# RECIPROCAL-WEDGE TRANSFORM: A SPACE-VARIANT IMAGE REPRESENTATION

by

Frank C. H. Tong

B.Sc. Chinese University of Hong Kong 1983

M.Sc. Simon Fraser University 1987

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the School
of
Computing Science


SIMON FRASER UNIVERSITY
August 1995

# APPROVAL

**Name:** Frank C. H. Tong

**Degree:** Doctor of Philosophy

**Title of thesis:** Reciprocal-Wedge Transform: A Space-variant Image Representation

**Examining Committee:** Dr. Veronica Dahl
Chair

---

Dr. Ze-Nian Li (Thesis Advisor)
Associate Professor, Computing Science

---

Dr. Brian V. Funt
Professor, Computing Science

---

Dr. Tom Calvert
Professor, Computing Science

---

Dr. Kamal Gupta (Internal Examiner)
Associate Professor, Engineering Science

---

Dr. Steven L. Tanimoto (External Examiner)
Professor, Computer Science
University of Washington

**Date Approved:** Aug. 16, 1995

**PARTIAL COPYRIGHT LICENSE**

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

**Title of Thesis/Project/Extended Essay**

Reciprocal-Wedge Transform : A space-variant Image Representation

**Author:** _____
(signature)

Frank C.H. Tong
(name)

Aug. 17, 95
(date)

# Abstract

The problems in computer vision have traditionally been approached as recovery problems. In active vision, perception is viewed as an active process of exploratory, probing and searching activities rather than a passive re-construction of the physical world. To facilitate effective interaction with the environment, a foveate sensor coupled with fast and precise gaze control mechanism becomes essential for active data acquisition.

In this thesis, the Reciprocal-Wedge Transform (RWT) is proposed as a space-variant image model. The RWT has its merits in comparison with other alternative foveate sensing models such as the log-polar transform. The concise matrix representation makes it enviable for its simplified computation procedures. Similar to the log-polar transform, the RWT facilitates space-variant sensing which enables effective use of variable-resolution data and the reduction of the total amount of the sensory data. Most interestingly, its property of anisotropic mapping yields variable resolution primarily in one dimension. Consequently, the RWT preserves linear features and performs especially well on translations in the images.

A projective model is developed for the transform, lending it to potential hardware implementation of RWT projection cameras. The CCD camera for the log-polar transform requires sensing elements of exponentially varying sizes. In contrast, the RWT camera achieves variable resolution with oblique image plane projection, thus

alleviating the need for non-rectangular tessellation and sensitivity scaling on the sensing elements. A camera model making use of the available lens design techniques is investigated.

The RWT is applied to motion analysis and active stereo to illustrate the effectiveness of the image model. In motion analysis, two types of motion stereo are investigated, namely, longitudinal and lateral motion stereo. RWT motion stereo algorithms are developed for linear and circular ego motions in road navigation, and depth recovery from moving parts on an assembly belt. The algorithms benefit from the perspective correction, linear feature preservation and efficient data reduction of the RWT.

The RWT imaging model is also shown to be suitable for fixation control in active stereo. Vergence and versional eye movements and scanpath behaviors are studied. A computational interpretation of stereo fusion in relation to disparity limit in space-variant imagery leads to the development of a computational model for binocular fixation. The unique oculomotor movements for binocular fixation observed in human system appears natural to space-variant sensing. The vergence-version movement sequence is implemented for an effective fixation mechanism in RWT imaging. An interactive fixation system is simulated to show the various modules of camera control, vergence and version. Compared to the traditional reconstructionist approach, active behavior is shown to be plausible.

# Acknowledgements

My foremost gratitude goes to my thesis advisor, Dr. Ze-Nian Li, for his constant support and encouragement. I have learned many things from Ze-Nian during the course of my working with him. I have learned from his persistence and industriousness as a researcher. However, I admire most his knowledge and vision.

My deepest gratitude also goes to Dr. Brian Funt. I thank him for introducing me to the area of computer vision. His inspiring suggestions have always been most valuable. I would also like to thank Dr. Tom Calvert for being on my advisory committee. I am grateful for his generosity with his time and comments. My thanks also go to Dr. Kamal Gupta. He is my professor, and he is also my friend. His thoroughness in reviewing my thesis is much appreciated.

I also owe my gratitude to Dr. Steven Tanimoto. I feel grateful to him for being my external examiner. He has been very generous with both his time and helpful comments. Steve is very knowledgeable in the area. His acceptance of my thesis makes me feel I have accomplished something valuable.

I would like to express my appreciation to Dr. Woshun Luk. His constant concern and encouragement are much appreciated. I am also thankful to Gray Hall for help with the proof-reading.

My thanks also go to many of the graduate students. In particular, I would like to

thank Graham Finlayson for the interesting and inspiring discussions. Carlos Wong and Xao Ou Ren shared the same office with me. I thank them for the refreshing chats that kept me going even in the most boring days.

I also thank the entire staff of the Computing Science department. We are lucky to have a crew of supporting staff who are so friendly and helpful. They indeed have made a viable environment throughout my stay.

I owe all my accomplishments to my parents. They worked so hard to raise a family of eight, yet they still supported us through school. It was not easy for them. Finally, and by no means least, I want to acknowledge the support of my wife, Mimi Kao. This thesis could not be possible without her caring and encouragement.

# Contents

# List of Figures

# Chapter 1

# Introduction

During the last three decades, many significant advances have been accomplished in computer vision. Many problems, on the other hand, still remain too hard to solve. In view of the limitations of the existing methodologies, researchers have been striving for more effective approaches. In the recent years, various active approaches have been developed and leading to promising results. The essence of these approaches lie in the interactability of an active agent with the visual environment.

In the past, the issues in computer vision research have largely been related to reconstruction of the physical world. The general belief was that the visual information flows from low-level to high-level processing. Once the world and its properties have been recovered from the images, high-level visual tasks can then be performed [Mar82]. However, since the low-level task of extracting useful visual information by itself is either intractable or demanding excessive amount of computation, it is not surprising that the research for subsequent visual processes for the higher level tasks have not shown much success. In one of the most effective perceptual systems, the human vision system, we do not just see, we look and actively interact with the visual

environment [Baj88]. Certain problems are only solvable with constant replenishment with visual information of the world and interactive search and exploration of the environment [AWB88, Bal91].

The lack of vision systems that can perform in real-time limits computer vision to the domains of image understanding based on static analysis. Oftentimes, the camera is pointed at a preset angle, and the image data are acquired passively. The bulk of computer vision is then conducted off-line, trying very hard to recover the physical circumstances (color, shape, depth, surface, etc.) of the imaged world. Subsequent visual tasks such as object recognition, shape and structure modeling, etc. then follow.

With the advances of high performance and massively parallel computers, real-time or near real-time performance have been achieved for some vision problems. Emphasis on interactive visual processing is no longer impractical. Problems once deemed unsolvable can now be performed with guided search by interactive probing and verification.

Questioning the reconstructionist approach [Mar82], a collection of related paradigms offered under various names such as active, animate, responsive, task-based, behavioral and purposive vision have recently been proposed which draw heavily on active probing and search, and emphasize on behavioral interaction. Collectively, these various paradigms are categorized as active vision methodologies.

## 1.1 Active Vision and Foveate Sensors

Active vision has been advocated by many researchers [AWB88, Baj88, Bal91, Tso92, SS93]. They argue that perception is not a passive process, but rather an active process of exploratory, probing and searching. An active visual system differs from a

passive system in its purposive interaction with the world. Some interesting results in active vision include smart sensing using multiresolution images in a pyramid [Bur88], fixation for 3-D motion estimation [Bal91, FA93], active stereo using focus, vergence control [AA93, KB93], and purposively adjusting multiple views for 3-D object recognition [KD94, GI94].

It has been argued that foveate sensors are central to the sensing mechanism of an active vision system because they are economic and effective when coupled with active control. Research into anthropomorphic space-variant resolution sensors now receives much attention. The human visual system has a special saccadic behavior of quickly directing the focus of attention to different spatial targets [Yar67, Car77]. A foveate sensor coupled with fast and precise gaze control form the distinctive feature of the sensing mechanism of an active agent. In nature, human retina has a fovea which is a small region (1-2°) near the optical axis. The foveal resolution is superior to the peripheral resolution by orders of magnitude [Car77]. A design of this kind realizes an economic structure of sensor hardware supporting simultaneously a wide visual field and local high acuity.

The study of Schwartz [Sch77] shows that the cortical image of the retinal stimulus resembles a log-polar conformal mapping. Sandini and Tagliasco [ST80] argue that the retina sensor offers a good compromise among large visual field, acceptable resolution, and data reduction. The log-polar transform is defined as $\mathbf{w} = (\log r, \theta)$ [WC79], where $r$ and $\theta$ are the polar coordinates of the original Cartesian image. By exploiting the polar coordinates, it simplifies centric scaling and rotation as the transformations now become shift operations in the $\log r$ and $\theta$ dimensions, respectively. As shown by Sandini and Dario [SD90], the scaling and centric rotational invariances of the log-polar transform make it a useful tool for 2-D object recognition. The transform is

also shown to be effective for estimation of time-to-impact from optical flow [TS93]. However, there is a major drawback with the log-polar transform. That the image patterns of linear structures and translational movements are distorted into streamlines of log-sine curves [WC79] adversely complicates the analysis of these common phenomena in computer vision.

## 1.2 Reciprocal-Wedge Transform

In this thesis, the *Reciprocal-Wedge Transform (RWT)* is proposed.[1] The RWT exhibits nice properties for computing geometric transformations owing to its concise matrix notation. As with the log-polar, the RWT supports space-variant sensing. As expected, the space-variant sampling facilitates efficient data reduction. In particular, the resolution variation is anisotropic, predominantly in one dimension. Consequently, the RWT preserves linear features in the original image. This renders the transform especially suitable for vision problems that are related to linear structures or are translational in nature, such as line detection, linear motion and stereo correspondence. In the later chapters, it will be shown that vision systems for parts inspection in automated manufacturing and vehicle navigation in road driving benefit from the anisotropic space-variant RWT representation.[2]

The capacity for parallel processing and the accessibility of multiple resolutions have made the pyramid model a widely adopted structure for fast image processing and parallel computational modeling for various visual processes. Burt popularized the pyramid architecture with his work in Gaussian pyramidal image encoding scheme

[1] This part of work has been published in [TL93, TL95].
[2] The result has also been published in [TL94].

[Bur84]. Tanimoto, Pavlidis [TP75], Cantoni, Levialdi [CL86] and Uhr [Uhr87] represent some of the early works. The power promised by pyramid architectures has drawn researchers into implementation of the hardware image pyramids. To date, the Image Understanding Architecture [WB91] represents the most ambitious project on a large scale three-dimensional pyramid architecture. The implementation of the two-dimensional pyramid architecture [ELT+92] offers cost-effectiveness and versatility both in iconic [LZ93] and functional [Li91] pyramidal mappings. It is shown in this thesis that a fast generation of RWT image can benefit from the parallelism and hierarchical linkage of the pyramidal architecture. In particular, the rectangular image space can be mapped to the two-dimensional pyramidal structure of the SFU hybrid pyramid in a way that exploits the more abundant computing power in the bottom of the pyramid for foveal processing.

A projective RWT model is developed in [TL93, TL95] which lends itself to a potential hardware implementation of the RWT projection cameras. A prominent problem of that rudimentary camera model is the requirement of focusing on a deep image plane along the optical axis. In this thesis, a new hardware camera model is proposed which overcomes the focus problem by using a lens focusing the non-paraxial non-frontal image onto an orthogonally placed RWT plane.

Many previous efforts have been made in developing new camera systems for computer vision applications. In general, these systems provide convenience and improvements in speed and/or quality, especially for special purposes imaging, e.g., stereopsis, space-variant sensing, etc. Teoh and Zhang [TZ84] described a single-lens camera for stereopsis. Two fixed mirrors and a rotating mirror are used to obtain stereo images in two snapshots. Because only one lens is needed, the camera calibration problem is alleviated. Goshtasby and Gruver [GG93] presented a single-lens single-shot stereo

camera which offers faster image acquisition and hence has potential to be used in dynamic scenes. Hamit [Ham93] reported on a near-fisheye CCD camera which provides an alternative to variable-resolution imagery. A fisheye lens is used to acquire 180° hemispherical field of view. Electronically, any portion of the view can be flattened and corrected, thus enabling zooming in on any areas of interest.

The prototype CCD camera for the log-polar transform [VdSKC+89, KVdS+90] comprises concentric rings of different widths on the sensor chip. The space-variant sampling is essentially achieved by using sensing elements of highly non-uniform size and non-rectangular shape. Special hardware is designed to read out signals from the circular CCDs. A special scaling technique is also needed to obtain roughly the same sensitivity from all the cells in the structure. A small fovea of uniform resolution at the center is fabricated to overcome the singularity of the log-polar transform at $r = 0$ and to provide higher resolution.

As the RWT camera is based on a projective model, the spatially varying resolution is achieved from the projection of the scene on an oblique image plane. The RWT camera has improved on certain drawbacks of the log-polar sensor. First, variable sampling is not a requirement of the sensor circuit. Therefore, an ordinary sensor array of rectangular tessellation and uniform grid size which is cheaper to fabricate can be used. Also shown in the later chapter, the singularity problem is eliminated by projecting the central fovea in the conventional frontal orientation.

## 1.3 Motion Stereo in RWT Domain

One of the first applications of the RWT is a simple road navigation system. It demonstrates that the perspective distortion of the road image is readily corrected by

the variable resolution of the RWT, enabling a more efficient search of the reduced data for the road direction.

The RWT is also shown to be applicable to stereo vision for depth recovery. One of the difficult problems in stereo vision is correspondence [MP79]. Once corresponding points in the pair of images are identified, their disparity values can be calculated and used to recover the depth. This thesis shows the application of the RWT to the correspondence process in *motion stereo* [Nev76]. Two types of motion stereo are discussed, namely longitudinal and lateral motion stereo. In both cases, the properties of the anisotropic variable resolution and linear features in the RWT domain are exploited to yield efficient space-variant resolution algorithms which work on the much reduced image data. The difficult and computationally expensive correspondence problem in both motion stereo cases is effectively reduced to an easier problem of finding collinear points in the epipolar planes, which is later solved by a voting algorithm for accumulating multiple evidence.

## 1.4 Active Fixation using RWT Sensor

Since the primary motive for space-variant sensing is its application in active vision, this thesis also studies the applicability of the RWT model in *fixation control* in active stereo. In a common mode of stereo vision, the left and right cameras are pointed at the angles converging at a point which is referred as the *point of fixation.* This approach has the advantage that the object at the point of fixation has a zero disparity, and the disparities of the other objects in the scene are measured relative to it. The approach allows visual computations to be done using relative algorithms which are simpler than strategies that use egocentric coordinates [Bal91]. In binocular stereo,

fixation facilitates estimation of depth from vergence [AA93]. When both cameras are converged at the same point, the cameras are rotated and their optical axes intersect. From the triangulation geometry of the baseline camera separation and the rotation angles, it is possible to determine the vergence angle and the 3-D location of the fixation point.

Psychological studies reveal that the eye movements involved in stereo fixation include both *vergence* and *version* movements[Car77]. When we shift our fixation from one point to another, vergence control is initiated to bring both eyes converged at the right depth. The versional movement, which is a synchronized panning of both eyes, is interleaved in between the vergence cycle to recenter both retinas at the new fixation point.

We view such a fixation mechanism as natural in space-variant sensing. Stereopsis is most effective in the Panum's area [Ogl64]. In light of the fact that sensing space is space-variant, we argue that it is both logical and functional to assume the Panum's area to be a narrow region near the fovea and the deep region at the periphery. In Chapter 6, a binocular RWT sensor is shown to support a space-variant Panum's area as well. When using the RWT as a foveate sensor, the vergence/version model for stereo fixation is naturally employed. A process of three stages — a version interleaved between two vergences — is implemented in a fixation system. A high-level intelligence component initiates the fixation shift. Based on the peripheral and foveal disparities, the vergence component performs the first and second vergence movements. The version component pans the two binocular cameras according to the image position of the target.

Functioning of the fixation system as a whole is demonstrated in a scanpath exercise of performing binocular visual exploration of an office environment. For demonstration purposes, a simplistic heuristic decision is adopted to evaluate the scanpath in which the next fixation is chosen to be the unexplored area with the most disparate image points. From the execution record, the system is shown working with the various inter-component interaction that lead successfully to the consequential gaze transfers.

## 1.5 Thesis Overview

The organization of the rest of the thesis is as follows. Chapter 2 presents a survey on the existing results in the related areas. Chapter 3 introduces the RWT model and its properties. A pyramidal architecture for mapping the RWT image space is also presented. Chapter 4 delineates the projective model and the potential camera implementation. Chapter 5 describes application of the RWT in road navigation. Applications of the RWT in two motion stereo cases and preliminary test results using real-world images are discussed. Chapter 6 studies the applicability of the RWT in binocular fixation. For demonstration, a scanpath experiment is done with simplistic heuristics. Chapter 7 presents the conclusions and discusses the potential extensions for future research.

# Chapter 2

# Survey

## 2.1 Active Vision

The ability to combine vision with behavior is vital to achieving robust, real-time perception for a robot interacting with a complex, dynamic world. In the paradigm of active vision, vision does not remain as a static analysis of passively sampled image data. Instead, it is understood in the context of the visual behaviors that the system is engaged in.

Traditionally, computer vision has been treated as to solve the problem of deriving an accurate 3-D description of the scene and recovering the properties of the imaged objects. The general idea is that if we could reconstruct the world, we would be able to perform various tasks such as recognizing the objects, navigating through the environment and avoiding obstacles. A vision system should comprise various modules that recover specific descriptions of the scene from the images. A methodology was developed for analyzing visual modules. In Marr's formulation of computer vision [Mar82], visual processing is realized in three levels: (1) computational theory, (2)

algorithms and data structures, (3) implementation. Much research was then devoted to the study and development of various modules [Hor86, AS89] and the integration of them [AS89].

Many researchers see the reconstructionist methodologies too stringent for practical real-time machine vision. Despite that ample mathematical theories describing various modules have been published, there is still a lack of successful visual systems. Common problems like structure from motion, in which one wishes to reconstruct the shape and 3-D motion of a moving object from its images, turn out to be very hard. However, Aloimonos [Alo90] demonstrated that we can achieve many highly non-trivial visual tasks in navigation without solving the general structure from motion problem. Ballard in [Bal91] argued that many visual behaviors may not require elaborate categorical representations of the 3-D world.

The structure and function of eye movements in the human visual system reveal the fundamental difference between an active agent (human) and a passive system (electronic camera). The human eye is distinguished from a camera because it possesses a fovea which supports very high sensor density. The fovea is in a small region near the optical axis. It has a diameter of one to two degrees of visual angle, representing less than 0.01% of the entire visual field. The foveal resolution is superior to the peripheral resolution by orders of magnitude. A design of such features an economic structure of sensor hardware supporting simultaneously a large field of view and local high acuity. In a study by Sandini and Tagliasco [ST80], they showed a gain of 30 : 1 in visual coverage with a logarithmic sensor distribution simulating the retinal structure.

With the small fovea in a large visual field, the human visual system is equipped with the saccadic behavior for quickly directing the fovea to different spatial targets.

An earlier systematic study of saccadic eye movements was done by Yarbus [Yar67]. Subjects given specific tasks related to a picture showed different scanning patterns as attempting to solve the visual problem at hand. The results are consistent with the reports from the other studies [Not70, NS71a, NS71c]. These observations reveal that eye movements, coupled with the foveate retina structure, are driven actively by the problem-solving behaviors to explore the visual world.

### Animate vision

Ballard [Bal89, Bal91] used the term *animate vision* for their behavioral perspective to active vision. In their perspective, vision is understood in the context of visual behaviors that the system is engaged in. One important feature of animate vision is gaze control. Gaze control is the mechanism for directing the fovea at a specific spatial target. Traditionally, visual systems work in isolation, solving ill-posed problems under conditions with many degrees of freedom. In the animate perspective, the gaze is controlled actively. The visual processing is interlinked with the sensory-motor behaviors. For example, one can use physical search to look for the desired object in the scene. A moving camera under ego-motion provides additional constraints on the imaging process [AWB88]. The blurring introduced by ego-motion while fixating can isolate the object being attended from the background. Similarly, one can exploit the near zero disparity produced in binocular vergence [CB92]. With the ability to fixate targets in the world, one can work with the object-centered coordinates which has the advantage of being invariant with respect to the observer's motion. Moreover, simpler approaches using relative algorithms become feasible.

### Purposive and qualitative vision

Aloimonos et al. [Alo90] study vision in a purposive manner. Problems should be formulated in relevance to the task at hand versus being solved in an abstract general principle leading to development of a module for the whole class of problems. In purposive thinking, computer vision is not studied by itself, but in the context of a big process in which vision is used as help. A vision system thus is defined according to the task as a collection of processes each of which is to solve a particular subtask related to the original visual problem. Very often, these subtasks are simple enough that they require only a qualitative decision from the visual process. Robust methods using the approaches of qualitative techniques are applicable. In [AH90], Aloimonos described the design of the Medusa system that can perform complex tasks without reconstructing the world.

### Active sensing

As Bajcsy [Baj88] pointed out, we do not just see, we look. Our pupil is adjusted to the level of illumination, our eyes are focused, converged or diverged to fixate the target. We even move our head or change our position to get a better view of the object. Perceptual activities are exploratory, probing and searching. The term "*active sensing*" is defined as a problem of control applied to the data acquisition process which is adaptive to the current state of the data interpretation and the goal of the task. A visual system in this perspective encompasses local and global models of sensing. The local models describe the physics and noise of the sensors, the processes of signal processing and data reduction mechanisms that are applied on the image data. The global models represent the feedback connections, how individual modules interact, and characterize the overall performance of the system. Control strategies

are devised based on how much the process is data-driven (bottom-up) and how much a priori knowledge is required (top-down). Krotkov [Kro89, KB93] demonstrated an active system using the sensor models of cooperative focus, vergence and stereo.

## 2.2 Log-polar Transform

### 2.2.1 Logarithmic mapping from retina to cortex

Study of topographical mapping of receptor peripherals onto the cerebral cortex started quite early. Five decades ago, Polyak [Pol41] suggested the existence of a mathematical projection of the retina on the cortex based on the anatomy of the visual cortex. Since then, a large volume of empirical data on the retinotopic mappings has been collected. Schwartz [Sch77] cleverly summarizes the data and produces an elegant mathematical form for the retinotopical mapping.

Using relatively crude recording techniques, early workers such as Talbot and Marshall [TM41] and Apter [Apt45] established the initial understanding of the cortical projection of the retinal stimuli. Subsequent work making use of more refined and sophisticated measuring techniques detailed the knowledge of the various sensory mappings. In view of these surface mappings, Arbib [Arb72] was led to characterize the brain as a layered somatotopically organized computer. In addition to all these predecessors, Daniel and Whitteridge [DW61] conducted extensive investigation and provided a wealth of quantitative data for analysis. They observed that, in the cortical mapping, the magnification factor from retina to cortex is symmetric in all radii but tapered off in a inverse relationship with the eccentricity. Mathematically, it is

$$M(w) \propto \frac{1}{\|z\|}$$

where $M$ is the magnification, $w$ is the cortical coordinates, and $z$ is the retinal coordinates, whereas $\|z\|$ measures the eccentricity from the foveal point on the retina. As the cortical magnification is a differential quantity, Schwartz [Sch77] inverted the derivative and yielded a mathematical function which describes the retinotopic mapping in an analytical manner:

$$w = \ln(z) \ . \tag{2.1}$$

Denote $z$ as a complex variable $r\, e^{i\phi}$, $w$ in eq. (2.1) will be $\ln r + i\phi$. Expressed in real variables, the mapping is popularized in its log-polar formulation, a semi-logarithmic mapping of the polar coordinates:

$$w = (u, v) = (\ln r, \phi) \ . \tag{2.2}$$

The discovery of log-polar structure of the retinotopic mapping is not due to coincidental observation. In fact, other researchers have reported experimental data supporting the log-polar conclusion. Allmann and Kaas [AK72, AK74, AK76] conducted tests on both the secondary and medial visual areas, and the inferior pulvinar region. They showed plots of log-spirals in the receptive field when stimuli along straight line trajectories across these visual areas were inflicted. In addition, discoveries of Hubel and Wiesel [HW74] about the hypercolumn modeling of the striate cortex are consistent with the log-polar mapping from the radial lines of receptor cells to the parallel columnar structure in the striate cortex.

### Log-polar transform for image processing

The strength of the log-polar mapping is revealed in its role in form invariant image analysis. Researchers have recognized the perceptual functioning of log-polar mapping in its form invariance property in size and rotation [Fun77, Sch77, Sch80]. For

example, we do not have problem in recognizing a familiar face, whether it is near or far from us. Although the retinal stimuli are very different, the cortical projection is affected only to the degree of a single translation. The reasoning is delineated as follows. Suppose the retinal image is magnified by a factor $k$, the point $z$ is taken to the point $z'$. The cortical mapping $w$ will become $w'$, and the change in the cortical image is no more than a translation.

$$w' = \ln z' = \ln(k \cdot z) = \ln z + \ln k = w + \ln k$$

In their work [WC79], Weiman and Chaikin used the properties of logarithmic mapping in image processing and computer graphics. When the curvilinear logarithmic grid is used in place of the conventional rectilinear Cartesian coordinate lattice, the mathematical expressions for geometric transformations are greatly simplified.

Magnification and rotation of image patterns are the common operations in image processing and display. As these operations involve matrix multiplications on the homogeneous coordinate representation of the image points, they often demand a lot of CPU time and normally represent the bottleneck in the total computation. Weiman and Chaikin [WC79] demonstrated the useful property that translation in the logarithmic space yields magnification and rotation in the Cartesian space. Suppose the image data in the logarithmic space is shifted $k$ units to the right and $\phi$ units upward, the global translation to every point $w$ is $w + k + i\phi$. The effect in Cartesian space can be seen by taking each point $z$ to $z'$ such that

$$z' = e^{w+k+i\phi} = e^k \cdot e^w \cdot e^{i\phi} = e^k \cdot z \cdot e^{i\phi} \ . \tag{2.3}$$

It is apparent in eq. (2.3) that the modulus of each image point $z$ is multiplied by $e^k$ and the argument is incremented by $\phi$. The entire image is therefore magnified by a factor of $e^k$ and rotated through an angle $\phi$.

Weiman and Chaikin [WC79] also discussed the conformal property of the log-polar mapping. Write the mapping as $z(w)$ and its derivative as $z'(w)$. The fact that the derivative exists yields the Taylor's series expansion:

$$z(w) \simeq z'(w_o)(w - w_o) + z(w_o) \ . \tag{2.4}$$

Eq. (2.4) indicates a localized effect of a magnification by $\|z'(w_o)\|$, a rotation by $\arg z'(w_o)$, and a translation by $z(w_o) - w_o \cdot z'(w_o)$. Thus, if the image pattern involves grid cells in a small neighborhood, the shape of the pattern is virtually undistorted. Weiman and Chaikin argued that the property is desirable because operators which are rotationally symmetric such as Laplacian and smoothing operators retain their applicability. Dwelling on the property, Funt et al. [FBT93] demonstrated their result of color constancy computation in the log-polar transplant of the corresponding Cartesian version.

Despite the fact that the log-polar mapping has these desirable properties, Weiman and Chaikin [WC79] show that the image pattern and its directional quantities (such as first-order derivatives) will suffer scale and rotational changes. This renders image registration problems difficult once the key pattern for registering the image is not in fixation. Hence, it is not surprising that stereo correspondence becomes extraordinarily complicated in the log-polar domain [GLW92]. The RWT model presented in this thesis not only does not obscure stereo correspondence, but also simplifies the disparity computation to a restricted operating range.

Another disadvantage of the log-polar mapping with respect to the RWT model is that it complicates image translation. It is always desirable to be able to represent straight lines in the log-polar coordinates. Nevertheless, straight lines in the rectilinear Cartesian lattice cut through the log-polar curvilinear grid. The result is a set of successive logarithmic sine and cosine curves which render the computation

for translation extremely difficult (Figure 2.1). On the contrary, the RWT preserves linear structures and is thus suitable for processing image translations. In this thesis (also in [TL94, TL95, LTR95]), the applicability of the RWT to problems in motion stereo is demonstrated.

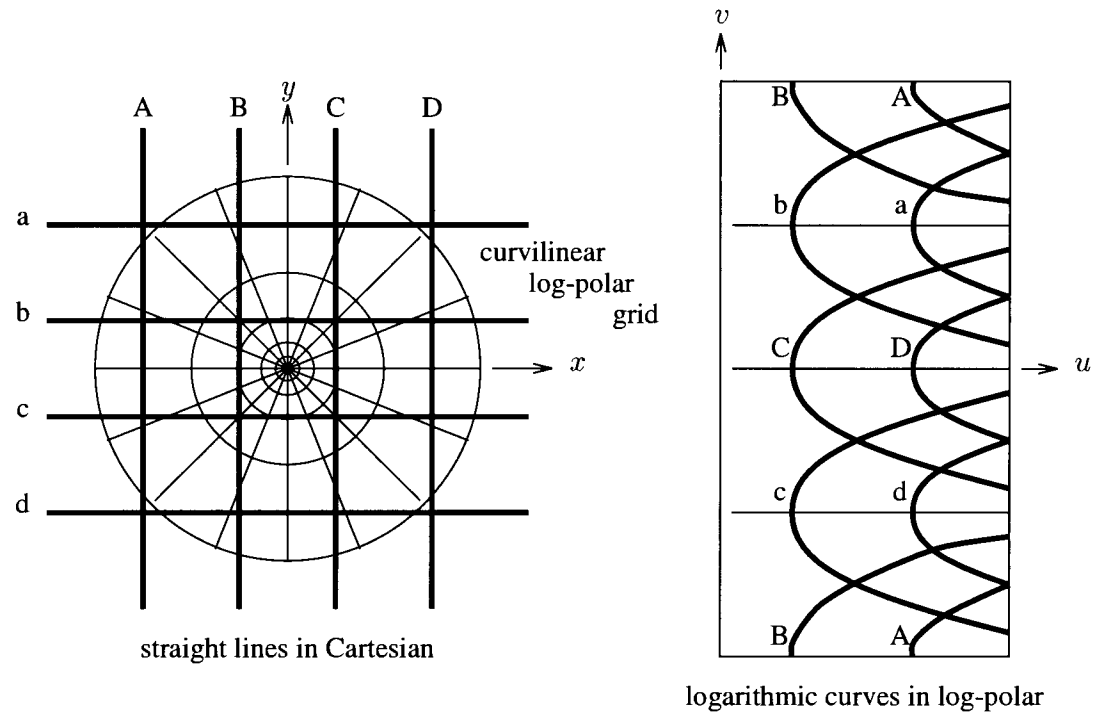



Figure 2.1: Images of straight lines under the logarithmic mapping.

### Considerations for logarithmic singularity

In [Sch80], Schwartz addressed the problem of log-polar mapping due to its divergence at the zero point. He proposed a linear function of eccentricity for the logarithmic mapping as the revised version of eq. (2.1):

$$w = \ln(z + a) \ . \tag{2.5}$$

The Taylor's series expansion of eq. (2.5) in the vicinity of $z = 0$ is equal to

$$\ln(z + a) \simeq \ln a + \frac{z}{a} .$$

As illustrated, the map is essentially linear for small $z$. The magnification factor is constant. For large $z$, the mapping is close to the complex logarithm. This new formulation of the retinotopic mapping supports a smooth map from a linear foveal representation to a complex logarithmic para- and peri-foveal surround. With appropriate choice of the linear constant $a$, Schwartz [Sch80] was able to achieve a good agreement of his model to the published data of the retinotopic mappings in a number of primate species. Design considerations on the number of pixels, the field radius and the shift parameter $a$ are investigated in [RS90]. The complex logarithmic sensor offers a good space complexity of about 1/50 the pixels of a uniform-resolution sensor while matching the field width and foveal resolution quality of the latter.

Problems of singularity at the zero point occur in our RWT formulation as well. In one of the variants to the RWT, the similar strategy of shifting the origin by a constant $a$ is adopted to cope with the divergence at the singularity.

### 2.2.2 The retina-like sensor

The retinotopic mapping has been implemented in a CCD array. Collaborated effort has been put together by the University of Pennsylvania, DIST in Italy and IMEC in Belgium to realize a prototype design of the retina-like CCD sensor called Retina [SD90, VdSKC+89, DBC+89]. The sensor comprises three concentric areas, each consists of 10 circular rows whose radii increase with eccentricity. 64 photosensitive sites are etched on each circle. The element size increases from $30 \times 30\ \mu m^2$ for the inner circle to $412 \times 412\ \mu m^2$ for the outer one. For design simplicity (in contrast

to [RS90]), the center of the chip is filled with 104 sensing elements measuring 30 × 30 $\mu m^2$. The elements are placed in a orthogonal pattern achieving the maximum resolution but uniform pixel size for the central fovea.

Complications arise because the sensors have to be read out in circular CCDs. Radial shift registers are devised to transport the charge from these circles. Special attention is devoted to obtain uniform sensitivity from the cells of variable sizes. Notably, in our RWT sensor, the problems due to circular CCDs are alleviated because rectangular tessellation is employed for the sensor array. The optical design rather than the variable sensor tessellation produces the space-variant resolution.

### 2.2.3 Space-variant sensing

As Bajcsy comments [Baj92], the nature of the information for visual processing changes in active vision. We no longer assume high quality data across the visual field, nor do we try to build a model of the world in one step. Instead, we adopt the role of active observer, moving the cameras around to gather information in interaction with the visual world. However, the cost of using foveate sensors is high since the new image space often requires re-adapting our vision tools from the Cartesian domain.[1] The gain is a drastic reduction in the data. Retina has a hundred times fewer pixels than a standard television camera. It also benefits from its form invariance functioning. Its use in active vision brings about a new and promising direction in visual processing.

In [ST80], Sandini and Tagliasco demonstrated the advantages of using anthropomorphic sensing features in operations in man-oriented environments. In robotics, because visual processing is normally performed for specific tasks, computer resources

[1] Although the differential and some other local operators have valid conformal transplants in the log-polar domain, in most cases, the image processing tools and vision algorithms (e.g. geometric transformations, stereo correspondence, etc.) indeed require re-definition of their meaning and usage in the new image space.

are normally employed to eliminate the irrelevant information in the acquired images. Thus, data reduction at the sensor level would support the efficiency and economy of visual processing. In their simulation, an efficient scheme involves a retina-like sensor which when directed to the attended field acquires a good amount of information about the relevant objects while achieving a preliminary reduction outside the fovea. A reduction ratio of about 30:1 was demonstrated in sample images of an industrial environment and a painting by Caravaggio. We dwell on the data reduction property of our RWT images as well. A reduction ratio in the order of 90% is also achieved in the application of our RWT to road vehicle navigation problems [TL94].

