a controversial problem since there is evidence for both use of this constraint [100] and for existence of multiple matches [101].

2.6.5 Compatibility

If point A in the right image matches point B in the left, the point B matches point A. Some researchers calculate right and left image disparities independently and than check for compatibility across the field to eliminate false matches. Figure 2.6 shows valid and invalid matches across two lines schematically where circles and arrows represent pixels and matches, respectively.

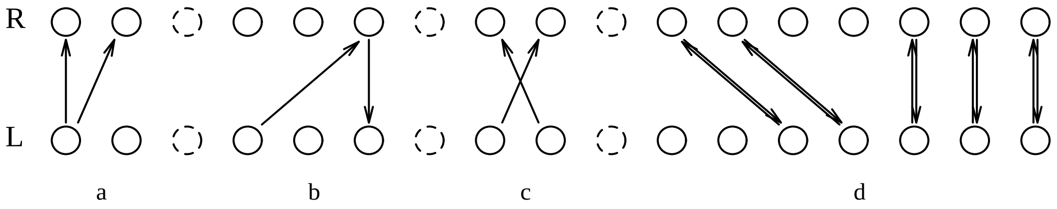

Figure 2.6. Matches between rows R and L violating a) the uniqueness constraint, b) the compatibility constraint and c) the orderness constraint. d) A valid matching field with 2 occluded pixels in the row R.

2.6.6 Epipolarity

Affine transformations are applied to the images such that the epipolar lines are collinear with image rows. The determination of the epipolar line reduces the search space to one-dimension, while the alignment with image rows greatly simplifies the search.

2.6.7 Limited Disparity Range

In accordance with Panum's fusional area, the disparity range in which a match is searched for is determined a priori. Sometimes, even when the epipolarity constraint is used, a small vertical disparity range is allowed to compensate for inexact registration.

2.7 Strategies

Once matching primitives are decided and constraints are set, we face a very large problem. A multi-dimensional space is to be searched for (in some sense) the best solution that satisfies all constraints. Since to visit all states for the best solution is impractical, if not impossible, we need to employ heuristics to reach the best or at least a good solution.

2.7.1 Multi-Channel Analysis

The existence of different band-pass frequency channels in the vertebrate visual cortex led some researchers to use frequency filters in stereo algorithms. Gaussian smoothing and Gabor-like filters are mostly used in band-pass filtering.

As the channel gets coarser (low-frequency), the size of the required masks gets larger, so the computational cost of the filters increases. An equivalent and simpler method is to smooth the image using a Gaussian kernel and to subsample it successively [102]. This way, a Gaussian image pyramid with various resolutions is formed. Usually a spacing of one octave between the channels is used which leads to resolutions of half of the finer channel (i. e., 256x256, 128x128, 64x64). A more rapid way to form the image pyramid is image consolidation that replaces four adjacent pixels with one pixel having the intensity of average of the four pixels [103]. Consider an n by n stereo pair with disparity range m. If integer disparity values are used there are n^2m possible solutions to the problem, while the number of possible solutions in the coarser channel is $(1/8)n^2m$. The accuracy of the result is half of the coarser channel. But we can use this result to constrain the solution in the next finer channel. This strategy is called coarse-to-fine analysis and is very popular in stereo research [16] [77] [42] [16] [93] [38] [103] [66] [40] [69] (See Figure 2.7). Besides the computational savings, this method generally leads more accurate final results. The disadvantage of the method is the spreading of any error in a coarse level to finer levels. Also, this method assumes spectral continuity. Coarse-to-fine analysis can be applied in a continuum of frequency scale, rather than separate channels. This approach is explained in the next subsection. The alternative multi-channel approach to coarse-to-fine analysis is to process each channel independently and to combine subsequently [78] [73].

2.7.2 Relaxation

All photometric and geometric constraints may be expressed in a global cost function, then the minimum of this function is searched using either a deterministic or stochastic relaxation rule in an iterative fashion. Barnard [68] defined an energy function as follows. Let $R_{i,j}$ and $L_{i,j}$ denote intensity of left view and right view pixels, respectively, and $d_{i,j}$ be the disparity at $R_{i,j}$. Then,

$$E = \sum_i \sum_j \left(\|R_{i,j} - L_{i,j+d_{i,j}}\| + \lambda \| \bigtriangledown (d_{i,j}) \| \right) \qquad (2.36)$$

Here the first term is the photometric constraint and the second term is the smoothness constraint. Barnard used simulated annealing (SA) to find a near-minimum solution. SA is a stochastic technique for optimization which takes its flavour from statistical mechanics [104]. A control parameter which is called

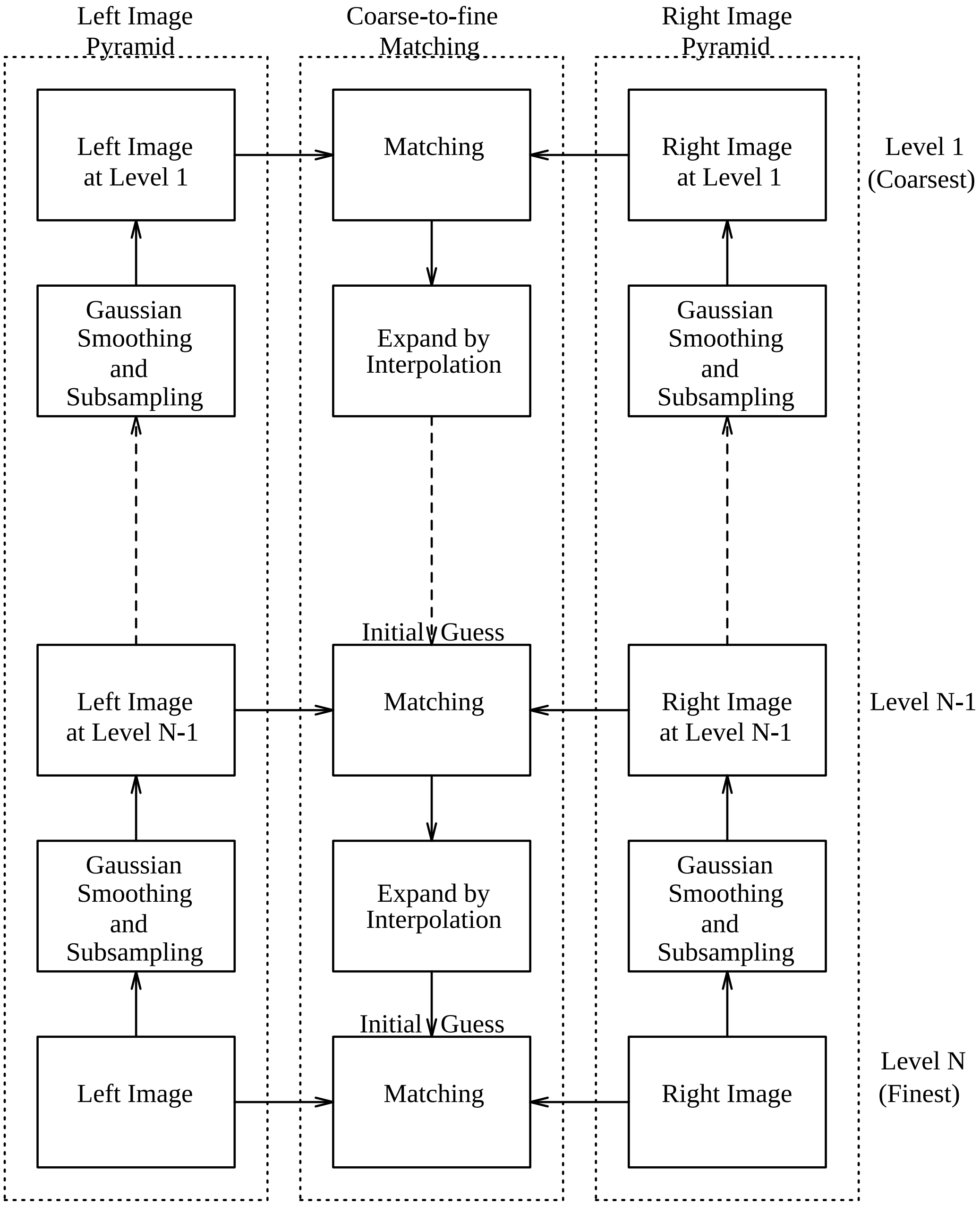

Figure 2.7. Coarse-to-fine control strategy.

temperature is gradually lowered as the network relaxes according to a stochastic update rule. Although finding a near global minimum solution is guarantied, this method is extremely expensive in terms of computation. Jordan and Bovik [69] used SA with a coarse-to-fine strategy and were able to reduce the computation several times but it is still expensive.

Hopfield-like neural networks are also used excessively in finding minimum of stereo energy functions [89] [70] [90] [36]. Marr and Poggio [89] did not give an energy function explicitly, though it is easy to find a Lyapunov function for their network. Instead they give a network structure and an iteration rule for the activities in the neural network as

$$C^{t+1}(x,y;d) = \sigma \left\{ \sum_{x',y';d' \in S(x,y;d)} C^{t}(x',y';d') - \epsilon \sum_{x',y';d' \in O(x,y;d)} C^{t}(x',y';d') + N^{0}(x,y;d) \right\} \tag{2.37}$$

where $\sigma()$ is a threshold function, ϵ is the inhibition constant, $N^0(x,y;d)$ is the correlation of point (x,y) in the right image with the point $(x-d,y)$ in the left image, $S(x,y;d)$ and $O(x,y;d)$ are the excitatory and inhibitory neighborhood which correspond to the smoothness and uniqueness constraints, respectively. Zhou and Chellappa [70] introduced a binary Hopfield network to minimize an energy function which used intensity derivatives for matching and smoothness constraint. Both neural networks performed well on random-dot stereo images, but in both approaches the smoothing term caused problems near depth discontinuities. To avoid these problems, line processes introduced in [105] are used in the stereo energy function:

$$E = \sum_i \sum_j \left(\|(R_{i,j} - L_{i,j+d_{i,j}})(1-l_{i,j})\| + \lambda \| \bigtriangledown (d_{i,j})\|(1-l_{i,j}) + \tau l_{i,j} \right) \tag{2.38}$$

where $l_{i,j}$ is a binary process which indicates a depth discontinuity at pixel (i,j) when it is 1. Photometric constraint as well as smoothing constraint is broken when $l_{i,j}$ is 1, so that occluded regions where a low-cost match cannot be found are also modelled. τ is the cost of a line-process element which is necessary to avoid an excessive number of line processes. Yuille [90] used a Hopfield-style analog network to minimize the above function. Although the neural network was successful for similar formulations of surface interpolation and motion analysis problems, it yielded poor results for stereo case. Yuille states that this bad behaviour of the network is due to the complicated structure of the energy function that has lots of local minima.

The discontinuities of depth generally correspond to intensity and texture edges. Gamble and Poggio use intensity edges to break smoothness [106]. Similarly, Toborg [91] obtained successful results using a Hopfield Network when he formulated the stereo energy function together with motion analysis and edge detection. Binary Hopfield network is also used for feature-based correspondence [36]. The interest points obtained from natural images by the Moravec operator are matched under the smoothness and uniqueness constraints. The network matches most of the points correctly after about 1000 iterations. Later Joshi and Lee [107] applied elastic nets, that were first introduced by Durbin and Willshaw [108] as a mechanism to establish ordered neural mappings, to the same problem. Their energy function is

$$E = -K \sum_i \ln \sum_j \phi(p'_{i,j}, K) \tag{2.39}$$

where

$$\Phi(p, K) = \exp \frac{-p^2}{2K^2}, \tag{2.40}$$

$$p'_{i,j} = \sqrt{\gamma x^2 + y^2},\ 0 \leq \gamma \leq 1, \tag{2.41}$$

$$x = l_{j,x} + d_{j,x} - r_{i,x}, \tag{2.42}$$

$$y = l_{j,y} + d_{j,y} - r_{i,y}, \tag{2.43}$$

and $r_{i,x}$ and $r_{i,y}$ are horizontal and vertical coordinates of i'th right point, respectively. $l_{j,x}$ and $l_{j,y}$ are similarly defined for left points and $d_{j,x}$ and $d_{j,y}$ are horizontal and vertical disparities of r_i, respectively. The parameter K is like the temperature in simulated annealing and lowered gradually but a deterministic update rule is used: the steepest gradient in the energy function is followed but the shape of the energy function is changed as K changes. Later, Leloğlu [109] used elastic nets in intensity-based dense stereo correspondence with smoothness constraint:

$$\begin{aligned} E \quad = \quad & -\alpha K \sum_i \sum_j \ln \sum_n \Phi(n - d_{i,j}, K) \| R_{i,j} - L_{i,j+n} \| \\ -0.25 \quad & ((d_{i,j} - d_{i,j-1}) + (d_{i,j} - d_{i-1,j})) \end{aligned} \tag{2.44}$$

Note that the energy function reduces to that of Barnard when the control parameter K goes to zero (see Equation 2.36). Another form of "deterministic annealing" is mean field annealing (MFA) which is applied to stereo correspondence by several researchers [39] [41]. MFA is similar to simulated annealing but is typically 50 times faster [41]. In simulated annealing, MFA and elastic nets, the control parameter can be related to coarse-to-fine analysis, because when the control parameter is large, fine details are smoothed out or ignored and as the control parameter gets smaller, finer details are taken into account. Two other similar approaches are graduated non-convexity algorithm of Blake and Zisserman [110] and scale space signal matching of Witkin et al. [111]. In graduated non-convexity algorithm, convex approximations to the true energy function are generated with decreasing levels of smoothness. In scale space tracking, the minimum of the energy function is followed as the scale (see Eqn. 2.3) is decreased. Later Whitten [112] applied scale space tracking to deformable sheets.

Another relaxation type is the spreading of constraints. The best and most unambiguous matches are accepted first and they constrain more ambiguous ones. Lloyd [43] first matches strong edges, then matches others, because a weak edge may be missing in one of the images. Sherman and Peleg [61] pair best-matching contours first, then constrain the neighbors of these matches through a support function. Kim and Bovik [34] match high-interest points and propagate those disparities along contours.

An interesting neural network approach to stereopsis is the work of Khotanzad et al. [113]. They train a feed-forward neural network with back-propagation learning rule imposing only the epipolarity constraint. The uniqueness and continuity constraints are automatically learned and coded by the neural network and the net outperformed the Marr-Poggio network.

2.7.3 Dynamic Programming

Dynamic programming is a search technique which always finds the minimum-cost path without visiting all nodes of a tree (See [114]). The stereo problem can be translated into a dynamic programming problem as follows: Consider two corresponding scanlines from two images under the epipolarity constraint. The points where almost vertical edges cross these rows are taken as matching primitives and all possible matches form the nodes. Using the orderedness constraint and a cost function based on the intervals between edgels, that is the cost of passing from one node to another, the minimum-cost path is found [43].

When the matching is performed independently on each row, a very valuable source of information, connectivity of edges, is squandered. Baker [115] uses this information in a cooperative procedure to detect false matches after processing each line independently while Ohta and Kanade [64] use it by forming nodes as vertically connected edges in a 3-dimensional search space. Cox et al. [67] use intensity as the matching primitive so they find the minimum-cost path through a 2-dimensional grid formed by the absolute differences of two scanlines. Boyer et al. [62] and Matthies [72] also use dynamic programming in stereo correspondence problem.