Yeshurun and Schwartz [YS89] exploited multiple fixations when building the representation of a scene through scanning using the log-polar sensor. Since resolution depends on the eccentricity, an image pattern has the highest resolution when the fixation point is placed close to it. They placed several fixation points $p = p_1, \cdots, p_n$ in different spots and produced frames with different resolution for the same image pattern. Their blending scheme then uses the "best" of each view to reconstruct the composite image. As the unified image of the scene is extracted from successive fixations, an attention algorithm is required to locate the fixation point for best information at each step. Yeshurun and Schwartz used the curvature of the contours in the scene as the criterion for fixation point "attractor". They showed that their algorithm exhibited a good convergence rate.

In our later example of binocular visual exploration, multiple fixations are devised to scan different objects in the scene. We adopt a similar strategy in determining our attention algorithm. Sizable objects lying away from the current fixation depth are considered the fixation point attractors.

### 2.2.4 Form invariant image analysis

Another thrust in exploiting the log-polar structure in visual processing capitalizes on the form invariance properties of the mapping. Sandini and other researchers carry these invariance properties to a great length in their applications in object recognition and motion analysis [SD90]. In the recognition task, Sandini and Dario matched the cortical map of the scene image against a pre-stored template. Because of the form invariance properties, one template for each object suffices irrespective of size and rotation. In another experiment, the observer is in ego-motion along its optical axis towards an object. The divergent optical flow in the retinal coordinates becomes globally consistent flow parallel to the horizontal in the cortical image. Detection of such global translation is greatly simplified. Earlier work by Jian et al. [JBO87] also exploits the convenient horizontal image motion in the log-polar mapping when computing depth from motion stereo. With the logarithmic mapping performed with respect to the focus of expansion, matching across frames is appreciably restricted to horizontal search windows. In [TS90], the advantage is reflected in the error analysis of depth from motion computation. Although the flow magnitude increases from the fovea to periphery in the retinal image, it is reduced to similar magnitude in the log-polar coordinates. The same accuracy is achieved throughout the field while the number of pixels to be processed is minimized. Young [You89] combined the use of both the Cartesian image and the log-polar map in object recognition. The method calculates the autocorrelation of the scene image to produce a position independent description of the object. Log-polar mapping of the result is essentially unaffected by the size and rotation variance.

In all applications, precise fixation on the pattern is required. This poses a limitation on the use of log-polar structure for eccentric stimuli processing. Problems such

as binocular fusion are complicated [GLW92]. The RWT provides an alternative to the log-polar transform for handling problems of eccentric image analysis. This thesis shows the use of RWT in disparity computation and binocular fixation.

## 2.3 Binocular Fixation

### 2.3.1 Stereopsis

Stereopsis results from the fact that each of a pair of eyes views the three-dimensional world at a slightly different vantage point. Consequently, the images falling on the retinas of the two eyes are slightly out of alignment from each other, giving rise to the phenomenon of binocular parallax. As the parallax is directly related to the spatial location of the object in relation to the two eyes, the re-alignment of the retinal images yields the sensation of the three-dimensionality of the world. In machine vision, cameras are used in place of the eyes. The parallax is measured in disparity between the two camera images. Exploiting the triangulation geometry in stereo imaging, Marr and Poggio [MP76] showed that depth information is recoverable from the disparity computation.

Stereopsis is one of the most studied areas in computer vision. Computer algorithms computing the stereoscopic disparity can be dated back to Marr and Poggio's work [MP76]. Disparities are computed as displacement of edge pixels between the left and right images. Matching for the corresponding but displaced edge pixels in the two images is a difficult problem. Marr and Poggio posed stereo correspondence as a minimization problem. Constraints for smooth surface and unique matches are imposed on the matching process. Other contributors to the area of research include [Gri85, MF81, BJ80a, BF82, OK85, Li94b, TL91].

Researchers have been attempting to develop computer algorithms for accurate disparity computation that will reconstruct the three-dimensional world from the stereo pair of images. Notwithstanding the persistent efforts of many fine researchers, the stereo correspondence problem still remains one of the difficult problems to be solved. The difficulty is perhaps due to the ambitious goal of total reconstruction of the physical world. Psychological studies in human visual perception have shown that many visual tasks are indeed exploratory in nature [Baj88, Bal91, AWB88]. This thesis, therefore, adopts the active perspective to stereo vision rather than the reconstructionist point of view.

### 2.3.2 Fixation

Although our fovea covers only some ten-thousandth of the visual field, we manage to achieve a vision as good as it would be if most of our retina were packed with the foveal receptors. The strategy is to have our eyes continually on the move, pointing the fovea at whatever we wish to see. Binocular stereo requires that both foveae simultaneously converge at the object of interest — a process called *binocular fixation* — to maximally exploit the foveal acuity for depth perception.

In human vision, the binocular fixation is accomplished by two components — *version* and *vergence* [Car77]. The version component is the *conjugate* movements of the eyes by which the gaze is transferred from one place to another, whereas the vergence movement, which converges the eyesight upon the new fixation point, is purely *anti-conjugate.*

### Version

Version is the conjugate movement of the eyes. Version movements are similar in amplitude and direction in the two eyes, and thus obey Hering's principle of "equal innervation" [Her68]. Pure version occurs when the gaze is transferred under zero disparity from one object to another. It requires that the two eyes maintain their convergence while panning synchronously at the same angle in the same direction.

Version is the fast saccadic movement of the two eyes. In fact, the movement is so fast that there is no time for visual feedback to guide the eye to its final position. Sometimes, the magnitude of the velocities can rearch more than 700° $s^{-1}$ for large amplitudes [Car77]. The duration of complete movement increases with increasing amplitude. For saccades larger than 5°, the duration is roughly given by 20 - 30 ms plus about 2 ms for every degree of amplitude [DC01, Hyd59, Rob64]. A rate of three saccades per second is normally observed in common visual problem solving [Bal91].

### Vergence

While pure version is associated with gaze transfer under zero disparity, pure vergence occurs when the lines of sight of the two eyes are converged or diverged under symmetric disparity. The vergence movement is initiated when the gaze is shifted from a distant object to a near one or vice versa. It is anti-conjugate in that the two eyes are rotated by the same amounts but in *opposite* directions. Contrary to version which is saccadic, vergence movements are visual guided and relatively slow.

As the version component is characterized by ballistic displacement, the vergence movement is quite a different behavior. In response to a step change in disparity, after some 160 ms latency time, the eyes move smoothly and comparatively slowly to their final positions [RW61]. The whole movement takes nearly 1 sec to complete.

The vergence system is believed to operate with intrinsic negative feedback because the movements are executed extremely accurately, in the sense that the final position of the eyes is within at most a minute or two of the vergence required for reducing the disparity to zero.

### 2.3.3 Oculomotor model

The strict division into pure version and pure vergence has led to the notion of an *oculomotor map* of visual space [Car77, Lun48]. Such a map is shown in Figure 2.2. It has the coordinates based on lines of equal version and lines of equal vergence. The latter (potentially called *isophores*) correspond exactly with the Vieth-Müller circles, which are a series of circles passing through the nodal points of each eye. They represent the fixations of equal disparity when the lines of sight are parallel. The lines of equal version, which could be called *isotropes*, form a series of rectangular hyperbolas whose center is the midpoint of the interocular base-line. Fixation shift from one point to another can be resolved into its versional and vergence components along these orthogonal coordinates.

A similar pattern of eye movements is seen when a subject shifts his gaze from one object to another [Yar57]. It starts with a slow symmetric vergence movement. A conjunct saccadic version is then superimposed in the middle of the course to bring the cyclopean axis in line with the target while the vergence movement is proceeding to completion. The sequence is shown in Figure 2.3.

To effect good vision over the entire visual field, it is essential to be able to direct the fovea at the objects of interest at various visual angles over the field. Gaze control, which is manifested in various patterns of eye movements, is an area of research in human perception. When a human subject is accomplishing a visual task, a scanpath

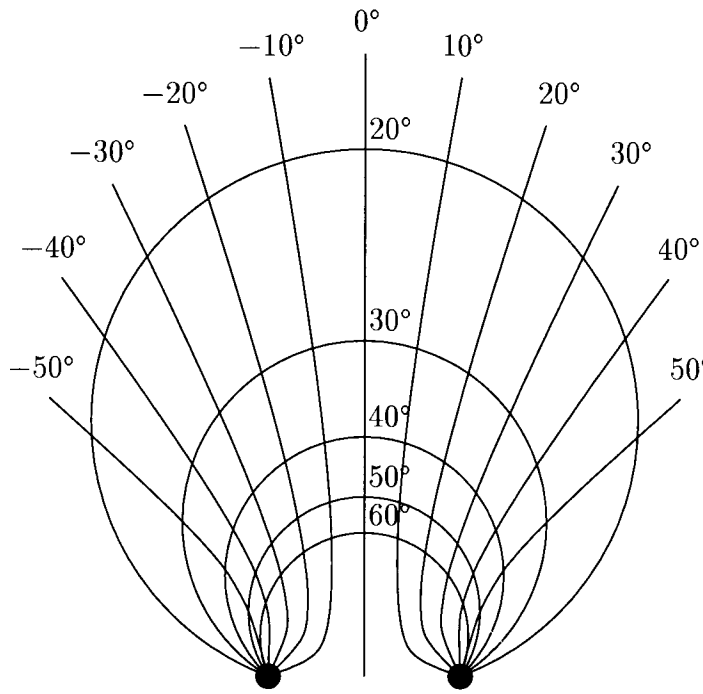



Figure 2.2: The oculomotor map of visual space.
The space coordinates are marked by lines of equal version (isophores) and lines of equal vergence (isotropes). The circular arcs are isophores and the rectangular hyperbolas are isotropes ([Car77, after [Lun48]]).

of eye fixations is normally observed to direct the gaze to a selection of objects in the scene to collect the necessary visual information. Extensive research by Yarbus [Yar67] demonstrates the goal-specific nature of scanpaths. In [NS71b], Noton and Stark postulated that memory of a pattern is formed in a sequence interleaved with eye movements during the recognition process. Eye movement is also shown to be critical for cognition. In Zinchenko and Vergiles's experiments [ZV72], subjects were found to be unable to solve many of the visual problems if they were not allowed to move their eyes.

In this thesis, a computational model for binocular fixation is investigated. It leads to the development and implementation of a fixation model for space-variant

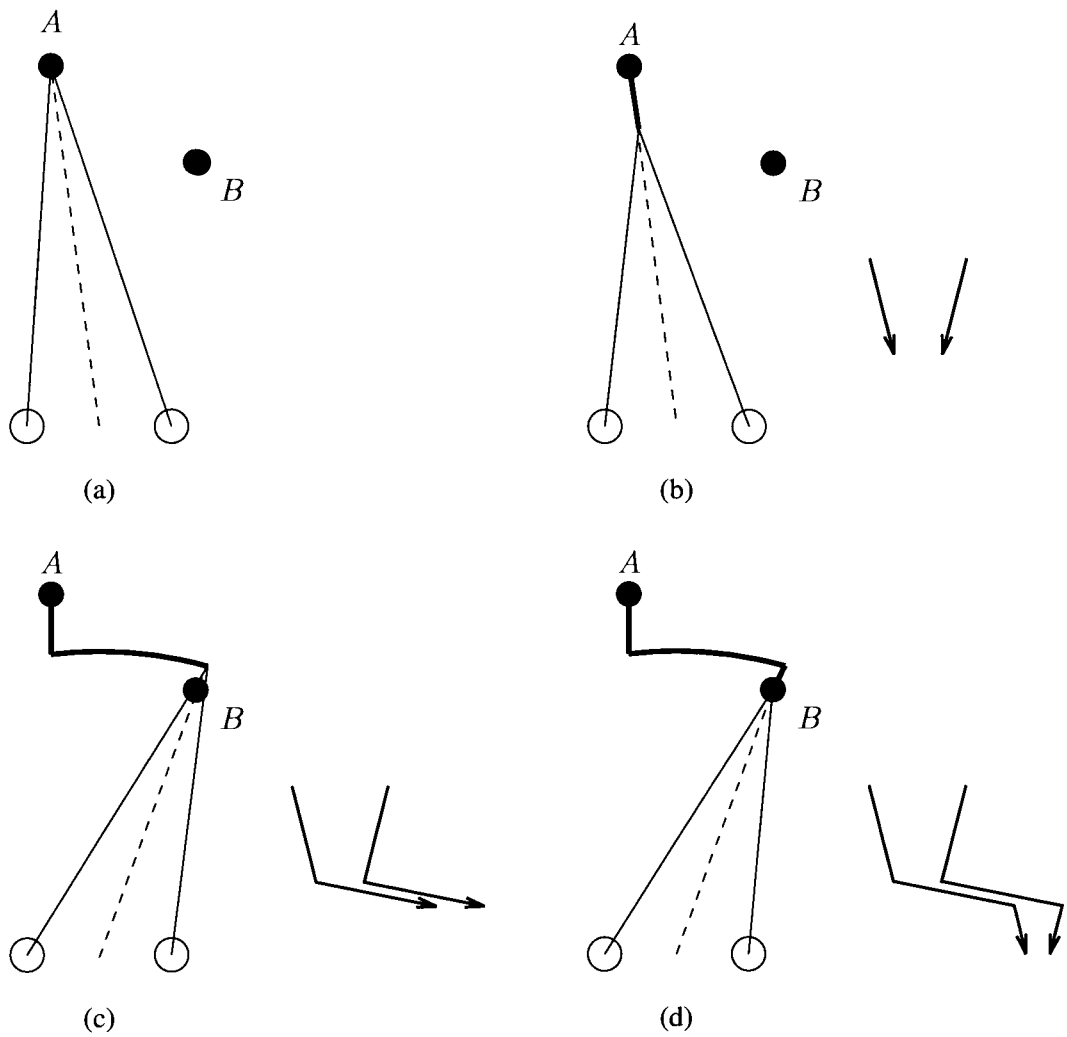



Figure 2.3: The sequence of events in a mixed version and vergence movement. The thick line on the left in each diagram shows the locus traced out by the point of fixation. The time course of the movement is shown on the right.

sensing using RWT. A scanpath experiment, inspired by the eye movement research, demonstrates the correct performance of our fixation system.

## 2.4 Advances in Stereo Verging Systems

In active visual following, the target is maintained at the center of the visual field, i.e., its retinal slip is minimized. In their experiments with the Rochester head, Coombs and Brown [CB92] studied the gaze holding problem in a dynamic environment. Binocular cue is used for vergence control. Once the cameras converge on the target, the near-zero disparity filter can isolate the target's image from the other scene objects. Smooth pursuit then keeps the target centered by tracking the centroid of the zero-disparity filtered window. Binocular disparity is used as a visual cue to vergence error in the cameras' vergence control. Disparity is computed using the cepstral filtering technique introduced in [BHT63]. A peak in the power cepstrum indicates the disparity which is then converted to the vergence angle.

Gaze control comprises both gaze holding and shifting. In active stereo, fixation is shifted from one point of attention to another. In our RWT fixation system, fixation is carried out in the stages of peripheral vergence, saccadic version and foveal vergence. This latter stage addresses the same issues as Coombs's vergence control. However, a simple correlation on foveal features is shown to be sufficient in our case.

Stereo problems are greatly simplified in verging systems because vergence control allows redistribution of the scene disparities around the fixation point, thus reducing the disparities over an object of interest to near zero. Olson [Ols93] presented a simple and fast stereo system that is suitable for the attentive processing of a fixated object. In view of the narrow limits of the Panum's area, the fusible range is thought to be a

privileged computational resource that provides good spatial information about the fixation point. Assuming vergence control, Olson's stereo algorithm capitalizes on a restricted disparity range. It gains from the slack demand for computation and allows selective processing via disparity filtering. The disparities are examined in multiple scales so that the system does not lose track of the rest of scene even though fixation is attended to the target of interest.

The Panum's area in Olson's system [Ols93] is a fixed narrow band around the Vieth-Müller circular horopter. Empirical data [Fis24, AOG32] indicate a spatially varying Panum's area. Our RWT Panum's area resembles the empirically observed one. The narrow Panum's region near the fovea is focused on the fixated target while the deep Panum's area in periphery is attended to the rest of the scene.

Vergence is guided by stereo disparity. Stereo correspondence, paradoxically, is difficult without fixation. An approach is to use other visual cues in cooperation with stereo disparity in guiding the binocular vergence.

Pahlavan, Uhlin and Eklundh [PUE93] developed their machine fixation model after the fixational behaviors in human vision. The vergence component in their KTH head-eye system is dealt with in accommodative and disparity aspects respectively. The accommodative vergence is driven by focusing which is measured with the gray-level variance. Correspondence is detected by calculating the normalized correlation on the centrally symmetric positions between the left and right images. The blur and disparity stimuli are then integrated to realize a cooperative effect on both accommodation and vergence of their KTH head. Incorporated with a stabilizing process with symmetric version movement, the vergence system was demonstrated with an experiment of real-time dynamic tracking of a moving person.

Krotkov and Bajcsy [Kro89, KB93] developed and implemented the idea of cooperative ranging in their agile stereo camera system [KSF88]. Accommodation and vergence alone are weak depth cues [Gra65, Gog61]. Krotkov's system demonstrates the reliability in ranging upon fusion of the focusing and stereo vergence components. Initially, a focusing procedure computes the gross depth of the target scene feature from the master camera. Based on that result, the vergence angle is calculated to servo the fixation of both cameras on the target. Then execution is split into two paths. One path performs stereo ranging with verification by focusing. The other performs focus ranging. The operating windows on both cameras are related by the disparity predicted from the focused depth. Improved reliability is successfully demonstrated by sensor fusion at the level of data acquisition. This form of cooperation exhibits visual behaviors analogous to human accommodative-convergence and convergence-accommodation at various steps.

Grimson at al. [GLROK94] used color in cooperation with stereo cues. In their work, they demonstrated how focus of attention is used to support the high level task of efficient object recognition. Color is used for fast indexing to the region of interest. Its use is combined with stereo cues to yield the disparity of the selected region. By converging the cameras accordingly, attention is directed to it. A second stereo matching within a narrow disparity range completes the figure/ground segmentation to un-clutter the scene for object recognition. The rationale is that both correspondence and model matching would be significantly impeded if the scene were cluttered.

Abbott and Ahuja [AA93] took integration of visual cues to great length in their University of Illinois Active Vision System. Complementary strengths of different cues are exploited in integration via active control of camera focus and orientation,

as well as aperture and zoom settings, thus coupling image acquisition and surface estimation dynamically and cooperatively in an active system. The idea agrees with the active approach of intelligent data acquisition [Baj85]. Two phases are involved in the process, namely fixation selection and surface reconstruction. Fixation selection is posed as an optimization problem that seeks to minimize large camera movements and develop the surface description outward from the current fixation, favoring the unexplored area. Based on Sperling's energy model [Spe70], the surface reconstruction is formulated to optimize among different cues of focus, disparity, surface smoothness. The objective function also includes the image contrast and disagreement among the cues and fixations. By selecting fixations to extend smoothly the evolving surface map, their implementation produces dense depth information for a deep and wide visual field.

Our active stereo ranging also employs the idea of active, intelligent data acquisition. Fixation favors conspicuous objects in the periphery. The range information is evolved to more accurate levels from different fixations.

## 2.5 Non-frontal Imaging

In our binocular verging system, the RWT cameras represent a non-frontal imaging device since the sensor surface is not assumed to be in a conventional frontal orientation. In our camera for imaging the road scene in a vehicle navigation problem [TL93], a horizontal sensor plane offers the RWT a spatially varying resolution that offsets the perspective distortion. This thesis will present a more elaborate non-frontal camera model for RWT space-variant imaging in Chapter 4.

Although not aimed to achieve space-variant sensing, Krishnan and Ahuja [KA94]

developed a non-frontal camera model for ranging using focusing. The non-frontal imaging geometry is exploited in the way that varying image distance from the optical center to the sensor plane occurs at different viewing angles. When the camera is panned across the scene, an object will be imaged at different angles. At one of these viewing angles during the course of panning, the image distance will be just right to produce a sharp and focused image of the object.

In Krishnan and Ahuja's camera, the sensor plane is equipped with three degrees of freedom. It can be translated, and rotated in two axes. Making use of the positioning and orientation of the sensor plane, up to three object points in the scene can be focused simultaneously. When the camera is swept across the scene, a series of images are generated. Each point in the scene will be imaged in focus at one instance or another. Therefore, the image series can then be analyzed to determine the sharply focused regions, the union of which will produce a composite focused image of the scene in a wide and deep field.

The camera can be used to obtain range from focusing as well. When the focus criterion function (such as [Kro89, LG82]) reaches its maximum for a scene point, the parameters such as the pan angle, the objective lens' focal length and the sensor's position and orientation are used to determine the range value using the range from focus methods [Pen87, EL93, KA93]. Problems of variation in the registered brightness and perspective warping are corrected at different imaging positions.

## 2.6 Directions in Active Vision Research

The National Science Foundation Active Vision Workshop held in 1991 set out the directions in active vision research [SS91]. The attendees laid down five major research

areas include attention, foveate sensing, gaze control, eye-hand coordination, and integration of vision with robot architectures.

This research fits in the picture because the RWT developed here provides a model for foveate sensing. Motion stereo is studied in this sensing model and the fixation mechanism for an RWT binocular system is presented. The system is suitable for research into scanpath behaviors in attentive processing. It also promises applications in vision-based tasks for situated robots.

# Chapter 3

# Reciprocal-Wedge Transform

## 3.1 The Mathematical Model

The Reciprocal-Wedge transform (RWT) was proposed as an alternative model for space-variant sensing [TL93]. The RWT maps a rectangular image into a wedge-shaped image. Spatially varying resolution is achieved as the smaller end of the wedge is sampled with fewer pixels than the wider end is. Mathematically, the RWT is defined as a mapping of the image pixels from the $x$-$y$ space to a new $u$-$v$ space such that

$$u = 1/x\ ,\quad v = y/x\ . \tag{3.1}$$

The lady's image in Figure 3.1 is used to illustrate how the Cartesian coordinates are mapped back and forth[1] to the RWT domain. The transformed image in Figure 3.1(b) shows a wedge-shape in an inside-out fashion because of the scaling effect of the $x$ reciprocal. Note the blurring at the periphery of Figure 3.1(c). In Figure

[1]Singularity occurs in the transform at $x = 0$ (the center strip). A variant of the RWT, which will be discussed in Section 3.1.2, was used in Figure 3.1 to cover the whole image including the center region.

3.1(d–f), the grid image is a template used to demonstrate the variable resolution of the transform. It is the differential magnification ratio across the width of the image that facilitates the continuously changing scale of image resolution from the center to the periphery.

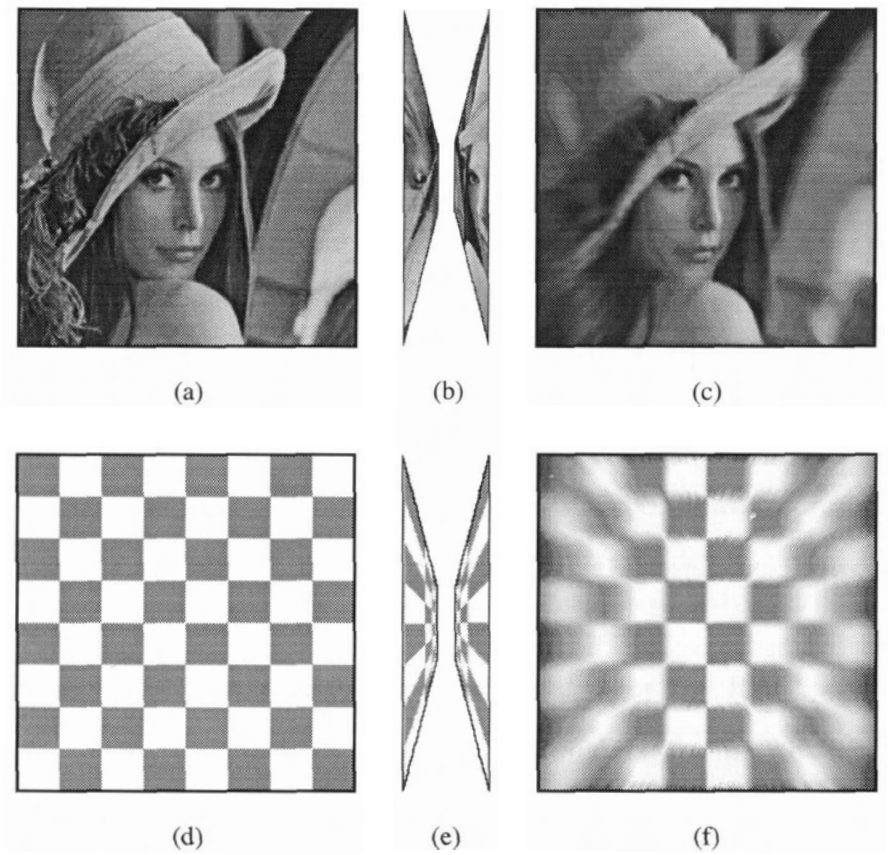



Figure 3.1: The Reciprocal-Wedge transform.
(a) The lady's image. (b) The RWT image shows two inside-out wedges. (c) The image when transformed back to the Cartesian domain. (d) A rectangular grid. (e) The RWT image. (f) The grid transformed back to illustrate the resolution varying from the center to the periphery.

### 3.1.1 Matrix notation

A concise representation for the transformation is derivable using the matrix notation. Adopting the homogeneous coordinates, the RWT defined in eq. (3.1) can be formulated as a cross-diagonal matrix of 1's, and the transformation can be computed as matrix operations.

$$\mathbf{T} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \mathbf{T}^{-1} , \tag{3.2}$$

$$\mathbf{w} = \mathbf{T} \cdot \mathbf{z} , \quad \mathbf{z} = \mathbf{T}^{-1} \cdot \mathbf{w} .$$

where $\mathbf{T}$ is the transformation matrix, $\mathbf{z} = [x\ y\ 1]^t$ and $\mathbf{w} = [u\ v\ 1]^t$. To elaborate,

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/x \\ y/x \\ 1 \end{bmatrix} \simeq \begin{bmatrix} 1 \\ y \\ x \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} .$$

The sign"$\simeq$" means equality within the homogeneous coordinate representation.

It is interesting to observe that the inverse of $\mathbf{T}$ is $\mathbf{T}$ itself, i.e., both the forward and backward transformations have the same matrix form.

The concise matrix notation yields an advantage for the RWT. Coupling their geometric transformation matrices with the RWT matrix, geometric transformations in the RWT domain become rather straightforward. If $\mathbf{M}$ is the transformation matrix in the $x$-$y$ space and $\mathbf{M}'$ is the corresponding matrix in the $u$-$v$ space, then

$$\mathbf{M}' = \mathbf{T} \cdot \mathbf{M} \cdot \mathbf{T}^{-1} .$$

Using rotation, translation and scaling as examples, it is well-known that the respective matrices $\mathbf{M}$ are:

$$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

Since both $\mathbf{T}$ and $\mathbf{T}^{-1}$ are cross-diagonal matrices of 1's (eq. (3.2)), their effect on $\mathbf{M}$ involves only row and column interchange. Thus, the respective matrices for the RWT domain can easily be derived as:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ t_y & 1 & 0 \\ t_x & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_x \end{bmatrix}.$$

Figure 3.2 shows the direct application of the transformation matrices in the RWT domain. In Figure 3.2(a), the matrices are applied directly in the $u$-$v$ space. For visual apprehension, the $x$-$y$ representation of the transformed results is reconstructed in Figure 3.2(b) to demonstrate the effects of the three matrices.

### 3.1.2 Remedy to singularity

The singularity of the RWT exists at $x = 0$, i.e., $u = 1/0 = \infty$ and $v = y/0$. Two remedies to the problem are proposed: patching and shifting.

Assuming the origin of the $x$-$y$ space is at the center of the image, the *patching* method provides an expedient fix to the singularity problem. The method excludes a strip of width $2\delta$ at the center, where $x$ value is zero or near zero, from the range of the RWT. The center strip from the original uniform-resolution image is then used to patch up the two wedge images from the RWT.[2] It is convenient in many cases to use a uniform-resolution model for the fovea because it is essential to maintain a high acuity within the extent of the fovea for most visual behaviors. Besides, the rich repertoire of existing computer vision techniques could be used for foveal processing.

The *shifting* method is an alternative way of fixing the singularity problem. It is to introduce a shift parameter $a$ in the RWT.[3] This variant formulation is called

[2]The log-polar transform also has the singularity problem at $r = 0$. A uniform-resolution patch at the center of the image is constructed in the prototype camera [VdSKC+89, KVdS+90].

[3]A similar shift parameter is also used in log-polar transform to the same effect [Sch80, RS90].

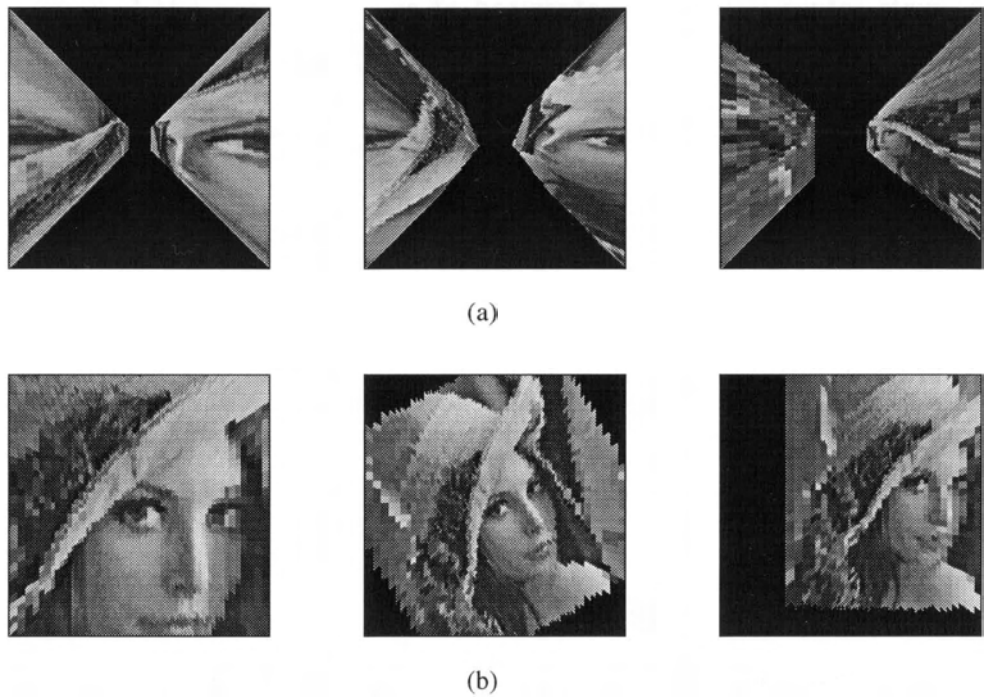



Figure 3.2: Geometric transformations on $u$-$v$ images.
(a) Direct application of the scaling, rotational and translational transformations on the $u$-$v$ lady's image. (b) The $x$-$y$ representation of the transformation results are reconstructed for visual apprehension of the effects of the scaling, rotation and translation.

*Shifted Reciprocal-Wedge Transform (S-RWT)*[4] [TL93].

$$u = 1/(x+a)\ , \quad v = y/(x+a)\ . \tag{3.3}$$

Both the forward and backward transformations for the S-RWT remain the same cross-diagonal matrix (eq. (3.2)) except the additional parameter $a$.

$$\mathbf{T} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & a \end{bmatrix}\ , \quad \mathbf{T}^{-1} = \begin{bmatrix} -a & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}\ .$$

[4]In fact, S-RWT has been used for the transformation in Figure 3.1 to take care of the singularity inherent in the original RWT equations (eq. (3.1))

The effect of the parameter $a$ is to horizontally shift the center strip (and the rest of the image) away from $x = 0$, or equivalently, shift the $x$ axis in the Cartesian image. The parameter $a$ should be of opposite sign for the left and right halves of the Cartesian image, i.e., the two halves of the image are respectively shifted in opposite directions. There is an advantage with the inclusion of the shift parameter in the S-RWT. As the space-variant resolution in RWT is caused by the $x$-reciprocal function (eq. (3.1)), the use of $a$ on $x$ in eq. (3.3) allows adjustment on the speed of changing scale of the resolution from fovea to periphery. Meanwhile, since $a$ leads only to a horizontal shift in the Cartesian image, for simplicity we can still use eq. (3.1) for the RWT for analysis of its properties.

It is not difficult to see that a combination of both patching and shifting can be adopted to take advantage of both techniques. Each of the single techniques can then be viewed as a special case where either $\delta = 0$ or $a = 0$. Our camera design in Section 4.4.2 will readily accommodate all these choices.

### 3.1.3 The RWT View-of-World

We now examine the effects of the forward and backward RWT. (The patching method is used for illustration in Figure 3.3. The S-RWT or the combination of the patching and shifting methods would yield similar results.)

Figure 3.3(a) depicts the effect of the forward RWT ($\mathbf{T}$). Excluding the strip of width $\delta$, one half of the rectangular $x$-$y$ image is turned into a wedge in an inside-out fashion because of the scaling effect of the $x$ reciprocal. Figure 3.3(b) shows the reassembled version which comprises the two halves of the RWT image and the center patch for the purpose of visual apprehension. The reassembled version is also referred as the *bipolar* representation of the RWT image because the origins for the left and

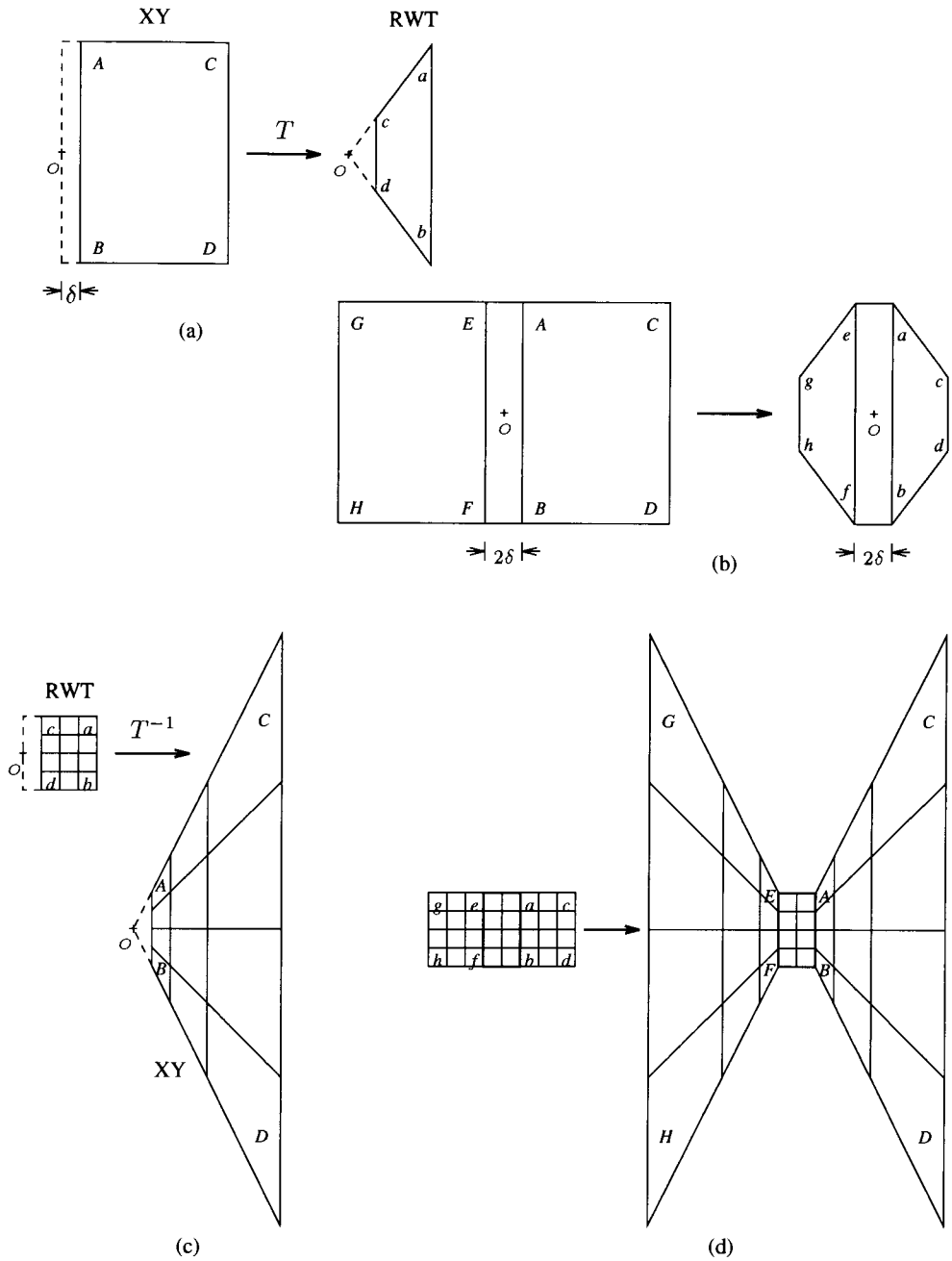



Figure 3.3: The RWT View-of-World.
(a) Forward RWT on a half-image. (b) A rectangular $x$-$y$ image is turned into a bipolar RWT image with a center patch. (c) Backward RWT on a half-image. (d) A rectangular RWT image corresponds to the RWT View-of-World in the $x$-$y$ domain.

right half-spaces are independently flipped to the two antipodes. As shown, the two pieces of the wedges have been properly flipped before the merging.