2.7.4 Hybrid Approaches

Feature-based approaches yield accurate but sparse solutions while area-based approaches give dense results at the expense of computational complexity. The idea of hybrid approaches is to use both methods together to compensate each other's weaknesses. In the work of Lim and Prager [41], the results of an edge-based stereo algorithm are used to improve the minimum energy solution obtained by mean field annealing of an area-based energy function similar to

$$E = \sum_i \sum_j \left(F(d_{i,j}) + \alpha \sum_{n \in N} (d_{i,j} - d_n) \right) \tag{2.45}$$

where $F(d_{i,j})$ is the matching cost based on intensities and N is a neighborhood of (i,j). That energy function is modified as follows at points where a disparity value d_e is available from the edge-based algorithm:

$$E = \sum_i \sum_j \left(F(d_{i,j}) + F(d_{i,j})\psi|d_{i,j} - d_e| + 2\alpha \sum_{n \in N} (d_{i,j} - d_n) \right) \tag{2.46}$$

The results obtained by the help of edge-based results are qualitatively better. Yuille [39] use feature-based and intensity-based terms in one energy function. Watanabe and Ohta [116] implemented a stereo vision system where three parallel modules are integrated through cooperative processing (See Figure 2.8). Point-based matching module uses maximum correlation value as the best match. Interval-based module is the same as the intra-scanline algorithm of Ohta and Kanade [64]. The segment-based module uses the length and orientation of edge segments and intensity values in their neighborhood. A module gives the results of its matching upon request of another module with confidence values of

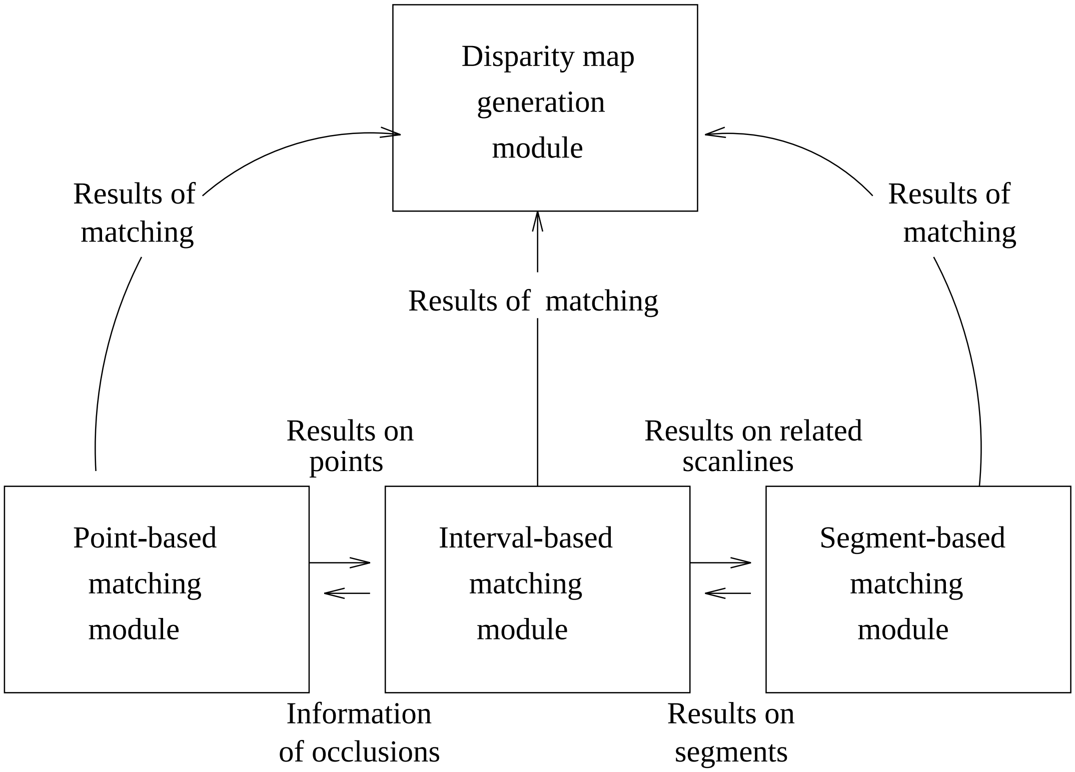
Disparity map generation module
Results of matching
Results of matching
Results of matching
Results on points
Results on related scanlines
Point-based matching module
Interval-based matching module
Segment-based matching module
Information of occlusions
Results on segments

Figure 2.8. Cooperative integration of three stereo modules

the matching. They also present quantitative results of the improvement when compared to the results obtained by operation of each module alone.

2.7.5 Using Other Sources of Information

It is well known that human visual perception owns its power to integration of information from a variety of sources such as motion, shading etc. Computer vision maturing in each of such methods now is in the way of building more comprehensive vision systems integrating those modules.

Fusing motion and stereo was considered by a number of researchers [117] [118] [119]. If we know the disparity field or optical flow for a sequence of stereo images, it is easier to compute the other one. Besides, the discontinuities of optical flow are generally also depth discontinuities. So, in general, one of them is computed first and is used to guide the other. On the other hand, Toborg and Hwang [91] calculated stereo disparity, optical flow and intensity contours simultaneously and cooperatively. They demonstrated the effectiveness of integrating visual modules on synthetic images.

Other visual cues used with stereopsis include shape-from-shading [120]

[121] [122] [123], shape-from-texture [124] and sonar [125]. Also, active systems which seek for useful additional information by controlling camera parameters are used more and more frequently [126] [127] [128] [129] [10].

Another useful source of information is additional images. See [2] for a review of trinocular stereo and [130] for a recent work where several images are used.

CHAPTER III

THE STEREO SYSTEM

3.1 An Overview of the System

The stereo system developed in this thesis employs coarse-to-fine strategy which is shown schematically in Figure 2.7. This hierarchical structure brings two advantages: increased accuracy and faster processing. The block diagram of the system at only one level of the hierarchy is depicted in Figure 3.1. First, the stereo image pair is Gaussian smoothed and subsampled to halve the resolution until reaching the coarsest level. After formation of the image pyramid, multi-resolution disparity analysis begins at the first, that is, the coarsest level. Each level, except the coarsest one, uses the disparity field supplied by the coarser channel as initial guess, so, each level, except the finest one, supplies its output to the finer level. The coarse-to-fine strategy carries the risk of spreading any error in a coarse level to higher levels. To avoid such problems, system tries not to make an early judgement: a disparity value is accepted only in existence of convincing evidence. Otherwise, determining the disparity is left as a job for higher levels.

In each level, a pixel-wise matching module, which uses normalized cross-correlations as matching primitives, is used first. In this module the goodness of a match is determined by using the value of the correlation and using a neighborhood support function. Only unambiguous matches are accepted and these accepted matches constrain their neighbors through uniqueness and orderedness constraints. The pixels at which the ambiguity is not resolved after several iterations are left unknown.

In the subsequent module, a sub-pixel disparity surface reconstruction is attempted with detection of discontinuities and with interpolation of disparity at unknown pixels. Here, sub-pixel means that disparity can have non-integer

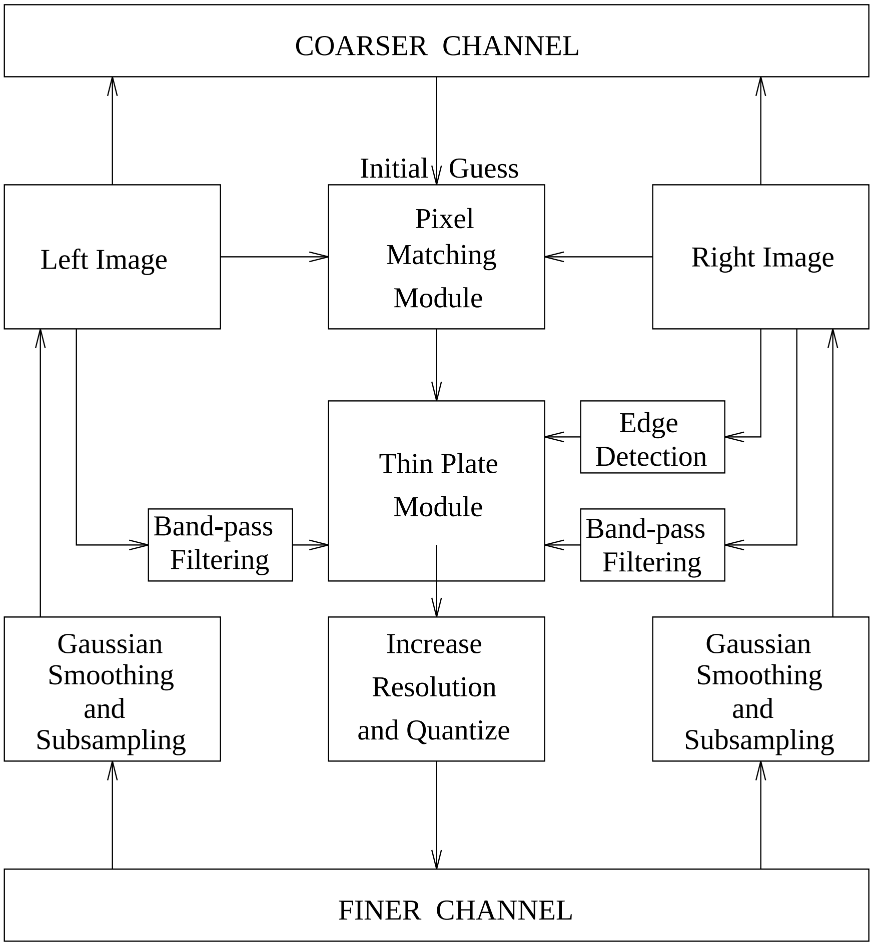

Figure 3.1. Block Diagram of the Stereo System at one Level

values; the disparities are still estimated only at pixel positions. The detection of depth discontinuities is guided by the intensity edges so an edge detection module is included. For subpixel matching, phases of band-pass filtered images are used as matching primitives, so a WFP filter module is employed. Very large unknown regions are not interpolated in the thin plate module and left unknown.

In the following subsections each block is explained in detail.

3.2 Forming the Gaussian Image Pyramid

The Gaussian smoothing before subsampling is necessary to avoid aliasing. The 1-D Gaussian function

$$g(x, \sigma_x) = \frac{1}{2\pi\sigma_x^2} e^{-\frac{x^2}{2\sigma_x^2}} \tag{3.1}$$

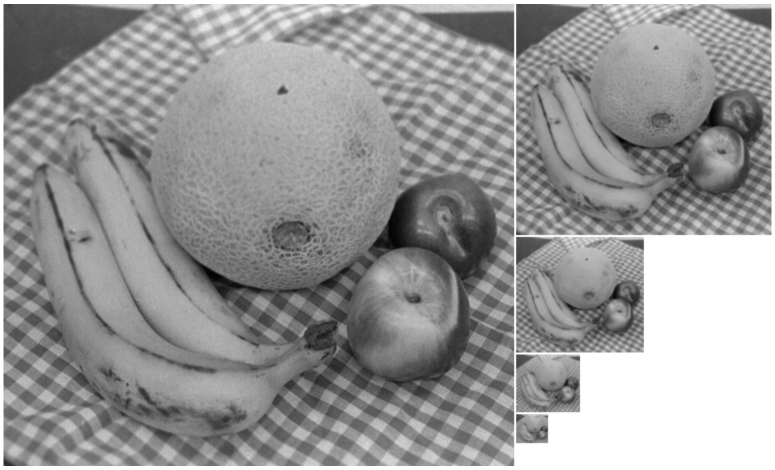

Figure 3.2. Image Pyramid of Right Image of Fruits Stereo Pair

has the Fourier transform

$$G(u, \sigma_x) = \frac{1}{2\pi\sigma_u^2} e^{-\frac{u^2}{2\sigma_u^2}} \tag{3.2}$$

where $\sigma_u = 1/2\pi\sigma_x$ and u is in *cycles/pixel*. We want to remove frequencies above $0.5\,cycles/pixel$ and above so we set $\sigma_u = 0.25$, that is, the frequency response is 13.5% of its maximum at $u = 0.5\,cycles/pixel$ or equivalently $\sigma_x = 2/\pi$. The 2-D Gaussian is separable to two 1-D Gaussians, so we convolved the image with the kernel

$$[\ 0.184 \quad 0.632 \quad 0.184\ ] \tag{3.3}$$

vertically and horizontally before subsampling. An image pyramid formed by this process on right image of the Fruits stereo pair (See subsection 3.9.1) is shown in Figure 3.2. For a hardware implementation of the stereo algorithm, VLSI chips which construct the image pyramid are available [131].

3.3 Pixel Matching

The pixel matching stage uses normalized cross-correlations for matching which is defined as

$$C(i, j, d) = \frac{\left(\sum_{x=-2}^{2}\sum_{y=-2}^{2} R(i+d+x, j+y)\; L(i+x, j+y)\right)^2}{\sum_{x=-2}^{2}\sum_{y=-2}^{2} R^2(i+d+x, j+y) \sum_{x=-2}^{2}\sum_{y=-2}^{2} L^2(i+d+x, j+y)}. \tag{3.4}$$

The correlation size is chosen as small as 5x5 to reduce computational complexity and to be able to match areas with large disparity gradient. Time required for

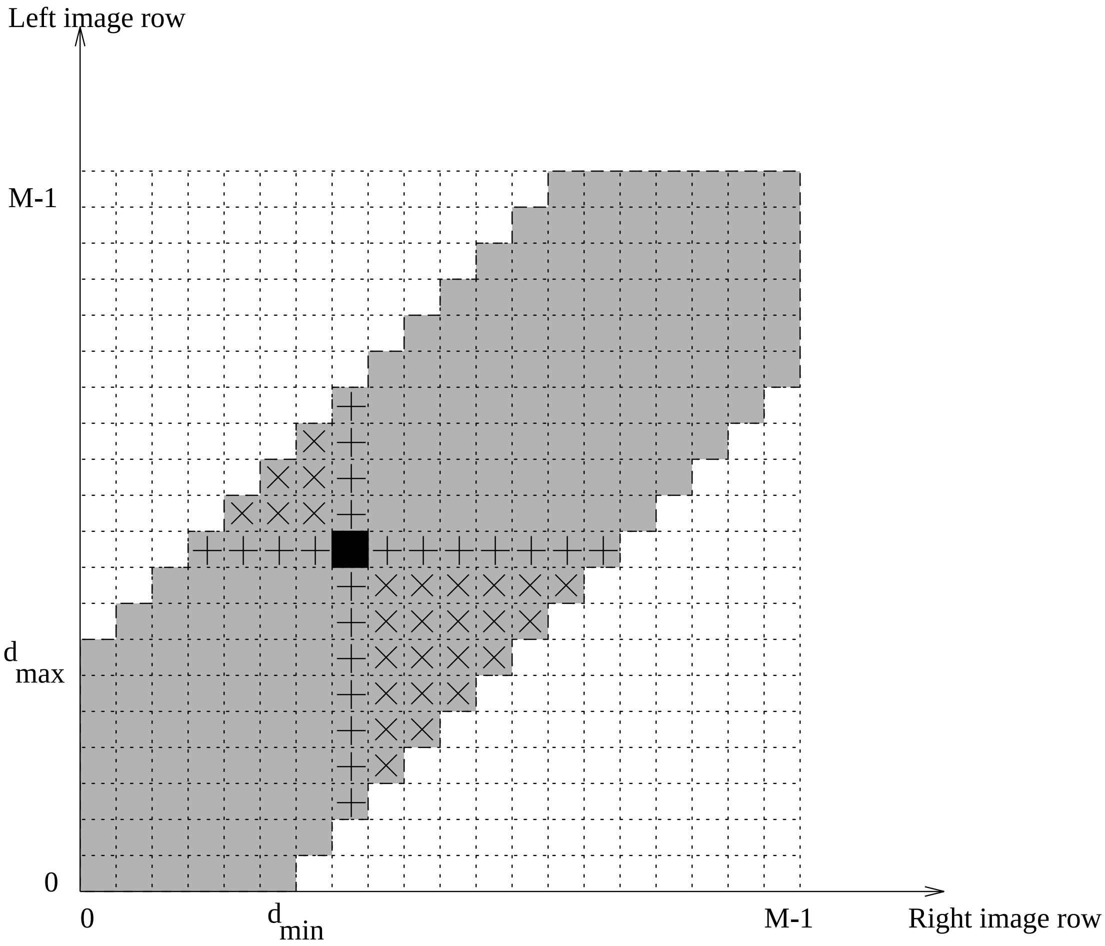

Figure 3.3. The matches forbidden by an accepted match due to uniqueness and orderness constraints

calculating correlations can be reduced dramatically by employing a VLSI chip designed for this purpose [132].