Figure 3.3(c) depicts the backward RWT ($\mathbf{T}^{-1}$). Since $\mathbf{T} = \mathbf{T}^{-1}$, Figure 3.3(c) and 3.3(a) appear similar, except that the RWT images in both cases are much smaller because of the data reduction. Nevertheless, Figure 3.3(c) reveals that an RWT rectangular region corresponds to a wedge-shape area in the $x$-$y$ space. Figure 3.3(d) shows the complete mapping including the center patch, the resulting image in the $x$-$y$ space is the RWT *View-of-World (VOW)*. The RWT-VOW is the effective space-variant view from a RWT camera using square/rectangular sensing elements. The center part (fovea) of the VOW obtains the highest resolution, which drops rapidly towards both sides (periphery).

Figure 3.4 illustrates how images in the Cartesian coordinates are mapped to the RWT domain, and then mapped back. The lady's image in Figure 3.4(a) is the original image (resolution 400 $\times$ 200) in the $x$-$y$ space. The combination method is applied where $\delta = 5$ and $a = 30$. The transformed image in Figure 3.4(b) shows the two wedges. The image is reduced to approximately 10% of its original size. Figure 3.4(c) shows the bipolar representation of the RWT image. Note the nice feature that the bipolar image is continuous at the two borders of the patch. Figure 3.4(d) is the restored lady's image. The blurring at the periphery is due to the inevitable (and desirable) loss of details after the image was reduced by the RWT.

In Figure 3.4, a grid image is also provided to clearly demonstrate the extent of the spatially varying resolution produced by the transformation. A continuously changing scale of resolution from the center to the periphery across the width of the image is supported.

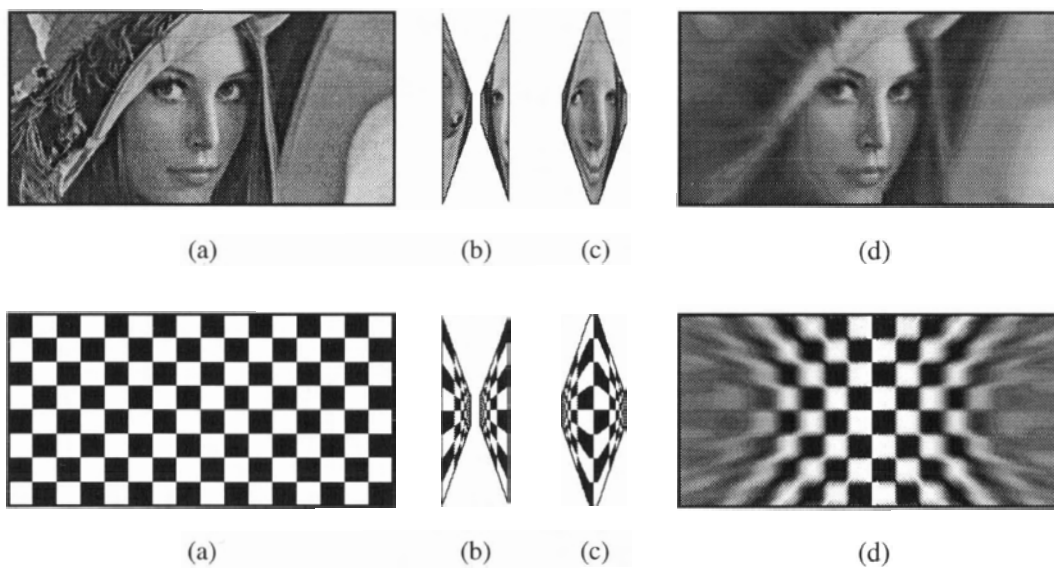



Figure 3.4: The Reciprocal-Wedge transform under the RWT VOW.
(a) The original image. (b) The RWT image shows two inside-out wedges. (c) The bipolar RWT image including the center patch. (d) The restored image when transformed back to the Cartesian domain.

## 3.2 Transformation on Linear Structures

Exploiting the polar coordinate representation, the log-polar transform gracefully simplifies the computation of centric transformations. Rotation and scaling about the origin become operations along the $\log r$ and $\theta$ axes. However, the polar coordinate representation adversely obscures other geometric patterns. For instance, linear structures get mapped to complicated patterns of logarithmic sine curves. Since both linear features and translations are very common in image analysis, this seems to be a major drawback of the log-polar transform.

RWT, on the contrary, does not employ the polar coordinates. It does not perform as well in computation of centric transformations. However, linearity of lines in the $x$-$y$ domain is preserved over the transformation. Furthermore, we argue that the RWT does not complicate curves in general. If a curve is represented with a polynomial, the degree does not change after the transformation.

### 3.2.1 Preservation of linear features

Lines exhibit interesting properties in the RWT. In fact, the following transformation dual ($L_{xy}$ and $L_{uv}$) of a line can be derived:

$$L_{xy} : y = m \cdot x + c \ , \quad L_{uv} : v = c \cdot u + m \ . \tag{3.4}$$

Given $L_{xy}$, the equation for $L_{uv}$ is readily obtained by substituting $x$ and $y$ in $L_{xy}$ with $1/u$ and $v/u$ respectively. It is obvious that the transformed structure $L_{uv}$ is also a line, which implies that the linearity of the line is preserved.[5] It is interesting to note that the values for the slope and intercept are interchanged between the transformation dual. Inferring from that, parallel lines with the same slope in $x$-$y$ will be

[5]Linear features are also preserved in the S-RWT. A line $L_{xy} : y = m + c$ is mapped to a line $L_{uv} : v = (c - ma)u + m$.

mapped to $u$-$v$ lines concurrent at the same $v$-intercept. Inversely, lines concurrent at the same $y$-intercept will form parallel lines in the $u$-$v$ domain.

**Extension to curves.** Let a curve in $x$-$y$ be denoted as:

$$\sum_{i=0}^{n}\sum_{j=0}^{n-i} a_{i,j}x^i y^j \;=\; 0\ .$$

By substituting $1/u$ for $x$ and $v/u$ for $y$, and rewriting the indices, the polynomial in $u$-$v$ becomes:

$$\sum_{i=0}^{n}\sum_{j=0}^{n-i} a_{(n-i-j),j}u^i v^j \;=\; 0\ .$$

This shows that the degree of the polynomial is preserved over the transformation. The shape of the curve may be different in the transform domain as the coefficients have been interchanged. For instance, a circle in $x$-$y$ would be mapped to an ellipse in $u$-$v$. (It would be a hyperbola or parabola should the circle be transversed by the $y$-axis.) The significance is that the RWT does not complicate curve patterns. In comparison, after the log-polar mapping, the resulting curve no longer keeps its polynomial form. One disadvantage is that undesirable complexity is introduced when problems of shape analysis or image data modeling are dealt with.

### 3.2.2 Line detection using the Hough transform

The Hough transform [DH72] provides a powerful tool for feature detection. The technique is most effective for line detection [TL92]. The preservation of the linearity of lines over the RWT implies that line detection using the Hough transform would be as simple in the RWT as in the Cartesian domain. With the switching between the slope and intercept parameters (eq. (3.4)), the vote patterns in the Hough space for the Cartesian and the RWT images form an interesting dual of reflection about

the main diagonal of the Hough space. (See Figure 3.5(c)).

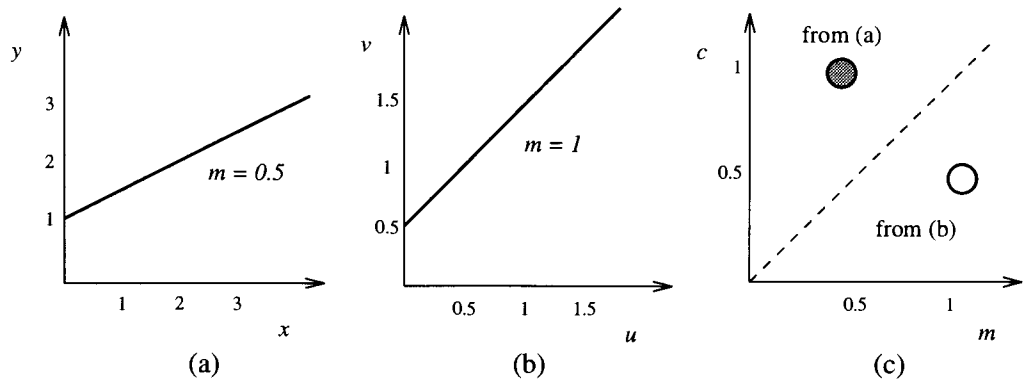



Figure 3.5: The duality relationship of linear structures in the RWT. (a) A line in the $x$-$y$ domain with a slope 0.5 and the intercept 1. (b) The dual in the $u$-$v$ domain. The slope is 1 and the intercept is 0.5, inversely. (c) The Hough space showing the peaks from (a) and (b) respectively. They form a reflection about the main diagonal.

## 3.3 Anisotropic Space-Variant Resolution

Like the log-polar transform, the RWT facilitates space-variant sensing which enables effective use of variable-resolution data and the reduction of total amount of the sensory data. Nevertheless, the variable resolution supported is anisotropic.

The essence of the RWT is the introduction of the reciprocal transformation. The variable resolution is primarily embedded in the $x$ dimension. It yields a grid whose resolution is variable for different $x$'s, but uniform along the $y$ dimension for any fixed $x$. The result is an anisotropic space-variant resolution, which is evident from the wedge-shaped grid in Figure 3.1(e).

The anisotropy can also be inferred from the partial derivatives of the RWT expressions from eq. (3.1):

$$\|\partial(u,v)/\partial x\| \;=\; \|(\partial u/\partial x, \partial v/\partial x)\| \;=\; \sqrt{(1+y^2)}/x^2\,, \qquad (3.5)$$

$$\|\partial(u,v)/\partial y\| \;=\; \|(\partial u/\partial y, \partial v/\partial y)\| \;=\; 1/x\ , \tag{3.6}$$

where $\|\cdot\|$ denotes the vector norm. Eqs. (3.5) and (3.6) show that the pixel resolution does not vary in the same manner for different directions. The grid width in the $x$ direction (for a fixed $y$) is mapped to a size diminishing in reciprocal of $x^2$. In the $y$ direction, the grid height is mapped by a function of $1/x$ to a uniform size independent of the $y$ value. Furthermore,

$$\partial(u,v)/\partial(x,y) \;=\; \begin{vmatrix} \partial u/\partial x & \partial v/\partial x \\ \partial u/\partial y & \partial v/\partial y \end{vmatrix} \;=\; \begin{vmatrix} -1/x^2 & -y/x^2 \\ 0 & 1/x \end{vmatrix} \;=\; -1/x^3\ .$$

Hence, the absolute value of the above Jacobian determinant is

$$|\partial(u,v)/\partial(x,y)| \;=\; 1/x^3, \tag{3.7}$$

which indicates that the area of a pixel is reduced by a factor of $1/x^3$ after the RWT.

On the contrary, the log-polar transform provides an isotropic variable resolution. The grid when mapped to the log-polar image changes size in the same scale in all directions. Sampling along the radial direction, the rate of change of the pixel resolution is

$$\|\partial w/\partial r\| \;=\; |d(\log r)/dr| \;=\; 1/r\ .$$

The area of a pixel thus diminishes isotropically in the rate of $1/r^2$.

The log-polar transform benefits from its conformal mapping. As the differential and local operators have valid conformal transplants in the log-polar domain, the related image processing tools and vision algorithms can also be available for processing in the log-polar domain with minimum overhead [WC79]. Compared to the log-polar, the RWT is neither conformal nor isotropic. The RWT, however, can also benefits from its matrix representation. Its matrices facilitate convenient mapping of linear transformations from the Cartesian to the RWT coordinates. As a result,

the established linear transformations for the Cartesian image processing are readily applicable in the RWT as well. (The application of geometric transformations on the RWT images has been demonstrated in Section 3.1.1.)

In the log-polar, the isotropic mapping facilitates the form invariance properties for centric patterns. In the RWT, it is the anisotropic mapping that enables the directionally biased RWT variable resolution. The directional variable resolution does not only benefit linear feature processing, but is also generally suitable for problems of translational in nature, such as motion stereo and binocular disparity computation.

Hence, the anisotropic mapping of the RWT makes it distinguished from the log-polar transform. On one hand, it is comparable to the log-polar for its space-variant resolution and data reduction. On the other hand, it is complementary to the log-polar for its suitability for linear transformations, lines and translations.

## 3.4 Pyramidal Implementation

The capacity for parallel processing and versatility of multiple resolutions have made the pyramidal architecture a widely adopted structure for fast image processing and parallel modeling for various visual processes. The Image Understanding Architecture [WB91] is an ambitious project on a three-dimensional pyramidal architecture. However, the two-dimensional pyramids have their advantages of cost-effectiveness and flexibility. The SFU hybrid pyramid [ELT+92] is a heterogeneous system offering the versatility in both iconic [LZ93] and functional [Li91] pyramidal mappings. In this section, a pyramidal implementation on the SFU pyramid for fast generation of RWT images is presented.

### 3.4.1 Pyramidal mapping

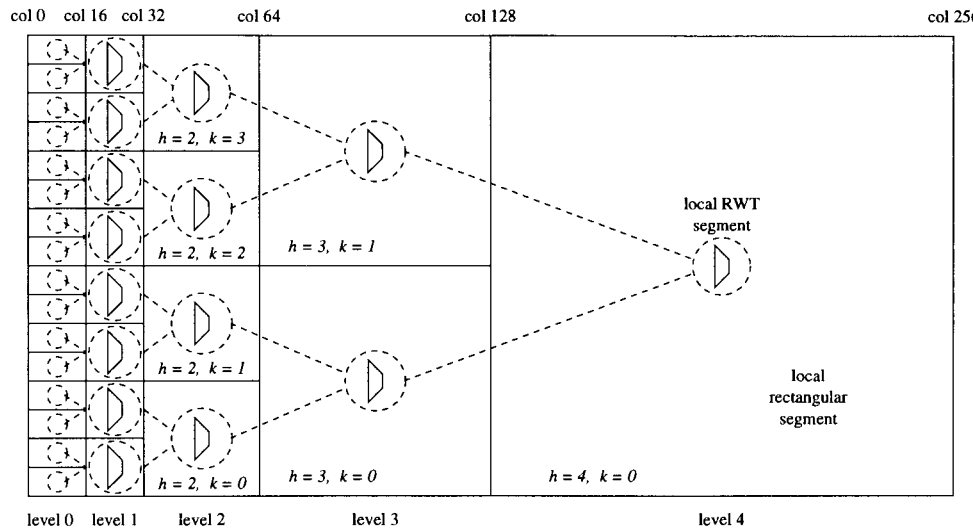



Figure 3.6: Mapping the image space to the pyramid.

The 2-D image space is conveniently mapped to the 2-D pyramidal structure in a way that exploits the more abundant computing power in the bottom of the pyramid for the image fovea. As an example for illustration, let us assume the entire image has a size of 1024 pixels across and 256 pixels down. The center strip of 32×256 is the image fovea. The two half images are thus 512×256 each with a 16×256 strip for the fovea. Here, the RWT singularity is handled by using the patching method. The patching method was chosen without particular preference. As a matter of fact, the shifting method is equally implementable.

In a global view, the rectangular image space is mapped to the pyramid as shown in Figure 3.6. In the figure, the left half of the image is shown, and our discussion on the implementation will be based on the left half-image only. Since the right half-image is symmetrically mapped to the pyramid in the same way, its implementation is exactly the same. The SFU pyramid has 63 T-800 transputer nodes inter-connected

in a binary tree. Together they form a two-dimensional pyramid of 6 levels. For illustration, a simpler version of 5 levels are used in the explanation. The dotted circles and arcs in Figure 3.6 are showing the 5 levels of the pyramid. The bottom level is reserved for uniform-resolution processing for the image fovea. In the figure, they are the level 0, and are not participating in the RWT image generation.

The pyramid nodes and their corresponding image blocks are labelled with the $h$ and $k$ indices. The $h$ index is related to the level number of the node, and the $k$ index is the node's position within the level. Refer to Figure 3.6 for the $k$ ordering of the nodes within different levels.

The pyramidal algorithm can be described in 2 steps. First, A pyramidal reduction process transfers the image segments up the pyramid from the bottom level. At each node, the image segment is reduced to half the resolution of the one from below. In the second step, each node performs the transformation to obtained an RWT segment for the local node. When the pyramid program is running, these 2 steps are actually pipelined together.

### 3.4.2 Pyramidal reduction

The rectangular image is loaded onto the pyramid from the bottom level up. Our mapping algorithm partitions the rectangular image space into segments of size of two's powers, and distributes the segments to the pyramid nodes in the way depicted in Figure 3.6. The segments, however, are stored in reduced resolution at each level. When the image is first loaded onto the level 0 nodes, each of the nodes gets a ribbon of 512×8. These foveal nodes then retain a block of 16×8 as the uniform-resolution fovea. The rest of the ribbon is then passed to the parent at level 1.

From level 1 on, the nodes are in the variable resolution region. They are involved

in the RWT image generation. Now, each level 1 node merges the 2 8-ribbons from its children to generate a 16-ribbon. A 16×16 block is retained as a local segment. This segment will get mapped to the RWT image in the later step. The merge operation can be formulated as follows:

$$A^1{}_{i,j} = \begin{cases} A_1{}^0{}_{i,j+16} & \text{if } i < 8 \\ A_0{}^0{}_{i-8,j+16} & \text{otherwise} \end{cases}$$

where $A^h{}_{i,j}$ is the image segment at the level $h$, and $A_k{}^{h-1}{}_{i,j}$ is the $k^{th}$ child the level $h-1$. It is the right child when $k = 0$ and the left child when $k = 1$. Figure 3.7 presents a graphical description of this step. The segment of the level $h$ node is a merged version from both children at the level $h-1$.

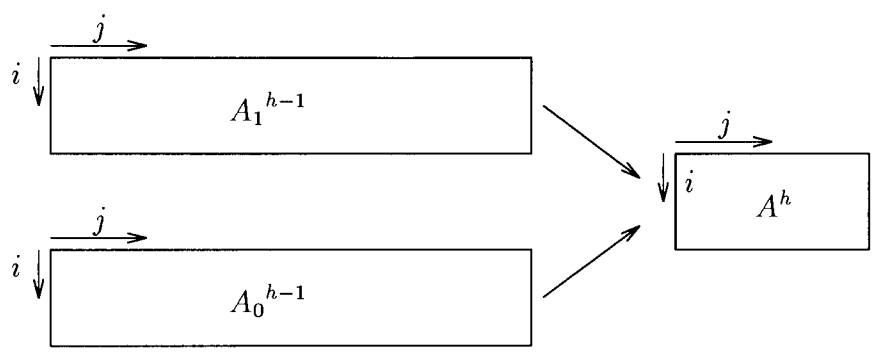



Figure 3.7: The pyramidal reduction step.

From level 2 on, every node takes 2 16-ribbons from its children, reduce-merges them into a 16-ribbon at half resolution. Again, a 16×16 block is retained for the local segment (Figure 3.7). The reduce-merge operation can be formulated as follows:

$$A^h{}_{i,j} = \begin{cases} A_1{}^{h-1}{}_{2i,2j+16} & \text{if } i < 8 \\ A_0{}^{h-1}{}_{2i-16,2j+16} & \text{otherwise} \end{cases}$$

At the end of this image distribution phase, each node has a 16×16 local segment of the original image. These segments have different resolutions according to their

levels in the pyramid. The segment at level $h$ represents a $16 \cdot 2^{h-1}$ square portion of the original image. Of course, this portion is stored in a reduced resolution at the size of 16×16.

### 3.4.3 Local RWT transformation

Having received its local segment, each node can perform a local RWT transformation on its local data, producing a segment of the entire RWT image. Before deriving the algorithm for the local RWT, we have to clarify the local image coordinates and how they are related to the global ones.

Right now, the 16×16 local block is indexed by $i$, $j$ for the rows and columns as shown in Figure 3.8. First of all, we set the local origin at the midpoint of the left edge, and the 2 axes as $x$ and $y$, like that in Figure 3.8. The local coordinates are not specified by $(x, y)$ where $x$ ranges from 0 to 15 and $y$ ranges from -8 to 7.

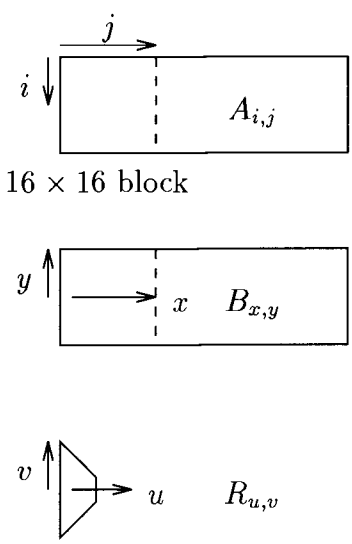



Figure 3.8: The RWT transformation step.

Let us denote the global coordinates with $(\tilde{x}, \tilde{y})$. From the recursive structure of

the pyramidal reduction, the local coordinates $(x, y)$ can be related to the global $(\tilde{x}, \tilde{y})$ as follows:

$$\tilde{x} = 2^{h-1} \cdot x + 2^{h-1} \cdot 16 \ , \tag{3.8}$$

$$\tilde{y} = 2^{h-1} \cdot y + 2^{h-1} \cdot 8 \cdot (2k - 2^{m-h} + 1) \ . \tag{3.9}$$

where $m$ is the height (maximum level) of the pyramid.

Let $(u, v)$ be the coordinates for the global RWT image. The global transformation is (eq. (3.1))

$$u = \frac{16^2}{\tilde{x}} \ , \quad v = \frac{16 \cdot \tilde{y}}{\tilde{x}} \ .$$

Since it is desirable to have the resolution of the RWT image be continuous with the foveal resolution at the boundary between the two, scale factors have been put in the above equations to adjust the $u$-$v$ resolution. By eqs. (3.8) and (3.9),

$$u = \frac{1}{2^{h-1}} \cdot \frac{16^2}{x + 16} \ , \tag{3.10}$$

$$v = \frac{16 \cdot y}{x + 16} + \frac{16 \cdot 8 \cdot (2k - 2^{m-h} + 1)}{x + 16} \ . \tag{3.11}$$

Alternatively, the global transformation in eqs. (3.10) and (3.11) can be performed in 3 simple operations for easy implementation.

$$\text{Local transform :} \quad u = \frac{16^2}{x + 16} \ , \quad v = \frac{16 \cdot y}{x + 16} \ , \tag{3.12}$$

$$v \text{ shearing :} \quad \frac{(2k - 2^{m-h} + 1)}{2} \cdot u \ , \tag{3.13}$$

$$u \text{ scaling :} \quad \frac{1}{2^{h-1}} \ . \tag{3.14}$$

Let $A_{i,j}$ be the local image segment, $B_{x,y}$ be the remapped image of $A_{i,j}$ with the origin set to the center row and the $y$ axis is upright. $R_{u,v}$ is the local RWT segment. Figure 3.8 indicates the relationships among the 3 coordinate systems. By eqs. (3.12-3.14), the transformation can be formulated as in the following 4 steps:

1. Move the axes: $B_{x,y} = A_{x,8-y}$
2. Local transform: $R_{u,v} = B_{\frac{16^2}{u}-16,\frac{16v}{u}}$
3. $v$-shear: $R_{u,v} = R_{u,v-\frac{(2k-2^{m-h}+1)}{2}u}$
4. $u$-scale: $R_{u,v} = R_{2^{h-1}u,v}$

At the end, each node will have its local RWT segment as illustrated in Figure 3.6.

# Chapter 4

# Camera Model

## 4.1 The RWT Projective Model

Figure 3.1(b) appears like the view of a picture from a grazing angle. In fact, one could regard the RWT as a projection of an image on a plane perpendicular to it. Examine the perspective projection in which the three-dimensional $XYZ$ space is projected onto the two-dimensional $Z$-$Y$ plane at $X = 1$ (Figure 4.1). Let the three-dimensional point be $(X, Y, Z)$ and the projection be $(Z', Y')$.

$$Z' = Z/X \ , \quad Y' = Y/X \ . \tag{4.1}$$

Now, the equations in (3.1) can be made equivalent to those in eq. (4.1) if the terms $x, y, 1, u, v$ in (3.1) are unified with the $X, Y, Z, Z', Y'$ in (4.1), respectively. In that sense, the RWT described by eq. (3.1) can also be viewed as a perspective reprojection in which the original image is on the $X$-$Y$ plane at $Z = 1$, and it is projected onto the $Z$-$Y$ plane at $X = 1$.[1]

[1]For simplicity, both focal lengths have been chosen as 1 in the above discussion. In general, the two images planes are at $Z = f$ and $X = f'$. As a result, the projective model will yield $u = 1/x \cdot f \cdot f'$ and $v = y/x \cdot f'$, which differ from eq. (3.1) by constant factors.

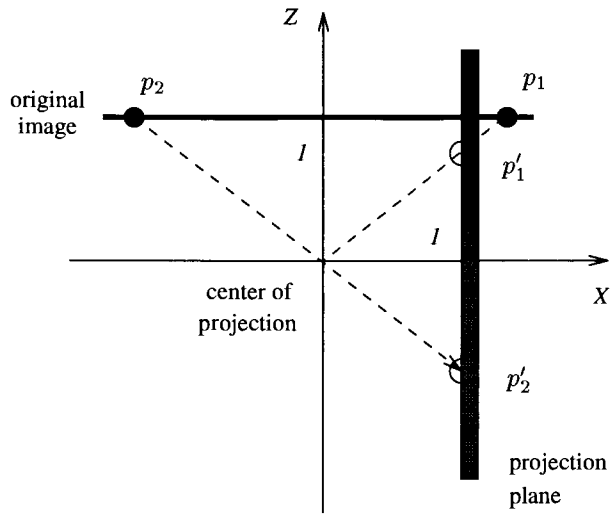



Figure 4.1: A perspective projection model.
The original image is placed on the $X$-$Y$ plane at $Z = 1$. It is reprojected onto the $Z$-$Y$ projection plane at $X = 1$. The pixels $p_1$ and $p_2$ are projected to $p'_1$ and $p'_2$, respectively.

It is interesting to notice that potentially the RWT could be implemented in hardware. Since the RWT image can be considered as another perspective projection of the Cartesian image onto an orthogonal projection plane, in a simplistic point of view, we can cascade the two processes into one. Figure 4.2 illustrates the idea. The sensor is fitted directly on the RWT projection plane mounted sideways. Thereby, the rays from the imaged objects strike directly onto the RWT sensor plane. The sensor plane is installed in two half-planes, the left and right ones, respectively, for the convenience of taking care of objects on each side of the optical axis.

The RWT camera can use a uniform sensor, which is cheap to fabricate. Space-variant sensing is realized by the oblique perspective projection on the sensor plane. The same sensor area on the plane yields variable area coverage of the visual field depending on the angle of projection. As shown in Figure 3.3, rectangular $x$-$y$ images are turned into wedge-shaped RWT images. A rectangular RWT sensor array inversely

corresponds to a wedge-shaped $x$-$y$ image — providing a foveate view-of-world. In fact, one can also alter the position, orientation or even the shape of the sensor plane to produce different space-variant sensors.

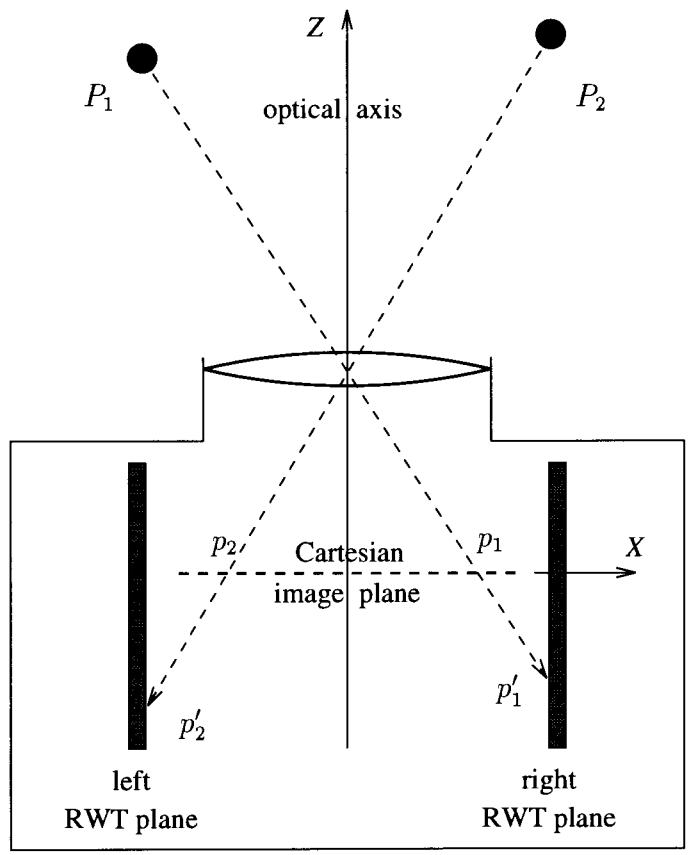



Figure 4.2: A rudimentary RWT projection camera.
The camera has its sensor placed on the left and right horizontal planes. Instead of forming an image on the frontal focal plane, lights from $P_1$ and $P_2$ passing through the lens are further projected onto the sideway-positioned RWT sensor planes to form images $p'_1$ and $p'_2$.

## 4.2 Non-Paraxial Focusing

The above discussion delineates a rudimentary idea of the RWT camera design. A prominent problem of Figure 4.2 is the necessity of focusing on a deep image plane in parallel to the optical axis. As shown in Figure 4.3 the object forms a sharp image on the focal plane normal to the optical axis. However, upon further projection onto the RWT plane which is positioned sideways and off-axis, the rays diverge, casting a blurred image on the RWT plane.

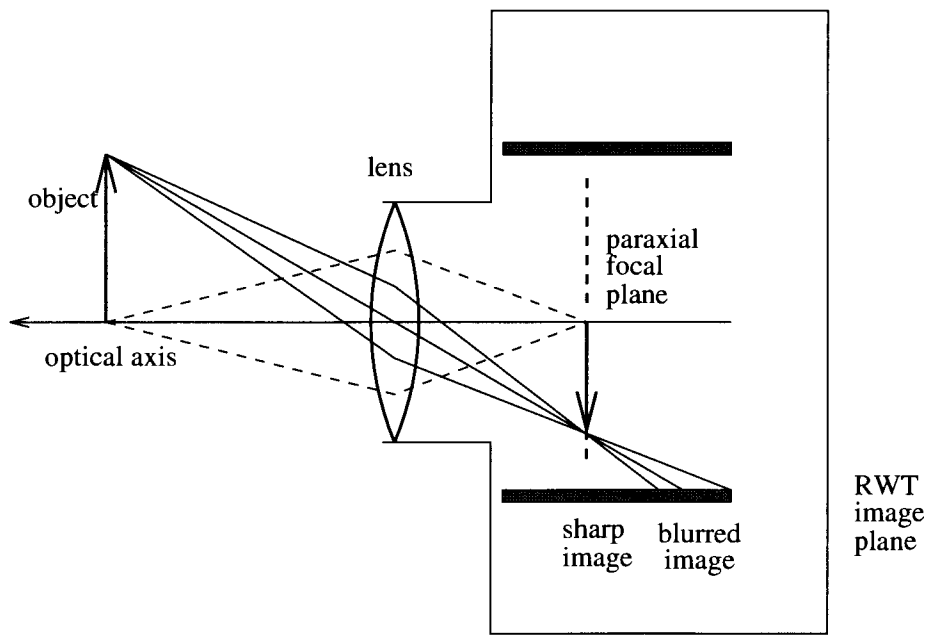



Figure 4.3: The focusing problem of the sideway-positioned RWT projection plane.

In general, it is difficult to get a focused image on an off-axis plane. Ordinary cameras have paraxial focal planes, i.e., only objects that are near the optical axis will form focused images on the focal plane. The pair of object and image points are called *conjugate points* and the planes through these points are the *conjugate planes.* This is true only under paraxial conditions. However, the RWT image plane in Figure 4.3 is located off-axis, and the condition for lens focusing is not paraxial.

### 4.2.1 The RWT lens

In addition to its off-axis position, the RWT image plane also assumes a non-frontal orientation like the one in the Krishnan's camera [KA94]. An optical condition of such non-frontal projection is known as the Sheimpflug condition [Bro65]. It occurs between tilted object and image planes (as shown in Figure 4.4). In fact, the projective model of the RWT can be achieved as non-frontal focusing between orthogonal conjugate planes.

Consider an image plane oriented at an angle to the optical axis of a lens. Without loss of generality, the problem is dealt with in the 2-D $Z$-$X$ plane. The result can be readily extended to the 3-D $XYZ$ space. Let the optical axis be aligned with the $Z$ axis, and the lens be on the $X$ axis. The normal of the image plane is on the $Z$-$X$ plane. The resulting configuration is as shown in Figure 4.4. A point on the image plane is denoted as $P_I(Z_I, X_I)$, and its conjugate object point is $P_O(Z_O, X_O)$. In the 2-D $Z$-$X$ space, the image plane is a line. Let us denote it with the equation $X_I = m_I \cdot Z_I + c_I$, where $m_I$ and $c_I$ are the slope and $X$-intercept of the line. As $P_O$ and $P_I$ are related by the lens refraction formula and are collinear along the principal ray which travels through the optical center of the lens, the conjugate relationship between the object and image plane can be derived by solving the equations involving the lens formula, the principal ray geometry and the plane equation.

$$\text{Lens formula}: \quad -\frac{1}{Z_O} + \frac{1}{Z_I} = \frac{1}{f}\ , \tag{4.2}$$

$$\text{Principal ray}: \quad \frac{X_O}{Z_O} = \frac{X_I}{Z_I}\ , \tag{4.3}$$

$$\text{Image plane}: \quad X_I = m_I \cdot Z_I + c_I\ . \tag{4.4}$$

Resolving for $X_O$ and $Z_O$, a linear equation is obtained.

$$X_O = \left(m_I + \frac{c_I}{f}\right) \cdot Z_O + c_I \ . \tag{4.5}$$

Generalized to 3-D, eq. (4.5) states that the objects which form focused images on the image plane are themselves on a plane as well. If denoted by $X_O = m_O \cdot Z_O + c_O$, the object plane is related to the image plane by

$$m_O = m_I + \frac{c_I}{f} \ , \quad c_O = c_I \ . \tag{4.6}$$

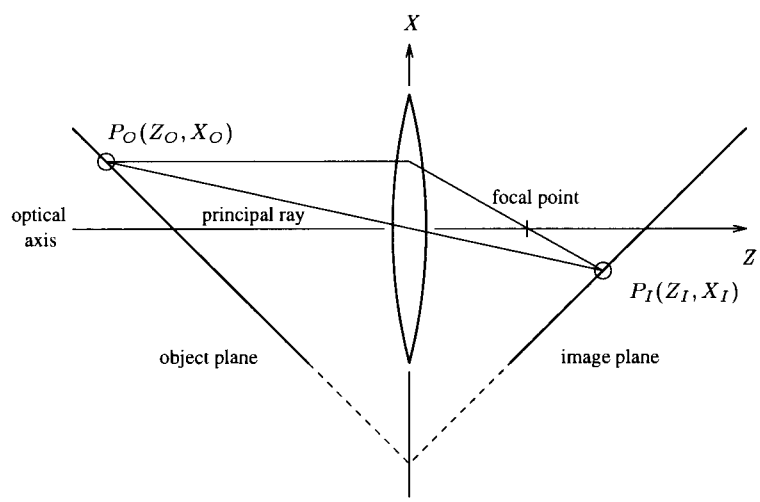



Figure 4.4: Optical principle in tilted plane focusing.

Note that the lens formula of paraxial focusing is still employed in the derivation of the off-axis conjugates. In fact, we are using the first-order lens formula (eq. (4.2)) in obtaining the simple linear relationship between the off-axis conjugates. The shortfall of the paraxial approximation in off-axis focusing is manifested in various kinds of lens aberrations which has to be compensated by a careful lens design [HZ74]. In Section 4.5, a choice of the periscopic lens design is made to that end.

An interesting case involves non-frontal focusing between two orthogonal conjugate planes. It realizes the projective model of the RWT transformation discussed in

Section 4.1, and successfully overcomes the focusing problem which would render the simple camera model in Figure 4.2 impractical. In the new camera model, the orthogonal conjugate planes are realized with a lens system constructed as in Figure 4.4. Both the object and image planes are at 45° to the $X$ axis. If $-m_O = m_I = \tan 45° = 1$, from eq. (4.6),

$$c_O = c_I = (m_O - m_I) \cdot f = -2f \ .$$

It means that the planes are installed at 90° to each other. They are arranged symmetrically on both sides of the lens. For the lens having a focal length $f$, the planes intersect at $2f$ below the lens.

The lens is re-drawn in Figure 4.5. Herein, the lens system is rotated by 45° to ease the distance computation for the next step. Now, the object plane is the vertical plane and the image plane is the horizontal one. The lens is located at the origin of the $XYZ$ coordinate space. For simplicity, the normal distance of the planes to the lens is again assigned a unit value. This makes the lens parameters consistent with the projective model in Figure 4.1, thus realizing the RWT transformation as defined in eq. (3.1). If the normal distance is not equal to one, then all the other distance measurements will simply be scaled by constant factors, as explained in Section 4.1.

From the geometry, $\vec{OP_O}$ is $(X_O, Y_O, 1)$, and $\vec{OP_O}$ is collinear with $\vec{OP_I}$. Therefore,

$$\vec{OP_I} \ = \ (1 \ , Y_I \ , \ Z_I) \ = \ -\vec{OP_O} \cdot \frac{1}{X_O} \ = \ \left(1 \ , \ \frac{Y_O}{X_O} \ , \ \frac{1}{X_O}\right) \ . \qquad (4.7)$$

This can be denoted as:

$$u \ = \ Z_I \ = \ \frac{1}{X_O} \ , \ \ v \ = \ Y_I \ = \ \frac{Y_O}{X_O} \ , \qquad (4.8)$$

which illustrates that such an projection between orthogonal planes through the origin achieves the RWT transformation.

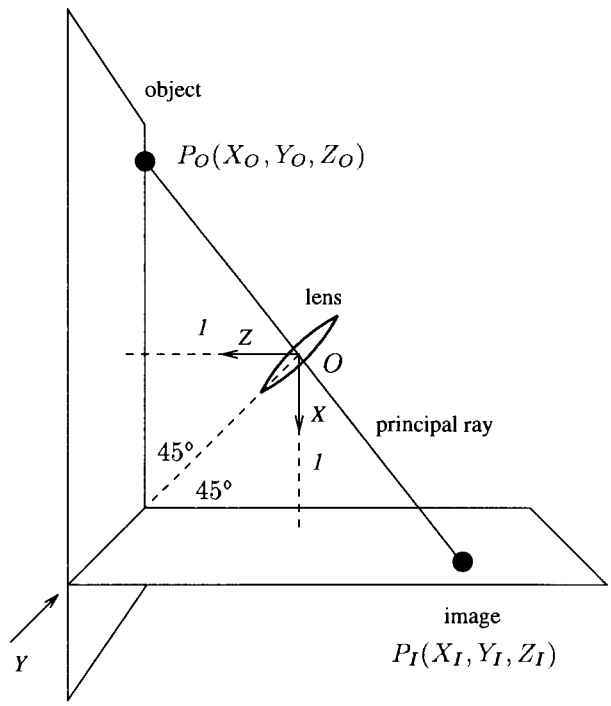



Figure 4.5: The prototype RWT lens.
The object and image planes are at 45° symmetrically on both sides of the lens. The normal distance of the planes from the lens is assigned a unit value. The principal ray from $P_O$, traveling through the optical center $O$ to $P_I$ is shown.

## 4.3 Projecting the Singularity

Similar to the singularity problem at $x = 0$ in eq. (3.1), the projective model from Figure 4.2 also fails for points near the $Z$ axis. In this section, the patching and shifting methods discussed in Section 3.1.2 are employed as the practical fix in the design for the camera. The following proposes three techniques, namely the *U-plane projection*, the *V-plane projection* and the *displaced-center projection*. The U-plane projection implements the patching method whereas the latter two provide alternative techniques for implementing the shifting method.

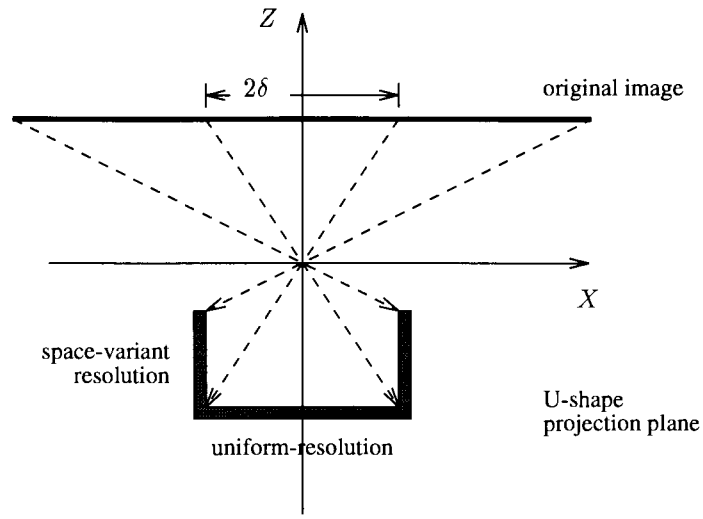



Figure 4.6: U-plane projection.
The center region of width $2\delta$ forms a uniform-resolution projection at the bottom portion of the U-plane, whereas the peripheral regions are projected in space-variant resolution onto the sideway-positioned arms of the U-plane.

### 4.3.1 U-plane projection

The patching method provides an expedient fix to the singularity problem. It excludes the center strip of width $2\delta$ of the original image from the space-variant mapping. The uniform-resolution data for the strip is used directly to patch up the two wedge images from the transform.

In the projective model, the method corresponds to two different projection strategies for the center strip and the peripheral region respectively. Figure 4.6 shows a U-shape projection plane implementing the two projections. The center region of the original image is projected normally onto the frontally oriented portion of the U-plane, producing a normal uniform-resolution image. The regions to the sides in the original image are projected as illustrated in Figure 4.1, forming the RWT projections on the sideway-positioned arms of the U-plane.

The advantages of using uniform-resolution model for the fovea were discussed in

Section 3.1.2. In fact, the U-plane model supports a seamless joint between the central rectangular fovea and the peripheral wedge-shaped regions. On the contrary, the spatially uninterrupted connection between the fovea and periphery is not supported in the log-polar implementation [VdSKC+89]. The square grid for the fovea and the ring structures for the log-polar periphery cannot simply be patched together.

### 4.3.2 V-plane projection

The shifting method discussed in Section 3.1.2 has been formulated in eq. (3.3) as the S-RWT. The following shows that the S-RWT can also be implemented with a V-plane projection.

Figure 4.7 depicts the V-plane projection. The two projection planes in Figure 4.2 are joined to form a V in this figure. The left arm of the V forms the projection plane for the right half of the original image and the right arm of the V is the projection plane for the left half. The singularity problem disappears because the center region of the original image gets projected to a $u$ position on the V-plane. It can be observed that the orientation of the V arms is not as steep as that of the sideway-positioned projection plane in Figure 4.1. A less drastic space-variant resolution should be expected. In fact, it can be shown that such a V-plane projection implements the space-variant resolution of the S-RWT of eq. (3.3).

Since the projection occurs independently on each side of the image, without loss of generality we examine the projection from the right side of the original image onto the left arm of the V. Figure 4.8 shows the ray diagram of the projection. A point $P$ on the original image is projected to $Q$ on the projection plane. $O$ is the center of projection, and $E$ is the origin of the $x$-$y$ space. To be consistent with the S-RWT formulation in eq. (3.3), the origin of the $u$-$v$ space $F$ is defined as the point of

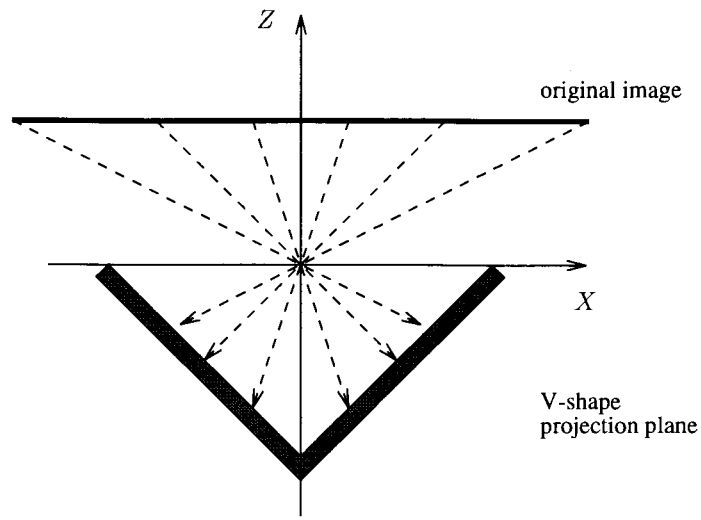



Figure 4.7: V-plane projection.
The left arm of the V forms the projection plane for the right half of the original image and the right arm of the V is for the left half. The singularity problem is resolved, and space-variant resolution is effected on both projection planes.

projection when $x = \infty$ and $y = 0$.

From the geometry in Figure 4.8, $\angle PRO = \angle ROF = \theta$.

$$\overline{OF} = \overline{RF} = \frac{r}{2\cos\theta} , \tag{4.9}$$

$$\overline{RE} = r\cos\theta . \tag{4.10}$$

From the similar triangles,

$$\frac{\overline{RE} + x}{\overline{RF} + u} = \frac{\overline{OF}}{u} . \tag{4.11}$$

Using (4.9) and (4.10) in (4.11),

$$u = \frac{(\frac{r}{2\cos\theta})^2}{x + (r\cos\theta - \frac{r}{2\cos\theta})} = \frac{f}{x + a} \tag{4.12}$$

by letting $f = r/(2\cos\theta)$, and $a = r\cos\theta - r/(2\cos\theta)$.

Imagine the vertical dimension in/out of the paper. It defines the $y$ coordinates on the image plane and the $v$ coordinates on the projection plane. Again, from the

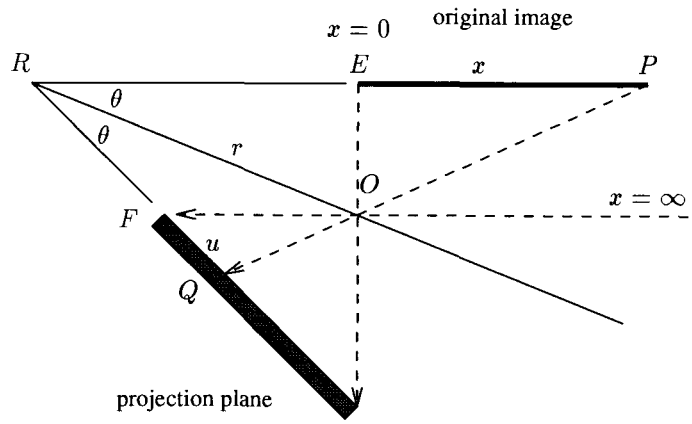



Figure 4.8: Geometry of the V-projection from $P$ to $Q$.

similar triangles,

$$\frac{\overline{PQ}}{\overline{OQ}} = \frac{\overline{PR}}{\overline{OF}} = \frac{x + r\cos\theta}{\frac{r}{2\cos\theta}} , \tag{4.13}$$

$$\frac{\overline{PQ}}{\overline{OQ}} = \frac{\overline{PO} + \overline{OQ}}{\overline{OQ}} = \frac{\overline{PO}}{\overline{OQ}} + 1 = \frac{y}{v} + 1 . \tag{4.14}$$

Combining (4.13) and (4.14),

$$v = \frac{\frac{r}{2\cos\theta} y}{x + (r\cos\theta - \frac{r}{2\cos\theta})} = \frac{fy}{x + a} . \tag{4.15}$$

From (4.12) and (4.15), we conclude that the $u$ and $v$ coordinates from the V-plane projection are effectively computing the S-RWT as defined in eq. (3.3) within a constant factor $f$.

### 4.3.3 Displaced-center projection

Alternatively, the S-RWT can be implemented with the displaced-center projection technique. The inspiration is from the shift parameter in eq. (3.3). The parameter $a$ is hinting at a shift on the $x$-origin when comparing eq. (3.3) with the formulation of the RWT in eq. (3.1). As Figure 4.1 is the projective model of eq. (3.1), a natural

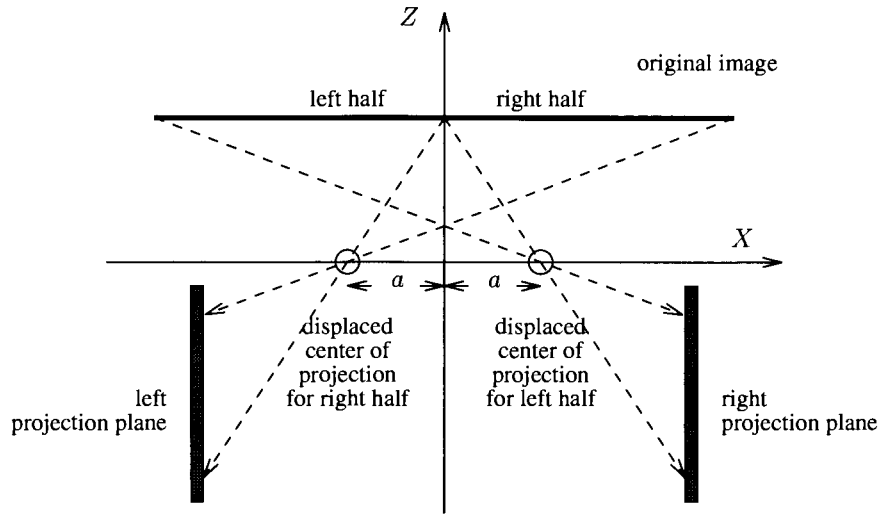



Figure 4.9: Displaced-center projection.
The centers of projection of the two half images are displaced away from the origin by $a$. Effectively, the right(left) half of the image appears to have been shifted by $+a(-a)$ upon its projection through the displaced center of projection.

implementation of eq. (3.3) could be one like Figure 4.1, but modified by shifting the $X$-origin or, relatively, by displacing the center of projection.

Figure 4.9 illustrates the displaced-center projection method. The center of projection for the right half of the original image is displaced by $-a$. Effectively, the right half image appears to have been shifted by $+a$ upon its projection through the displaced center of projection onto the left projection plane. Similarly, the center of projection for the left half of the original image is displaced by $+a$, causing the data to be shifted by $-a$ upon its projection onto the right projection plane.

As both the V-plane and the displaced-center projection methods are able to implement the S-RWT to the same effect, either one of them can be used in place of the other. As a matter of fact, the displaced-center projection has advantages over the V-plane projection method. First, it offers a more natural interpretation of eq. (3.1). Second, the displaced-center method does present an easier implementation

of the S-RWT. Consider that altering the parameter $a$ in the S-RWT would change the parameters $r$ and $\theta$ in the equations (4.12) and (4.15). This may involve adapting the lens focusing in the V-plane projection (Figure 4.8).

In our design of the prototype RWT camera (Section 4.4.2), a combined use of the patching and shifting methods is implemented to support the flexibility in dealing with the singularity problem. The U-plane and the displaced-center projection techniques are employed for the patching and shifting methods respectively.

## 4.4 A Prototype RWT Camera

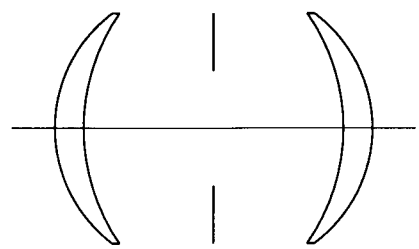
The RWT is implemented as a lens projection between orthogonal planes under the Sheimpflug condition [Bro65] discussed in Section 4.2.1. The lens focusing is modeled with the first-order paraxial approximation. Practical considerations of various kinds of lens aberrations become an issue when performing the actual design of the lens.

### 4.4.1 Periscopic lens design

As mentioned above, the first-order paraxial approximation is used to its advantage for deriving the off-axis focusing. The lens aberrations thus encountered in off-axis focusing are compensated with a careful lens design. In the RWT lens design, the projection between 90° planes imposes a stringent requirement on the lens performance. Light rays reflected off the intermediate screen normally strike the RWT lens at a wide-angled oblique incidence. Lens aberrations are adverse under such conditions. As an initial attempt, we have chosen to use the periscopic lens as the candidate for the RWT lens. The design data is generally available [Kin78]. The periscopic lens has a symmetrical configuration of two meniscus-convex lens positioned on both sides of

a central stop (Figure 4.10).

The periscopic lens has the advantages that it has little distortion and lateral color aberrations [Kin78]. Coma aberration can be ignored as it will be corrected automatically by the symmetry of the lens configuration. Moreover, the field curvature can be flattened by selecting the appropriate stop diameter. The periscopic lens is shown to be necessary in our simulation tests. It will be shown later that when an ordinary simple lens is used, the field curvature aberration causes poor focusing.




| | | Curvature($cm^{-1}$) | Separation($cm$) | Refractive Index |
|---|---|---|---|---|
| front-meniscus | outer | 0.4734 | | |
| | | | 0.1861 | 1.5233 |
| | inner | 0.3358 | | |
| | | | 0.9170 | |
| stop | | | | |
| | | | 0.9170 | |
| rear-meniscus | inner | -0.3358 | | |
| | | | 0.1861 | 1.5233 |
| | outer | -0.4734 | | |

Figure 4.10: The periscopic lens and the lens design data.
The effective focal length of the periscopic lens is 70 mm and the stop diameter is 8.87 mm.

### 4.4.2 Design of the RWT camera

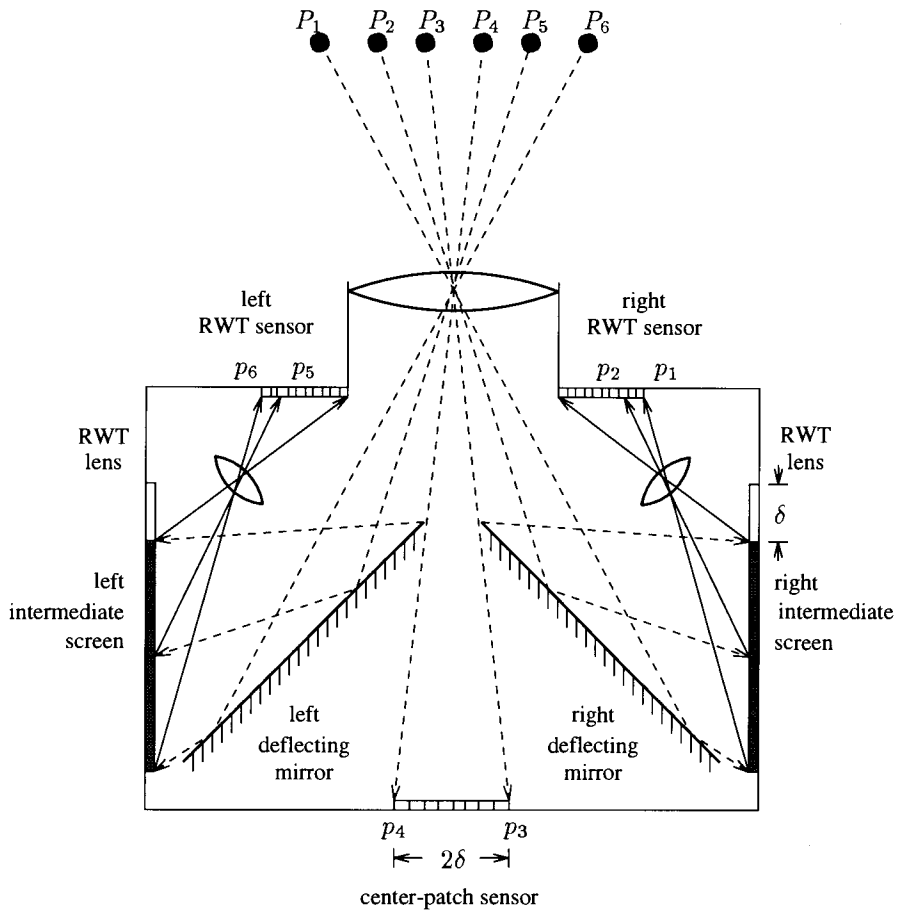



Figure 4.11: The RWT camera model

The camera objective lens projects the image on the two intermediate screens on either side through the deflecting mirrors. The RWT lenses then focus the images from the screens onto the orthogonal RWT image planes. A center slot is opened between the deflecting mirrors enabling uniform-resolution projection onto the bottom wall. The full RWT image comprises segments from the RWT sensors on either side merged with the center patch from the bottom.

The design of the camera is based on the model from the previous section. The light rays are split into left and right halves by using the deflecting mirrors. A similar setup of image splitting mirrors is also used in a stereo camera system by Teoh and Zhang [TZ84], except that the deflecting mirrors used in our camera split the visual field in the middle. The left field is projected onto the right screen and the right field to the left screen. This facilitates the implementation of the two half planes in RWT to take care of objects on each side of the visual field.

The two intermediate screens on either side of the camera play the role of the object planes for the RWT lenses. Each RWT lens projects from the respective screen onto an orthogonally located RWT image plane at the front wall of the camera. The RWT transformation is realized by projection between these two orthogonal planes.

A practical consideration is how to take care of the singularity of the RWT at $x = 0$. The patching and shifting techniques as discussed in Section 3.1.2 are employed in this design. In Figure 4.11, a center slot is opened between the two deflecting mirrors. Objects near to the optical axis ($P_3$ and $P_4$) are now projected to the center of the bottom wall of the camera ($p_3$ and $p_4$). As in the ordinary cameras, the image at the center is a uniform-resolution projection.

To implement the shift parameter in eq. (3.3), a shift by $a$ on the $x$-$y$ images from both intermediate screens needs to be performed. However, the shift can also be realized by relative repositioning of the RWT lens. The lens is required to be positioned at the $XYZ$ origin in Figure 4.5. Moving the lens and the projection plane along the $X$ axis in relation to the object plane effectively achieves the shifting operation on the object's $X_O$ coordinate. A shift by $a$ on $X_O$ thus causes eq. (4.8) to realize the S-RWT which is defined in eq. (3.3). Practically, the RWT lens-sensor units on either side of the camera in Figure 4.11 can be adjusted up and down in the

diagram to implement the shift.

Now, three segments of the RWT image are formed at three locations. Merging of the three pieces (left RWT, center patch, right RWT) will yield a connected image like the bipolar RWT image shown in Figure 3.3. Note that the sensors are delimited in the way that the RWT image is continuous over the boundary between the left and the center segments, and also between the center and the right segments. Proven technologies from the 3-chip color cameras can be employed to deal with the problems of synchronization and alignment among the three sensors.

Further design considerations for perfecting the camera design require deeper understanding of optical instruments. For example, a practical concern about the use of the intermediate screens would be the weakness of the resulting irradiance at the RWT sensor planes after the diffuse reflection by the intermediate screens. The light energy entering the camera through the field objective lens will get dispersed in all directions due to the diffuse reflection by the screens. Consequently, only a small portion of the energy will be collected by the RWT lenses and get projected onto the sensor planes. Ultra-sensitive CCD sensors may be needed for recording the dim image when it arrives at the end of the optical path.

Another concern is the diffraction effects caused by the center slot. When the slot gets smaller, the diffraction effects become more eminent. Special measures may be required to alleviate the diffraction or the foveal patch sensor should be mounted on the side-wall alongside the screens to eliminate the need for the center slot altogether.

Despite all these detailed design considerations, the model depicted in Figure 4.11 is used to illustrate the basic principles of the optical construction which shows the unique RWT projection and the implementation of the S-RWT and foveal patch. Any practical design could be developed based on this basic model. In fact, camera design

from this model would be appealing on three counts. First, the RWT obtains space-variant resolution by using oblique projection between orthogonal planes. It does not require sensing elements of variable sizes to achieve variable resolution. The main advantage of using uniform sensors is thus realized. Second, the rectangular shape of the sensors allows merging of the sensors to deliver a connected bipolar RWT image. Third, the design accommodates a flexible implementation of both methods of S-RWT and foveal patch. The foveal patch is adjustable by varying the aperture of the center slot between the two deflecting mirrors. The shift parameter for the S-RWT can be adjusted by shifting the RWT lens-sensor units.

## 4.5 Optical Simulations

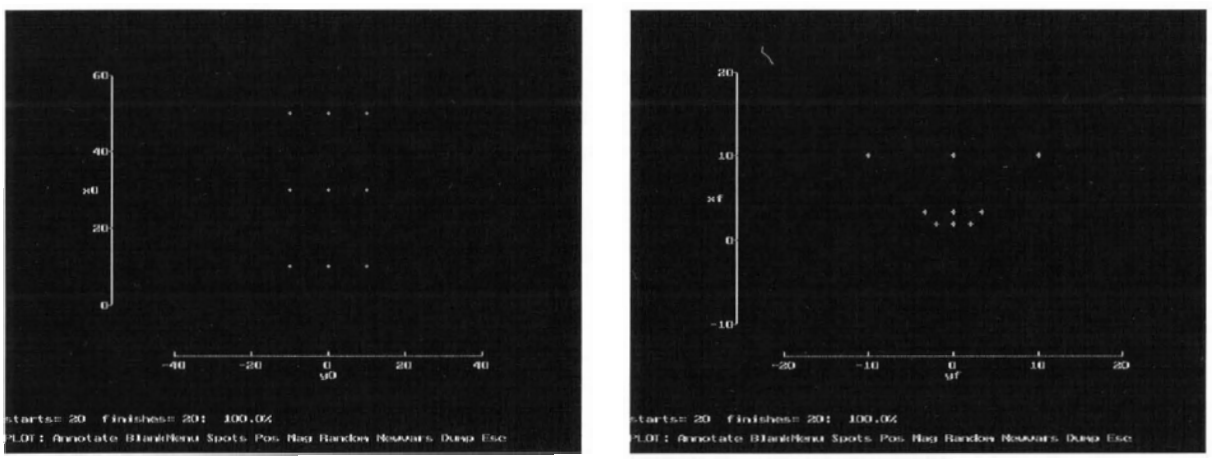
Before a hardware prototype is built, the design for the RWT camera has been tested using an optical ray tracing simulation software. The Beam Optical Ray Tracer[2] is used to provide a test environment in which the optics of lens refraction is simulated. Since the optics of uniform-resolution projection for the center patch is well proven in the conventional cameras, and since the optical path comprising the camera objective through the intermediate screens is primarily adopted from the design of ordinary stereo cameras, our tests are conducted mainly on the optics of the RWT lens projection.

Our first experiment uses nine grid points placed on the object plane as shown in Figure 4.12(a).[3] Pencils of rays radiated from the points are propagated through

[2]Beam Optical Ray Tracer is the product of Stellar Software at Berkeley, CA, U.S.A., copyright 1990.

[3]Because it is easier to generate (or obtain) rectangular $x$-$y$ images, they are used in this simulation. As a result, the generated RWT images are of the wedge shape. A real hardware RWT camera will have a rectangular sensor array and the corresponding view of the world will be of the wedge shape as pointed out in Section 3.1.3. Since $T = T^{-1}$, the simulation result is equally valid.

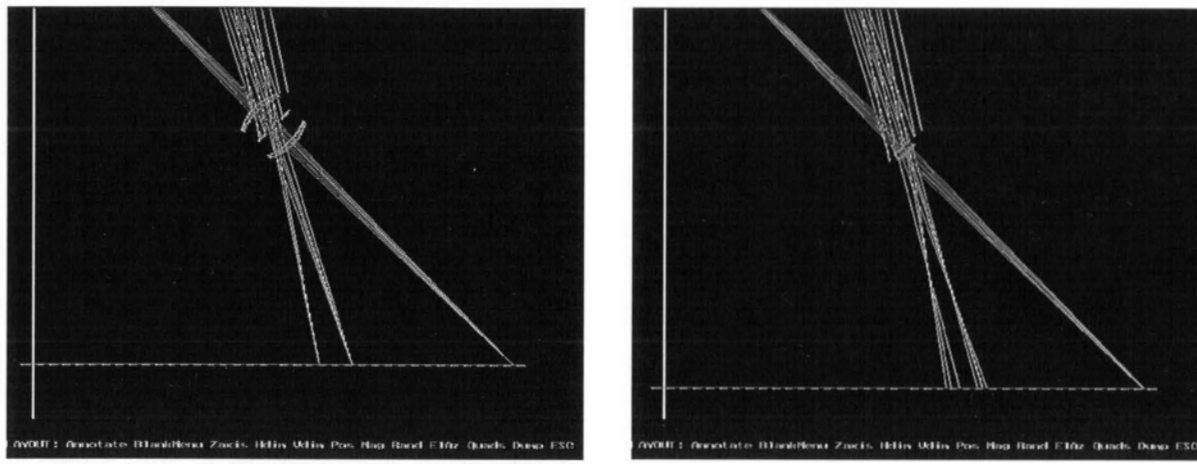
the lens and converged onto the image plane. The refraction process is simulated. Figure 4.12(b) plots the images of the focused grid points. The distinctive wedge shaped pattern can be recognized. The ray diagram from the Beam Optical Tracer is drawn in Figure 4.13(a). From the diagram, it can be observed that good focusing is achieved. The reported error (standard deviation) of the landing position of different rays from the same grid point is below 0.02 cm. For appreciation of the periscopic design for the RWT lens, a comparison is made between a simple biconvex lens and the periscopic lens. Figure 4.13(b) clearly shows the adverse blurry condition arising from the lens aberrations.



(a) (b)

Figure 4.12: Focusing test with nine grid points.
(a) Nine grid points on the object plane. (b) The focused image as viewed on the RWT plane.

In Figure 4.14, a dense grid is placed on the object plane and the projected pattern on the RWT image plane is obtained. This test reveals the accuracy of the lens in performing the RWT transformation. The error measured against the computed RWT image is very small — $rms$ error = 0.038 cm.