To spread the constraints, a mask, an MxNxD size binary array where M and N are width and height of each image respectively and D is the maximum allowed range, is used. Each entry of this array corresponds to a match and when it is set to 1 indicates that this match is forbidden. The shaded area in Figure 3.3 is the set of possible matches among the pixels of two lines from the right and left images which are limited by user-supplied maximum and minimum disparity values. When a match is accepted (the black square), matches corresponding to "+"s and "x" are marked as forbidden matches in the mask, since they violate the uniqueness constraint and order reversal constraint, respectively.

When a match is supplied from a coarser channel, things go a little bit different. Since the initial guesses are supplied by the thin plate module, one may

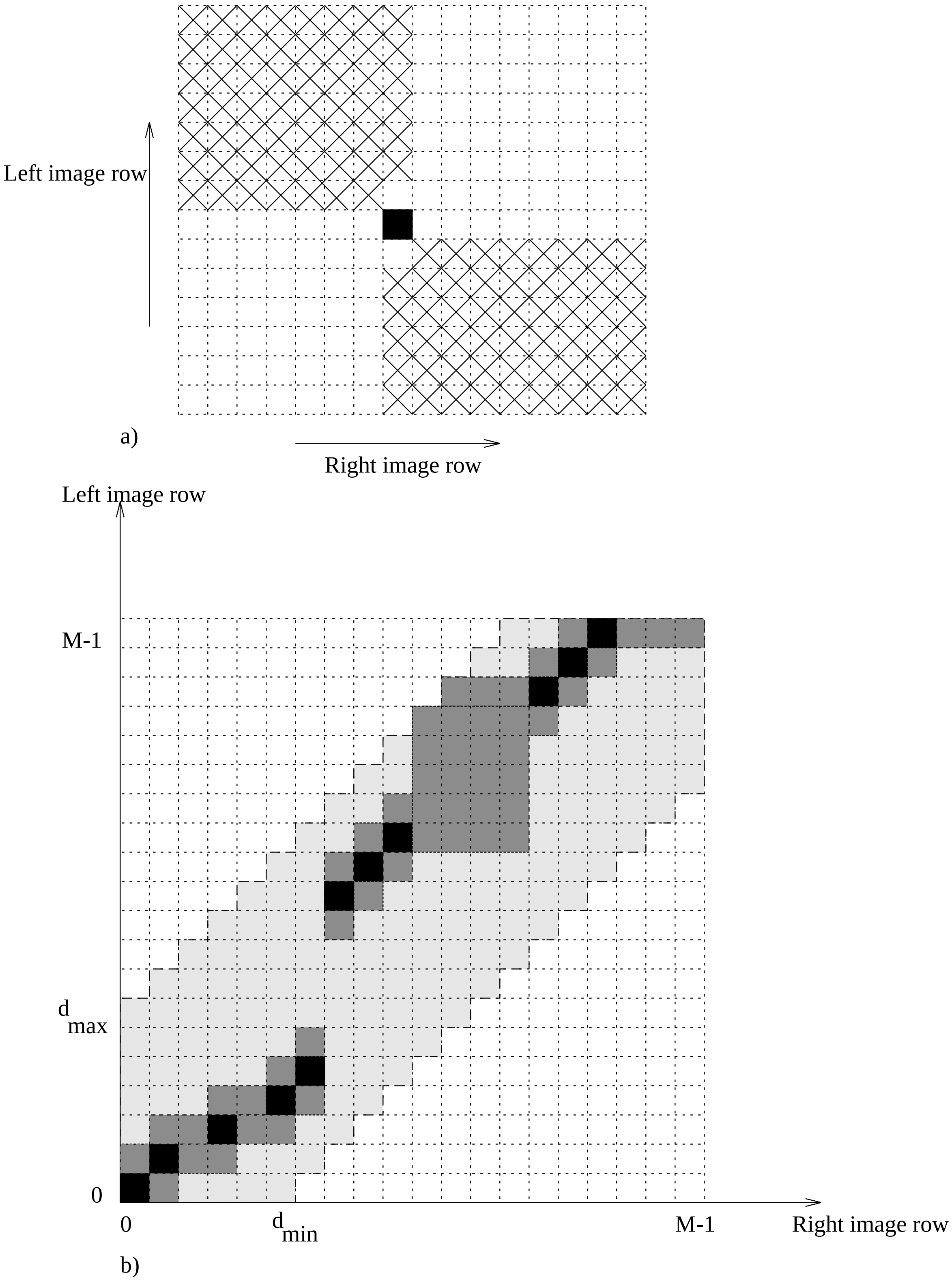

Figure 3.4. a) The matches forbidden by a match supplied by the coarser channel, b) the matches forbidden by a set of matches supplied by the coarser channel.

be tempted by the idea that these matches are exact. However, this is not always true because of several reasons. Firstly, depth discontinuities that are small in the coarser level are sometimes not detected and the disparity field is smoothed. Secondly, any error in disparity is doubled due to the increase in resolution. Finally, false matches may distort the thin plate. To anticipate for ambiguity in initial guesses we allow ± 1 pixel disparity around these matches which results in the forbidden area pattern shown in Figure 3.4 a). The shaded area in Figure 3.4 b) shows allowed matches when the matches indicated as black are supplied from the coarser channel. The correlations corresponding to matches forbidden by initial guesses are never calculated, that is one of the reasons why coarse-to-fine strategy reduces computational complexity. At any level, most unambiguous matches are accepted first. These matches constrain their neighbors and some of the ambiguous potential matches become less ambiguous. These matches are accepted and they further constrain others. The algorithm is as follows:

1. Let support threshold, T_s and correlation threshold, T_c take large values.

2. Find maximum correlation disparity at every point of the right image among unmasked correlations above the threshold T_c.

3. Do above for the left image. Those disparities which are not consistent in both sets are discarded (Agreement between views). Remaining matches are written in a temporary array.

4. Calculate support for each candidate disparity and discard those having support belove T_s.

5. Check for any violation of uniqueness and order reversal constraints among candidate matches. In case of violation, discard both matches.

6. Accept remaining matches in the temporary array as true and mark corresponding forbidden areas in the mask.

7. Decrease T_c and T_s according to a predetermined schedule and go to Step 2.

3.3.1 Calculation of Support

In acceptance of a match, the support it collects from its neighborhood is a very important metric. If a neighbor of the pixel is an accepted match, it delivers

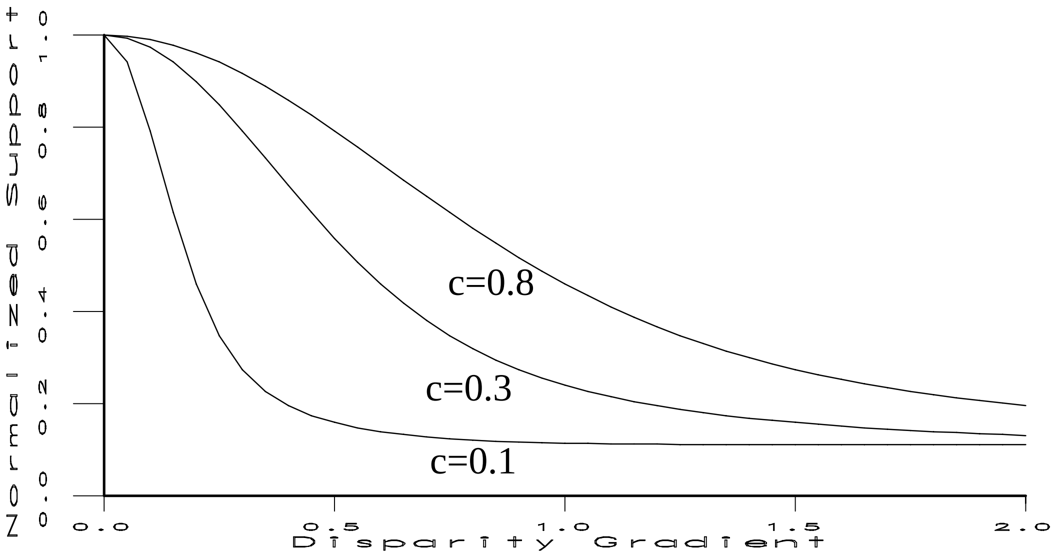

Figure 3.5. Support of a plane as a function of disparity gradient.

a certain support, else the candidate disparity in the temporary array is used. From neighboring cites where both fail, no support is provided. Intuitively, we expect that as the neighbor gets farther it should deliver less support since, the probability that it belongs to a surface different in depth, increases. Similarly, those points that are close in disparity should supply more support. A function which satisfies both constraints is the support function of Prazdny (See Eqn. 2.34). Ignoring the normalisation constant, the support a disparity $d(i,j)$ collects is

$$S(i,j) = \sum_{x=-K}^{K} \sum_{y=-K}^{K} \lambda_{i+x,j+y} e^{\frac{(d_{i,j}-d_{i+x,j+y})^2}{2c^2(x^2+y^2)}} \tag{3.5}$$

where $\lambda_{i',j'}$ is 1 if there is an accepted or candidate disparity value available at point (i',j') and 0 otherwise. In our stereo system, $M = 4$ and $c = 0.3$ values are used. In Figure 3.5 the support collected from a plane of disparities passing through (i,j) is plotted against the disparity gradient for various values of c. The choice of $c = 0.3$ roughly corresponds to a gradient limit of 0.5. The computational complexity of the support function is greatly reduced by employing appropriate look-up tables.

3.4 Edge Detection

For detecting intensity edges Canny operator with hysteresis threshold is used which is available as a part of HVision Image Processing Package from

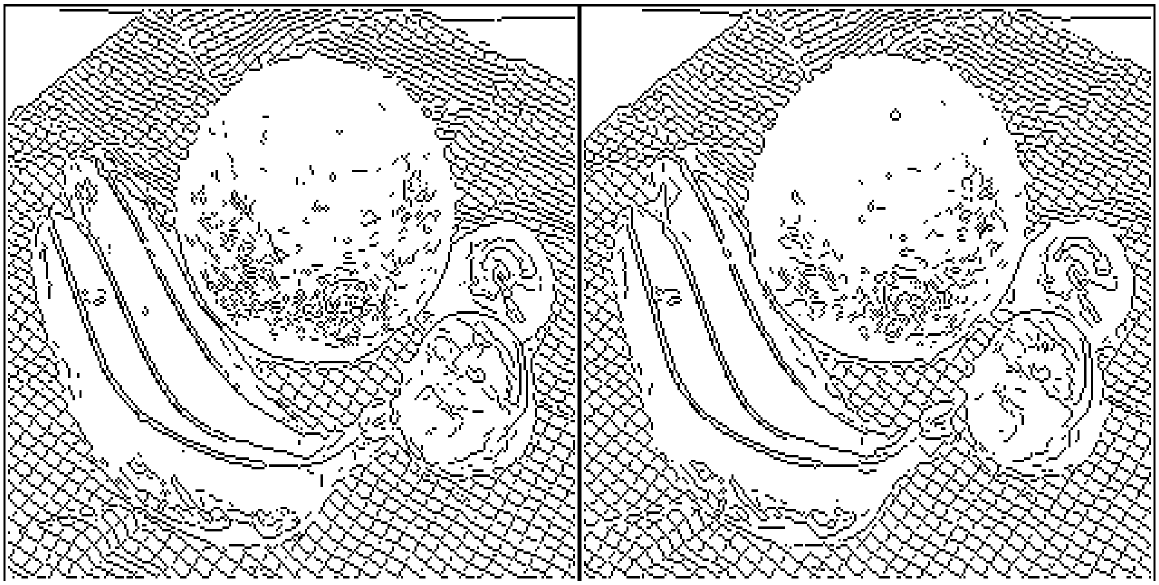

Figure 3.6. Canny edges of fruits stereo pair

Harvard Robotics Laboratory. Since the direction of edges is not supplied by the program, they are calculated separately as $\theta = \arctan(I_H/I_V)$ at edgel positions where I_H and I_V are horizontal and vertical gradients obtained by convolving the image with the kernels

$$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad and \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}, \tag{3.6}$$

respectively. Edges detected by the Canny operator on fruits stereo pair (232x256) with the default parameters of the system is shown in Figure 3.6.

3.5 Band-Pass Filtering

Westelius [80] lists the requirements for a band-pass filter to be used for phase matching, as follows:

- The filter must not be sensitive to DC.
- The impulse response of the filter must span a phase range of $[-\pi, \pi]$ without any wrap around.
- The phase must increase monotonically.

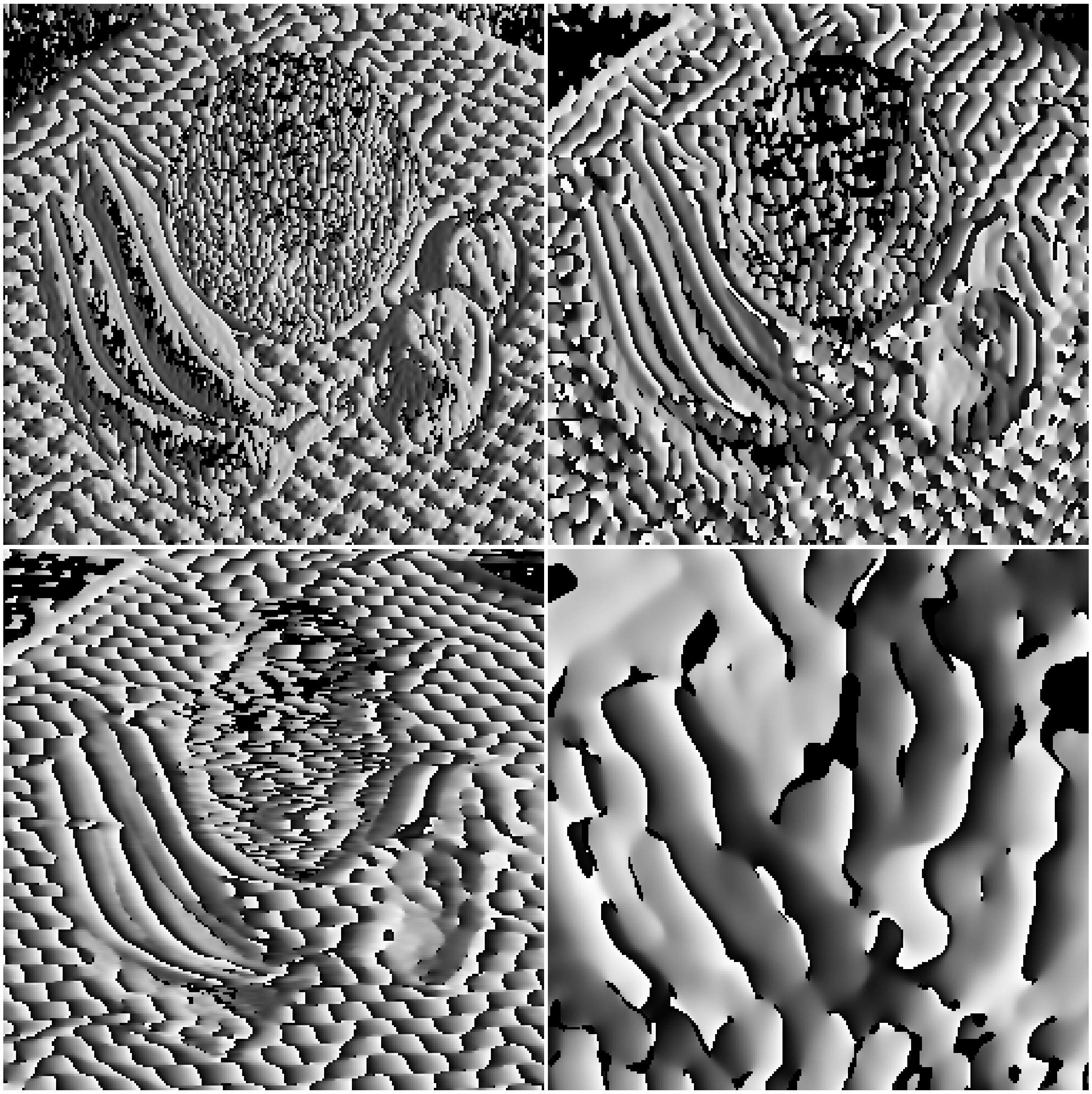

Figure 3.7. Right part of fruits stereo pair filtered with WFP filters of period 4, 8, 16, and 32 pixels.