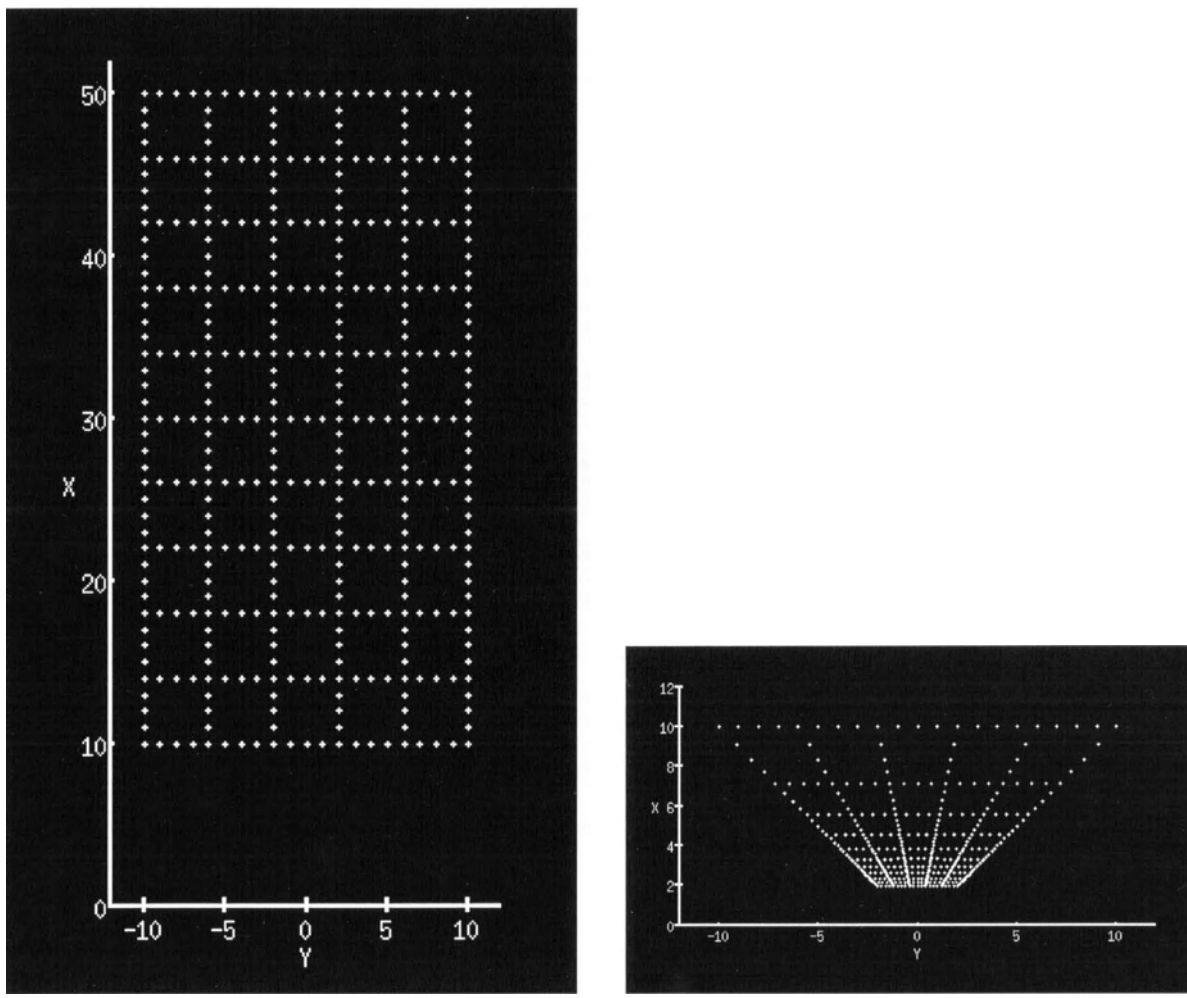


(a) (b)

Figure 4.13: Ray diagrams showing the lens focusing.
(a) Good focusing is achieved with the use of periscopic design for the RWT lens. (b) Poor focusing arises from lens aberrations with the use of a simple biconvex lens. The ray diagram reveals the intolerable field curvature aberration.

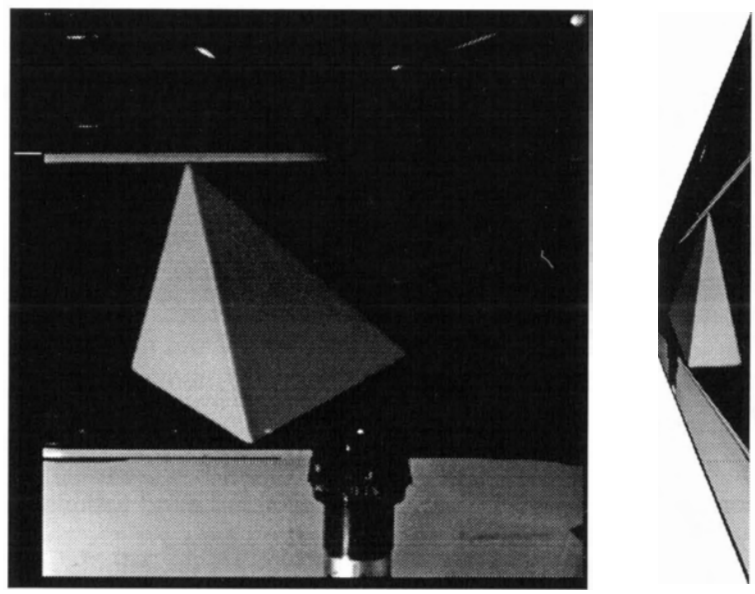
Our experiment is concluded with a test on real image data. The assembly belt image from the motion stereo experiments in [TL94] is used.[4] Figure 4.15(a) is an image of the assembly belt scene from the intermediate screen of the RWT camera (it looks just like a normal uniform-resolution image), and Figure 4.15(b) is the RWT image achieved through simulation on the RWT lens.

[4]In some applications, it is better to use only one half of the image where $x > 0$ (and hence $u > 0$). In this test case the origin is located at the middle of the left border in the original $x$-$y$ assembly belt image. A small patch ($\delta = 15$) at the left side of the $x$-$y$ image is not transformed.




(a) (b)

Figure 4.14: Accuracy test on focusing using a dense grid.
(a) A grid placed on the object plane. (b) The projected image as viewed on the RWT plane. The *rms* error measured against the computed RWT image is 0.038 cm.



(a) (b)

Figure 4.15: Focusing test using real data.
(a) The belt image as view on the intermediate screen. (b) The image on the RWT plane.

# Chapter 5

# Applications of RWT Mapping

## 5.1 RWT Imaging in Road Navigation

In the problem of road following, an efficient search for road features can be effected with the variable resolution offered by the RWT.

Different approaches have been devised for road detection in various experimental autonomous land vehicles. An area-based voting scheme based on the Hough transform is applied to compute the direction of the road in the CMU Navlab [THKS88]. In the VaMoRs project [DM92], visual features of the road edges are detected based on the "Gestalt" hypothesis under adverse situations of shadows and absent lane-markings.

Both methods search over the perspective images for road features. The drawback is that the nearby section of the road gets overly attended whereas the far side toward the horizon is disproportionately under-sampled. Arguably, this differential scale of attention to detail is not suitable for driving on the road. One has to pay sufficient attention to a reasonable distance to see the general direction of the road, while at

the same time remaining aware of the road segment immediately ahead.

Lotufo, *et al.* [LMD+90] presents the *plan-view transformation* method for road navigation. The original perspective road image is projected to a grid inclined by a pan angle $\theta$, which is chosen so that the road edges are nearly parallel to the boundaries of the grid. It is also reported that the new images are of a reduced size (typically by a factor of 32).

### 5.1.1 Perspective inversion by RWT

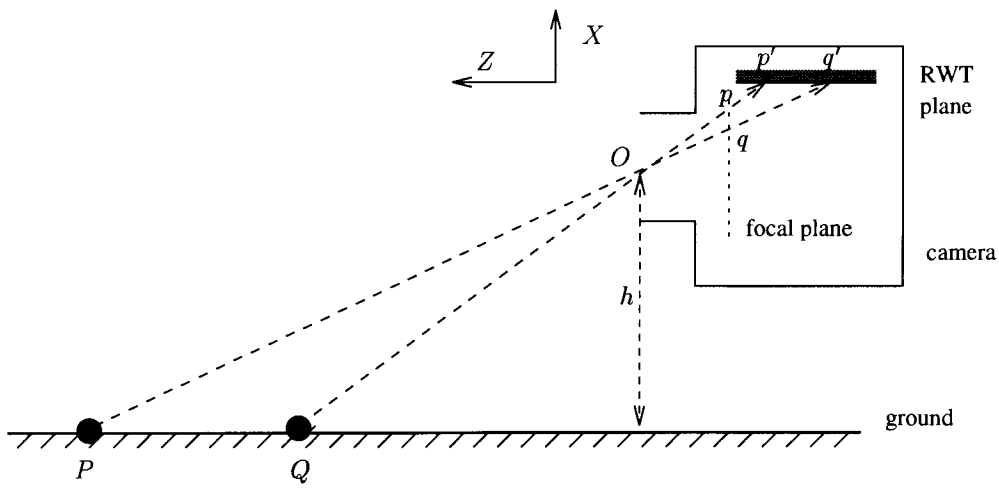



Figure 5.1: Perspective inversion effected by the RWT projection.
The image size $(p - q)$ varies with the position of the segment $(P, Q)$, whereas the size of the RWT image $(p' - q')$ does not.

Effectively, the RWT re-samples the image to a variable resolution which counterbalances the differential scale of details in the perspective projection. Figure 5.1 illustrates the perspective inversion. With the road $(P, Q)$ on the ground projected onto a horizontal plane in the RWT camera, the projection $(p', q')$ does not change in magnification, and the perspective projection is practically inverted.

$$p' - q' = f/h \cdot (P - Q) \; .$$

The road in the RWT image appears as though it were from an aerial view (Figure 5.3(b)). However, an important difference is that the RWT camera is pointing toward the front which is vital to driving. Moreover, the nearby section is sampled at a much reduced resolution. The overall data volume can be greatly reduced to achieve a comparable performance.

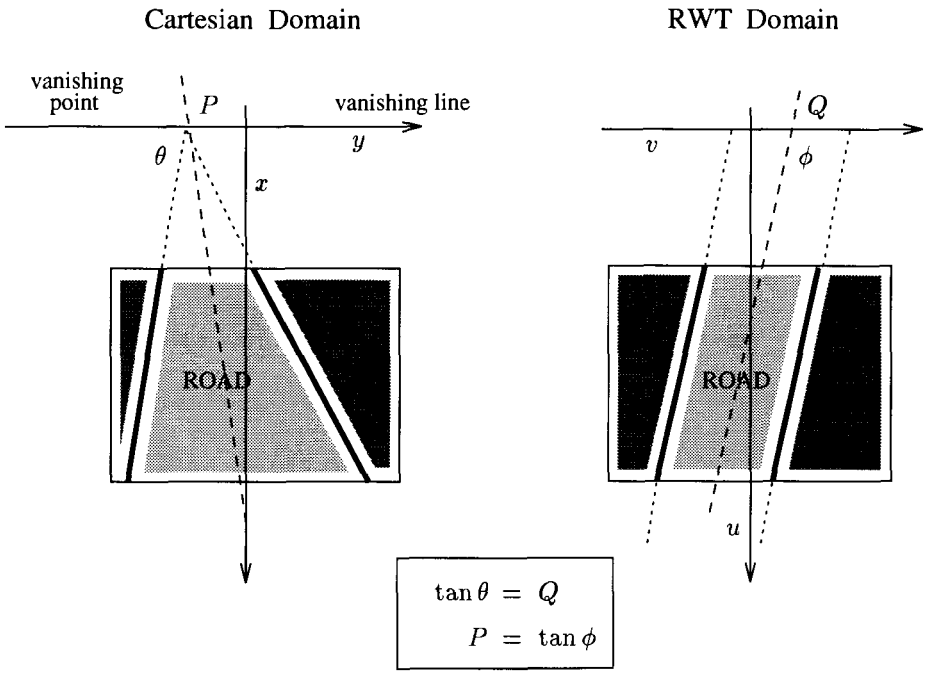



Figure 5.2: The RWT dual of the road image.
The vanishing point in the Cartesian domain just becomes the direction in the RWT domain and vice versa.

In the CMU Navlab project [THKS88], a road is perceived as converging at one point on the vanishing line, and is parameterized by $P$ and $\theta$ (Figure 5.2). As discussed in section 3.2.2, the detection of converging lines at the vanishing point for finding the road direction can be carried out by detecting parallel road boundaries in the RWT image. The technique of the Hough transform is equally applicable. While edge tracking is employed to calculate the geometric model of the road in the VaMoRs

project, the RWT image benefits by eliminating the variable search ranges for road features in the near and distant sections of the road. In all cases, the RWT supports an efficient representation of the road image as the data volume is greatly reduced by its spatially variable sampling.

### 5.1.2 Results

Figure 5.3(a) is a synthetic image of a road scene. The image has a resolution of $128 \times 256$. Figure 5.3(b) is its RWT image of size $32 \times 128$. The remote section of the road has retained its resolution whereas the excessive information at the near side is suppressed. The total area of the search region is significantly reduced. The direction of the road is detected using the Hough method described above. In the RWT image, the road direction is detected as $\tan\phi = -22$, yielding the position of the vanishing point $P$ in the original Cartesian image.

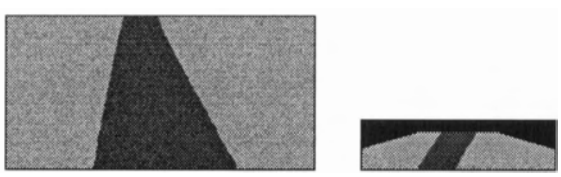



Figure 5.3: The synthetic image of a road scene.
(a) The road image of resolution $128 \times 256$. (b) The RWT image of size $32 \times 128$.

## 5.2 Depth from Ego Motion

### 5.2.1 Motion stereo

Okutomi and Kanade pointed out in [OK93] that the distance between the pair of cameras in stereo vision greatly affects the precision and error rate of the correspondence process. A short baseline will provide less precision; whereas, a longer baseline will result in a higher error rate due to false matches. To alleviate the dilemma, they proposed the *multiple-baseline stereo* method wherein different baselines are generated by lateral displacements of a camera.

Consider a manufacturing environment with intelligent robots working on assembly lines where the belts are moving at constant speed. Multiple snapshots of the moving objects on the belt can be taken in a rapid succession by a single camera. The controlled belt movement provides the necessary stereo disparity. Moreover, it can guarantee that the disparity occurs only along the epipolar lines. This method is called *Motion Stereo* [Nev76]. Its greatest advantage is the simplicity in camera control and calibration. Suppose the camera is looking down the $Z$ direction, i.e., its optical axis is the $Z$-axis. We call the above moving belt situation *lateral motion stereo* where objects move on a $Z = Z_0$ plane, perpendicular to the $Z$-axis. Another type of motion stereo is *longitudinal motion stereo* in which objects move along the $Z$ direction, such as when an autonomous vehicle travels along a highway.

Bolles, Baker, and Marimont [BBM87] proposed a technique of epipolar-plane image analysis for determining structure from motion. It was pointed out that for straight-line and constant-speed camera motions, simple linear structures will be formed on the epipolar-planes (Figure 5.4), where the slope of these lines indicates the depth of the feature points.

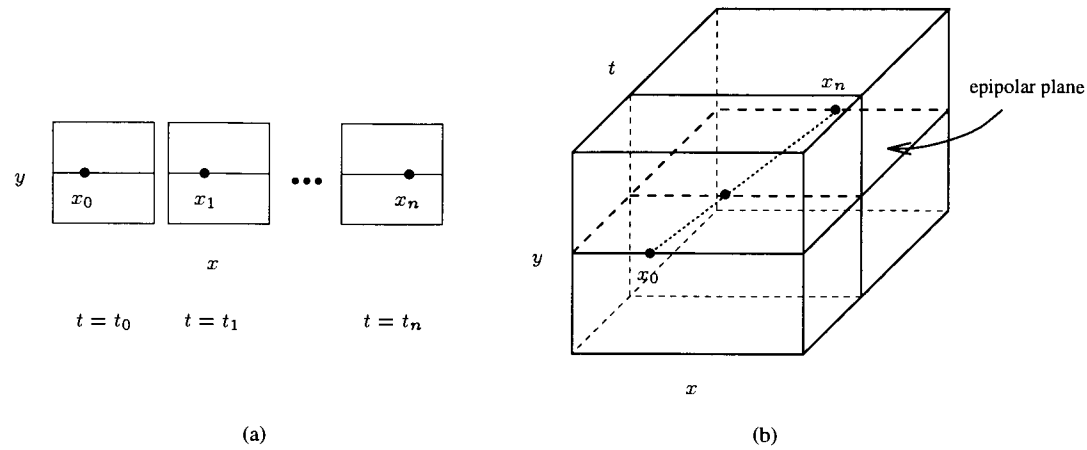



Figure 5.4: Epipolar-plane image analysis.
(a) A feature point moves along the epipolar line in the $x$-$y$ plane with a constant speed. (b) A linear locus is formed on the epipolar ($x$-$t$) plane in the $xyt$ space.

This section presents the adaptation of epipolar-plane analysis for depth recovery using RWT images from motion stereo sequences. The longitudinal motion stereo and lateral motion stereo will be examined in Section 5.2.2 and Section 5.2.3. In Section 5.2.4 a voting scheme for searching the collinear points on the epipolar plane in both motion stereo cases will be discussed.

### 5.2.2 Longitudinal motion stereo

Depth recovery in autonomous vehicle navigation provides an example of the longitudinal motion stereo in which the relative object movement is along the $Z$ direction at a constant speed. Figure 5.5(a) illustrates a point moving from position $P_0(X_0, Y_0, Z_0)$ at $t_0$ to position $P_1(X_0, Y_0, Z_1)$ at $t_1$. The $x$-coordinates of its projections on the ordinary $x$-$y$ image plane are $x_0$ and $x_1$. The corresponding images on the RWT $u$-$v$ plane are $u_0$ and $u_1$. As shown, the focal lengths are $f$ and $f'$ respectively. For simplicity (and with the deviation of a constant factor), it is assumed that $f = f' = 1$.

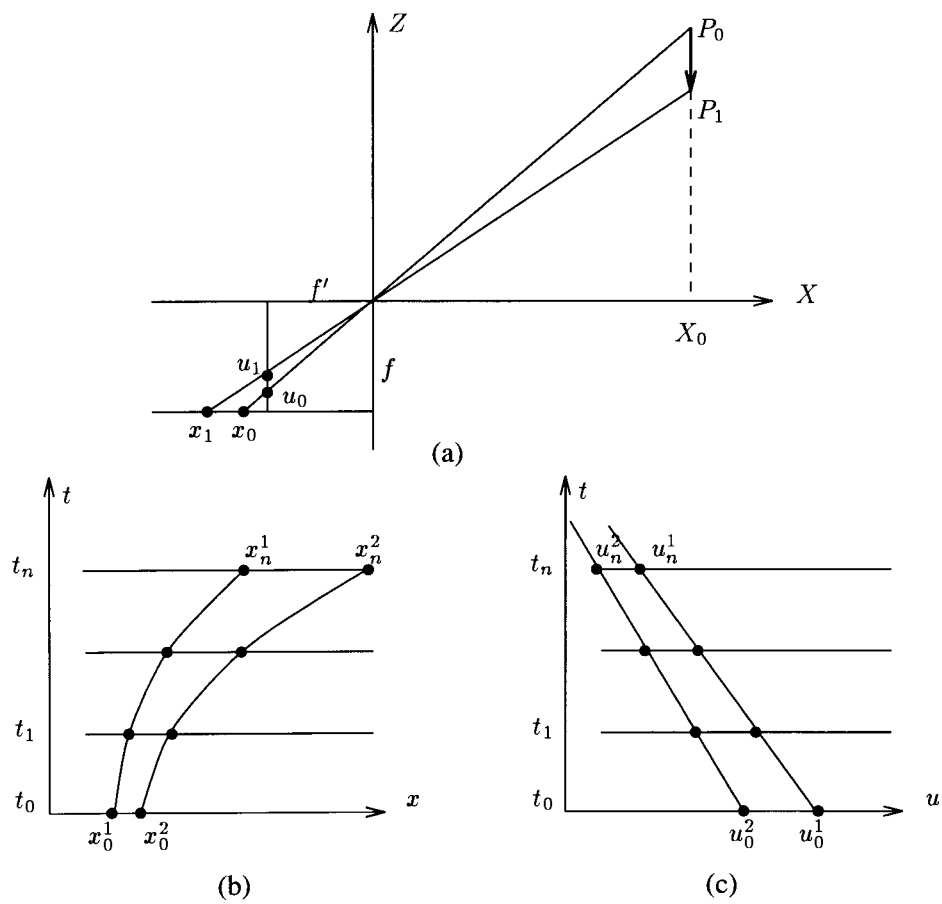



Figure 5.5: Longitudinal motion stereo.
(a) Imaging the longitudinal motion. (b) The $x$-$t$ plane from ordinary longitudinal motion stereo images. (c) The $u$-$t$ plane after the RWT.

From similar triangles,

$$\frac{x}{1} = \frac{X}{Z} .$$

Since there is no change in $X$, $X = X_0$,

$$\frac{dx}{dt} = -\frac{X_0}{Z^2}\frac{dZ}{dt} = -\frac{CX_0}{Z^2} \propto \frac{1}{Z^2} , \tag{5.1}$$

where $C = \frac{dZ}{dt}$ is the known constant speed.

If multiple images of the longitudinal motion stereo are used, then $x_0^k$, $x_1^k$, ...and $x_n^k$ are a sequence of corresponding points for the point $P^k$ at $t = t_0$, $t = t_1$, ...and $t = t_n$ in the $x$-$t$ epipolar plane. As shown in Figure 5.5(b), their locus is nonlinear (a curve), which is implied by eq. (5.1).

It can be shown that the reciprocal function used in the RWT happens to counterbalance the above nonlinearity. From Figure 5.5(a)

$$\frac{u}{1} = \frac{Z}{X} .$$

It follows,

$$\frac{du}{dt} = \frac{1}{X_0}\frac{dZ}{dt} = \frac{C}{X_0} ,$$

or

$$\frac{dt}{du} = \frac{X_0}{C} .$$

Therefore, $u_0^k$, $u_1^k$, ...and $u_n^k$ in the $u$-$t$ epipolar plane are collinear points, and the slope of their connecting line is the constant $\frac{X_0}{C}$. Moreover, the line equation is

$$t = \frac{X_0}{C}u + T ,$$

where T is the $t$-intercept. Since at $t = t_0 = 0$, $u = u_0$; and $u_0 = \frac{Z_0}{X_0}$, it can be derived that

$$T = -\frac{X_0}{C}\frac{Z_0}{X_0} = -\frac{Z_0}{C} .$$

This result immediately turns the problem of depth recovery in the longitudinal motion stereo into a simpler problem of detecting lines in the $u$-$t$ plane, where $t$-intercepts are proportional to *depth* of the point $P^k$.

### Extension to ego motion

In the following, the longitudinal stereo model is extended to a general case of ego motion in which the vehicle is moving on the $Y$-$Z$ plane with an axial velocity $\dot{s}$ and a rotational speed $\dot{\theta}$. Such a model typifies the road driving motion in which the vehicle is curving along the road. Within a short time span, the vehicle motion can be satisfactorily approximated with a circular course; that is, changes in both $\dot{s}$ and $\dot{\theta}$ over the time span of investigation are assumed to be negligible.

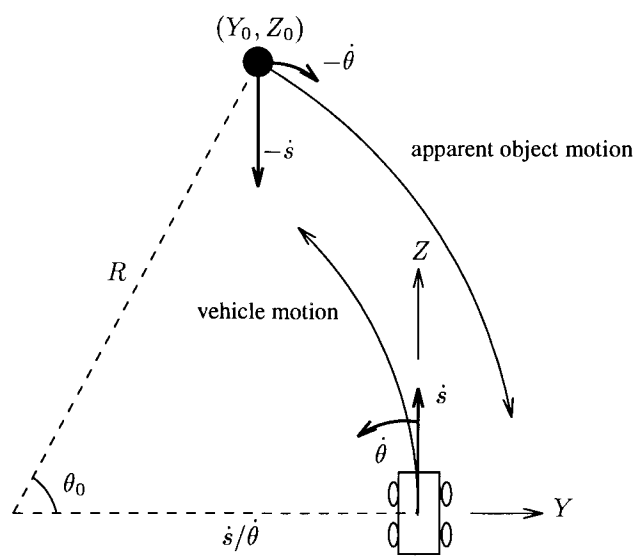



Figure 5.6: Motion of an object in relation to the vehicle.
In the world coordinates, the vehicle is traveling at an axial velocity of $\dot{s}$ and rotational speed of $\dot{\theta}$, describing a circular path of radius $\dot{s}/\dot{\theta}$. In the viewer-centered coordinates of the vehicle driver, the object is moving in the opposite direction with the same speed. It also appears to move on a circular trajectory at the same center.

Assuming the vehicle is moving in an otherwise static world, the apparent motion

of the world in the view of the vehicle driver is a composite motion of axial translation $-\dot{s}$ and centric rotation $-\dot{\theta}$. Take the vehicle driver as the center of reference and align the $Z$ axis with the direction of travel as depicted in Figure 5.6. At time $t$, the position of the object is at $(X, Y, Z)$.

$$\begin{aligned} \dot{Y} &= Z\dot{\theta} \ , \\ \dot{Z} &= -Y\dot{\theta} - \dot{s} \ . \end{aligned}$$

Solving the differential equations,

$$Y = R\cos(\theta_0 - \dot{\theta}t) - \dot{s}/\dot{\theta} \ , \tag{5.2}$$

$$Z = R\sin(\theta_0 - \dot{\theta}t) \ . \tag{5.3}$$

The form of the equations indicates a circular path for the object's apparent motion. $R$ is the radius of the circular path and the center is at $Y = -\dot{s}/\dot{\theta}$, $Z = 0$. At $t = t_0 = 0$, the object is at the position $(Y_0, Z_0)$, where $Y_0 = R\cos\theta_0 - \dot{s}/\dot{\theta}$ and $Z_0 = R\sin\theta_0$. Hence, $\theta_0$ is the arc angle on the circular path at which the object is initially located.

From Figure 5.5(a), the mapping from $(Y, Z)$ to $(u, v)$ is

$$u/f' = Z/X \ , \quad v/f' = Y/X \ . \tag{5.4}$$

Apply the mapping on eq. (5.2-5.3). The image motion on the $u$-$v$ plane now is

$$\begin{aligned} u &= \frac{f'}{X} R\sin(\theta_0 - \dot{\theta}t) \ , \\ v &= \frac{f'}{X} R\cos(\theta_0 - \dot{\theta}t) - \frac{f'}{X}\frac{\dot{s}}{\dot{\theta}} \ . \end{aligned}$$

Let $r = \frac{f'}{X}R$ and $a = \frac{f'}{X}\frac{\dot{s}}{\dot{\theta}}$. The $u$-$v$ motion equations can be rewritten as

$$u = r\sin(\theta_0 - \dot{\theta}t) \ , \tag{5.5}$$

$$v = r\cos(\theta_0 - \dot{\theta}t) - a \ . \tag{5.6}$$

Apparently, the $u$-$v$ motion is along a circular trajectory with the radius $r$, and the center of curvature is at $u = 0$ and $v = -a$ (see Figure 5.7(a)).

Use $\omega$ for the arc distance measured from the $v$ axis along the circular trajectory as shown in Figure 5.7(a). The advantage of using $\omega$ is that it shows a linear relationship with $t$ (see Figure 5.7(b)).

$$\omega = r(\theta_0 - \dot{\theta}t) . \tag{5.7}$$

One useful property of using the $\omega$-$t$ line is the readily computable extrapolated $t$-intercept. Putting $\omega = 0$ in eq. (5.7), the $t$ indicates the time in which a point comes to the $v$ axis. This time measure yields the *time-to-contact.*[1]

$$t = \theta_0/\dot{\theta} . \tag{5.8}$$

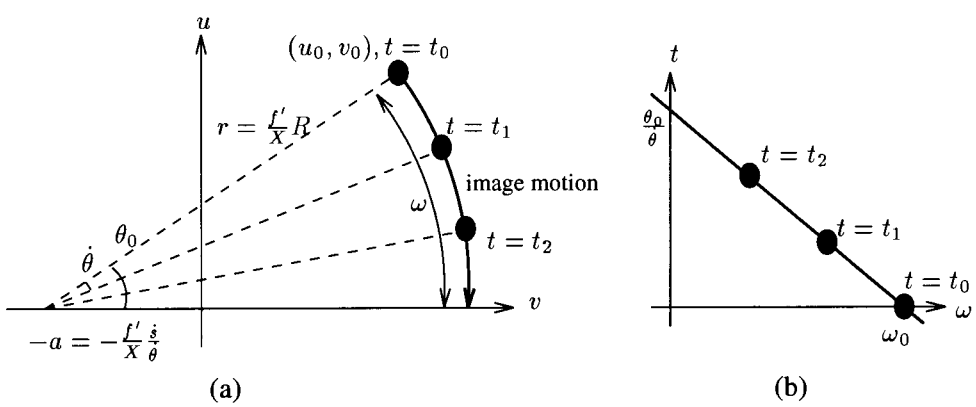



Figure 5.7: Image motion in $u$-$v$.
(a) The image motion is a circular arc centered at $-a = -f'/X \cdot \dot{s}/\dot{\theta}$ and has a radius of $r = f'/X \cdot R$. The initial arc angle is $\theta_0$ for the point position $(u_0, v_0)$ at $t = t_0 = 0$. The point is approaching at an angular speed $\dot{\theta}$. (b) When the arc length $\omega$ is measured against $t$, it shows a linear relationship. The $t$-intercept is $\theta_0/\dot{\theta}$.

The $x$-$y$ uniform resolution image represents the perspective projection of the driving scene. The reciprocal function used in the RWT counterbalances the perspective

[1]The intuitive interpretation of the time-to-contact is the time that the observer takes to come into contact with the plane in which the object resides.

nonlinearity and yields a linear mapping of the road surface (eq. (5.4)). The linear mapping enables the preservation of the circular image motion as corresponding to the original vehicle motion. Such is not the case in the $x$-$y$ image because of the perspective distortion

$$x/f = X/Z , \quad y/f = Y/Z ,$$

which results in a complicated movement on the $x$-$y$ plane.

A search algorithm can be devised to find the circular trajectories on the $u$-$v$ plane as described in eqs. (5.5-5.6). When visualized in the 3-D $uvt$ space, the circular trajectory becomes a helical curve. The search is essentially a problem of fitting the helical model to the $uvt$ data. Nevertheless, the search space is much restricted by exploiting the constraints due to the simple vehicle motion. The helical trajectory in $uvt$ has no more than two degrees of freedom even if none of the constants $f', \dot{s}, \dot{\theta}$ are known a priori. From eqs. (5.5-5.6), the center of the helix is on the $v$ axis. Choose $a$ for the position of the center in eq. (5.6). The radius $r$ and the arc length $\omega$ for each feature point $(u, v, t)$ from the RWT image sequence can be determined from eqs. (5.5-5.7). The helical trajectory of the point $(u_0, v_0)$ corresponds to a straight line passing through $\omega_0$ in the $\omega$-$t$ projection (Figure 5.7). Now, choose a value for the line slope such that the line passing through $\omega_0$ would fit to the $\omega$-$t$ projection of the feature points. The best fitted line over different values of $a$ yields the best solution to the helical trajectory of $(u_0, v_0)$. By eq. (5.8), the $t$-intercept of the $\omega$-$t$ indicates the time-to-contact.

The model of the longitudinal motion stereo for linear vehicle motion is a special case of this general model for ego motion. When $\dot{\theta} \to 0$, $\dot{s}/\dot{\theta}$ approaches $\infty$ and so do $r$ and $a$ in eq. (5.5-5.6). The circular trajectory therefore approaches a line along the $u$ direction, and the arc length $\omega$ now directly corresponds to the $u$ coordinate. In

the general model, the $t$-intercept in $\omega$-$t$ (eq. (5.8)) indicates the time-to-contact for an object if the vehicle motion were to prevail. In the special case of linear vehicle motion, the time-to-contact conveniently gives a direct measure of the depth of an object as well.

$$t = \frac{\theta_0}{\dot{\theta}} = \frac{R\theta_0}{R\dot{\theta}} = \frac{Z_0}{-\dot{s}} \propto Z_0 .$$

### 5.2.3 Lateral motion stereo

This section uses the example of the moving assembly line mentioned earlier. For simplicity, we first assume that the belt moves in the $X$ dimension in the 3-D space. Its projected movement on the $x$-$y$ plane is therefore along the $x$ direction only. For a point $x^k$ at $y = y^l$, $x_0^k$, $x_1^k$, ...and $x_n^k$ is a sequence of corresponding points at $t = t_0$, $t = t_1$, ...and $t = t_n$ in the $x$-$t$ epipolar plane from the original (ordinary) lateral motion stereo images, where the epipolar lines are horizontal (Figure 5.8(a)). When the speed of the belt is constant and images are taken at equal intervals, $x_0^k$, $x_1^k$, ...and $x_n^k$ fall on a single line in the $x$-$t$ plane, and $\frac{dx}{dt} \propto$ *disparity* $d$. Hence, the correspondence problem in the lateral motion stereo is equivalent to a problem of finding collinear points in the $x$-$t$ epipolar plane. Since the disparity is inversely proportional to the actual depth in the 3-D scene, it follows that $\frac{dt}{dx} \propto$ *depth* of the point $x^k$.

After the RWT, the epipolar line corresponding to $y = y^l$ remains a line in the $u$-$v$ space, and $v = y^l u$. The new epipolar line is generally at an angle with respect to the $u$-axis passing through the origin. We denote the distance between the point $w(u, v)$ and the origin as $\omega$. For a point $(x, y^l)$ on $y = y^l$,

$$\omega = \sqrt{u^2 + v^2} = \sqrt{1 + (y^l)^2}/x . \tag{5.9}$$

The epipolar plane for the lateral motion stereo becomes the $\omega$-$t$ plane as shown in

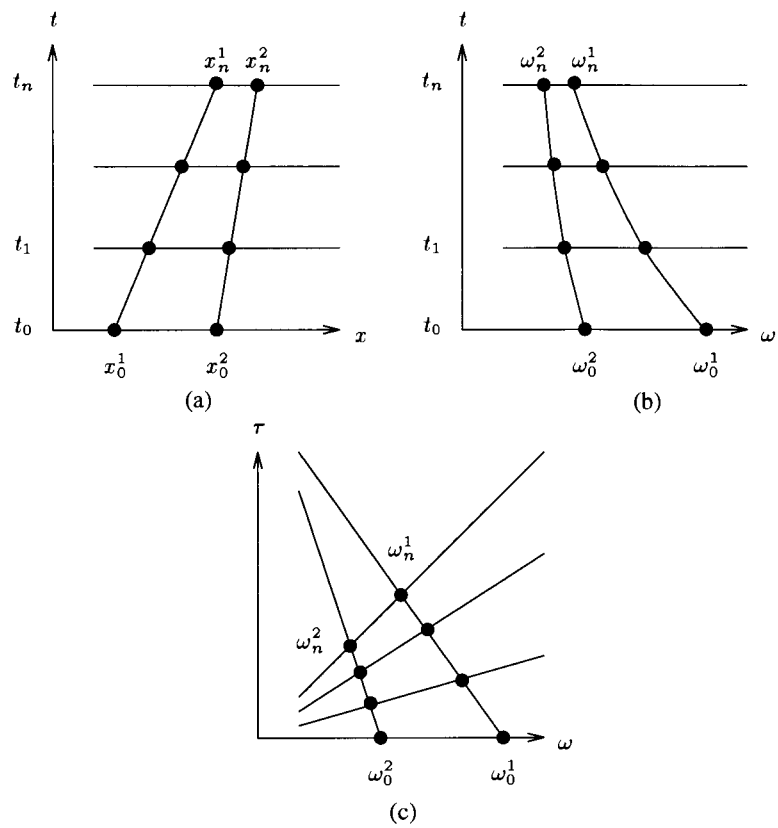



Figure 5.8: Epipolar planes in lateral motion stereo.
(a) The $x$-$t$ plane from ordinary lateral motion stereo images. (b) The $\omega$-$t$ plane where $\omega = \sqrt{1 + (y^l)^2}/x$. (c) The $\omega$-$\tau$ plane where $\omega = \sqrt{1 + (y^l)^2}/x$ and $\tau = t/x$.

Figure 5.8(b). Note the new sequence of the corresponding points $\omega_0^k$, $\omega_1^k$, ... and $\omega_n^k$ do not generally fall on a single line in the $\omega$-$t$ plane.