- The filter should allow only positive frequencies. This is necessary to obtain a monotonically increasing phase.
- The singular points in the output of the filter must cover as small an area as possible.
- The size of the filter must be as small as possible to keep the computational cost small.

Among many possible complex band-pass filters, WFP is preferred for the system. M (see Eqn. 2.21) is chosen as $2\pi/\omega_0$ so that the window covers one period of

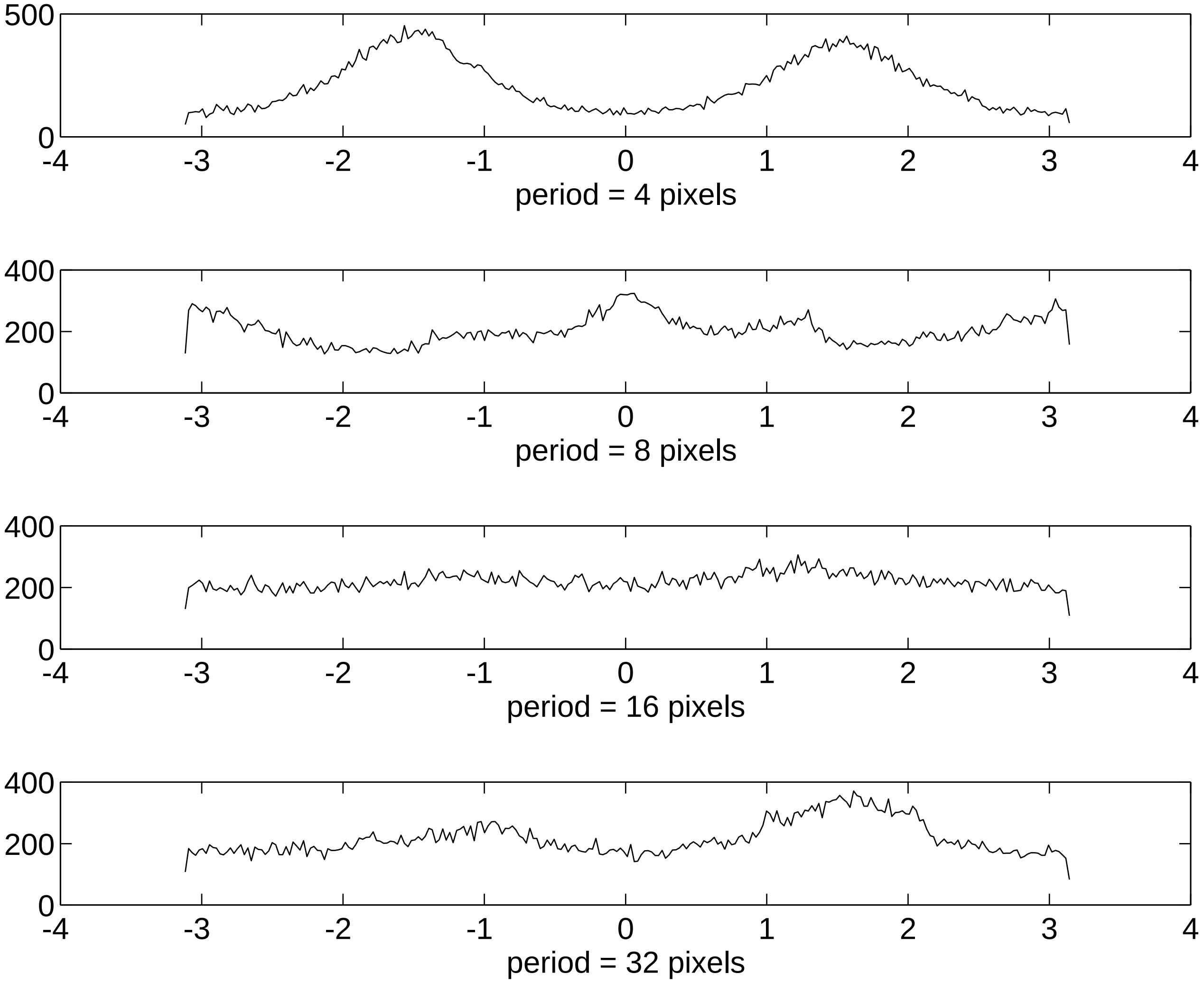

Figure 3.8. Histograms of phase images for various filter sizes.

$e^{j\omega_o x}$. The filter has no DC but has some response in negative frequencies. On the other hand, the spatial support of the filter is small and this property is very desirable for stereo matching. Figure 3.7 shows right part of fruits pair filtered with WFP filters of period 4, 8 , 16 and 32 pixels. Figure 3.8 shows the histogram of the phase images where it is clearly seen that the filter has some bias towards $\pm\pi/2$.

The output of band-pass filters can be used for other purposes as well, like edge detection [133], texture analysis [134] and motion analysis [135], in a more comprehensive vision system.

3.6 Thin Plate Module

The functions of this module can be classified as follows:

- To interpolate for regions where the former module could not find any match: sometimes the information in higher frequencies may resolve ambiguities.

- To locate depth discontinuities accurately and explicitly: A binary line process indicating depth discontinuities is determined. Since this process is guided by intensity edges, which likely correspond to the location of depth discontinuities, the accuracy is increased.

- To supply more accurate initial guesses to the upper level and to have subpixel accuracy at the output of finest level.

In pixel-matching module, any accepted match is meaningful in both images: it corresponds to a certain pixel in each image. But since we use non-integer values in thin plate module we need to create a dominant eye. The right image disparities supplied from the pixel matching module is chosen to fit the sheet.

The best disparity field $d(i,j)$ and line processes $l(i,j)$ are found as the minimum of the following energy function:

$$\begin{aligned} E \;=\; & (\alpha/4)\,\{(1-l_{i,j+1})\,(d_{i,j}-d_{i,j+1})(1-l_{i,j-1}) + (d_{i,j}-d_{i,j-1}) \\ & \quad + (1-l_{i+1,j})\,(d_{i,j}-(d_{i+1,j}) + (1-l_{i-1,j})\,(d_{i,j}-d_{i-1,j}))\}^2 \\ & - \beta \sum_{k=0}^{C-1} cos(\phi^{r,k}_{i,j} - \phi^{l,k}_{i+d_{i,j},j}) \\ & + \gamma\, h_{i,j}\, \{max(0, |d_{i,j}-o_{i,j}| - R)\}^2 \\ & + \delta\, l_{i,j} \\ & + \varepsilon\, e_{i,j}\,(1-l_{i,j})\,|\nabla d_{i,j} \cdot \vec{n}|. \end{aligned} \tag{3.7}$$

The constants α, β, γ, δ and ε reflect the relative importance of various terms of the energy function and their optimum values are determined experimentally. Ignoring the binary field $l(i,j)$ which indicate the existence of a discontinuity or an occlusion at position (i,j), the first term reduces to

$$\alpha\,\{(d_{i,j}-(d_{i,j+1}+d_{i,j-1})/2) + (d_{i,j}-(d_{i+1,j}+d_{i-1,j})/2)\}^2 \tag{3.8}$$

which is nothing but the squared magnitude of sum of the second derivatives of the disparity field $d_{i,j}$ at two orthogonal directions. This term imposes smoothness on the reconstructed surface, but the line processes help circumvent the subversive effects of smoothing near discontinuities of disparity field.

The second term penalizes the phase differences between the matched images which are expected to be very small for perfectly matched images. Here, $\phi^{x,k}$ denotes the phase of band-pass filtered image x in the k'th channel. The subsection following this one is devoted to explanation of this term due to its importance.

The third term tries to minimize the difference between the disparity field and the disparity values supplied by the pixel matching module, $o_{i,j}$'s. $h_{i,j}$ is unity if such a value is available at point (i, j). Since $o_{i,j}$'s are quantized to integer values, a range R above and below these values are not penalized.

The last two terms are included to detect discontinuities. A discontinuity $l(i, j)$ is assigned unit cost δ to prevent an excessive number of discontinuities. The last term is the cost for disparity gradient $\nabla d_{i,j}$ which is calculated using kernels defined in Eqn. 3.6 and contributes the energy function only when there is an intensity edge at (i, j), that is, $e_{i,j}$ is 1. $\vec{n}$ is the unit normal vector to the edge; the dot product favors discontinuities when the directions of the intensity edge and that of the disparity gradient are similar.

To find the minimum of this function we perform gradient descent. Initially, $d(i, j)$'s are assigned available $o_{i,j}$ values. After every several iterations some $d(i, j)$'s are interpolated where $o_{i,j}$'s are not available. Actually, there are some flags which indicate if there are any $d(i, j)$ value and any $\phi^{r,l}_{i,j}$ values at (i, j), but they are not shown in the energy function for the sake of simplicity and clarity.

The equations governing the gradient descent are:

$$\begin{aligned}
\frac{\partial E}{\partial d_{i,j}} = \; & \alpha\,((1 - l_{i,j+1})d_{i,j+1} + (1 - l_{i,j-1})d_{i,j-1} \\
& \quad +(1 - l_{i+1,j})d_{i+1,j} + (1 - l_{i-1,j})d_{i-1,j} \\
& \quad -(4 - l_{i,j+1} - l_{i,j-1}1 - l_{i+1,j} - l_{i-1,j})\, d_{i,j}) \\
& - \beta \sum_{k=0}^{C-1} sin(\phi^{r,k}_{i,j} - \phi^{l,k}_{i+d_{i,j},j})(\phi^{l,k}_{i+d_{i,j}+1,j} - \phi^{l,k}_{i+d_{i,j},j}) \\
& + \gamma\, h_{i,j}\, 2\, max(0, |d_{i,j} - o_{i,j}| - R)\, sgn(d_{i,j} - o_{i,j}) \qquad (3.9)
\end{aligned}$$

and

$$\frac{\partial E}{\partial l_{i,j}} = \delta - \varepsilon\, e_{i,j} \left| \nabla d_{i,j} \cdot \vec{n} \right| . \tag{3.10}$$

In the second equation, the term resulting from the first term of Eqn. 3.7 is ignored. Since $l_{i,j}$ is binary, it is set to 1 if Eqn. 3.10 is positive.

It is well known that gradient descent is not successful on non-convex energy surfaces, especially on very complicated ones like that defined in Eqn. 3.7, when applied directly. However, the initial value supplied, that is, the output of the pixel-matching module, is very close to the global minimum of the function. As a result, it is very unlikely that the gradient descent sticks into local minima. By employing the pixel matching module, the necessity of using computationally expensive methods, like simulated annealing, mean field annealing or graduated non-convexity, is avoided.

3.6.1 The Phase Term

It is easy to show that the correlation of two signals $l[n]$ and $r[n]$,

$$C_{rl}[x] = E\left\{ l[n]\, r^{*}[n-x] \right\} \tag{3.11}$$

can be written in terms of the discrete Fourier transforms of the two signals, $L[k]$ and $R[k]$, as

$$C_{rl}[x] = \mathcal{F}^{-1}\left\{ L[k]\, R^{*}[k] \right\} \tag{3.12}$$

where

$$L[k] = \frac{1}{N} \sum_{n=0}^{N-1} l[n] e^{-jk(2\pi/N)n} \tag{3.13}$$

and

$$R[k] = \frac{1}{N} \sum_{n=0}^{N-1} r[n] e^{-jk(2\pi/N)n}. \tag{3.14}$$

Note that $L[k]$ and $R[k]$ can be interpreted as WFP filtered versions of $l[n]$ and $r[n]$, respectively. If the correlation is replaced with phase correlation, we obtain

$$P_{rl}[x] = \mathcal{F}^{-1}\left\{ \frac{L[k] R^{*}[k]}{|L[k]|\, |R^{*}[k]|} \right\} \tag{3.15}$$

As a measure of the similarity of two signals $l[n]$ and $r[n]$, we need the quantity

$$Real\left\{ P_{rl}[0] \right\} = \sum_{k=0}^{N-1} cos(\phi^{l,k} - \phi^{r,k}) \tag{3.16}$$

where $\phi^{l,k}$ and $\phi^{r,k}$ are the phases of $L[k]$ and $R[k]$, respectively. The right-hand side of this equation is very similar to the phase term of our energy function with one major difference. The "channels" have constant absolute bandwidth, so, the relative bandwidth decreases as the frequency gets higher. This has two disadvantages. Firstly, if there exists even a small disparity gradient, the frequencies of signals are scaled differently and, as a result, the phase differences corresponding to different frequencies of the signals are considered. Secondly, since the amplitude spectrum of natural images generally decays with $1/f$ [24], the low-frequency channels carry much energy when compared to high-frequency channels. To overcome these disadvantages we use constant relative bandwidth channels instead of constant absolute bandwidth channels.

3.6.2 Interpolation

During the stage of interpolation, unknown disparities whose at least 4 neighbors out of 8 are known (matched by the pixel matching module or interpolated before) are assigned the average of the known neighboring disparities. The "at least 4 out of 8" rule lets only convex areas to be interpolated. Note that this stage is not a blind interpolation since interpolated disparities are later fine tuned by the membrane model using the smoothness constraint and phase values.

3.7 Detection of Occlusions

Due to the difficulty in formulating occlusions in the energy function, they are detected independently. If there exists two disparity values $d(i,j)$ and $d(i,j+k)$ such that

$$d(i,j) - d(i,j+k) \geq k-1, \quad k>1, \tag{3.17}$$

then $l(i,j+1),\ldots,l(i,j+k-1)$ are set to 1 due to occlusion.

3.8 Increasing the Resolution of Disparity

Before supplying the output of the thin plate module to the pixel-matching module of the higher channel, the number of pixels must be doubled in both directions. The disparities at inserted pixels are obtained by averaging. If any neighbor of such a pixel is unknown or is a discontinuity, that point is also labelled as unknown. Finally, all the disparity values are multiplied by 2 and quantized to integer values. If the resulting integer disparity field contains pair of disparity values that violate uniqueness or orderness constraints, both disparities are discarded.

3.9 Experimental Results

In this section, the results obtained by the system on several stereo image pairs are presented. Images from different domains are chosen so that powerful and weak sides of the algorithm are evinced.

3.9.1 Fruits Stereo Pair

This 232x256 sized monochrome image pair, which originates from Dr. W. Hoff, University of Illinois, contains many of the difficulties of stereo vision: repetitive texture (the table cloth), textureless areas (surface of the melon), physical difference across images (the fly on the melon in only the right image) and large fusion interval (-13 to 11). Figure 3.9 shows the outputs of both modules at 3 levels of the hierarchy. It is clearly seen that the resulting disparity field is satisfactory except small irregularities of depth discontinuities, false matches at upper part of the mellon and small false match patches at upper corners of the image which are all due to lack of texture. Figure 3.12 and Figure 3.11 are the occlusions and depth discontinuities detected, and 3D-rendering of the final result.

In Figure 3.10 the results obtained by a one-level algorithm is presented for demonstrating the effectiveness of coarse-to-fine analysis. The repetetive texture causes large false matching areas at upper right quarter of the image in one-level algorithm. Besides, a comparison of Table 3.1 and Table 3.2 shows the computational savings obtained by employing a coarse-to-fine strategy.

3.9.2 Pentagon Stereo Pair

This 512x512 sized pair (See Figure 3.13) originates from Prof. Takeo Kanade of Carnegie-Mellon University. This pair involves mostly frontoparallel surfaces within a small fusion interval (-8 to 10; note that the size of the images is twice that of Fruits pair), nevertheless accurate detection of discontinuities is difficult.

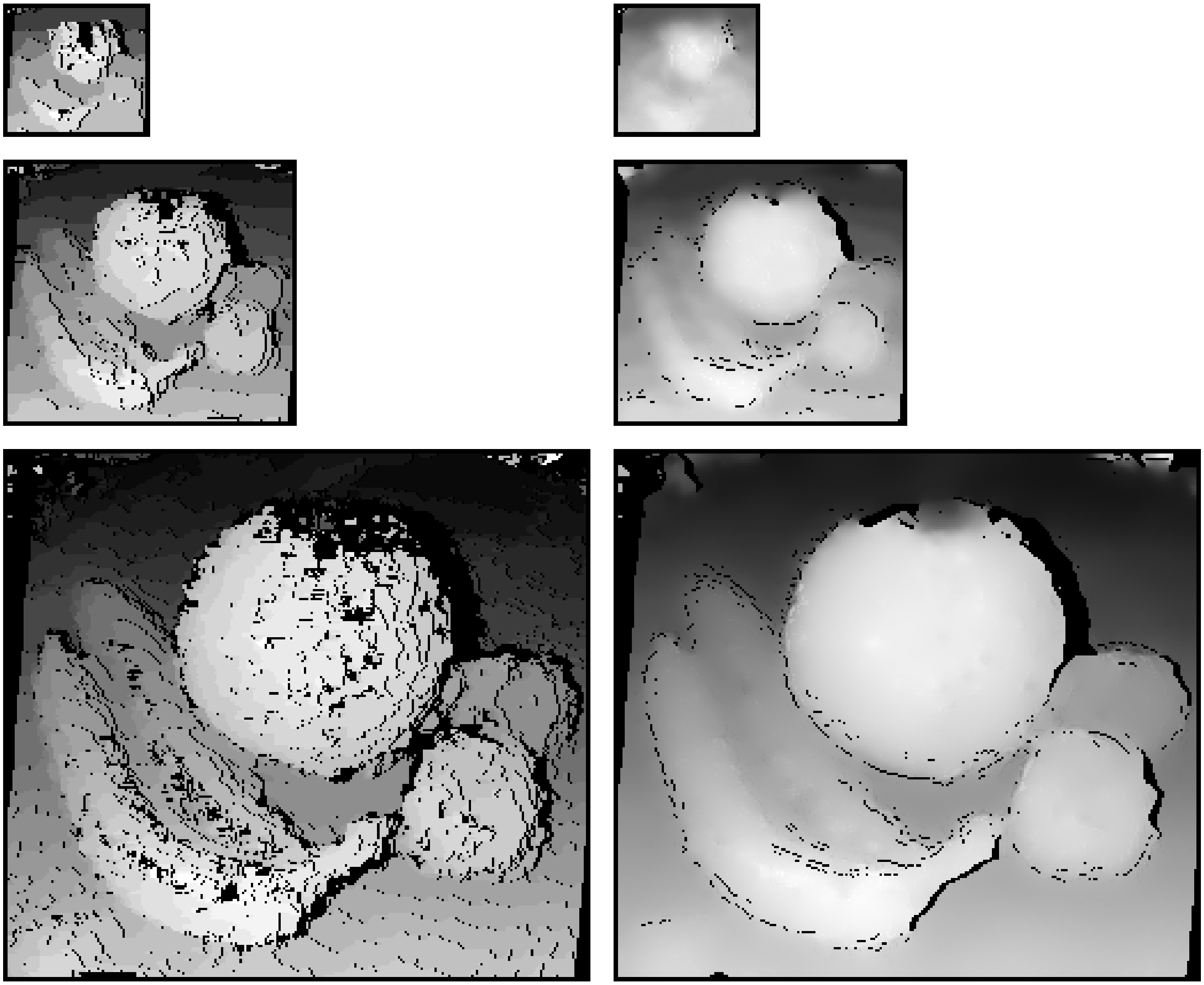

Figure 3.9. Results for Fruits stereo pair at 3 levels of the hierarchy. Left: Pixel matching module outputs. Right: Thin plate module outputs.

The results at three levels of the algorithm and a 3D-rendering of the final result are shown in Figure 3.14 and Figure 3.15, respectively. Table 3.3 show how computational complexity increases rapidly by image size.

3.9.3 Brutus Stereo Pair

This 188x144 sized by pair (See Figure 3.16) is obtained by clipping and vertically lowpass filtering the Brutus stereo pair from NEC Research Institute. The results obtained by a two-level hierarchical algoritms and associated processor user time vales are shown in Figure 3.17 and Table 3.4, respectively. Also a 3D-rendering of the final result is shown in Figure 3.18. Here, a weakness of the

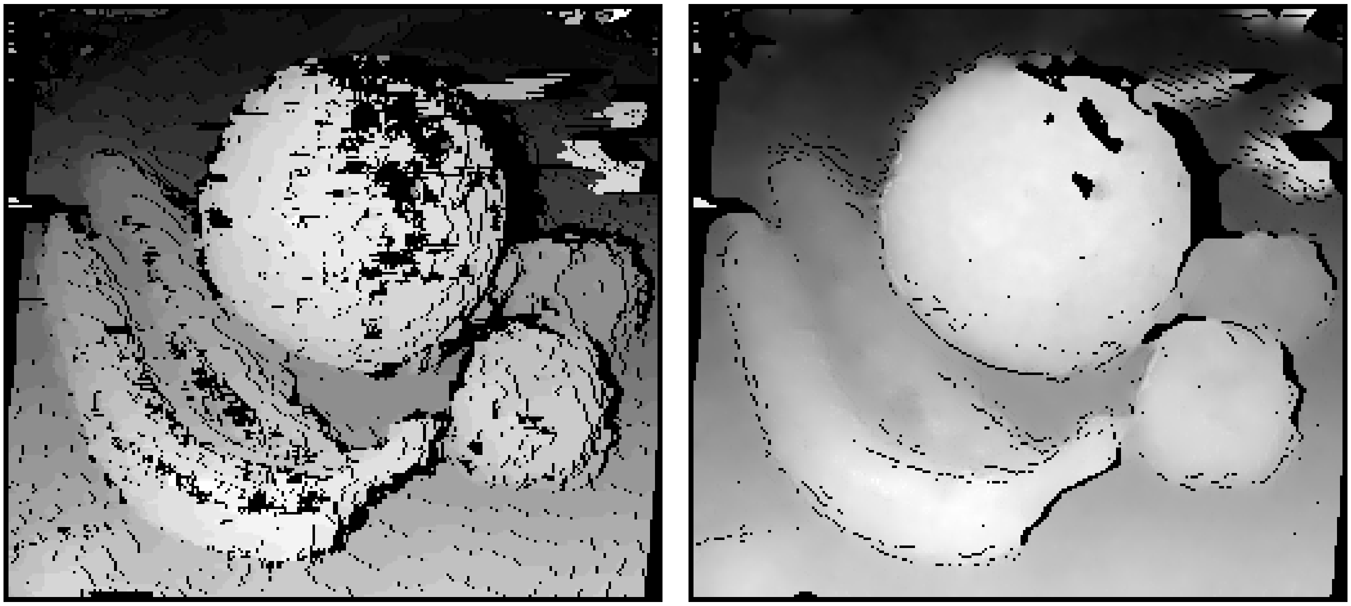

Figure 3.10. The final result of one-level algorithm on Fruits.

	58x64	116x128	232x256	Total
Gauss. Sm. and Sups.	0.6	2.6	-	3.2
Band-pass filt. (T=4 pixels)	0.8	2.8	14.8	18.4
Band-pass filt. (T=8 pixels)	1.4	5.6	21.2	28.2
Edge Detection	0.1	0.5	1.8	2.4
Pixel Matching Module	4.6	20.2	1:57.0	2:21.8
Thin Plate Module	4.0	17.1	1:08.6	1:29.7
Increasing Resolution	0.0	0.1	-	0.1
Total	11.5	48.9	3:43.4	4:43.8

Table 3.1. The user time spent for each module at each level of the hierarchy for Fruits stereo pair.

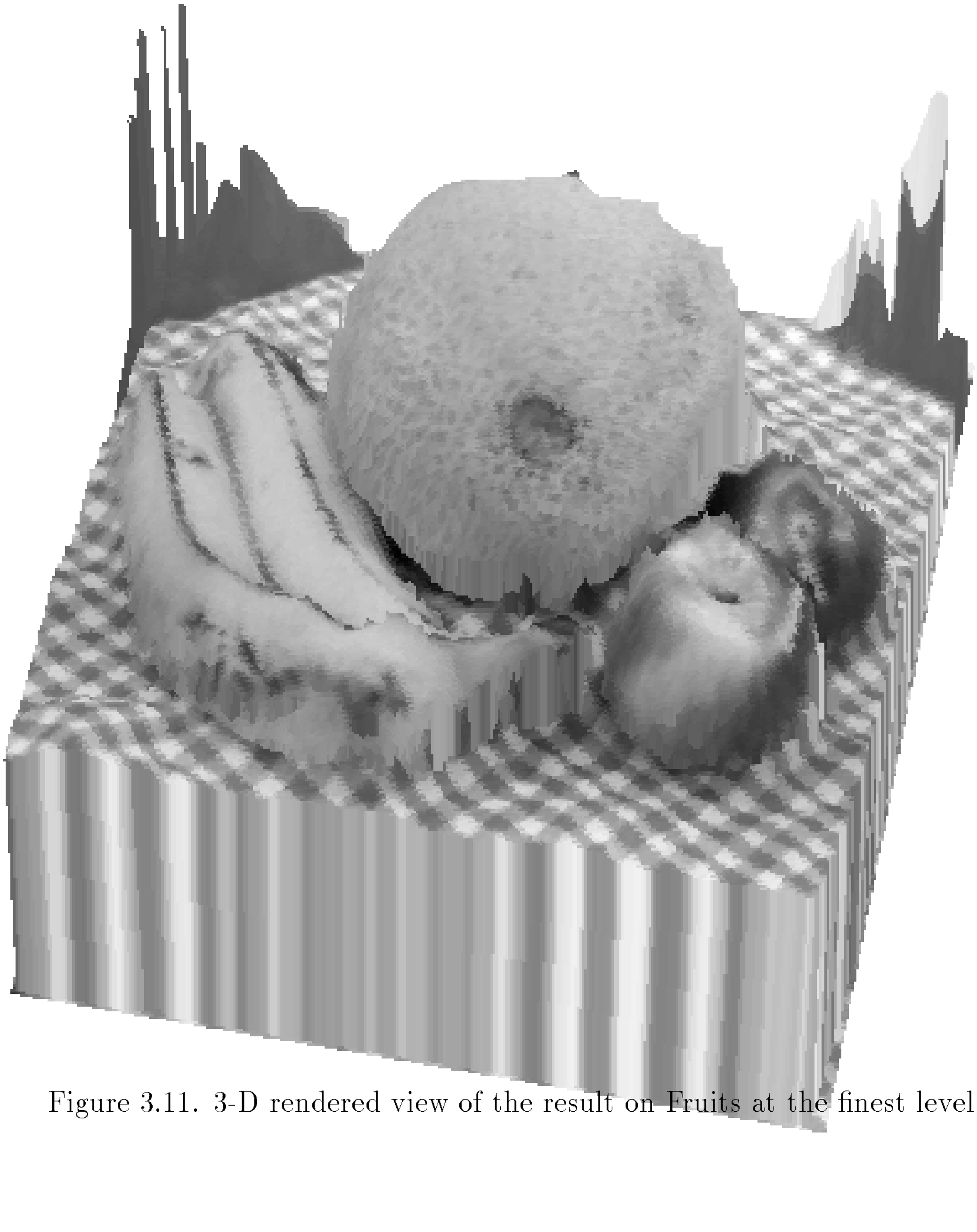

Figure 3.11. 3-D rendered view of the result on Fruits at the finest level.

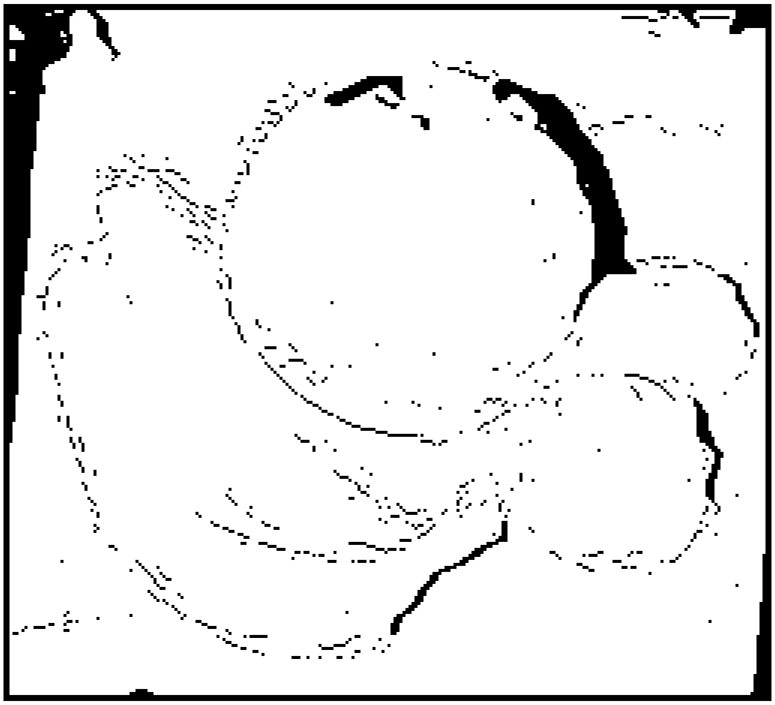

Figure 3.12. The discontinuities and occlusions for Fruits stereo pair.

	232x256
Gauss. Sm. and Sups.	-
Band-pass filt. (T=4 pixels)	14.8
Band-pass filt. (T=8 pixels)	21.2
Edge Detection	1.8
Pixel Matching Module	3:51.2
Thin Plate Module	1:19.5
Increasing Resolution	-
Total	5:48.5

Table 3.2. The user time spent for each module by one-level algorithm for Fruits stereo pair.