### Creation of a new $\omega$-$\tau$ epipolar plane

To recover the linearity, an additional transformation

$$\tau = t/x \tag{5.10}$$

can be applied to the variable $t$ which is similar to what is applied to $y$ in the RWT. The $x$-$t$ epipolar plane from lateral motion stereo images is now converted into a new $\omega$-$\tau$ epipolar plane of the RWT images (Figure 5.8(c)). The horizontal epipolar lines in the $x$-$t$ plane become concurrent epipolar lines converging toward the origin in the $\omega$-$\tau$ plane. The lines that connect the corresponding points also remain linear.

Suppose $L_{xt}$ is a line in the $x$-$t$ plane,

$$L_{xt} : \ t = m' \cdot x + c' \ .$$

Its transformation dual (derivable from eqs. (5.9, 5.10)) in the $\omega$-$\tau$ plane is $L_{\omega\tau}$:

$$L_{\omega\tau} : \ \tau = c'/\sqrt{1 + (y^l)^2} \cdot \omega + m' \ .$$

The slope $m'$ of line $L_{xt}$ becomes the $\tau$-intercept of $L_{\omega\tau}$ in the RWT motion stereo.[2] Instead of $\frac{dt}{dx} \propto depth$, it is now the case that the $\tau$-intercept of the line that connects the corresponding points in the $\omega$-$\tau$ plane is $\propto depth$ of the point $\omega^k$.

### Extension to any linear motion on $Z = Z_0$ plane

Although it was assumed above that the belt moves along the $X$ dimension only, this can be relaxed to any linear movement on a $Z = Z_0$ plane in the 3-D scene. The

[2]This is similar to the transformation dual in eq. (3.4), except the slope of $L_{\omega\tau}$ is not $c'$ because of the additional transformation on t (eq. (5.10)).

projected locus on the $x$-$y$ plane is the epipolar line $L_{xy}$ of which the slope $m$ and $y$-intercept $c$ are known parameters.

As shown by eq. (3.4), after the RWT the line $L_{xy}$ is transformed to the new line $L_{uv}$. Let $\zeta$ be the length of the line segment $L_{xy}$ from the $y$-axis to $(x, y)$,

$$\zeta = \sqrt{x^2 + (y - c)^2} = \sqrt{1 + m^2}\, x \ . \tag{5.11}$$

Apparently, $\zeta$ and $x$ have a linear relationship. A $\zeta$-$t$ epipolar plane (similar to the $x$-$t$ plane) can thus be constructed for the ordinary lateral motion stereo in which corresponding points are collinear. Let the line that connects the collinear points in the $\zeta$-$t$ plane be

$$L_{\zeta t} : \ t = m' \cdot \zeta + c' \ . \tag{5.12}$$

Now, let $\omega$ be the length of the line segment $L_{uv}$ from the $v$-axis to $(u, v)$,

$$\omega = \sqrt{u^2 + (v - m)^2} = \sqrt{1 + c^2}\, u \ .$$

Because $u = 1/x$, use eq. (5.11),

$$\omega = \sqrt{1 + c^2} \ / \ x = \sqrt{(1 + m^2)(1 + c^2)} \ / \ \zeta \ .$$

If we introduce a new parameter $\tau = t/\zeta$, then the line in the $\zeta$-$t$ plane (eq. (5.12)) will be converted into a line in the $\omega$-$\tau$ plane,

$$L_{\omega\tau} : \ \tau = (c'/\sqrt{(1 + m^2)(1 + c^2)}) \ \cdot \omega + m' \ .$$

In this way, the previous method for the lateral motion stereo can be extended to handle known linear motions on any $Z = Z_0$ plane.

### 5.2.4 Search in the epipolar plane

As described above, the correspondence problem in both the longitudinal and lateral motion stereo can be reduced to a problem of searching for collinear points in the

epipolar planes ($u$-$t$ plane for the longitudinal, $\omega$-$\tau$ for the lateral). Similar to the procedures for the Hough transform [DH72], a voting algorithm for accumulating multiple evidence can be developed. Without loss of generality, the search for linear motion on the $u$-$t$ plane in the longitudinal stereo will be used here to illustrate the method. (The extension to circular ego motion requires a somewhat different search, i.e., search for helical curves in the $uvt$ space. By introducing $\omega$ as the arc length, the problem was shown in Section 5.2.2 to be equivalent to finding collinear points on the $\omega$-$t$ plane. For efficiency, a slightly different search algorithm was suggested earlier in Section 5.2.2.)

In general, any point at $t = t_i$ can be paired with any point at $t = t_j$ $(j > i)$ to form a hypothetical line segment. Its intercept on the $t$-axis suggests a possible depth value which is inversely proportional to the disparity $d$. A 3-D $uvd$ voting space is created[3] and each hypothetical line will cast a vote at the position $(u, v, d)$ in the $uvd$ space. Since $n + 1$ collinear points can form $O(n^2)$ hypothetical lines and they will vote to the same $(u, v, d)$, a peak will be formed in the $uvd$ space which indicates the consensus on the correct disparity value for the point $(u, v)$. The line detection problem can thus be solved by this voting procedure followed by a peak-detection procedure.

On each $u$-$t$ plane at $t = t_i$ there are $k_i$ edge points, i.e., $u_i^1$, $u_i^2$, ...and $u_i^{k_i}$. A complete pairing of two possible end points at $t_i$ and $t_j$ will produce numerous hypothetical line segments and therefore clutter the $uvd$ voting space. The following heuristics are employed to improve the voting process:

**Use relatively long hypothetical voting lines.** Due to limitations of the image resolution there is always some error in the $u$-$v$ coordinates, especially at the

[3]Since the concerned depth in the scene can be very large whereas disparity $d$ usually has a small range, it is preferable to use $d$ for the voting space.

periphery of the RWT images. If the short hypothetical lines were to be used for voting, a small amount of error in the $u$-$v$ coordinates would result in relatively large errors in the calculation of the slope and intercept, and consequently the disparity values. A minimum length is therefore chosen to exclude the short voting lines.

**Specify a reasonable range for depth.** A range of concerned depth can be represented as $[T_{min}, T_{max}]$ to reduce the number of candidate pairs. The vertices $T_{min}$ and $T_{max}$ on the $t$-axis and the lower end point $u_i^k$ form a triangle which defines the search region for the possible pairing end point $u_j$.

### 5.2.5 Experimental results

#### Longitudinal motion stereo

A vehicle navigation example is used to illustrate the longitudinal motion stereo. Figure 5.9(a) shows a CMU image sequence of a road scene obtained from a driving expedition. Four frames of the 8-snapshot sequence (each has a size of 512 × 512 pixels) are shown to visualize the forward motion from driving. The RWT images of the motion sequence have been generated in software. The data reduction factor is over 90%. Figure 5.9(b) shows the RWT edge images.

Some implementation details should be followed when generating the RWT images. First, the $X$-axis in the world coordinate system is the vertical axis as indicated in Figure 5.1. Accordingly, the $x$-axis in the $x$-$y$ images and the $u$-axis in the RWT ($u$-$v$) images are the vertical axes in these images. Second, the model of our longitudinal stereo requires both the camera movement and its optical axis be along the $Z$-axis. According to this simple model, the FOE (Focus of Expansion) is always at the center of the $x$-$y$ road images. When dealing with a FOE which is significantly off center because of intentional pan/tilt on the camera orientation, the FOE must be

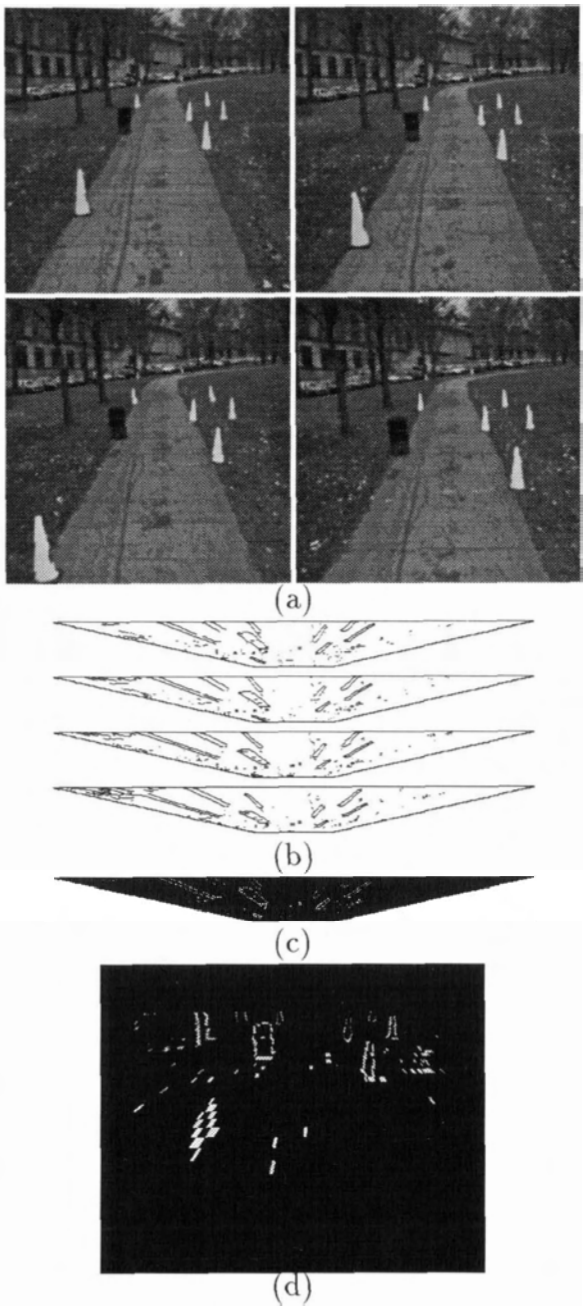



Figure 5.9: Depth computation using the RWT in linear motion.
(a) A sequence of a driving scene, only images 1, 3, 5, and 8 are shown. (b) Edge images from the above RWT images. (c) Gray-level coded depth map computed from all eight images. (d) The depth map transformed back to the $x$-$y$ space (uniform-resolution) for visual apprehension.

determined and used as the origin of the $x$-$y$ space for the RWT transformation. This is the situation in the CMU image sequence which apparently had the camera pointing slightly toward the ground.

Even under the best effort to align the camera orientation with the vehicle movement, the FOE could still be off center slightly. As a result, the epipolar line may not align perfectly with the $u$-axis. To accommodate the resulting error, the search region for collinear points used in the Hough method discussed in Section 5.2.4 has been relaxed accordingly. That is, instead of searching on an epipolar plane, a neighborhood of the plane was employed as the search region.

In the images in Figure 5.9(b), some portions of the trees and buildings are not shown, because they are either above the FOE or too close to the singularity ($x = 0$) line to be included. The rest of the scene is very well captured in all the RWT images. One can also observe the advantage of the variable-resolution imaging in this example as the excessive details in the near side of the road, which are not as relevant to the driving task, are averaged out in the coarse resolution periphery of the RWT images.

The algorithm described in Sections 5.2.2 and 5.2.4 has been implemented. The correspondence ambiguities are resolved successfully and good depth recovery results are obtained. Figure 5.9(c) shows the grey-level coded depth map. In Figure 5.9(d), the RWT depth map is transformed back to the uniform-resolution $x$-$y$ space so that the relationship to the original road image can be better apprehended. Note that the depth values of the traffic cones, the trash can and the tree trunks are correctly resolved.

### Extension of longitudinal motion stereo to ego motion

A sequence of 20 motion images ($400 \times 494$) of a table scene was taken in the lab using the SFU hybrid pyramidal vision machine [LTR95]. Four of them are shown in Fig. 5.10(a). The camera was mounted on the NOMAD 200 mobile robot. The NOMAD was moving forward while turning left.

As before, the $X$-axis in the world coordinate system, the $x$-axis in the $x$-$y$ images and the $u$-axis in the RWT ($u$-$v$) images are the vertical axes.

By calibrating the camera it is determined that the $Y$-$Z$ plane on which the camera makes the circular movement is slightly below the whiteboard. In this way, the $y$-axis (where $x = 0$) on the $x$-$y$ image is determined. The center of the axis is taken as the origin for the RWT. The whiteboard in the scene is above the origin which is not in the lower half of the $x$-$y$ image in consideration here. Fig. 5.10(b) shows the edge maps for the RWT images for the lower half of the table scene. As before, the top portion of the tape boxes and cup are excluded because they are too close to the origin. The effect of spatially variable-resolution sensing is apparent. In this case, the front edges of the table are in the periphery and compressed in the RWT images.

Since the projections of the movement in the $uvt$ space follow a helical curve, search is conducted along such possible curves in the 3-D $uvt$ space, which reduces the complexity of matching significantly. For a given $a$ and $(u_0, v_0)$, the helical trajectory is well-defined and incurs little ambiguity in possible matching candidates on the locus. After gathering the matching points, their arc length $\omega$ is calculated and used to derive the "time-to-contact". Fig. 5.10(c) is the grey-level coded map of time-to-contact in the RWT domain and Fig. 5.10(d) is the map in the original $x$-$y$ domain generated by an inverse RWT.

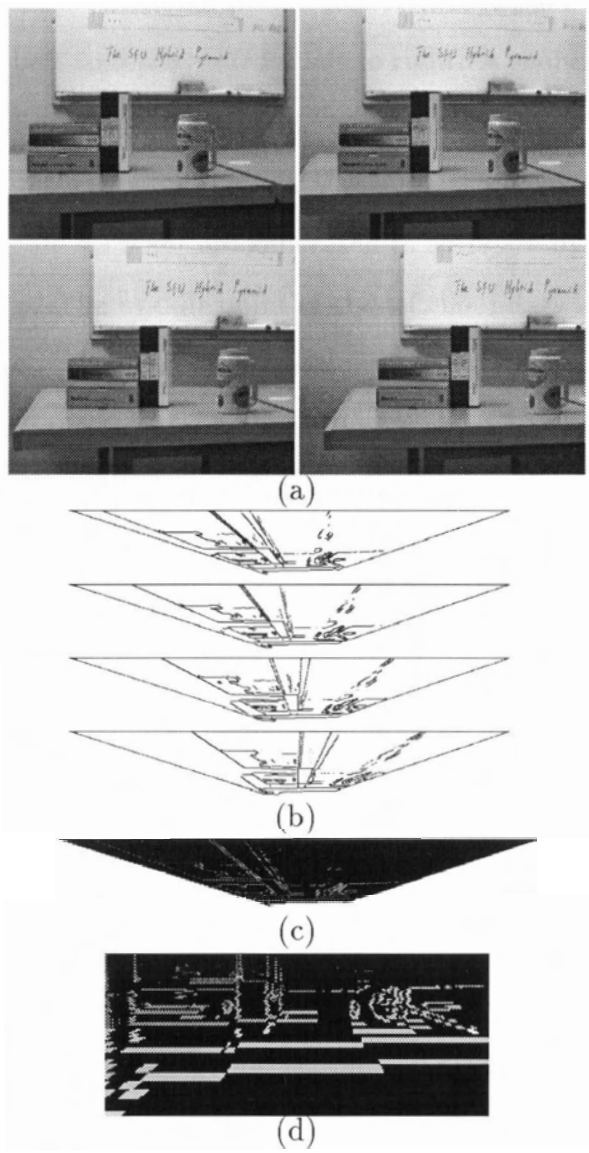
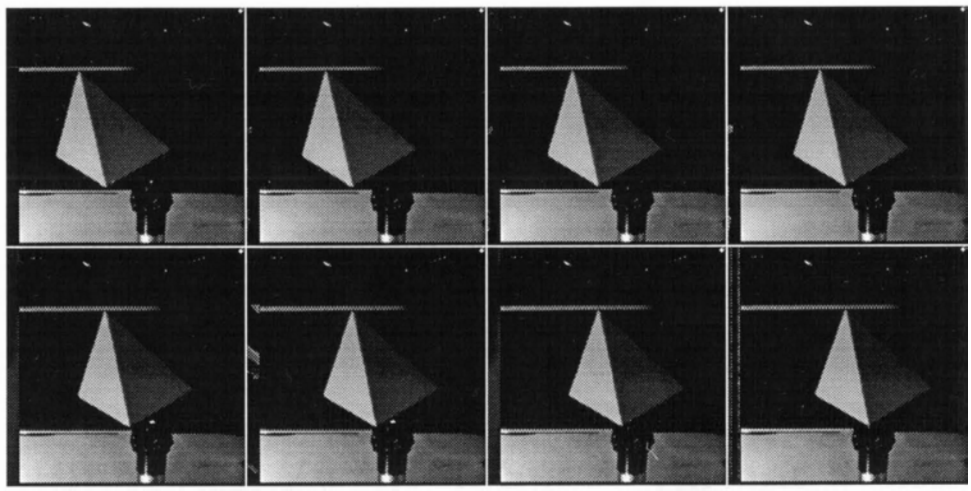



Figure 5.10: Analysis of ego motion.
(a) A dynamic sequence of an office scene, only images 1, 7, 13, and 19 are shown. (b) Edge images from the above RWT images. (c) Map of time-to-contact computed from all twenty images. (d) Map of time-to-contact transformed back to the $x$-$y$ space.

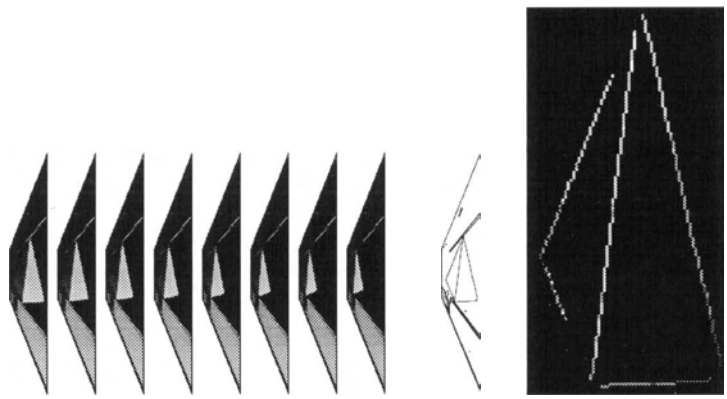
### Lateral motion stereo

For obtaining lateral motion stereo images in our lab, a pyramidal wooden block is placed on a conveyor belt that moves from left to right. A sequence of eight snapshots (each has a size of $512 \times 512$ pixels) from a conventional CCD camera is used in the experiment (Figure 5.11(a)) since the RWT camera is not available yet. As before, the RWT images (Figure 5.11(b)) are generated in software by mapping the original images from $x$-$y$ to $u$-$v$. The middle point of the left boundary of the $x$-$y$ image plane is used as the origin for this mapping. In our experiment, the area of the resulted RWT images is chosen at approximately 1/10 of the original images.

Gradient-based edge detection is first performed on the RWT images. Figure 5.11(c) shows the edge map from the first RWT image. Collinear points in the $\omega$-$\tau$ plane are detected and their $\tau$-intercept yields the depth and, indirectly, the disparity. The voting in the *uvd* accumulator space results in clusters yielding the correct disparity. Figure 5.11(d) displays the depth map. The result shows that most of the disparity changes along the edges of the pyramid are computed successfully.



(a)



(b) (c) (d)

Figure 5.11: Depth computation using the RWT in lateral motion stereo.
(a) Ordinary lateral motion stereo images of a pyramidal block on a moving belt. (b) Software-generated RWT images. (c) Edge map of the first RWT image. (d) Grey-level coded depth map for the pyramidal block from variable-resolution lateral motion stereo.

# Chapter 6

# Active Stereo

## 6.1 Binocular Vision in Space-variant Sensing

Experiments have shown that the human stereopsis accepts only a very limited range of disparities. The Panum's area forms a limited zone about the fixation point. Beyond the Panum's area, we can no longer fuse the stereo images. In computer vision, stereo correspondence is linked to the fusion of two disparate retinal images. The problem is formulated as computing the image disparity within an operating range.

Correspondence algorithms are normally incorporated with the various matching constraints to render the problem solvable. Uniqueness, continuity [MP76], and the figural continuity [MF81] are the commonly used ones. Burt and Julesz [BJ80b] conducted some experiments on fusion in the context of disparity gradient. An amendment to the previous understanding of Panum's fusional area was made. Binocular fusion occurs only when the disparity gradient does not exceed a critical value of $\sim 1$. Li [Li94a] generalized the notion of disparity gradient to subsume various constraints for stereo matching.

After more than three decades of intensive research in stereo vision, the computational framework for stereopsis from uniform resolution images has been relatively well-established. The link to psychological vision is that correspondence is computed as the fusional result, and the disparity yields the 3-D percept. As the methods devised are mostly for accurate recovery of the image disparity, the process can be considered as computing the foveal fusion in the domain of space-variant sensing. However, the structure and functional objective of the peripheral vision are distinguished from those of the foveal processing. The issues of peripheral fusion have not received much attention. This may be in part due to the lack of research in anthropomorphic sensors. With the invention of the space-variant sensor [VdSKC+89], the issues related to active stereo has received attention in recent years.

In this chapter, we shall investigate the Panum's fusion in the context of space-variant binocular sensing. Specifically, the computational view of the Panum's fusional area in the space-variant RWT sensing space will be studied, and a model of the fixation mechanism in an RWT binocular system will be presented.

### 6.1.1 Panum's fusional area

Objects on the horopter form stereo images on the corresponding retinal elements in the two eyes. Images of zero disparity as such are perfectly fusible, and are seen single. Panum (1861) showed that zero disparity is not the necessary condition for singleness [Ogl64]. An image on one eye would fuse with a similar image on the retina of the other eye within a small area about the corresponding point.

Consider the zero disparity case. Suppose the eyes are fixating an object $P_{Horopter}$ (Figure 6.1). $P_{Horopter}$ is on the horopter. The object forms zero disparity images in the two eyes, and thus is seen single. Another object $P_{Inner}$ is located to the inner

side of $P_{Horopter}$. As $P_{Inner}$ is moved towards the viewer, at a certain point, one will no longer be able to fuse the images and start to see double. Similarly, the object $P_{Outer}$ to the outer side of $P_{Horopter}$ is seen double when it is sufficiently away from the horopter. This type of doubling is known as physiologic diplopia. The images produced are said to be crossed disparate, and uncrossed disparate, respectively. The interval between $P_{Inner}$ and $P_{Outer}$, where no doubling is seen, defines the limits of the Panum's fusional area.

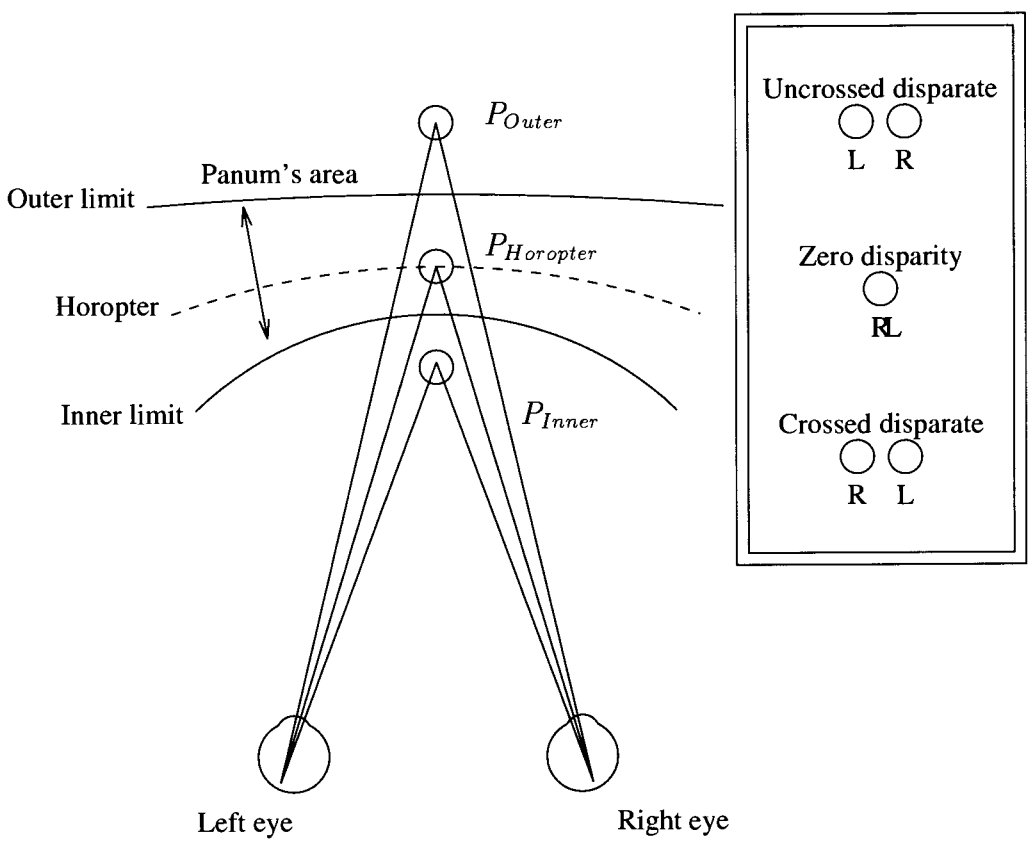



Figure 6.1: Panum's fusional area.
Within a region about the horopter, disparate images are fused despite their images not falling on the corresponding retinal elements. To the outer side of the Panum's area, uncrossed disparate images are seen. Objects to the inner side yield crossed disparate images.

In human vision, it is functional to address only a small range of disparities near the fovea because thereby one can filter out the irrelevant visual information and concentrate on the object of interest about the fixation point. One wants to keep a visual account, albeit coarse, of the environment in the visual periphery, because monitoring of the general environment is crucial for smooth ego motion and fast response to impending activities. Quantitative studies by Fischer (1924) and Ames (1932) yield data that plot out the size of the Panum's area at different visual angles [Ogl64]. Fender and Julesz [FJ67] reported that binocular fusion occurs in regions vary from 6 min. of arc at the center of the visual field to 20 min. of arc at the peripheral angle of 6°.

An extended Panum's fusional area is perhaps ideal for accurate spatial perception of the scene. However, it is unrealistic because it represents too great a demand on the fusion process, as fusion would be expected to be performed over an excessive range of disparity. Olson [Ols93] believed that stereopsis plays an ecological role of privileged computational resource, like the fovea that provides information about fixated targets only. The severe limitation of the size of the Panum's area is seen as beneficial, since binocular single vision is focused on the fixated target while stimuli from the rest of the scene are largely filtered out as irrelevant.

While Olson compares the role of Panum's area to the functional value of the retinal fovea, we relate the rapid dilation of the Panum's area at the peripheral visual angles to the coarse sensor resolution at the retinal periphery. When one interacts with the environment, accurate foveal processing serves well for attentive inspection of the fixated target. However, general monitoring of the wide visual field is obviously important for detection of activities, smooth maneuvering and the spatial percept of the external environment.

In this thesis, a functional perspective that relates the spatial extent of the Panum's area to space-variant sensing resolution is adopted. A uniform resolution image does not meet the requirements of this fovea-periphery structure for visual processing. It has neither the sufficient resolution for foveal vision, nor the coarse resolution for peripheral processing. In particular, to achieve a deep sensing range at large eccentricity, the disparity calculation has to be carried out over an excessively wide range. This is because the uniform resolution image data contain information far too detailed for peripheral vision purposes. In this respect, the space-variant resolution is highly desirable. The RWT image is suitable for space-variant processing as it can support a good foveal resolution and, at the same time, a desired level of coarse resolution in the periphery. Furthermore, the RWT simplifies the disparity computation because its variable resolution is affected primarily in the horizontal dimension. The horizontal displacement inflicted in stereo images due to the binocular disparity is well captured in the RWT representation.

## 6.2 Computational Model for Binocular Fixation

### 6.2.1 Fusional range in RWT

In computer vision, the fusional range is computationally modeled by disparity limits. Olson and Coombs [OC91, CB92] perform real-time pursuit on a fixated object by running a near-zero-disparity filter on the stereo images. The verging system of Olson [Ols93] operates in a limited range of disparities ($-3$ to 3). Based on the studies of anomalous stereopsis [Ric71], Barnard [BF90, Bar90] computes image disparities at 3 values only, namely 1 for crossed disparate, 0 for near-zero disparity and $-1$ for uncrossed disparate images.

The significance of the variable extent of the Panum's area has not been attended to. We address the issue of variable Panum's area in relation to the space-variant retinal resolution. In particular, the RWT we develop in this thesis supports space-variant resolution. It also achieves a variable fusional region. In the following, a binocular system of RWT cameras is studied. We set up the projection equations and fed them to Maple V [Red94] (a numerical software for scientific computation) to obtain the plots of the disparity contours for the different fusional limits.

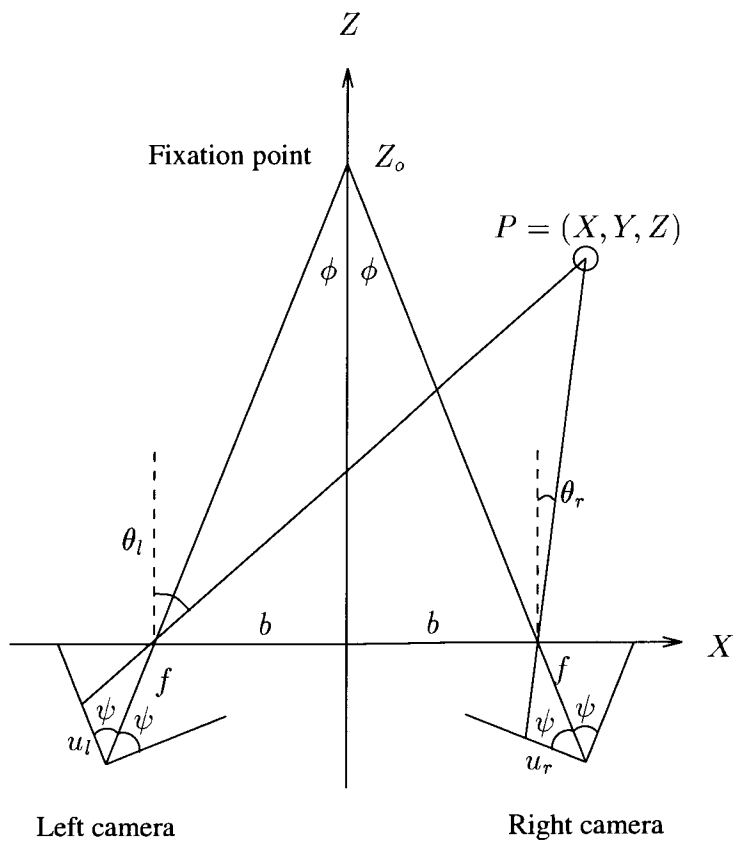



Figure 6.2: An RWT binocular system.

Figure 6.2 gives a schematic diagram of the RWT binocular system. The cameras are placed symmetrically on the two sides about the $Z$-axis, with their nodal points on the $X$-axis, and imaging the positive $Z$ half-space. Let $2b$ be the baseline separation

of the cameras. The focal length of the cameras is denoted by $f$, and the inter-projection-plane angle is $2\psi$. The cameras are fixating the point $Z_o$ on the $Z$-axis. Let $P$ be a point located at $(X, Y, Z)$; $u_l$ and $u_r$ are the RWT coordinates of the left and right images of $P$ respectively. Let the disparity be denoted by $d$. The triangulation geometry in Figure 6.2 yields the following equations:

$$\begin{aligned}
\frac{-u_l}{\sin(\theta_l - \phi)} &= \frac{f}{\sin(\theta_l - \phi + \psi)}\ , \\
\frac{-u_r}{\sin(\theta_r + \phi)} &= \frac{f}{\sin(\theta_r + \phi + \psi)}\ , \\
\tan(\theta_l) &= \frac{X + b}{Z}\ , \\
\tan(\theta_r) &= \frac{X - b}{Z}\ , \\
d &= u_l - u_r\ .
\end{aligned}$$

The system of equations are solved for $X$ and $Z$ at different disparity values, $d$. Without loss of generality, set $b = 200$, $f = 200$, $\phi = 45°$, and $Z_o = 6000$ (in 1/100 inch unit). The numerical values of $X$ and $Z$ are calculated for $d$ ranging from $-4$ to $+4$. Figure 6.3 plots the $(X, Z)$ coordinates for $d = 0, \pm 2, \pm 4$. Each of the curves represents a disparity contour of a particular $d$. All points on the same contour will form disparate images in the two RWT cameras with the disparity $d$. These contour are due to the specific imaging configuration of the RWT binocular system. However, the corresponding fusional region indeed exhibits the desired property of fovea-periphery variable extent. In this example, the fusional region at the peripheral angle of 36° is twice as deep as that at the central position.

Comparison of the RWT binocular system are drawn with the conventional uniform-resolution cameras. The model of a verging system of uniform-resolution cameras is

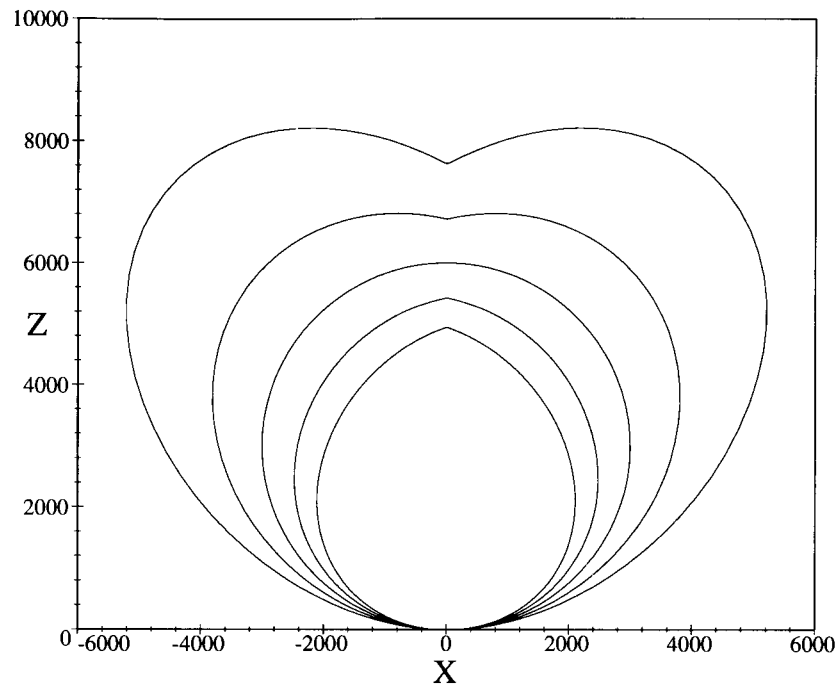



Figure 6.3: Disparity contours for the RWT binocular projection.
The plot is obtained by setting the baseline separation $2b = 400$, the focal length $f = 200$, and the inter-plane angle $\phi = 45°$. The cameras are converged at the fixation point of 6000 on the cyclopean axis. From the outermost contour to the innermost one, the disparity contours are plotted in the order of $d = +4, +2, 0, -2, -4$.

given in Figure 6.4. This time, the set of equations yielded read as follows:

$$\begin{aligned} \frac{-x_l}{f} &= \tan(\theta_l - \phi) \ , \\ \frac{-x_r}{f} &= \tan(\theta_r + \phi) \ , \\ \tan(\theta_l) &= \frac{X + b}{Z} \ , \\ \tan(\theta_r) &= \frac{X - b}{Z} \ , \\ d &= u_l - u_r \ . \end{aligned}$$

Again, the system of equations are solved in Maple V for $X$ and $Z$. Similarly, a plot

of the $(X, Z)$ coordinates is performed for $d = 0, \pm 2, \pm 4$, with the settings of $b = 200$, $f = 200$, and $Z_o = 6000$ (Figure 6.5).