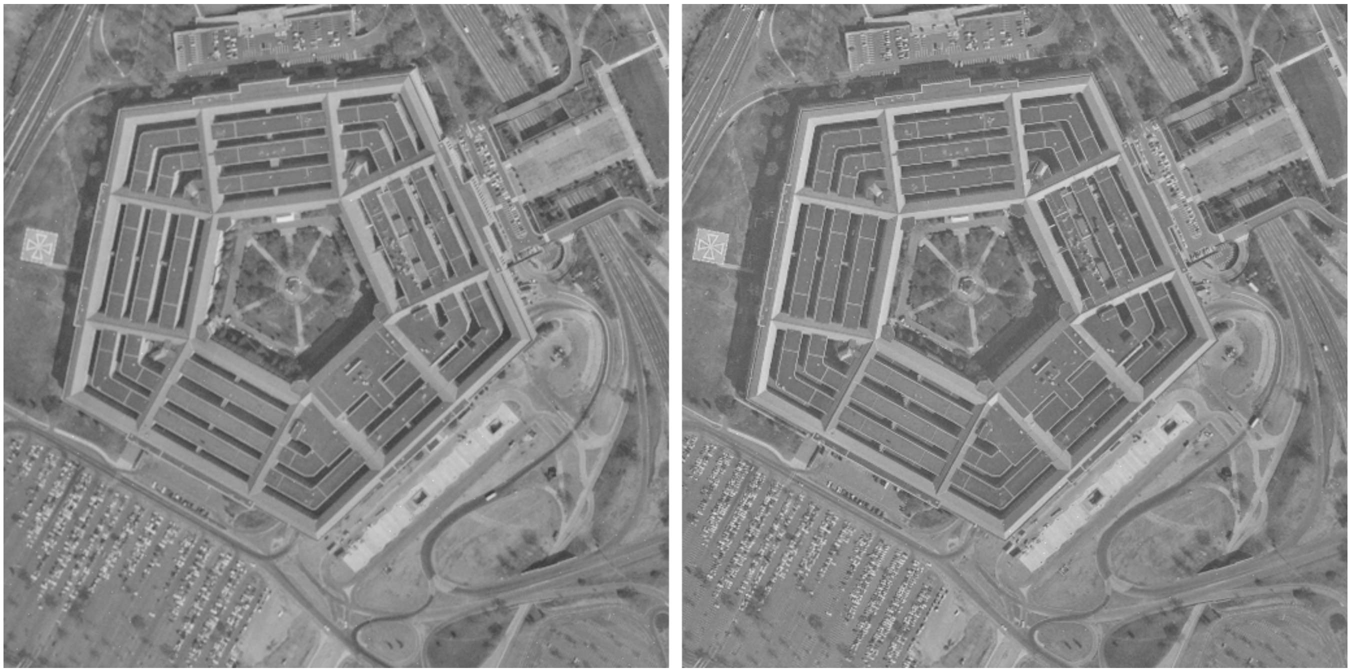

Figure 3.13. Pentagon Stereo Pair.

	128x128	256x256	512x512	Total
Gauss. Sm. and Sups.	2.8	11.2	-	14.0
Band-pass filt. (T=4 pixels)	4.0	15.8	1:03.4	1:23.2
Band-pass filt. (T=8 pixels)	6.0	24.0	1:34.8	2:04.8
Edge Detection	0.6	2.2	8.5	11.3
Pixel Matching Module	17.2	1:38.4	8:31.6	10:27.2
Thin Plate Module	20.6	1:28.9	6:01.9	7:51.4
Increasing Resolution	0.1	0.5	-	0.6
Total	51.3	4:01.0	17:20.2	22:12.5

Table 3.3. The user time spent for each module at each level of the hierarchy for Pentagon stereo pair.

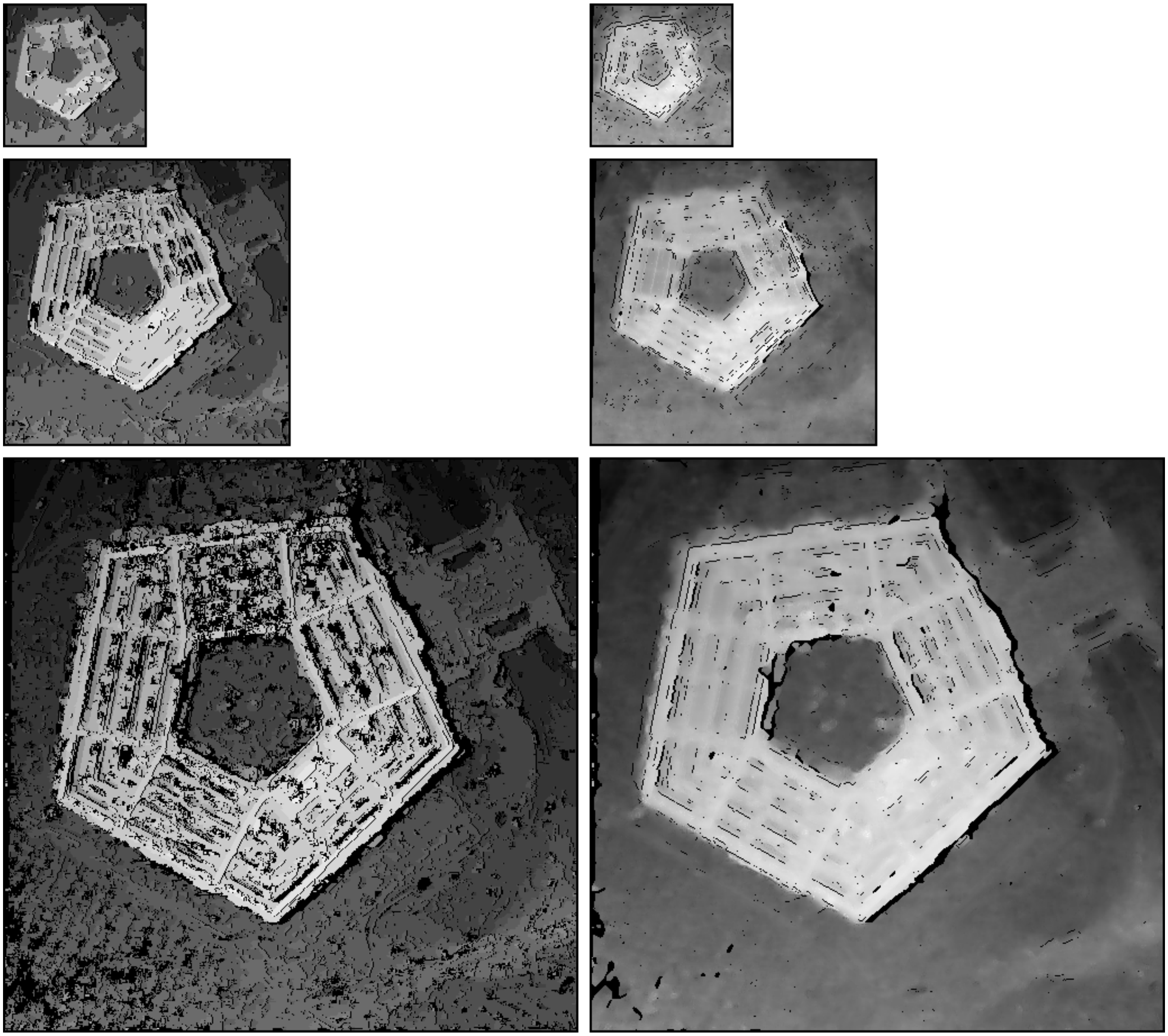

Figure 3.14. Results for Pentagon stereo pair at 3 levels of the hierarchy. Left: Pixel matching module outputs. Right: Thin plate module outputs.

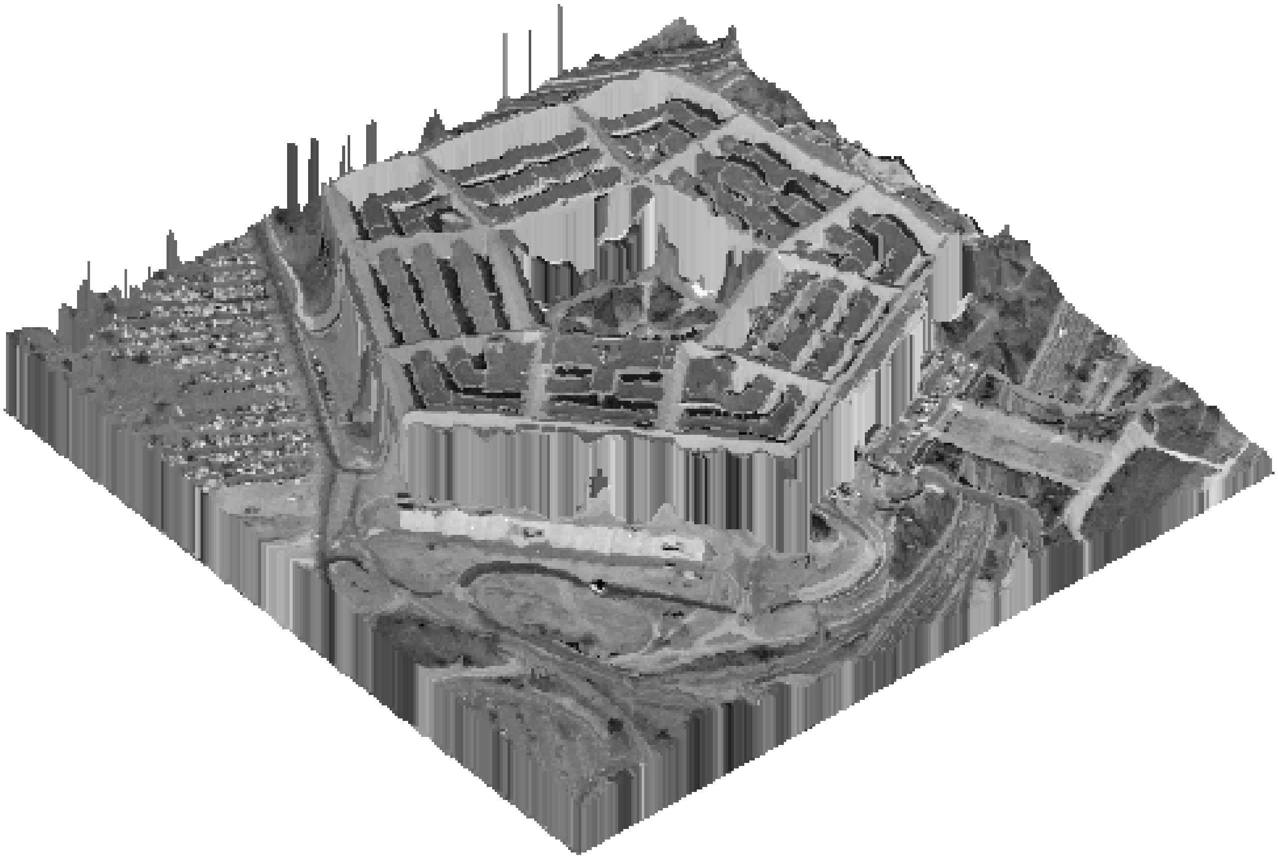

Figure 3.15. 3-D rendered view of the result on Pentagon at the finest level.

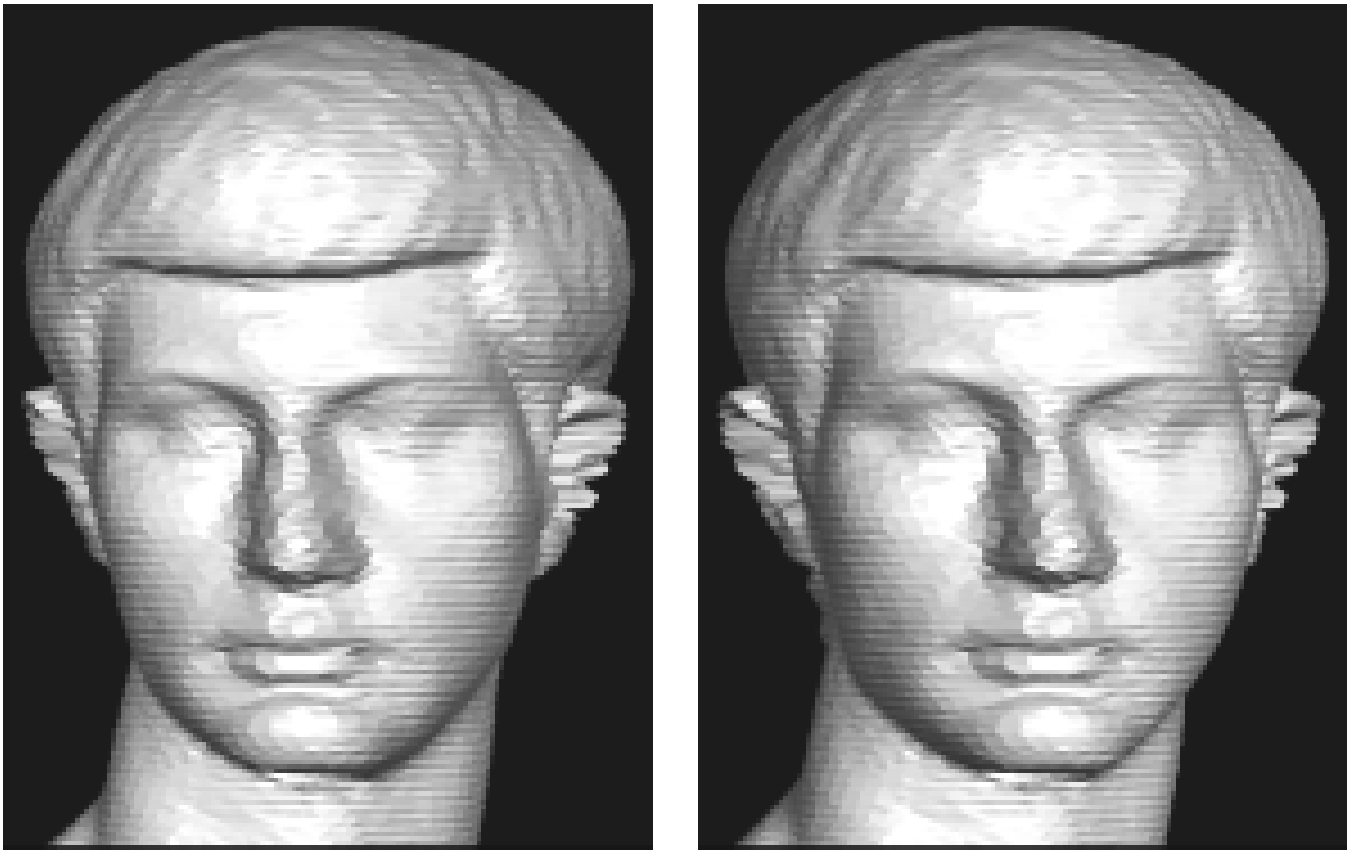

Figure 3.16. Brutus Stereo Pair.

algorithm can be seen: the disparity field spreads into textureless areas.