The graph shows that the desired fovea-periphery variable fusional region is not achieved in the uniform-resolution case. Inversely, the dimension of the fusional region decreases with eccentricity. With the set of settings in use, the fusional region is reduced to half at the peripheral angle of 36°. Apparently, it is not suitable for a peripheral field which is both wide and deep.

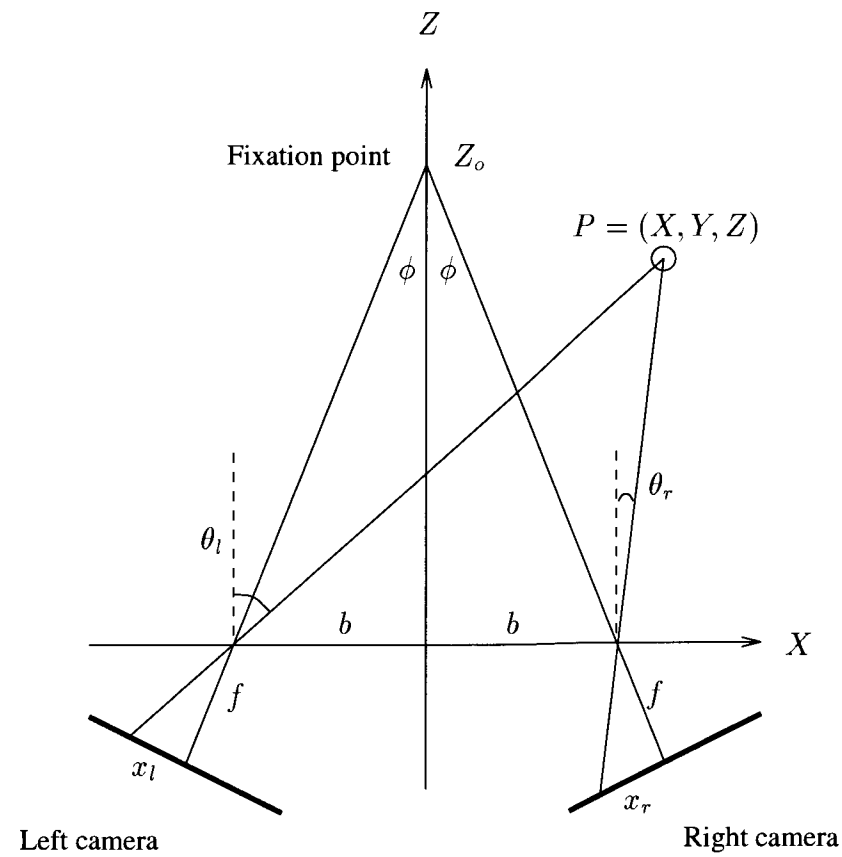



Figure 6.4: A verging system with uniform-resolution cameras.

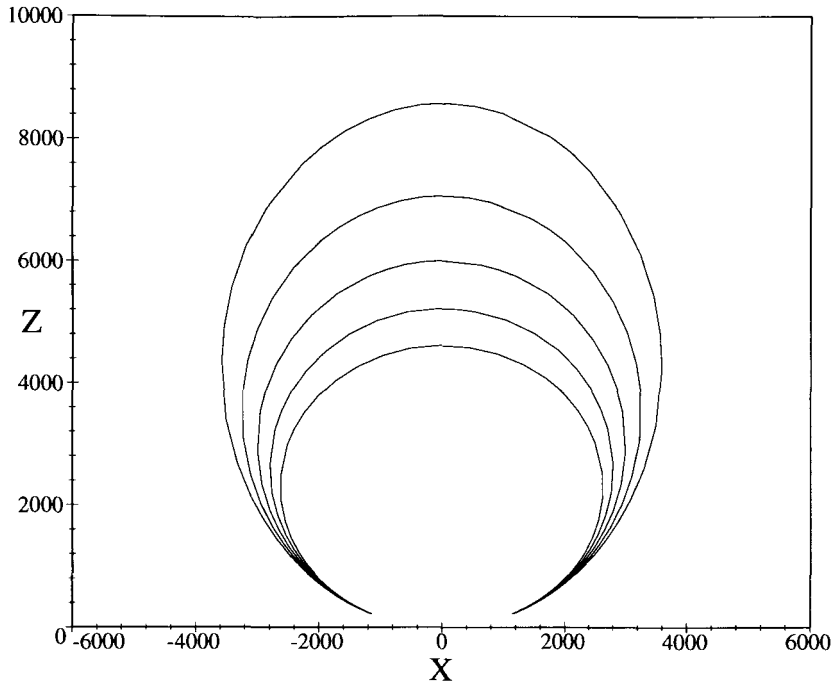



Figure 6.5: Disparity contours for uniform-resolution cameras.
The plot is obtained by setting the baseline separation $2b = 400$, the focal length $f = 200$, and the fixation distance $Z_o = 6000$. From the outermost contour to the innermost one, the disparity contours are plotted in the order of $d = +4, +2, 0, -2, -4$.

### 6.2.2 Fixation mechanism

Psychological studies have shown that the oculomotor mechanism for binocular fixation is effected by a mixed movement of vergence and version of the two eyes (Figure 2.3). In this thesis, we develop a computational model for the similar camera movement in relation to the computation with space-variant image resolution.

Experiments show that when one changes fixation to a nearer target point, the two eyes first undergo a symmetrical vergence to bring the fixation nearer to the target. In the middle of the vergence movement, a conjunctive saccade is superimposed to

swing the gaze in line with the target. The vergence then proceeds to completion in the final stage to bring the fixation accurately to the target.

If cameras of uniform sensing resolution were used, the binocular fixation process would be much simplified. Shifting from one fixation to another would involve calculating the exact image disparity and angular position of the target. The process could then be accomplished by generating independent pan-tilt movement to each camera, since it is possible to complete exact calculation for the target at the previous fixation. To assume such a retina, ignores all the problems ranging from hardware requirements to processing complexity. After all, to make such an assumption would beg the question of whether the fixation process was genuinely necessary to perception, since the high resolution sensory data of the scene is already available without the need for specific gaze control.

There is no doubt that uniform resolution cameras could hardly be supported. In fact, it is apparent that there is a strong relevance of space-variant sensor resolution to the unique camera movement for binocular fixation.

From the computational point of view, space-variant sensor resolution supports fusional area of variable size. This is because a disparity near the point of fixation yields refined and narrow depth range; whereas, the same disparity at the periphery corresponds to an coarse but deep depth range. Thus, the variable fusional area is not only functional, it also represents a logical structure in space-variant sensing.

The unique camera movement now becomes natural in a binocular system with space-variant sensors. Consider the case when the cameras are fixating an object $A$ in the scene, and is about to change gaze to a nearer one $B$ at periphery. $A$ is fixated in the fusional region at the fovea. $B$, although located in the periphery, is covered in a deeper fusional area. Computationally, the fusional area's limit is used to the

advantage for restricting the disparity range. Under the limited operating range for disparity, $B$'s disparity is readily resolvable even though its depth differs very much from that of the fixation. If the cameras were straightforwardly gazed at $B$ at this time, $B$ might become out of the fusional limit when it is brought into the foveal direction. The depth of $B$ would be difficult to calculate and the fixation would fail. A more effective mechanism is to have a first vergence to change the fixation distance so that $B$ is lying close to the horopter after the vergence. This also prepares for the versional movement so that when $B$ is brought to the foveal direction, it will still be imaged within the fusional limit. Next, based on the rough estimate of $B$'s visual angle, a pan movement is launched to direct both cameras to the direction near $B$. Now, $B$ is in a near-foveal direction, and located within the fusional limit. This is true owing to the first vergence. Finally, a second vergence can be executed to bring $B$ accurately into fixation.

Figure 6.6 summarizes the camera movement of a space-variant binocular sensor. As a matter of fact, it resembles the eye movements observed in the human visual system [Yar57] (Figure 2.3).

## 6.3 Binocular Fixation using RWT Images

RWT fits in the model described above. The RWT supports a space-variant sensing resolution. As we have discussed, the unique camera movement of binocular fixation is closely related to space-variant sensor resolution.

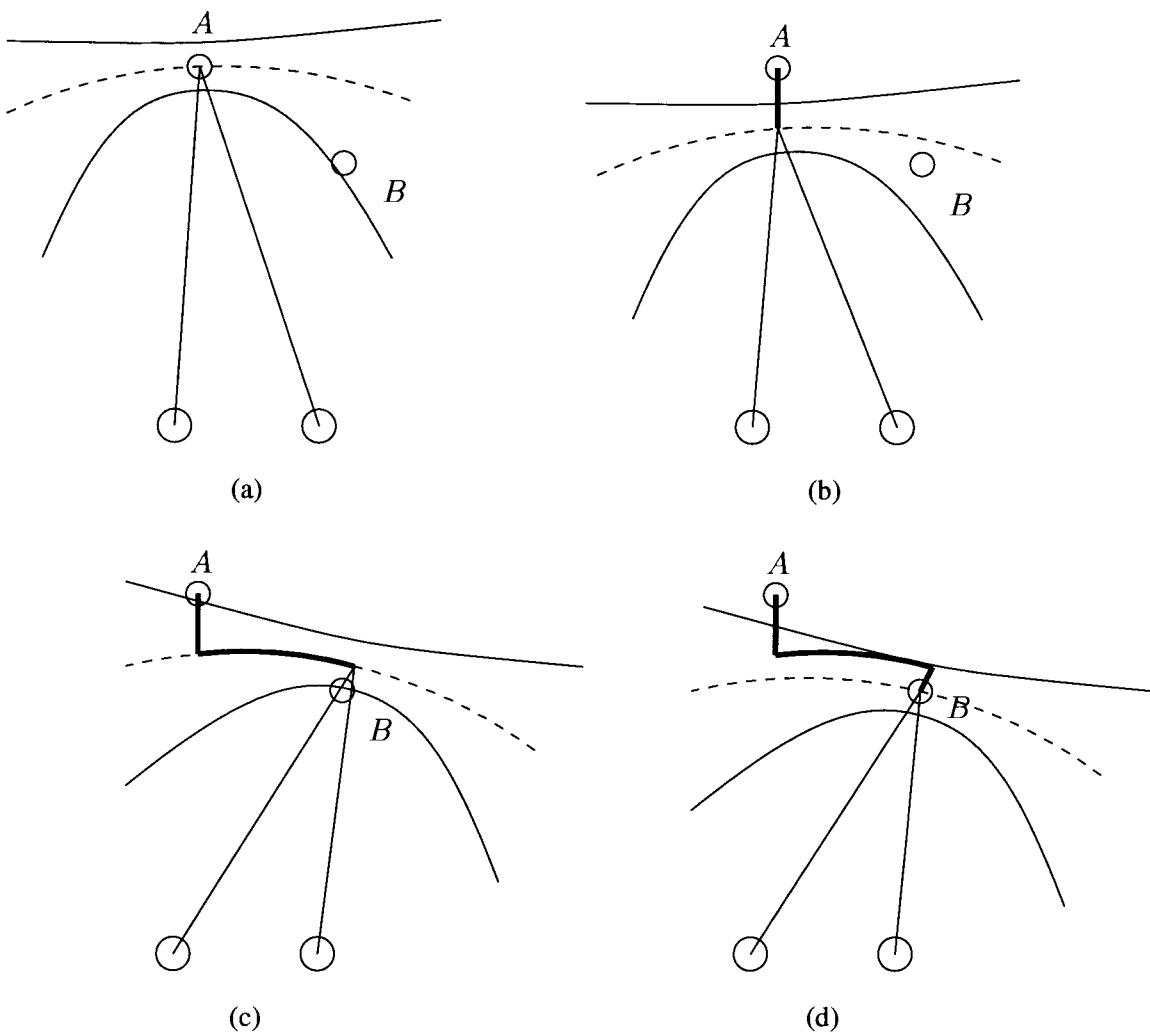



Figure 6.6: Ocular movement of space-variant binocular sensor.
(a) The cameras are fixating $A$. (b) First vergence brings the fixation point to close to $B$'s depth. (c) Version brings the cameras in line with $B$. (d) Second vergence, the cameras fixate precisely on $B$.

### 6.3.1 Disparity computation

Another property that renders RWT suitable for stereo vision is the anisotropy of its space-variant resolution. In stereo vision, the disparate images formed in the binocular cameras differ from each other by a horizontal displacement. It is this disparity that indicates the depth of the imaged object. In the conventional images, disparity is computed by correlation along the horizontal dimension. A rectangular pattern in the Cartesian image appears as shifted along the horizontal streamlines (Figure 6.7(a)). Recall from Section 3.2.1, the horizontal streamlines are mapped to radials in the RWT domain. Figure 6.7(c) shows the bipolar RWT image. The radial streamlines converge at the two antipodes on the $u$-axis. In the RWT image, the rectangular pattern is transformed into a wedged rectangle displaced along the radial streamlines.

Disparity computation may become very complicated in other schemes of image representation. In the log-polar model, horizontal streamlines are mapped to complicated log-sine curves (Figure 6.7(b)). The difficulty is at least two-fold. First of all, disparate images are not related in a linear structure any more. Search for stereo correspondence has to be conducted along these log-sine curves which are expensive to compute. In addition, the image pattern gets rotated and scaled while being translated along the log-sine curve. A complicated procedure is required to calculate the image motion in order to make it possible for a correlation operator to be used for the disparity computation [GLW92].

The anisotropic property of the RWT space-variant resolution effects the mapping primarily along the $x$ dimension only. The $y$ dimension is largely unaffected except by being scaled according to $1/x$. The verticals in the $x$-$y$ grids are invariantly mapped to verticals. The horizontals are mapped into radial lines. In spite of that the image

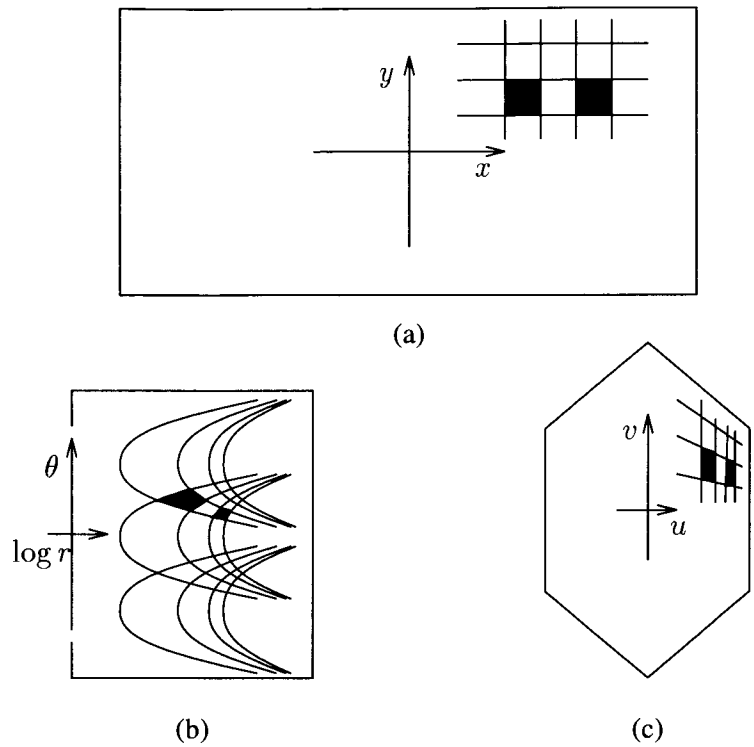



Figure 6.7: Disparity in different image representations.
(a) Disparity is manifested in horizontal translation in the Cartesian image. (b) Horizontal translation becomes a complicated image motion in the log-polar domain. (c) Horizontal translation is mapped to translation along the radial streamlines in the RWT image.

pattern gets scaled under space-variant resolution when translated along the grid lines, image rotation which occurs in log-polar transform is not inflicted in the RWT domain. The equations for correspondence in RWT domain do not contain rotational components. If $d$ is the image disparity, the left and right RWT image coordinates can be written as:

$$\begin{aligned} \text{Left image point} &: (\ u\ ,\ v\ )\ , \\ \text{Right image point} &: (\ u+d\ ,\ v+\frac{d}{u}\cdot v\ )\ . \end{aligned}$$

In the experimental tests, application of the correlation operator along the radial

streamlines yields good estimates of the RWT image disparities.

### 6.3.2 Fixation transfer

For simplicity, the correlation method is used as an operator for disparity computation. A windowed correlation is performed on the RWT stereo images within a limited operating range of disparity that corresponds to the space-variant fusional area.

In an RWT binocular system, when changing from the current fixation to another target at the visual periphery, the model for camera movement described in Section 6.2.2 is followed. A variable fusional area results from the space-variant pixel resolution. Upon changing gaze from the current fixation point to the next target, the target may be located well within the fusional limit at the periphery. A rough estimate for the target's peripheral disparity is calculated. The two cameras are then converged/diverged to reduce this disparity. This corresponds to the first vergence movement. Next, the cameras are panned to the viewing angle of the target to bring the target to the fovea of the RWT cameras for higher resolution imaging. This operation corresponds to the versional movement. The target now in the foveal direction of the cameras is likely imaged with a residual foveal disparity. Correlation is performed in the fovea. Based on the resulting disparity the cameras are converged/diverged to zero in on the target precisely. This movement corresponds to the second vergence.

Figure 6.8(a-d) shows a test on a computer simulation of the fixation process in an RWT binocular system. An office scene is originally imaged with a camera at two viewing positions. In the simulation, pan-tilt movements of the camera are simulated by centering the image at the appropriate pixel. Figure 6.8(a) shows the images corresponding to a fixation on the computer keyboard in the office scene. It shows the RWT images of the scene and the disparity map. These RWT images are the data

used in the actual computation. The Cartesian edge map is also shown here for the reader's apprehension of the disparate scene images and the camera orientations.

As the chair is located at a closer range to the cameras in relation to the keyboard (the current fixation point), it exhibits a non-zero disparity. The disparity value, however, is small as it is located in the periphery. The image disparities in this example are well within the fusional limit. The disparities are computed by applying a $3 \times 3$ windowed correlation over a range of $[-5, 5]$. The disparity results reveal different disparities for objects at different depth from the keyboard. The chair has a large crossed disparity whereas the magazine organizer on the desk shows a non-zero uncrossed disparity.

The fixation exercise in this test is to change the gaze from the computer keyboard to the chair. Three intermediate steps are involved. Initially, the cameras are fixated at the keyboard. A disparity of $-4$ is detected with the chair at a peripheral angle corresponding to $u = -72$ pixels. By the RWT inverse transformation, a $-4$ disparity at $u = -72$ is translated back to the Cartesian domain to a disparity of $-10$ pixels at $x = -101$. Should there be a real hardware camera control to the binocular system, a mapping function is required to map the $-10$ disparity to the disjunctive vergence angle that converge the cameras so that the peripheral disparity of the chair image becomes zero. In this exercise, the vergence is simulated by re-centering the left Cartesian scene image by 5 pixels to the right and the right scene image by 5 pixels to the left. The RWT images are then obtained from the Cartesian scene images for the new camera orientations as though they are from the real RWT cameras. Figure 6.8(b) now shows the result of the first vergence. The chair images at $u = -70$ are now well aligned as seen in the edge map in (b), and the disparities shown in the disparity map demonstrate that zero disparity is achieved with the chair images.

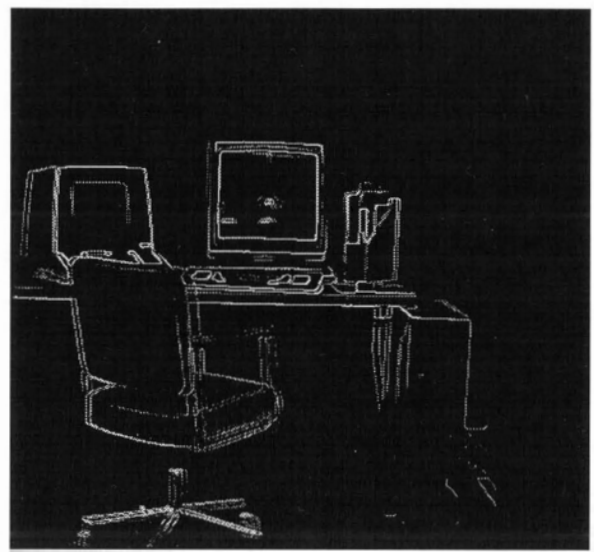
Next, the cameras are panned to the left for an angle corresponding to 72 pixels in the RWT domain. Again, this is accomplished by re-centering both Cartesian images by 101 pixels to the left. Figure 6.8(c) shows the result of this conjunctive versional movement. The chair images now come to the foveal region of the cameras. It is observable that the estimate for the peripheral disparity during the first vergence is not accurate enough for high resolution processing inside the fovea. The disparity map in (c) shows that the residual disparity in the chair images becomes apparent once they are shifted to the fovea. This foveal disparity, however, has a value well within the operating range of the fusional limit since the first vergence has already achieved a good approximation.

Figure 6.8(d) now takes the vergence to completion. The foveal disparity of the chair is computed. It is a small residual disparity of 1 pixel. The cameras are then diverged by an angle corresponding to 1 pixel in the RWT images. Carried out in simulation, the right Cartesian scene image is re-centered by 1 pixel to the right. The RWT is applied to obtain the new images as the result of the second vergence. The disparity map in (d) shows that the cameras are precisely fixating the chair in the fovea.

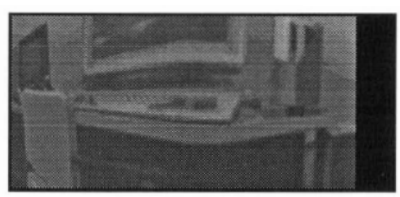
The RWT supports the fixation mechanism in an effective way. If fixation were performed on the conventional uniform-resolution image data, large disparities would have to be calculated. Eminent problems associated with large disparity, such as multiple ambiguous matches and slow computation have to be resolved.

### 6.3.3 A system view

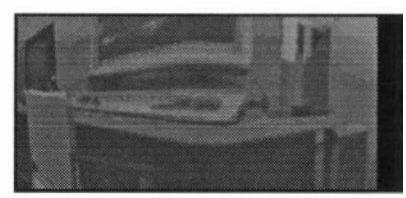
This thesis reports on the design and simulation of a system for the interactive fixation process described above. Figure 6.9 shows the system. It comprises the vergence



edge map



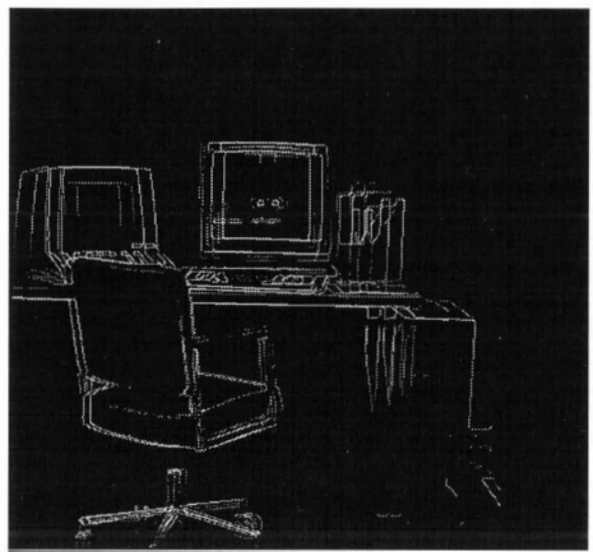
left RWT



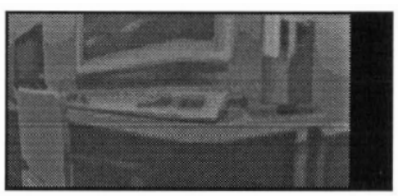
right RWT

disparity

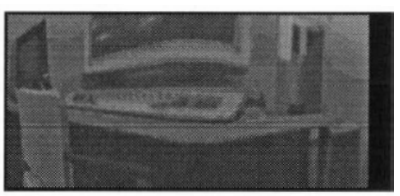
Figure 6.8: (a) Fixation sequence. Initially, fixation is on the computer keyboard.



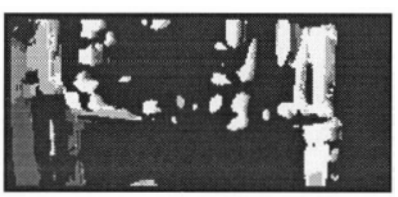
edge map



left RWT

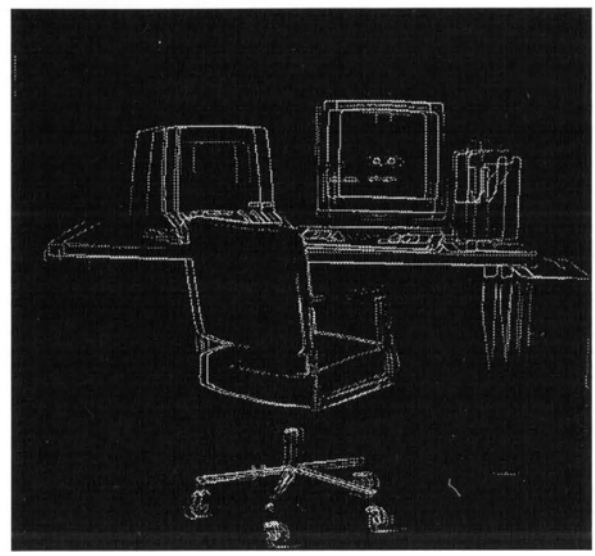


right RWT

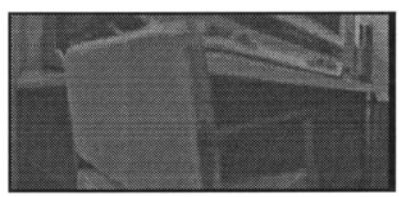


disparity

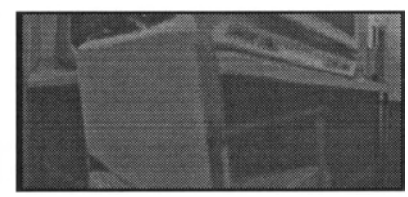
Figure 6.8: (b) First vergence. the peripheral disparity of the chair becomes zero.



edge map



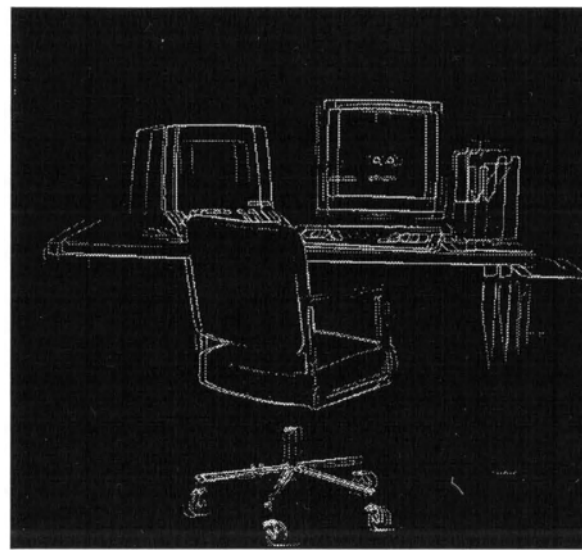
left RWT



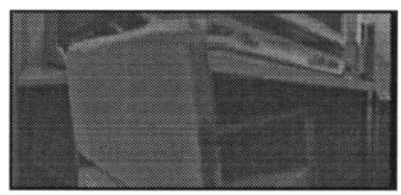
right RWT

disparity

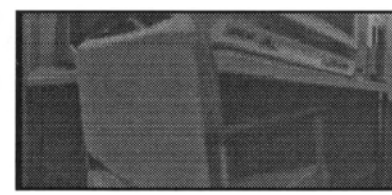
Figure 6.8: (c) Version. The chair is brought to the fovea.



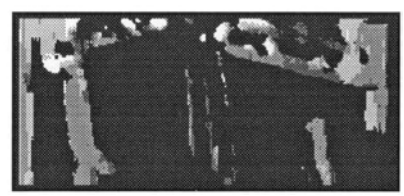
edge map



left RWT



right RWT



disparity

Figure 6.8: (d) Second vergence. Fixation is precisely on the chair.

and version components interfacing with the controller of the camera pan-tilt platform. The next fixation which initiates vergence and version oculomotor sequence is computed by the "where-next" component. Vergence is a slow and visually guided process. It is adjusted according to the disparity, thus completing the feedback loop.

The camera platform houses two cameras each of which has the two degrees of freedom for pan and tilt respectively. Examples of pan-tilt platforms can be found in the previous research [CB92, Kro89, AA93, PUE93] and the reports collected in [CBB93]. In our system, the cameras are RWT cameras which output RWT images of the scene directly. If ordinary cameras were used, the RWT images could be generated from the uniform-resolution images with a Reciprocal-Wedge transformation routine. The gaze angles for vergence and version are mapped to the mechanical movements of pan and tilt for individual camera. The version angle drives identical movements of pan and tilt for both cameras, whereas the vergence is split evenly into disjunctive convergence or divergence between the two cameras.

The component "where-next" represents the high-level intelligent process for selecting the next fixation point in the scene. The left and right RWT images are combined to yield a cyclopean image of the scene (for convenience the left image is used in our simulation). The "where-next" component searches in this cyclopean image for features of interest. In fact, the next-fixation computation is a highly involved process [Yar67]. Although this high-level intelligent process for computing the next fixation is an interesting topic for research, it is beyond the scope of this thesis. In the fixation exercise which involved shifting attention from the computer keyboard to the office chair, the next fixation (the chair) is actually typed in by hand. In the following demonstration of an active fixation system, simplistic heuristic criteria are used to show the usual scanpath behavior in binocular visual exploration.

Once the next fixation has been decided, vergence and version are initiated. Different strategies are employed when computing disparities in the foveal and peripheral regions. Area-based techniques are used in the peripheral regions and feature-based techniques are used in the foveal region. As image data are imprecise under the coarse resolution and reduced size in the peripheral regions, accurate localization of fine features is not expected. Area-based windowed correlation techniques matching image areas are more appropriate at the periphery. Inside the fovea, acute sensitivity is facilitated. More sophisticated feature-based techniques can be employed. Edge features are detected and matched with attributes such as edge orientation and gradient.

In Figure 6.9, two disparity modules are simulated, namely the peripheral disparity and foveal disparity described above. The former is used in the first vergence to eliminate the peripheral disparity. The latter is used in the second vergence to converge precisely on the target inside the fovea.

The position of next fixation is used to drive the versional movement. Synchronous panning motion is produced to swing the cameras in line with the target. Due to the coarse resolution in the periphery, the initial estimate for the magnitude of the panning motion is not able to put the fovea precisely on a feature of the target for foveal processing. The module for foveal-feature position detects the image features inside the fovea. A small adjustment is then initiated by the versional control to bring the target feature in line.

### 6.3.4 A scanpath demonstration

Scanpath is the sequence of fixation that one exercises during a visual scan. The scanpath behavior of the system is demonstrated in an experiment of binocular visual exploration. Although the cognitive modeling of scanpaths is a rigorous research

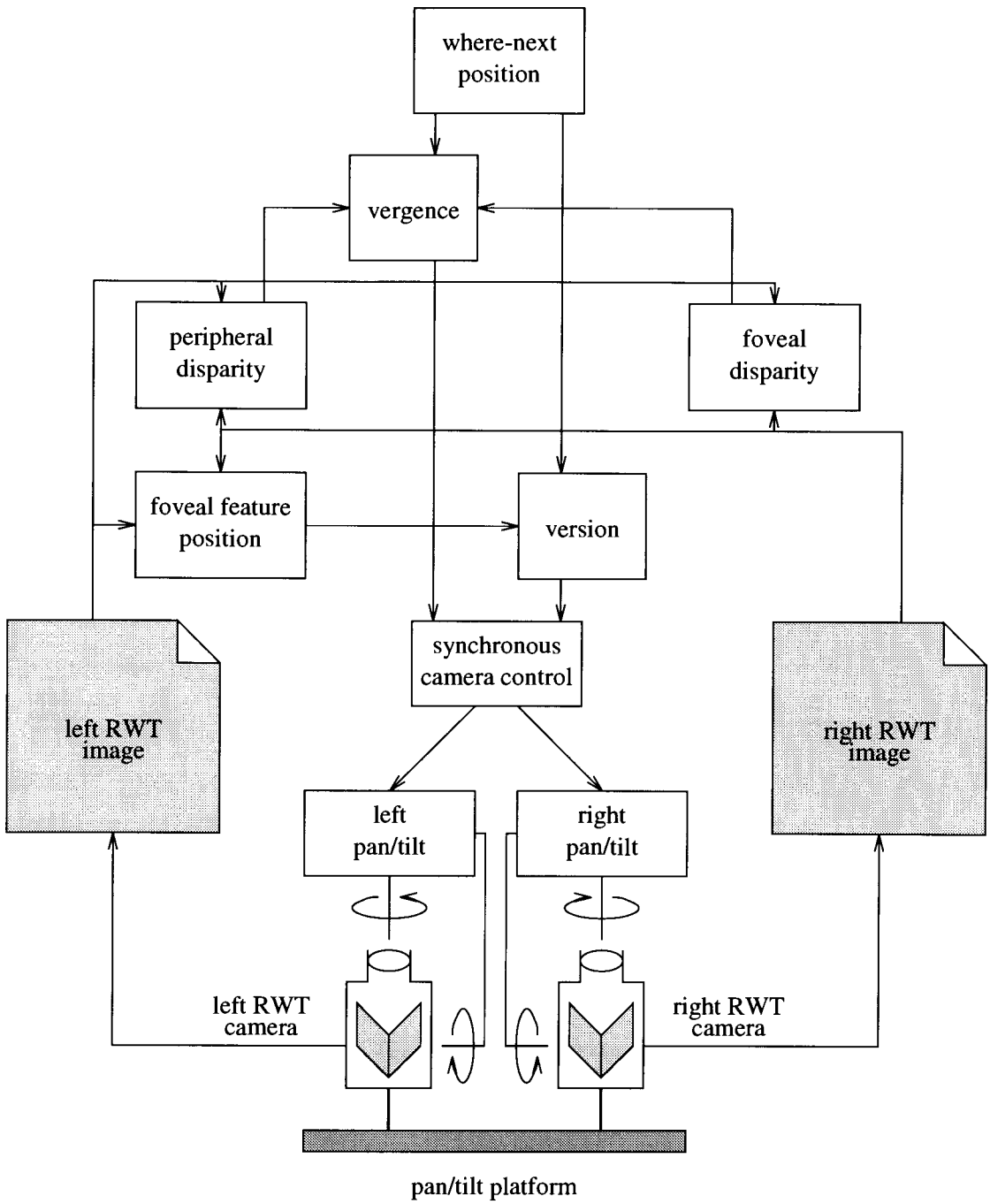



Figure 6.9: An interactive fixation system.

topic in psychology [Yar57, Yar67, NS71c, NS71b, SE81, Gou76], we do not delve into the issues raised therein. Instead, at each stop, simplistic heuristics are employed to determine the next fixation. The resulting scanpath is used to demonstrate the working of our fixation system.