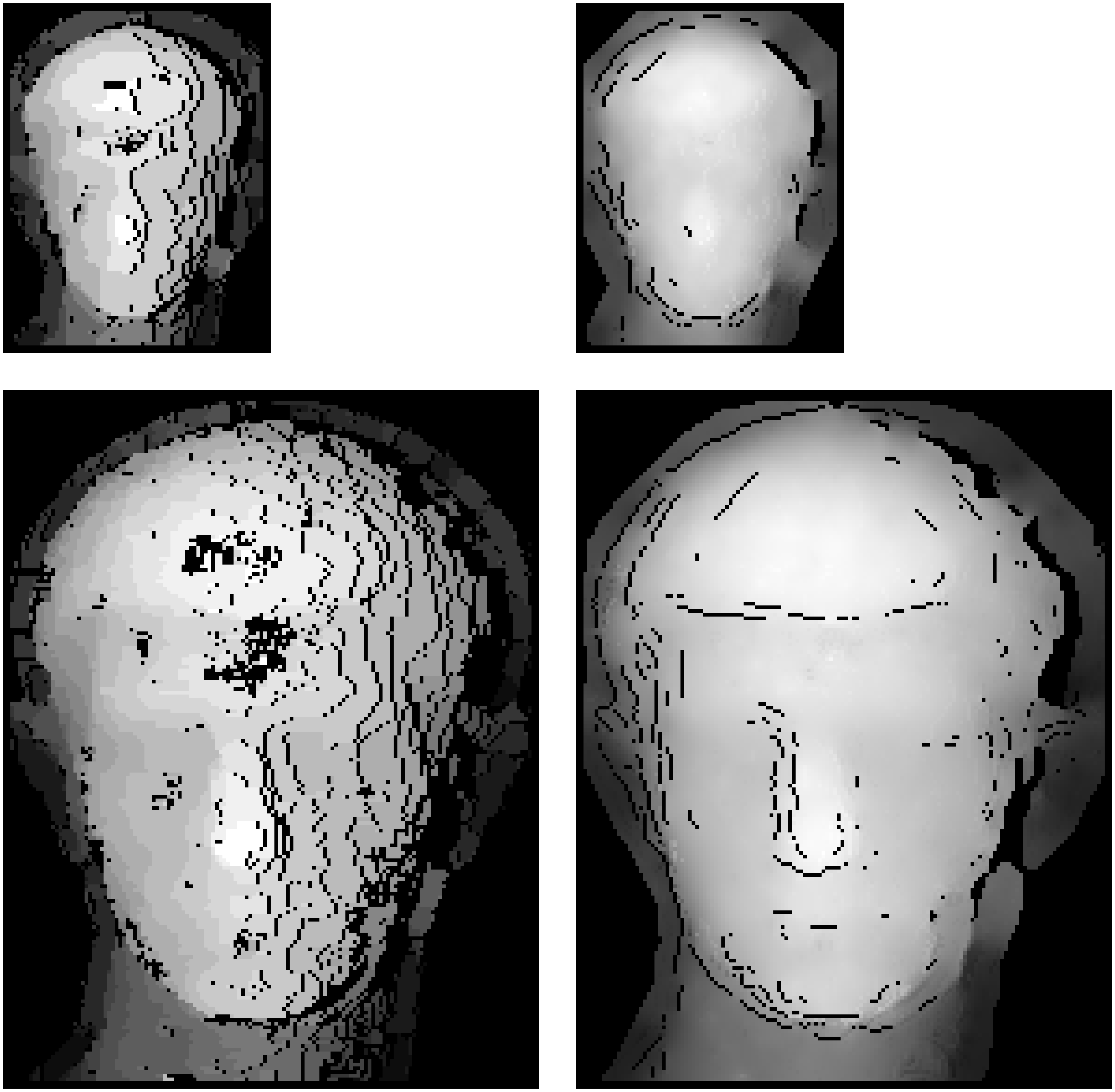

Figure 3.17. Results for Brutus stereo pair at 2 levels of the hierarchy. Left: Pixel matching module outputs. Right: Thin plate sheet module outputs.

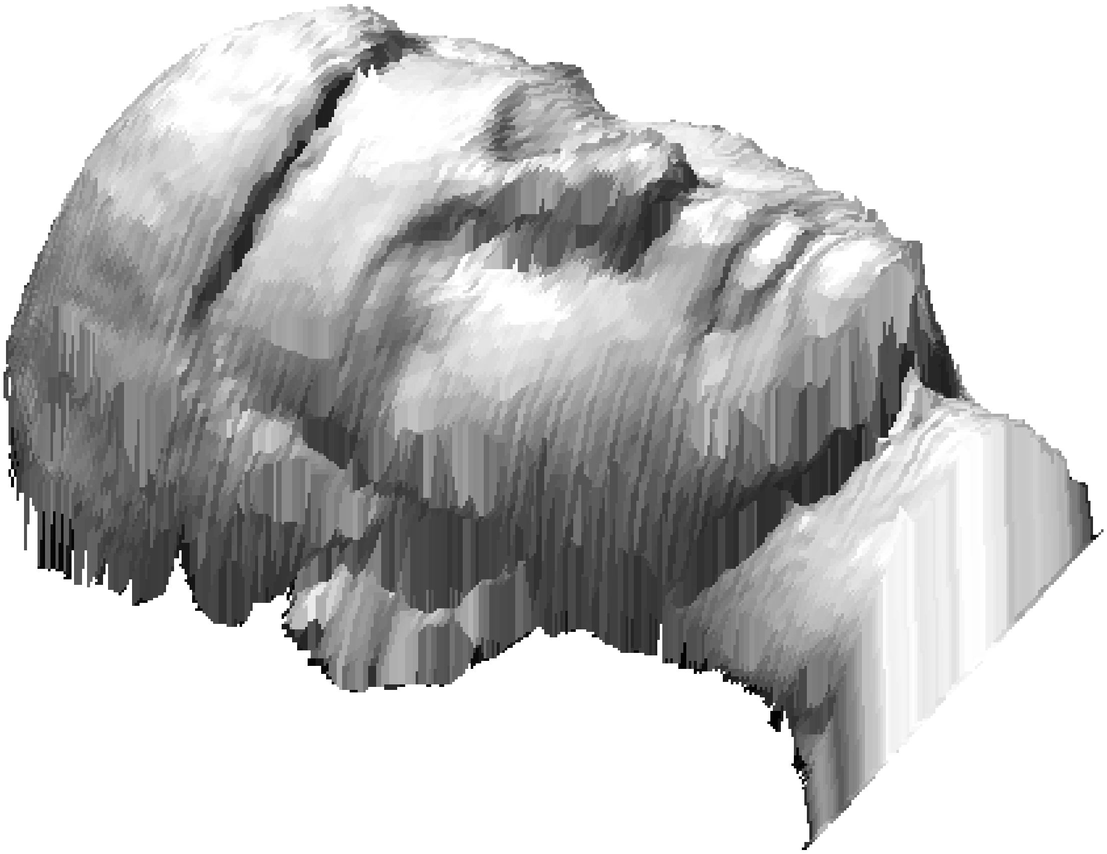

Figure 3.18. 3-D rendered view of the result on Brutus at the finest level.

	96x72	188x144	Total
Gauss. Sm. and Sups.	2.4	-	2.4
Band-pass filt. (T=4 pixels)	1.6	6.2	7.8
Band-pass filt. (T=8 pixels)	2.4	9.6	12.0
Edge Detection	0.2	0.9	1.1
Pixel Matching Module	9.9	46.2	56.1
Thin Plate Module	6.9	27.0	33.9
Increasing Resolution	0.0	-	0.0
Total	23.4	1:29.9	1:53.3

Table 3.4. The user time spent for each module at each level of the hierarchy for Brutus stereo pair.

CHAPTER IV

CONCLUSIONS

4.1 Conclusions

In this thesis, a stereo correspondence system which adopts a hierarchical structure is presented. In each scale, a pixel-matching module and a thin plate module are employed. The former utilizes normalized cross-correlations as matching primitives. Since the correlation size is chosen as small as 5x5, the associated computational cost is low. Besides, the problems of correlation due to disparity gradient and depth discontinuities are kept minimum. The ambiguities are resolved through a neighborhood support function and through spreading of constraints. The smoothness constraint is not used in this module, instead, a neighborhood support function, which works well even in the close neighborhood of depth discontinuities, is employed. The spreading of constraints works as follows: Only unambiguous matches are accepted as true and they constrain their neighbors by using uniqueness and orderness constraints which in turn cause some previously ambiguous matches to be unambiguous. The combination of the support function and spreading of constraints proves to be powerful: false matches are very rare.

The thin plate module tries to match subpixel primitives, which are phases of band-pass filtered images from several frequency channels, while interpolating small convex areas and detecting the discontinuities and occlusions explicitly. The piece-wise continuous surface is constrained by the disparities found by the pixel-matching module. The thin plate is expressed as a cost function whose minimum is searched using gradient descent method. This method generally sticks into local minima when applied to non-convex energy functions, especially to complicated energy functions like this one. However, guidance of pixel matches prevents this problem. Intensity edges detected by Canny operator helps detection of discontinuities.

The system obtains the final results within several minutes on a SPARCstation 2 and can be implemented in parallel hardware. Though it is not absolutely error-free, it succeeds on various kinds of images with a very small ratio of false matches. The problems associated with repetitive texture and with lack of texture are mostly solved by spreading of constraints and by coarse-to-fine strategy. Discontinuities and occlusions are generally detected, though the accuracy is not satisfactory.

Most of the known stereo algorithms are successful on certain kinds of images. The robustness of this system is shown on outdoor and indoor images with large disparity limits and with large disparity discontinuities: the ratio of false matches is small for all cases. Explicit detection of discontinuities and occlusions and subpixel resolution disparity map are two other outstanding properties of the system.

4.2 Directions For Further Research

Although the system developed is robust, it still makes false matches and it cannot detect discontinuities accurately. Some possible improvements on the existing system are as follows:

- Feature matching will probably increase the speed and accuracy of the system. Matching of features can easily be cooperatively integrated into the pixel-matching module.

- Fleet et al. [79] showed that the phase of band-passed filtered images behaves pathologically at and near discontinuities and proposed to detect such areas. Since the phase is not reliable in these areas, the suppression of the phase term there may improve the results.

- Detecting edges cooperatively or using active contours will lead to better discontinuity detection. Besides intensity discontinuities, texture edges and illusory contours can be taken into consideration. Also, edges detected as depth discontinuities can be tracked in scale space.

REFERENCES

[1] S. T. Barnard and M. A. Fishler, “Computational Stereo”, Computing Surveys, 14, 553 (1982).

[2] U. R. Dhond and J. K. Aggarwal, “Structure from Stereo - a Review”, IEEE Transactions on Systems, Man, and Cybernetics, 19, 1489 (1989).

[3] S. D. Cochran, “Surface Description from Binocular Stereo”, PhD thesis, School of Engineering, University of Southern California (1990).

[4] E. D. Castro and C. Morandi, “Registration of Translated and Rotated Images Using Finite Fourier Transforms”, IEEE Transactions on Pattern - Analysis and Machine Intelligence, 9, 700 (1987).

[5] A. Bani-Hashemi, “A Fourier Approach to Camera Orientation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15, 1197 (1993).

[6] L. G. Brown, “A Survey of Image Registration Techniques”, ACM Computing Surveys, 24, 325 (1992).

[7] R. Y. Tsai, “An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision”, In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 364 (1986).

[8] R. Tsai, “A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses”, In L. Wolff, S. Shafer, and G. Healey, editors, Radiometry – (Physics-Based Vision). Jones and Bartlett (1992).

[9] S. D. Blostein and T. S. Huang, “Error Analysis in Stereo Determination of 3-D Point Positions”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 9, 752 (1987).

[10] E. L. Grimson, “Why Stereo Vision is Not Always About 3D Reconstruction”, Technical Report AI Memo No. 1435, MIT Artificial Intelligence Laboratory (1993).

[11] D. H. Hubel, Eye, Brain and Vision, Scientific American Library, New York, USA (1988).

[12] V. Bruce and P. Green, Visual Perception: Physiology, Psychology and - Ecology, Lawrence Erlbaum Associates, Hove, UK (1990).

[13] L. Splillmann and J. S. Werner, editors, Visual Perception: the Neurophysiological Foundations, Academic Press, Inc., New York, USA (1990).

[14] Y. Yeshurun and E. L. Schwartz, “Cepstral Filtering on a Columnar Image Artchitecture: A Fast Algorithm for Binocular Stereo Segmentation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 759 (1989).

[15] B. Julesz, “Binocular Depth Perception of Computer Generated Patterns”, Bell Systems Technical Journal, 39, 1125 (1960).

[16] D. Marr, Vision, W. H. Freeman and Company, New York (1982).

[17] C. Schor, I. Wood, and J. Ogawa, “Binocular Sensory Fusion is Limited by Spatial Resolution”, Vision Research, 24, 661 (1984).

[18] B. Julesz, Foundations of Cyclopean Perception, The University of Chicago Press, Chicago (1971).

[19] M. J. Morgan and R. J. Watt, “Mechanisms of Interpolation in Human Spatial Vision”, Nature, 299, 553 (1982).

[20] D. R. Badcock and C. M. Schor, “Depth-Increment Detection Function for Individual Spatial Channels”, Optical Society of America, Journal A, 2, 1211 (1985).

[21] P. Burt and B. Julesz, “A Disparity Gradient Limit for Binocular Vision”, Science, 208, 615 (1980).

[22] D. H. Hubel and T. N. Weisel, “Receptive Fields, Binocular Interaction and Functional Artitecture in the Cat’s Visual Cortex”, Journal of Physiology, London, 160, 106 (1962).

[23] G. F. Poggio and B. Fischer, “Binocular Interaction and Depth Sensitivity in Striate and Prestriate Cortex of Behaving Rhesus Monkey”, Journal of Neurophysiology, 40, 1392 (1977).

[24] D. J. Field, “Relations Between the Statistics of Natural Images and the Response Properties of Cortical Cells”, Optical Society of America, Journal A, 4, 2379 (1987).

[25] P. Schiller, B. L. Finlay, and S. F. Volman, “Quantitative Studies of Single-Cell Properties in Monkey Striate Cortex. I. Spatiotemporal Organization of Receptive Fields”, Journal of Neurophysiology, 39, 1288 (1976).

[26] R. D. Freeman and I. Ohzawa, “On the Neurophysiological Organization of Binocular Vision”, Vision Research, 30, 1661 (1990).

[27] S. Marcelja, “Mathematical Description of the Responses of Simple Cortical Cells”, Optical Society of America, Journal A, 70, 1297 (1980).

[28] J. G. Daugman, “Two-Dimensional Spectral Analysis of Cortical Receptive Field Profile”, Vision Research, 20, 847 (1980).

[29] D. A. Pollen and S. F. Ronner, “Phase Relationships Between Adjacent Simple Cells in the Visual Cortex”, Science, 212, 1409 (1981).

[30] I. Ohzawa and R. D. Freeman, “The Binocular Organization of Simple Cells in the Cat’s Visual Cortex”, Journal of Neurophysiology, 56, 221 (1986).

[31] M. Nomura, G. Matsumoto, and S. Fujiwara, “A Binocular Model for the Simple Cell”, Biological Cybernetics, 63, 237 (1990).

[32] H. R. Wilson, R. Blake, and D. L. Halpern, “Coarse Spatial Scales Constrain the Range of Binocular Vision on Fine Scales”, Optical Society of America, Journal A, 8, 229 (1991).

[33] R. J. Watt, “Scanning from Coarse to Fine Spatial Scales in the Human Visual System After the Onset of a Stimulus”, Optical Society of America, Journal A, 4, 2006 (1987).

[34] N. H. Kim and A. C. Bovik, "A Contour-Based Stereo Matching Algorithm Using Disparity Continuity", Pattern Recognition, 21, 505 (1988).

[35] H. P. Moravec, "Towards Automatic Visual Obstacle Avoidance", In Proceedings of 5th International Joint Conference on Artificial Intelligence, 584 (1977).

[36] N. M. Nasrabadi and C. Y. Choo, "Hopfield Network for Stereo Vision Correspondence", IEEE Transactions on Neural Networks, 3, 5 (1992).

[37] R. Deriche and G. Giraudon, "A Computational Approach for Corner and Vertex Detection", International Journal of Computer Vision, 10, 101 (1993).

[38] J.-S. Chen and G. Medioni, "Parallel Multiscale Stereo Matching Using Adaptive Smoothing", In Proceedings of First European Conference oc Computer vision, 99, Antibes, France (1990).

[39] A. L. Yuille, D. Geiger, and H. Bültloff, "Stereo Integration, Mean Field Theory and Psychophysics", In Proceedings of First European Conference oc Computer vision, 73, Antibes, France (1990).