The experiment is conducted with the image data of the office scene in Figure 6.8. Initially, the fixation is set on the computer keyboard on the desk. The next point of interest is chosen based on three considerations. (1) It is a sizable object worth exploring. (2) It has the most disparate image in the current scene. (This drives the system to sweep the entire depth of the scene efficiently.) (3) It has not been explored in detail as yet so that the system would not come to the same object repeatedly. The heuristics are simple enough, yet work successfully in transferring the initial fixation from the computer keyboard to the magazines standing next to the monitor. As shown in Figure 6.10(a), the gaze is then changed to the chair, the computer terminal, and then to the roller wheels of the chair.[1]

The prime observation we emphasize from the outcome of this experiment is the successful working of the fixation system as a whole in implementing the fixation transfer mechanism at each fixation. For example, the initial fixation is on the computer keyboard (Figure 6.10(a-1)). The RWT disparity image in Figure 6.10(b-1) shows an extended area (325 pixels) of 2-pixel disparity occur at the position of $u = 51$ and $v = 33$ (corresponding to the magazines in the scene). The execution log of the simulation program indeed has recorded the following inter-component interactions that happened in the system.

As the "where-next" component evaluated the next fixation to $(51, 33)$, the fixation

[1]Perhaps, that the scanpath is comparable to a scan made by a human subject represents a side-result of this experiment. It may worth further exploration to search for heuristics for visual scanning.

transfer routine was initiated in the vergence and version components. The first vergence was effected by a vergence control to the camera for a divergence angle corresponding to a 2-pixel peripheral disparity at the position $(51, 33)$. Then the version component was initiated with a pan-tilt corresponding to 51 right and 33 up in the RWT coordinates (equivalent to 55 right and 46 up in the Cartesian coordinates). A foveal disparity then was evaluated to $-1$ pixel, causing the vergence component to launch the second vergence for a convergence angle corresponding to a 1-pixel foveal disparity. Finally, an edge feature was detected by the foveal feature component at a position 2 pixels to the left of the center. This resulted in a versional adjustment of 2 pixels, placing the fovea precisely on the edge feature (i.e., on the magazines). The result can be appreciated in Figure 6.10(a-2) which shows a dark edge of the magazines positioned right at the center of both stereo images. The process then continued with the "where-next" selecting position $(-90, 54)$ for the new fixation, and the fixation routine was repeated. Overall, the log records indicate the successful execution by the fixation system as a whole with correct interactions between the various components.

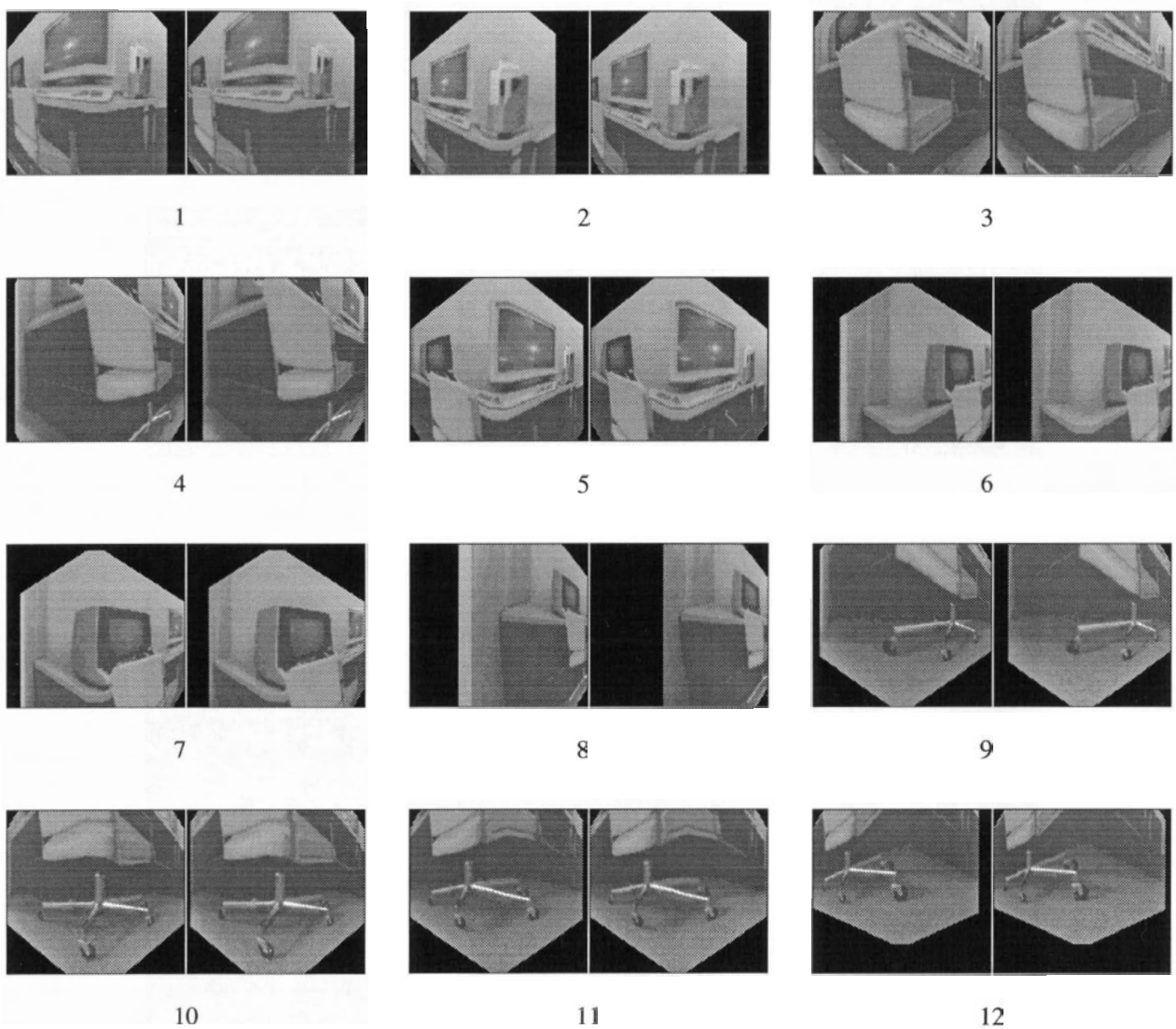



Figure 6.10: (a) Fixation sequence in binocular visual exploration of the office scene.

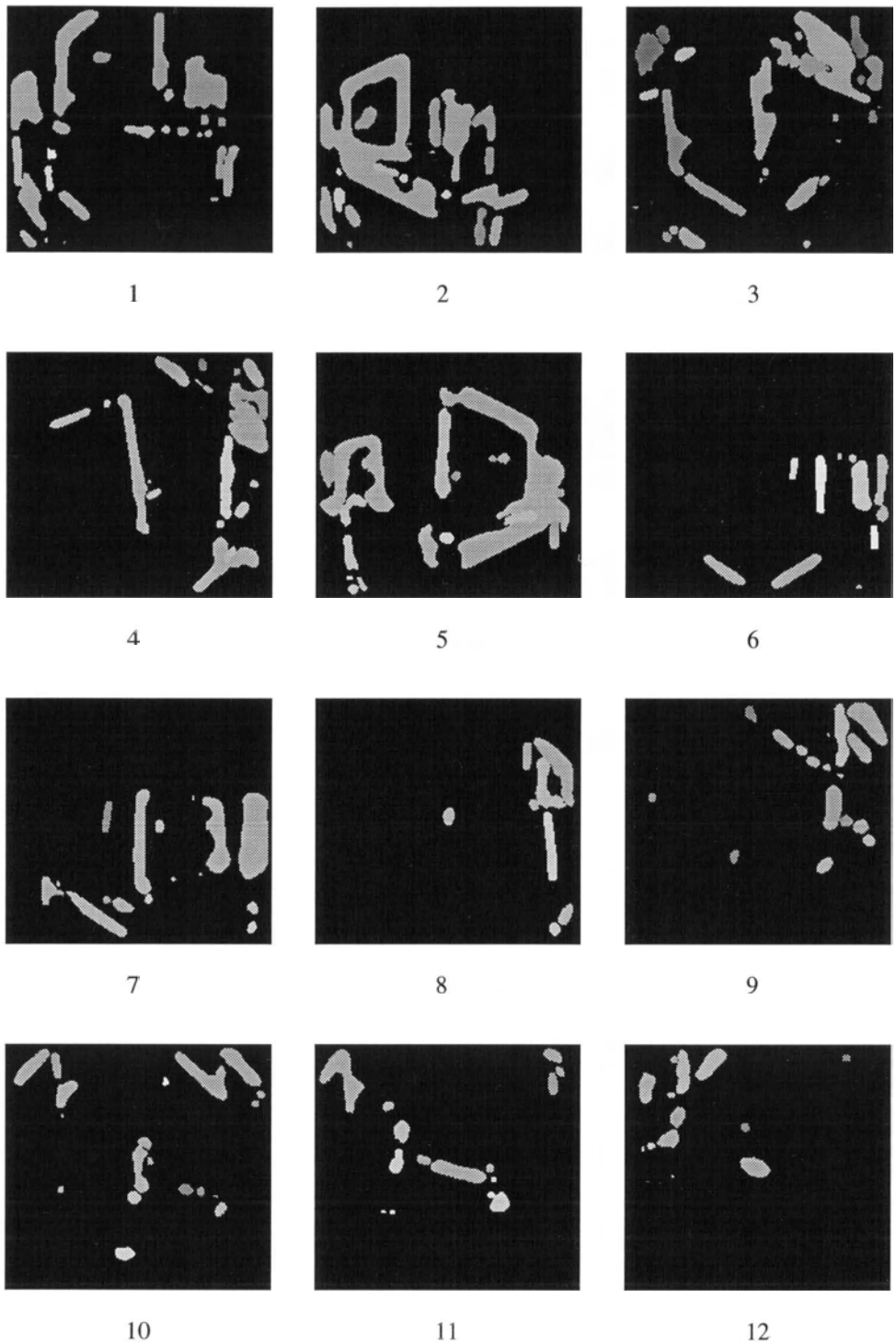



Figure 6.10: (b) Disparities in the RWT images.

# Chapter 7

# Conclusions and Discussion

Departing from the conventional reconstructionist approach, various active vision methodologies have recently been proposed which draw heavily on active probing and search, and emphasize behavioral interaction. One central issue in active vision is foveate sensing. Log-polar mapping has been developed by researchers as a space-variant sensor model for active data acquisition. In this thesis, I have developed an alternative image model called the Reciprocal-Wedge transform (RWT). This chapter summerizes the contributions and suggests some extensions for future research.

## 7.1 Contributions

1. I have developed the Reciprocal-Wedge transform (RWT) as an image model for space-variant sensing.

   The RWT is presented as an alternative model to the log-polar transform. Exploiting the polar coordinate representation, the log-polar does well on centric rotational and scaling transformations. It, however, complicates linear features

and translational transformations. Complementary to the log-polar, the RWT preserves linear features in the image, and its anisotropic variable resolution is suitable for directional space-variant sensing for many vision problems which are translational in nature, such as stereo and linear motion. A concise matrix representation is presented. Properties of the RWT in geometric transformations are described. A pyramidal algorithm for the RWT image mapping is presented. The pyramidal implementation realizes the fast generation of RWT image by exploiting the parallelism and hierarchical linkage of the pyramidal architecture.

2. A camera model is proposed. The optical problem of focusing has been rectified.

   The projective model for the transform leads to a simple RWT camera design. A prominent problem of the simple camera model is the requirement of focusing on a deep image plane along the optical axis. A new hardware camera model is proposed which realizes the RWT in real-time. The new model overcomes the focus problem by using a lens focusing the non-paraxial non-frontal image onto an orthogonally placed RWT plane. Unlike the log-polar sensor, the variable sampling is not a requirement of the RWT sensor circuit. Hence, an ordinary sensor array of rectangular tessellation and uniform grid size can be used which is much cheaper to fabricate.

3. The RWT is shown suitable for recovering depth in both the longitudinal and lateral motion stereo.

   The primary advantage of the proposed method of motion stereo using RWT images is its efficiency since the variable-resolution RWT images have a significantly reduced volume of data. The variable-resolution motion stereo offers more detail and precision in depth recovery at the fovea than at the periphery

of the RWT images, which seems to be natural. Its implication to active sensing appears to be direct.

4. The work of the longitudinal motion stereo is also extended to more general ego motion, especially circular movements (rotations).

   The RWT mapping is shown preserving the circular image motion as corresponding to the original vehicle motion, indicating that the RWT is applicable to general ego motions where world-centered coordinates are employed. This contrasts with the limitations of handling moiton in a viewer-centered coordinate system using the log-polar transform in which only the object at the center is nicely represented.

5. A computational model for binocular fixation is developed.

   The model provides a computational interpretation of the Panum's fusional area in relation to disparity limit in space-variant sensor space. The unique oculomotor pattern for binocular fixation observed in human system appears natural to space-variant sensing. The vergence-version movement sequence is implemented for an effective fixation mechanism in the RWT imaging. In addition, an interactive fixation system is presented to show the various modules of camera control, vergence, version and where-next work together.

## 7.2 Future research

This research does not stop here. It is important that the enthusiasm is maintained by on-going investigation in areas such as space-variant processing, gaze control, or active vision at large. Some suggestions are made in the following as extension of this

work or future directions related to other areas in a wider context.

1. From software to hardware implementation of the RWT.

   Presently, the RWT images are generated from the conventional CCD camera data using software. The slow speed does not meet the requirement of real-time space-variant sensing using RWT. Although the execution on pyramid machine can significantly speed up the process, it is desirable to have the camera model implemented in real hardware. The development of the camera model in this thesis is preliminary. Obviously, the delicate optics of the proposed camera could incur high cost, and the optical design of the RWT lens can be further enhanced. One such problem is that it requires a strong lens or else the camera could be bulky. An interesting feature is that the camera has the potential to implement an adjustable shift RWT with that the scale of space-variance can be adjusted. Presently, the camera model does not address these issues. Future research in these directions would certainly be contributive to the actual implementation of the camera.

   Before an actual hardware camera is available, a hardware video remapper can be an alternative. Weiman, in the work [WJ89], was using a video remapper for generating in video rate the log-polar map from the conventional CCD camera image. As a future research, issues of design and development of the hardware remapper algorithm can be investigated.

2. From restricted motion stereo to general ego-motion.

   The motion stereo models are restricted to longitudinal and lateral motion of the observer. When extended to ego-motion, circular ego-motion is modeled to approximate the course of general motion within a short time span. The immediate

extension can be an investigation into the general planar ego-motion. A more ambitious one would be the research into genuine 3-D ego-motion. Potential applications include the navigation problems for mobile robots whose motions are largely planar, or motion problems such as docking and maneuvering problems related to vision systems ranging from hand-mounted to aircraft-ridden ones.

3. From fixation to active vision.

   Binocular fixation fits in the general direction of active vision. The RWT supports a foveate sensor, and fixation provides the essential gaze control mechanism in an active system. Issues of other types of gaze control such as monocular gaze control for problems ranging from text processing to pattern locator or analyzer can be investigated. This thesis has touched slightly the problem area of camera movements and scanpath modeling. These problems have the potential of applications in attention and visual exploration in situated robots.

# Bibliography

[AA93] 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(10):1007–1029, 1993.

[AH90] J. Aloimonos and J. Y. Hervé. Correspondenceless stereo and motion: planar surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):504–510, 1990.

[AK72] J. M. Allmann and J. H. Kaas. A representation of the visual field in the inferior nucleus of the pulvinar in the owl monkey. *Brain Research*, 40:291–302, 1972.

[AK74] J. M. Allmann and J. H. Kaas. The organization of the second visual area (V-II) in the owl monkey: A second order transformation of the visual hemifield. *Brain Research*, 76:247–265, 1974.

[AK76] J. M. Allmann and J. H. Kaas. A representation of the visual field on the medial wall of occipital-parietal cortex in the owl monkey. *Science*, 191:572–575, 1976.

[Alo90] J. Aloimonos. Purposive and qualitative active vision. In *Proc. International Conference on Pattern Recognition*, pages 346–360, 1990.

[AOG32] Jr. Ames, A., K. N. Ogle, and G. H. Gliddon. Corresponding retinal points, the horoptor and size and shape of ocular images. *Journal of the Optical Society of America*, 22:538,575, 1932.

[Apt45] J. T. Apter. Projection of the retina on the superior colliculus of cats. *Journal of Neurophysiology*, 8:123–134, 1945.

[Arb72] M. A. Arbib. *The Metaphorical Brain.* Wiley, New York, 1972.

[AS89] J. Aloimonos and D. Shulman. *Integration of Visual Modules: An Extension of the Marr Paradigm.* Academic Press, Boston, 1989.

[AWB88] J. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1(4):333–356, 1988.

[Baj85] R. Bajcsy. Active perception vs. passive perception. In *Proc. Workshop on Computer Vision*, pages 55–59, October 1985.

[Baj88] R. Bajcsy. Active perception. *Proceedings of IEEE*, 76(8):996–1005, 1988.

[Baj92] R. Bajcsy. An active observer. In *Proc. DARPA Image Understanding Workshop*, pages 137–147, 1992.

[Bal89] D. H. Ballard. Behavioral constraints on computer vision. *Image Vision Computing*, 7(1), 1989.

[Bal91] D. H. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.

[Bar90] Stephen T. Barnard. Recent progress in cyclops: A system for stereo cartography. In *Proc. DARPA Image Understanding Workshop*, pages 449–455, 1990.

[BBM87] R.C. Bolles, H.H. Baker, and D.H. Marimont. Epipolar-plane image analysis: an approach to determining structure from motion. *International Journal of Computer Vision*, 1:7–55, 1987.

[BF82] S. T. Barnard and M. A. Fischler. Computational stereo. *Computing Surveys*, 14(4):554–572, December 1982.

[BF90] Stephen T. Barnard and Martin A. Fischler. Computational and biological models of stereo vision. In *Proc. DARPA Image Understanding Workshop*, pages 439–448, 1990.

[BHT63] B. P. Bogert, M. J. R. Healy, and J. W. Tukey. The quefrency analysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking. In M. Rosenblatt, editor, *Proc. Symposium on Time Series Analysis*, pages 209–243, New York, 1963. Wiley.

[BJ80a] Peter Burt and Bela Julesz. A disparity gradient limit for binocular fusion. *Science*, 208:615–617, 1980.

[BJ80b] Peter Burt and Bela Julesz. Modifications of the classical notion of panum's fusional area. *Perception*, 9:671–682, 1980.

[Bro65] E. B. Brown. *Modern Optics.* Reinhold Publishing Corp., 1965.

[Bur84] P. J. Burt. The pyramid as a structure for efficient computation. In A. Rosenfeld, editor, *Multiresolution Image Processing and Analysis*, pages 6–35. Springer-Verlag, 1984.

[Bur88] P. J. Burt. Smart sensing within a pyramid vision machine. *Proceedings of IEEE*, 76(8):1006–1015, 1988.

[Car77] R. H. S. Carpenter. *Movements of the Eyes.* Pion, London, 1977.

[CB92] David Coombs and Christopher Brown. Real-time smooth pursuit tracking for a moving binocular robot. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 23–28, 1992.

[CBB93] H. I. Christensen, K. W. Bowyer, and H. Bunke, editors. *Active Robot Vision: Camera Heads, Model Based Navigation and Reactive Control*, volume 6 of *Machine Perception and Artificial Intelligence.* World Scientific, 1993.

[CL86] V. Cantoni and S. Levialdi, editors. *Pyramidal Systems for Image Processing and Computer Vision.* Springer-Verlag, 1986.

[DBC+89] I. Debusschere, E. Bronckaers, C. Claeys, G. Kreider, J. Van der Spiegel, P. Bellutti, G. Soncini, P. Dario, F. Fantini, and G. Sandini. A 2D retinal CCD sensor for fast 2D shape recognition and tracking. In *Proc. 5th International Conference on Solid State Sensors and Transducers*, pages 25–30, Montreux, 1989.

[DC01] R. Dodge and T. S. Cline. The angle velocity of eye movements. *Psychology Review*, 8:145–157, 1901.

[DH72] R.O. Duda and P.E. Hart. Use of the Hough transform to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[DM92] E. D. Dickmanns and B. D. Mysliwetz. Recursive 3-d road and relative ego-state recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):199–213, 1992.

[DW61] P. M. Daniel and D. Whitteridge. The representation of the visual field on the cerebral cortex in monkeys. *Jounral of Physiology*, 159:203–221, 1961.

[EL93] John Ens and Peter Lawrence. An investigation of methods for determining depth from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:97–108, 1993.

[ELT+92] John Ens, Ze Nian Li, Frank Tong, Danpo Zhang, Stella Atkins, and Woshun Luk. A hybrid pyramidal vision machine for real time object recognition. In A. M. Veronis and Y. Paker, editors, *Proc. Fifth Conference of North American Transputer Users Group: Transputer Research and Application 5*, pages 90–103. IOS Press, 1992.

[FA93] C. Fermüller and Y. Aloimonos. The role of fixation in visual motion analysis. *International Journal of Computer Vision*, 11(2):165–186, 1993.

[FBT93] B. V. Funt, M. Brockington, and F. Tong. Conformal transplantation of lightness to varying resolution sensors. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 563–569, 1993.

[Fis24] F. P. Fisher. Fortgesetzte studien über binokularsehen (tschermak): III. experimentelle beiträge zum begriff der sehrichtungsgemeinschaft der netzhäute auf grund der binokularen noniusmethode. *Arch. f. d. ges. Physiol.*, pages 234–246, 1924.

[FJ67] D. Fender and B. Julesz. Extension of panum's fusional area in binocular stabilized vision. *J. of the Optical Society of America*, 57(6):819–830, 1967.

[Fun77] B. V. Funt. WHISPER: A problem-solving system utilizing diagrams and a parallel processing retina. In *Advance Papers of the Fifth International Joint Conference on Artificial Intelligence.* MIT, August 1977.

[GG93] A. Goshtasby and W.A. Gruver. Design of a single-lens camera system. *Pattern Recognition*, 26(6):923–937, 1993.

[GI94] K. D. Gremban and K. Ikeuchi. Planning multiple observations for object recognition. *International Journal of Computer Vision*, 12(2/3):137–172, 1994.

[GLROK94] W. E. L. Grimson, A. Lakshmi Ratan, P. A. O'Donnell, and G. Klanderman. An active visual attention system to play "where's waldo". In *Proc. ARPA Image Understanding Workshop*, pages 1059–1065, 1994.

[GLW92] N. C. Griswold, J. S. Lee, and Carl F. R. Weiman. Binocular fusion revisited utilizing a log-polar tessellation. In Linda Shapiro and Azriel Rosenfeld, editors, *Computer Vision and Image Processing*, pages 421–457. Academic Press, San Diego, 1992.

[Gog61] W. C. Gogel. Convergence as a cue to absolute distance. *Psychology*, 52:287–301, 1961.

[Gou76] John D. Gould. Looking at pictures. In Richard A. Monty and John W. Senders, editors, *Eye Movements and Psychological Processes*, pages 323–345. Lawrence Erlbaum Associates, New Jersey, 1976.

[Gra65] C. Graham. Visual space perception. In C. Graham, editor, *Vision and Visual Perception.* Wiley, New York, 1965.

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

[Ham93] F. Hamit. Near-fisheye CCD camera widens the view. *Advanced Imaging*, 8(3):50–53, 1993.

[Her68] E. Hering. *Die Lehre vom binocularen Sehen.* Engelmann, Leipzig, 1868.

[Hor86] B. K. P. Horn. *Robot Vision.* M. I. T. Press, 1986.

[HW74] D. H. Hubel and T. N. Wiesel. Sequence regularity and geometry of orientation columns in the monkey striate cortex. *Journal of Comparative Neurology*, 158:267–293, 1974.

[Hyd59] J. E. Hyde. Some characteristics of voluntary human ocular movements in the horizontal plane. *Am. J. Ophthalmol.*, 48:85–94, 1959.

[HZ74] E. Hecht and A. Zajac. *Optics.* Addison-Wesley, 1974.

[JBO87] Ramesh Jain, Sandra L. Bartlett, and Nancy O'Brien. Motion stereo using eqo-motion complex logarithmic mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(3):356–369, 1987.

[KA93] Arun Krishnan and Narendra Ahuja. Range estimation from focus using a non-frontal imaging camera. In *Proc. Eleventh National Conference on Artificial Intelligence*, pages 830–835, Washington, D.C., July 1993.

[KA94] Arun Krishnan and Narendra Ahuja. Obtaining focused images using a non-frontal imaging camera. In *Proc. ARPA Image Understanding Workshop*, pages 617–620, 1994.

[KB93] Eric Krotkov and Ruzena Bajcsy. Active vision for reliable ranging: Cooperative focus, stereo, and vergence. *International Journal of Computer Vision*, 11(2):187–203, 1993.

[KD94] K. N. Kutulakos and C. R. Dyer. Recovering shape by purposive viewpoint adjustment. *International Journal of Computer Vision*, 12(2,3):113–136, 1994.

[Kin78] R. Kingslake. *Lens Design Fundamentals.* Academic Press, 1978.

[Kro89] Eric Krotkov. *Active Computer Vision by Cooperative Focus and Stereo.* Springer-Verlag, 1989.

[KSF88] E. Krotkov, J. F. Summers, and F. Fuma. An agile stereo camera system for flexible image acquisition. *IEEE Transactions on Robotics and Automation*, 4(1):108–113, 1988.

[KVdS+90] G. Kreider, J. Van der Spiegel, et al. The design and characterization of a space variant CCD sensor. In *SPIE Vol. 1381 Intelligent Robots and Computer Vision IX: Algorithms and Techniques*, Boston, November 1990.

[LG82] G. Ligthart and F. C. A. Groen. A comparison of different aotofocus algorithms. In *Proc. Sixth International Conference on Pattern Recognition*, pages 597–600, October 1982.

[Li91] Ze Nian Li. Vision in pyramids — object recognition in real time. In *Proc. International Conference on CAD/CAM, Robotics, and FOF*, pages 344–349, 1991.

[Li94a] Ze Nian Li. Disparity gradient revisited. In *Int. Symp. on Information, Computer, and Control*, pages 468–473, 1994.

[Li94b] Ze Nian Li. Stereo correspondence based on line matching in Hough space using dynamic programming. *IEEE Transactions on Systems, Man and Cybernetics*, 24(1):144–152, 1994.

[LMD+90] R. A. Lotufo, A. D. Morgan, E. L. Dagless, D. J. Milford, J. F. Morrissey, and B. T. Thomas. Real-time road edge following for mobile robot. *Electronics and Communications Engineering Journal*, 2(1):35–40, 1990.

[LTR95] Ze Nian Li, Frank Tong, and Xao Ou Ren. Applying reciprocal-wedge transform to ego motion. In *Proc. IASTED International Conference on Robotics and Manufacturing*, pages 256–259, 1995.

[Lun48] R. K. Luneburg. *Mathematical Analysis of Binocular Vision.* Princeton University Press, Princeton, NJ, 1948.

[LZ93] Ze Nian Li and Danpo Zhang. Fast line detection in a hybrid pyramid. *Pattern Recognition Letters*, 14(1):53–63, 1993.

[Mar82] D. Marr. *Vision.* W. H. Freeman, 1982.

[MF81] J. E. W. Mayhew and J. P. Frisby. Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence*, 17:349–385, 1981.

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

[MP79] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proc. Royal Society of London, Series B*, 204:301–328, 1979.

[Nev76] R. Nevatia. Depth measurement by motion stereo. *Computer Graphics and Image Processing*, 5:203–214, 1976.

[Not70] D. Noton. A theory of visual pattern perception. *IEEE Transactions on System, Science and Cybernetics*, 6:349–357, 1970.

[NS71a] D. Noton and L. Stark. Eye movements and visual perception. *Scientic American*, 224(6):34–43, 1971.

[NS71b] D. Noton and L. Stark. Scanpaths in eye movements during pattern perception. *Science*, 171:308–311, 1971.

[NS71c] D. Noton and L. Stark. Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 11:929–942, 1971.

[OC91] T. J. Olson and D. J. Coombs. Real-time vergence control for binocular robots. *International Journal of Computer Vision*, 7(1):67–89, 1991.

[Ogl64] K. N. Ogle. *Researches in Binocular Vision.* Hafner, New York, 1964.

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

[OK93] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):353–363, 1993.

[Ols93] Thomas J. Olson. Stereopsis for verging systems. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 55–60, 1993.

[Pen87] Alex Paul Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:523–531, 1987.

[Pol41] S. Polyak. *The Retina.* University of Chicago Press, Chicago, 1941.

[PUE93] K. Pahlavan, T. Uhlin, and J. O. Eklundh. Dynamic fixation. In *Proc. 4th International Conference on Conputer Vision*, pages 412–419, Berlin, 1993.

[Red94] D. Redfern. *Maple Handbook: Maple V Release 3, 2nd Ed.* Springer-Verlag, 1994.

[Ric71] W. Richards. Anomalous stereoscopic depth perception. *J. of the Optical Society of America*, 61(3):410–414, 1971.

[Rob64] D. A. Robinson. The mechanics of human saccadic eye movements. *Journal of Physiology (London)*, 174:245–264, 1964.

[RS90] A. S. Rojer and E. L. Schwartz. Design considerations for a space-variant visual sensor with complex-logarithmic geometry. In *Proc. 10th International Conference on Pattern Recognition, volume II*, pages 278–285, Atlantic City, 1990.

[RW61] C. Rashbass and G. Westheimer. Disjuctive eye movements. *Journal of Physiology (London)*, 159:339–360, 1961.

[Sch77] E. L. Schwartz. Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. *Biological Cybernetics*, 25:181–194, 1977.

[Sch80] E. L. Schwartz. Computational anatomy and functional architecture of striate cortex: spatial mapping approach to perceptual coding. *Vision Research*, 20:645–669, 1980.

[SD90] G. Sandini and P. Dario. Active vision based on space-variant sensing. In *Proc. 5th International Symposium on Robotics Research*, pages 75–83, Tokyo, 1990.

[SE81] Lawrence Stark and Stephen R. Ellis. Scanpaths revisited: Cognitive models direct active looking. In Richard A. Monty Dennis F. Fisher and John W. Senders, editors, *Eye Movements: Cognition and Visual Perception*, pages 193–226. Lawrence Erlbaum Associates, 1981.

[Spe70] G. Sperling. Binocular vision: A physical and a neural theory. *American Journal of Psychology*, 83:461–534, 1970.

[SS91] M. J. Swain and M. Stricker. Promising directions in active vision. Technical Report TR CS 91-27, University of Chicago, 1991.

[SS93] M. J. Swain and M.(Ed.) Stricker. Promising directions in active vision. *International Journal of Computer Vision*, 11(2):109–126, 1993.

[ST80] G. Sandini and V. Tagliasco. An anthropomorphic retina-like structure for scene analysis. *Computer Graphics and Image Processing*, 14:365–372, 1980.

[THKS88] C. Thorpe, M. H. Hebert, T. Kanade, and S. A. Shafer. Vision and navigation for the Carnegie-Mellon Navlab. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(3):362–373, 1988.

[TL91] Frank Tong and Ze Nian Li. Backprojection for stereo matching using transputers. In *Proc. SPIE Symposium on Machine Vision Architectures, Integration, and Applications*, volume 1615, pages 373–385, Boston, MA, 1991.

[TL92] Frank Tong and Ze Nian Li. On improving the accuracy of line extraction in hough space. *International Journal of Pattern Recognition and Artificial Intelligence*, 6(5):831–848, 1992.

[TL93] Frank Tong and Ze Nian Li. The reciprocal-wedge transform for space-variant sensing. In *Proc. International Conference on Computer Vision*, pages 330–334, Berlin, 1993.

[TL94] Frank Tong and Ze Nian Li. Reciprocal-wedge transform in motion stereo. In *Proc. IEEE International Conference on Robotics and Automation*, pages 1060–1065, San Diego, 1994.

[TL95] Frank Tong and Ze Nian Li. Reciprocal-wedge transform for space-variant sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):500–511, 1995.

[TM41] S. A. Talbot and W. H. Marshall. Physiological studies on neural mechanisms of visual localization and discrimination. *American Journal of Ophthalmology*, 24:1255–1263, 1941.

[TP75] S.L. Tanimoto and T. Pavlidis. A hierarchical data structure for picture processing. *Computer Graphics and Image Processing*, 4:104–119, 1975.

[TS90] M. Tistarelli and G. Sandini. Estimation of depth from motion using an anthropomorphic visual sensor. *Image and Vision Computing*, 8(4):271–278, 1990.

[TS93] M. Tistarelli and G. Sandini. On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):401–410, 1993.

[Tso92] J. K. Tsotsos. On the relative complexity of active vs. passive visual search. *International Journal of Computer Vision*, 7(2):127–141, 1992.

[TZ84] W. Teoh and X. D. Zhang. An inexpensive stereoscopic vision system for robots. In *Proc. International Conference on Robotics*, pages 186–189, 1984.

[Uhr87] L. Uhr, editor. *Parallel Computer Vision.* Academic Press, 1987.

[VdSKC+89] J. Van der Spiegel, G. Kreider, C. Claeys, I. Debusschere, G. Sandini, P. Dario, F. Fantini, P. Bellutti, and G. Soncini. A foveated retina-like sensor using CCD technology. In C. Mead and M. Ismail, editors, *Analog VLSI Implementation of Neural Systems*, pages 189–211. Kluwer Academic Publishers, Boston, 1989.

[WB91] C. C. Weems and J. H. Burrill. The image-understanding architecture and its programming environment. In V. K. Prasanna Kumar, editor, *Parallel Architectures and Algorithms for Image Understanding*, pages 525–562. Academic Press, 1991.

[WC79] C. F. R. Weiman and G. Chaikin. Logarithmic spiral grids for image processing and display. *Computer Graphics and Image Processing*, 11:197–226, 1979.

[WJ89] C. F. R. Weiman and R. D. Juday. Tracking algorithms using log-polar mapped image coordinates. In David P. Casasent, editor, *Proc. SPIE Symposium on Intelligent Robots and Computer Vision VIII: Algorithms and Techniques*, pages 843–853, 1989.

[Yar57] A. L. Yarbus. Eye movements during changes of the stationary points of fixation. *Biofizika*, 2:698–702, 1957.

[Yar67] A. L. Yarbus. *Eye Movements and Vision.* Plenum, New York, 1967.

[You89] David Young. Logarithmic sampling of images for computer vision. In *Proc. 7th Conference on Artificial Intelligence and Simulation of Behavior*, pages 145–150, 1989.

[YS89] Y. Yeshurun and E. L. Schwartz. Shape description with a space-variant sensor: algorithms for scan-path, fusion, and convergence over multiple scans. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1217–1222, 1989.

[ZV72] V. P. Zinchenko and N. Y. Vergiles. *Formation of Visual Images: Studies of Stabilized Retinal Images (Translated by Consultants Bureau).* Plenum, New York, 1972.