[40] Y. C. Hsieh, D. M. McKeown, and F. P. Perlant, "Performance Evaluation of Scene Registration for Cartographic Feature Extraction", IEEE Transactions on Pattern Analysis and Machine Intelligence, 14, 214 (1992).

[41] K.-G. Lim and R. Prager, "Using Markov Random Field to Integrate Stereo Modules", Technical report, Cambridge University Engineering Department (1992).

[42] W. E. L. Grimson, "Computational Experiments with a Feature Based Stereo Algorithm", IEEE Transactions on Pattern Analysis and Machine - Intelligence, 7, 17 (1985).

[43] S. A. Lloyd, "A Dynamic Programming Algorithm for Binocular Stereo Vision", GEC Journal of Research, 3, 18 (1985).

[44] J. J. Jordan and A. C. Bovik, "Using Chromatic Information in Edge-Based Stereo Correspondence", CVGIP: Image Understanding, 54, 98 (1991).

[45] N. M. Nasrabadi, “A Stereo Vision Tecnique Using Curve-Segments and Relaxation Matching”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 14, 566 (1992).

[46] D. Marr and E. Hildreth, “Theory of Edge Detection”, Royal Society of London, Philosophical Transactions, Series B, 207, 187 (1980).

[47] J.-S. Chen, A. Huertas, and G. Medioni, “Fast Convolution with Laplacian-of Gaussian Masks”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 9, 584 (1987).

[48] J. J. Clark, “Authenticating Edges Produced by Zero-Crossing Algorithms”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 43 (1989).

[49] G. E. Sotak and K. L. Boyer, “The Laplacian-of-Gaussian Kernel: A Formal Analysis and Design Procedure for Fast Accurate Convolution and Full-Frame Output”, Computer Vision, Graphics and Image Processing, 48, 147 (1989).

[50] F. Ulupınar and G. Medioni, “Refining Edges Detected by a LoG Operator”, Computer Vision, Graphics and Image Processing, 51, 275 (1990).

[51] J. J. Clark and P. D. Lawrence, “A Theoretical Basis for Diffrequency Stereo”, Computer Vision, Graphics and Image Processing, 35, 1 (1986).

[52] A. P. Witkin, “Scale Space Filtering”, In Proc. of the Ninth International Joint Conf. on Artificiall Intelligence, 1019, Karlsruhe, Germany (1983).

[53] Y. Lu and R. C. Jain, “Behavior of Edges in Scale Space”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 337 (1989).

[54] D. Marr and T. Poggio, “A Computational Theory of Human Stereo Vision”, Royal Society of London, Philosophical Transactions, Series B, 204, 301 (1979).

[55] J. Canny, “A Computational Approach to Edge Detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 8, 679 (1986).

[56] R. Deriche, "Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector", International Journal of Computer Vision, 1, 167 (1987).

[57] J. Porrill, S. B. Pollard, J. B. Bower, J. E. W. Mayhew, and J. P. Frisby, "TINA: The Sheffield AIVRU Vision System", In Proceedings of 10th International Joint Conference on Artificial Intelligence, Milan, Italy (1987).

[58] R. Horaud and T. Skordas, "Stereo Correspondence Through Feature Grouping and Maximal Cliques", IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 1168 (1989).

[59] J. E. W. Mayhew and J. P. Frisby, "Psychophysical and Computational Studies towards a Theory of Human Stereopsis", Artificial Intelligence, 17, 349 (1981).

[60] G. Medioni and R. Nevatia, "Segment-Based Stereo Matching", Computer Vision, Graphics and Image Processing, 31, 2 (1985).

[61] D. Sherman and S. Peleg, "Stereo by Incremental Matching of Contours", IEEE Transactions on Pattern Analysis and Machine Intelligence, 12, 1102 (1990).

[62] K. L. Boyer, D. M. Wuescher, and S. Sarkar, "Dynamic Edge Warping: An Experimental System for Recovering Disparity Maps in Weakly Constrained Systems", IEEE Transactions on Pattern Analysis and Machine Intelligence, 21, 143 (1991).

[63] K. L. Boyer and A. C. Kak, "Structural Stereopsis for 3-D Vision", IEEE Transactions on Pattern Analysis and Machine Intelligence, 10, 144 (1988).

[64] Y. Ohta and T. Kanade, "Stereo by Intra-InterScanline Search Using Dynamic Programming", IEEE Transactions on Pattern Analysis and Machine Intelligence, 7, 139 (1985).

[65] M. Ito and A. Ishii, "Three-View Stereo Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, 8, 524 (1986).

[66] D. de Vleeschauwer, “An Intensity-Based, Coarse-to-Fine Approach to Reliably Measure Binocular Disparity”, CVGIP: Image Understanding, 57, 204 (1992).

[67] I. J. Cox, S. Hingorani, B. M. Maggs, and S. B. Rao, “Stereo Without Regularization”, Technical report, NEC Research Institute (1992).

[68] S. T. Barnard, “A Stochastic Approach to Stereo Vision”, In Proceedings of the Fifth International Conference on Artificial Intelligence, 676 (1986).

[69] J. J. Jordan and A. C. Bovik, “Using Chromatic Information in Dense Stereo Correspondence”, Pattern Recognition, 25, 367 (1992).

[70] Y. T. Zhou and R. Chellappa, “Stereo Matching Using a Neural Network”, In Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 940 (1988).

[71] M. Kass, “Computing Visual Correspondence”, In Proceedings of ARPA Image Understanding Workshop, 54, Arlington, VA, USA (1983).

[72] L. Matthies, “Stereo Vision for Planetary Rovers: Stochastic Modeling to Near Real-Time Implementation”, International Journal of Computer Vision, 8, 71 (1992).

[73] P. Fua, “A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features”, Machine Vision and Applications, 6, 35 (1993).

[74] S. D. Cochran and G. Medioni, “3-D Surface Description from Binocular Stereo”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 14, 981 (1992).

[75] B. Hotz, “Etude de Technique de Stéréovision par Corrélation”, Technical report, CNES, Toulouse, France (1991).

[76] M. Okutomi and T. Kanade, “A Locally Adaptive Window for Signal Matching”, International Journal of Computer Vision, 7, 143 (1992).

[77] H. K. Nishihara, “Practical Real-Time Imaging Stereo Matcher”, Optical - Engineering, 23, 536 (1984).

[78] T. D. Sanger, "Stereo Disparity Computation Using Gabor Filters", Biological Cybernetics, 59, 405 (1988).

[79] D. J. Fleet, A. D. Jepson, and M. R. M. Jenkin, "Phase-Based Disparity Measurement", CVGIP: Image Understanding, 53, 198 (1991).

[80] C.-J. Westelius, "Preattentive Gaze Control for Robot Vision", PhD thesis, Department of Electrical Engineering, Linköping University (1992).

[81] J. J. Weng, "Image Matching Using the Windowed Fourier Phase", International Journal of Computer Vision, 11, 211 (1993).

[82] M. Nomura, "A Model for Neural Representation of Binocular Disparity in Striate Cortex: Distributed Representation and Veto Mechanisms", Biological Cybernetics, 69, 165 (1993).

[83] T.-Y. Chen, W. N. Klarquist, and A. C. Bovik, "Stereo Vision Using Gabor Wavelets", In Proc. of the IEEE Southwest Symposium on Image Analysis and Interpretation, Dallas, Texas, USA (1994).

[84] M. R. M. Jenkin and A. D. Jepson, "Recovering Local Surface through Local Phase Difference Measurements", CVGIP: Image Understanding, 59, 72 (1994).

[85] D. Gabor, "Theory of Communication", Journal of IEE, 93, 429 (1946).

[86] A. Papoulis, Probability, Random Variables and Stochastic Process, McGraw-Hill, Singapore (1965).

[87] D. J. Fleet and A. D. Jepson, "Stability of Phase Information", IEEE Transactions on Pattern Analysis and Machine Intelligence, 15, 1253 (1993).

[88] C. D. Kuglin and D. C. Hines, "The Phase Correlation Image Alignment Method", In Proc. of the IEEE International Conference on Cybernetics and Society, 163, IEEE, New York, USA (1975).

[89] D. Marr and T. Poggio, "A Cooperative Computation of Stereo Disparity", Science, 194, 283 (1976).

[90] A. L. Yuille, "Energy Functions for Early Vision and Analog Networks", Biological Cybernetics, 61, 115 (1989).

[91] S. T. Toborg and K. Hwang, "Cooperative Vision Integration Through Data-Parallel Neural Computations", IEEE Transactions on Computers, 40, 1368 (1991).

[92] R. Mohan, G. Medioni, and R. Nevatia, "Stereo Error Detection, Correction and Evaluation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 113 (1989).

[93] W. Hoff and N. Ahuja, "Surfaces from Stereo: Integrating Feature Matching, Disparity Estimation, and Contour Detection", IEEE Transactions - on Pattern Analysis and Machine Intelligence, 11, 121 (1989).

[94] H. Maître and W. Luo, "Using Models to Improve Stereo Reconstruction", IEEE Transactions on Pattern Analysis and Machine Intelligence, 14, 269 (1992).

[95] K. Prazdny, "Detection of Binocular Disparities", Biological Cybernetics, 23, 93 (1985).

[96] J. P. Frisby and S. B. Pollard. Computational Issues in Solving the Stereo Correspondence Problem, 331. The MIT Press, Cambridge, Massachusetts, USA, (1991).

[97] T. Poggio, V. Torre, and C. Koch, "Computational Vision and Regularization Theory", Nature, 317, 314 (1985).

[98] S. B. Pollard, J. E. W. Mayhew, and J. P. Frizby, "PMF: A Stereo Correspondence Algorithm Using a Disparity Gradient Limit", Perception, 14, 449 (1985).

[99] R. Szeliski and G. Hinton, "Solving Random Dot Sterograms Using the Heat Equation", In Proceedings of Computer Vision and Pattern Recognition, 212 (1985).

[100] D. Weinshall, "Perception of Multiple Transparent Planes in Stereo Vision", Nature, 341, 737 (1989).

[101] S. B. Pollard and J. P. Frisby, "Transparency and the Uniqueness Constraint in Human and Computer Stereo Vision", Nature, 347, 553 (1990).

[102] P. J. Burt and E. H. Adelson, "The Laplacian Pyramid as a Compact Image Code", IEEE Transactions on Communications, 31, 532 (1983).

[103] A. Meygret, M. Thonnat, and M. Berthod, "A Pyramidal Stereovision Algorithm Based on Chain Points", In Proceedings of First European Conference oc Computer vision, 83, Antibes, France (1990).

[104] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing", Science, 220, 671 (1983).

[105] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721 (1984).

[106] E. B. Gamble and T. A. Poggio, "Visual Integration and Detection of Discontinuities: The key role of intensity edges", Technical report, MIT Artificial Intelligence Laboratory (AI Memo No. 970, 1987).

[107] A. Joshi and C.-H. Lee, "Elastic Nets and Stereo Correspondence", In Proceedings of IJCNN, 423, Beijing (1992).

[108] R. Durbin and D. J. Willshaw, "An Analogue Approach to the Travelling Salesman Problem Using an Elastic Net Method", Nature, 326, 689 (1987).

[109] U. M. Leloğlu, "Dense Stereo Correspondence Using Elastic Nets", In Proceedings of 2nd Turkish Symposium on Artificial Intelligence and Artificial Neural Networks, 289, Bosphorus University (1993).

[110] A. Blake and A. Zissermann, editors, Visual Reconstruction, MIT Press, Cambridge, MA, USA (1987).

[111] A. Witkin, D. Terzopoulos, and M. Kass, "Signal Matching through Scale Space", International Journal of Computer Vision, 1, 133 (1987).

[112] G. Whitten, "Scale Space Tracking and Deformation Sheet Models for Computational Vision", IEEE Transactions on Pattern Analysis and Machine - Intelligence, 15, 697 (1993).

[113] A. Khotanzad, A. Bokil, and Y.-W. Lee, "Stereopsis by Constraint Learning Feed-Forward Neural Networks", IEEE Transactions on Neural Networks, 4, 332 (1993).

[114] P. H. Winston, Artificial Intelligence, Addison-Wesley, London, UK (1984).

[115] H. H. Baker, "Depth from Edge and Intensity Based Stereo", Technical report, Stanford University Artificial Intelligence Laboratory, Stanford, CA, USA (AIM-347, 1982).

[116] M. Watanabe and Y. Ohta, "Cooperative Integration of Multiple Stereo Algorithms", In Proceedings of Third International Conference on Computer Vision, 476, International House Osaka, Osaka, Japan (December 4–7, 1990).

[117] K. M. Mutch, "Determining Object Translation Information Using Stereoscopic Motion", IEEE Transactions on Pattern Analysis and Machine Intelligence, 8, 750 (1986).

[118] A. M. Waxman and J. H. Duncan, "Binocular Image Flows: Steps Toward Stereo-Motion Fusion", IEEE Transactions on Pattern Analysis - and Machine Intelligence, 8, 715 (1986).

[119] L. Li and J. H. Duncan, "3-D Translational Motion and Structure from Binocular Image Flows", IEEE Transactions on Pattern Analysis and Machine Intelligence, 15, 657 (1993).

[120] C. M. Thompson, "Robust Photo-Topography by Fusing Shape-from-Shading and Stereo", PhD thesis, Massachusets Institute of Technology (1993).

[121] W. E. L. Grimson, "Binocular Shading and Visual Surface Reconstruction", Computer ViSsion, Graphics and Image Processing, 28, 19 (1984).

[122] H. H. Bülthoff and H. A. Mallot, "Integration of Depth Modules: Stereo and Shading", Optical Society of America, Journal A, 5, 1749 (1988).

[123] J. E. Cryer, P.-S. Tsai, and M. Shah, "Combining Shape from Shading and Stereo Using Human Visual Model", Technical Report CS-TR-92-25, University of Central Florida (1992).

[124] M. L. Moerdler, “The Integration from Stereo and Multiple Shape-from-Texture Cues”, In Image Understanding Workshop, 786 (1988).

[125] L. Matthies and A. Elfes, “Integration of Sonar and Stereo Range Data Using a Grid-based Representation”, In International Conference on Robotics and Automation, 727, IEEE (1988).

[126] N. Ahuja and A. L. Abbott, “Active Stereo: Integrating Disparity, Vergence, Focus, Aperture, and Calibration for Surface Estimation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15, 1007 (1993).

[127] D. J. Coombs, “Real-time Gaze Holding in Binocular Robot Vision”, PhD thesis, Department of Computer Science, University of Rochester (1992).

[128] E. Krotkov and R. Bajcsy, “Active Vision for Reliable Ranging: Cooperating Focus, Stereo, and Vergence”, International Journal of Computer Vision, 11, 187 (1993).

[129] A. Yuille and D. Geiger, “Stereo and Controlled Movement”, International - Journal of Computer Vision, 4, 141 (1990).

[130] M. Okutomi and T. Kanade, “A Multiple-Baseline Stereo”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15, 353 (1993).

[131] G. S. van der Wall and P. J. Burt, “A VLSI Pyramid Chip for Multiresolution Image Analysis”, International Journal of Computer Vision, 8, 177 (1992).

[132] Lane, Thacker, and Ivey, “A Correlation Chip for Stereo Vision”, In Proceedings of 5th British Machine Vision Conference, York, UK (1994).

[133] R. Mehrotra, K. R. Namuduri, and N. Ranganathan, “Gabor Filter-Based Edge Detection”, Pattern Recognition, 25, 1479 (1992).

[134] I. Fogel and D. Sagi, “Gabor Filters as Texture Discriminators”, Biological Cybernetics, 61, 103 (1989).

[135] D. J. Fleet and A. D. Jepson, “Computation of Component Image Velocity from Local Phase Information”, International Journal of Computer Vision, 5, 77 (1990).